Research Article

A Cost Function Approach to the Prediction of Passenger Distribution at the Subway Platform

Xiaoxia Yang1, Xiaoli Yang2, Zhenling Wang1, and Yuanlei Kang3

1Institute of Complexity Science, Qingdao University, Qingdao 266071, China
2College of Management and Economics, Tianjin University, Tianjin 264670, China
3CRRC Qingdao Sifang CO., LTD., Qingdao 266111, China

Correspondence should be addressed to Xiaoxia Yang; yangxiaoxia@qdu.edu.cn

Received 16 April 2018; Accepted 10 September 2018; Published 1 October 2018

Academic Editor: Juan C. Cano

Copyright © 2018 Xiaoxia Yang 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.

This paper proposes a cost function approach to predict the passenger distribution at the platform, which is contributed to provide services to the passenger safety and convenience and also an efficient use of the subway platform. According to the limited observation and field data collection of Beijing Xuanwumen subway station, passenger behaviors and basic attributes at the platform are analyzed. Based on the analysis and investigation, factors including the distance to the waiting area, the passenger density in the visual field, and the length of the waiting area occupied by passengers are put forward as important factors to affect the choice of the waiting area. The determination of the real-time choice defined by these factors is applied to model the passengers’ waiting area choice behaviors. Simulation experiments are run for the model calibration and validation combining with the collected field data. The results show that the passenger distribution which arises from the model is capable of keeping consistent with the actual distribution in the rough. The model is helpful for controlling how heavy carriages are congested and providing suggestions to optimize the layout of platform facilities.

1. Introduction

In most cities, subways have contributed to the reduction of surface traffic and a better utilization of land resources. However, they have given rise to complex passenger flows due to complex station geometries that are most pronounced during peak hours. The passenger dynamics relates to not only physical characteristics but also behavior habits [1]. During the last few decades, the interest in passenger/pedestrian dynamics has increased significantly for researchers from physics, psychology, computer science, etc. [2–5]. So far, research on passenger/pedestrian dynamics at public places mainly involves modeling, fundamental diagram, behavior characteristics, evacuation dynamics, route choice, etc.

Pedestrian modeling methods are basically investigated from both the microscopic level and the macroscopic level [6–8]. The microscopic pedestrian model mainly contains the cellular automata model [9–12], the social force model (SFM) [13, 14], the agent-based model [15, 16], and the game theory model [17, 18]. The microscopic cellular automata model is a grid-based discrete model [19], which is more suitable for pedestrian dynamics in the complex environment because of its simplicity and efficiency [20, 21]. The social force model is a continuous force-driven model, and its first application mainly focused on emergency evacuation from buildings [22, 23]. The agent-based model usually uses the virtual agents to develop the social structure; it provides an innovative perspective to study pedestrian dynamics [24]. Manley et al. [25] presented an agent-based model that can determine effects of changing the built environment on the overall evacuation time of passengers. The macroscopic pedestrian model focuses on studying the whole moving trend of the crowd described by the average speed, density, location, and time. The typical macroscopic model usually means the fluid dynamic model which treats the pedestrian flow as fluid or gas [26, 27]. Applications of the pedestrian simulations prompted a number of commercial evacuation software programs, such as EGRESS, EXODUS, SIMULEX, EXITT, WAYOUT, FDS+Evac, etc. [28, 29].
Generally, the validity of a pedestrian model is obtained by comparing the simulation results with the fundamental diagrams. Based on the field observation, some characteristics of walking behaviors are found and also represented by pedestrian models, such as the typical self-organization phenomena which refer to lane formation in the bidirectional flow [30], strips in the crossing flow [31], and “zipper” effects. Ensuring the safety of evacuation has been an important factor during the design of public infrastructure [32]. Pedestrian models have been applied to investigate evacuation problems from different perspectives, such as computing evacuation time and finding evacuation bottlenecks [33]. Route choice not only has a direct effect on trying to stay away from congested routes but also affects the evacuation time significantly [34].

Platform, as an important part of rail transport, usually has a relatively large and complicated passenger flow. Passenger behaviors at the subway platform consist of walking, waiting area choice, waiting for a train, and alighting and boarding. Much research has been done on alighting and boarding behaviors from data statistic and modeling, as well as the study on walking behaviors [35]. Seriani and Fernandez proposed a method to determine the effect of passenger traffic management in the alighting and boarding time by means of simulations and experiments [36]. Wu and Ma introduced a new classification method of the crowdedness level at the platform, considering passenger flow characteristics and boarding services [37]. Johansson et al. studied the waiting behaviors based on the SFM by introducing a series of extensions [22]. Basically, passenger distribution at the platform, resulting from the waiting area choice behavior, directly affects the congestion degree in carriages, the research on which can provide suggestions to optimize the layout of the platform facilities and thereby adjust passenger distributions. However, as the process of waiting area choice is chaotic, modeling this choice behavior is full of uncertain factors beyond our knowledge.

Existing studies on the waiting area choice behaviors are still comparatively few [38, 39] because a lot of investigation labor and time are required during the field data collection [40]. According to Szplett and Wirasinghe [41], the distribution of passengers at a platform is not uniform and changes with time. Wu et al. proposed a passenger distribution model based on the potential energy function for waiting areas [38]. However, the random characteristics of passenger distribution at the platform and also the detailed passenger walking dynamics of choosing a waiting area are not considered. The main contribution of this paper is that the proposed model combines the real-time distribution of passengers at the platform with the microscopic passenger movement dynamics for improving the simulation authenticity and accuracy. Given the initial passenger distribution at the platform and the traffic inflow, the proposed model can exhibit the behaviors of passengers’ waiting area choice in real time.

Different pedestrian models provide different levels of modeling characteristics and details. This paper chooses the SFM as the reference passenger driven model which can offer a good performance of reflecting pedestrian behaviors, so the combination of the waiting area choice model and the SFM can be capable enough to depict passengers’ searching behaviors at the platform, while at the same time also guaranteeing the accuracy.

The structure of this paper is as follows. Section 2 proposes a cost function approach to predict passenger distribution at the platform, in which the distance to the waiting area, passenger density in the visual field, the length of the waiting area occupied by passengers, and other factors are all considered. Also, the passenger driven model which refers to the very frequently applied SFM is introduced. Section 3 gives passengers’ basic attributes at Beijing Xuanwumen subway station, and calibrates and validates the proposed model in this paper based on the limited observation and field data at Beijing Xuanwumen subway station and Shanghai natural history museum station. Section 4 reviews the key discoveries.

2. Model

In this section, we define a cycle as the time gap between an arrival of two successive trains shown in Figure 1. During a cycle, the total time gap can be divided into dwell time and separation time of the train “k” before an arrival of the next train “k + 1”. Passenger behaviors such as searching behaviors, waiting behaviors, and alighting and boarding behaviors occur in each cycle. When passengers enter the platform during the separation time, they will first begin searching for a relatively appropriate waiting area which is not only within an acceptable range but also not very crowded yet; then waiting behaviors and alighting and boarding behaviors will occur. When passengers enter the platform during the dwell time, waiting behaviors may not be required if there is enough time for boarding and also enough space inside of the corresponding carriage. According to Wu and Ma [42],

![Figure 1: The diagram of a cycle.](image-url)
a platform mainly consists of two areas shown in Figure 2: circulating areas and waiting areas. Both waiting behaviors and alighting and boarding behaviors are mainly carried out in the waiting areas.

Subway stations are mostly in a relatively closed underground space, and pedestrians walking in these confined space often produce some different behavior habits from walking in the ordinary sidewalk. Firstly, passenger traffic at the platform varies significantly with an arrival of a train. Note that passenger traffic here means the total number of passengers who may enter or leave the platform. Passenger traffic is not continuously invariant but shows sudden increase or sudden decrease especially during dwell time and the initial stage of separation time. The sudden change of passenger traffic relates to the location of a station, time, the weather condition, large events, etc. The impact of a sudden increase of passenger traffic on various facilities at the station is very large especially around the stair/escalator, which may further result in the formation of a security risk. Secondly, passenger flow at the platform has the nonuniform characteristics in time and space. During the morning and evening peak hours of workdays, passenger traffic is particularly large which is several times higher than that at any other time of a day. Large differences in the passenger traffic for the areas of a platform with two different driving directions do exist. Take the platform of line 4 at Beijing Xuanwumen subway station, for example; passenger density at the platform with Anheqiao North direction which leads to the city center is obviously higher than that with the opposite direction during the morning rush hours, while the situation is just reversed during the evening rush hours. This is because people mainly work in the inner ring of the city and live in the suburbs.

According to Hoogendoorn and Bovy [43], pedestrian behavior has three levels which are, respectively, strategic level, tactical level, and operational level. Passenger travel purposes, bonded groups, etc. at the platform are assumed to be known in this paper which are all at the strategic level. This directly determines which side of the platform passengers choose. Note that the platform could be with the island type or the side type. When passengers enter the platform, the waiting area choice behavior affected by both personal factors and external factors is at the tactical level; this decision is mostly performed after some time which is needed to gather information. This paper assumes that passengers can get information from time to time and could make quick decisions once entering the platform. The novel contribution of our paper is to present a waiting area choice model. The force-driven equation for passengers’ walking in this paper is the SFM [23], in which both physical and motivation forces are considered; this is thus at the operational level.

2.1. Waiting Area Choice Model for Passengers. Through a large number of investigations at the subway platform and also analyzing the video data, we discover that waiting area choice behavior for passengers is more likely to be affected by the passenger density in the visual field, the distance to the waiting area, a cycle at the dwell time or separation time, large pieces of luggage, and other uncertain factors. Each influence factor can result in generating a part of cost. In this paper, we assume that a passenger makes continuous choices according to the real-time situations, and this decision-making considers all the above influence factors.

This paper first puts forward the concept of the expected cost \( C_{iw}^w(t, x, y) \) for the passenger \( i \) in the position \((x, y)\) to choose a waiting area \( w \) at time \( t \). When determining which waiting area, the choice is governed by the expected cost, which represents a possible minimum cost is an optimal waiting area. We express \( C_{iw}^w(t, x, y) \) as

\[
C_{iw}^w(t, x, y) = \sum_{k=1}^{3} C_{iw}^k(t, x, y) + \xi, \tag{1}
\]

where \( C_{iw}^k(t, x, y) \) is the expected cost caused by detailed factors to affect the selection of waiting areas, and \( \xi \) is an uncertain cost caused by some other uncertain factors beyond our knowledge which fits the normal function.

We introduce \( C_{1w}^w \), \( C_{2w}^w \), and \( C_{3w}^w \) as the expected costs resulting from the distance to the waiting area, the length of the waiting area occupied by passengers, and passenger density in the area from his/her current position to the corresponding waiting area shown in Figure 3, respectively.

We compute \( C_{1w}^w \) as

\[
C_{1w}^w(t, x, y) = \exp\left( \frac{\alpha_1 \cdot d_{iw}^w(t, x, y) \cdot \mu(t, x, y)}{\beta_1} \right). \tag{2}
\]

We define \( d_{iw}^w(t, x, y) \) as the distance from passenger's current position \((x, y)\) to the center point of the waiting area \( w \) at time \( t \). The influence degree of \( d_{iw}^w(t, x, y) \) on the waiting area choice can also be affected by the passenger density nearby. We hypothesize that variable \( \mu(t, x, y) \) represents this influence; it is given by (3). It can be interpreted as follows: when passengers enter the platform from the stair/escalator, and then observe a large number of pedestrians gathering in the waiting areas which are near the stair/escalator, they may prefer to choose a further waiting area.

\[
\mu(t, x, y) = \begin{cases} 
1 & \text{if } \rho(t, x, y) \leq \rho_0 \\
\frac{\rho(t, x, y)}{\rho_0} & \text{if } \rho(t, x, y) > \rho_0,
\end{cases} \tag{3}
\]
where \( \rho(t, x, y) = N/ \sum_{i=1}^{N} |A_i| \) is the passenger density nearby within the sector area with a radius \( R_n \) and is illustrated in Figure 4. Note that the density is calculated by the Voronoi diagram [44], \( N \) is the number of passengers in the sector, and \( A_i \) is the area of each Voronoi cell. Actually, the radius of a vision field \( R_v \) is usually larger than \( R_n \). We hypothesize a central angle of \( 2\theta = 170^\circ \). \( \rho_0 \) is determined according to the level of service [45] listed in Table 1. In this paper, we assume \( \rho_0 = 0.83 \text{ p/m}^2 \). When \( \rho_0 \geq 0.83 \text{ p/m}^2 \), passengers will feel uncomfortable and their mobilities will be restricted severely at these areas. At this time, some evacuation strategies will be adopted at the platform, for example, public broadcasting or staff guiding passengers to somewhere with relatively few people as shown in Figure 5.

\[
\beta_i \text{ is a sensitive positive parameter for scaling the expected cost } C_{ij}^w, \quad \alpha_i \text{ is an inertia positive parameter, which is affected by the passenger’s arriving time at the platform and whether or not they are carrying large pieces of luggage. Then, } \alpha_i \text{ is expressed by}
\]

\[
\alpha_i = \begin{cases} 
\frac{DT}{TL} & \text{if } (\text{dwell}(t) = 1 \text{ or luggage}(t) = 1) \text{ and } d^{iw} > d_0; \\
1 & \text{if others.}
\end{cases}
\]

Here, \( DT \) means dwell time which is defined in Figure 1, and \( TL \) refers to the time left in the dwell time period. \( \text{dwell}(t) = 1 \) represents a cycle at the dwell time in time instant \( t \), otherwise at the separation time. \( \text{luggage}(t) = 1 \) represents the passenger carrying large pieces of luggage, otherwise not carrying. When passengers at somewhere of the platform are informed of the coming of a train, they may prefer to choose a nearby waiting area with relatively few people. If passengers carry large pieces of luggage which can result in the walking speed reducing, the distance factor will be considered as a very important influence factor in this paper, and passenger would like to choose a nearer waiting area. In this paper, we hypothesize that \( d_0 \) is a positive constant.

The number of passengers in the waiting area is another important factor to affect the decision-making reflected by the variable \( C_{ij}^w(t, x, y) \). We define the expression of
The parameter $C^{'iw}(t, x, y)$ as a piecewise function according to a cycle at different time levels; it is given by

$$C^{'iw}(t, x, y) = \begin{cases} 
\beta_3 L^w_{after}(t) + \alpha_3 \frac{1}{L^w - L^w_{after}(t)} & \text{if } dwell(t) = 1; \\
\beta_2 L^w_{before}(t) + \alpha_2 \frac{1}{L^w - L^w_{before}(t)} & \text{if } dwell(t) = 0.
\end{cases} \tag{5}$$

$L^w_{after}(t)$ and $L^w_{before}(t)$ are the length of the waiting area occupied by passengers at the dwell time and the separation time, respectively. According to our observation, the length of queue becomes shorter with an arrival of a train. $L^w$ is the physical length of the waiting area $w$, which is determined according to the structure of the platform. In this paper, we only consider the situation of $L^w \geq L^w_{after}(t)$ for the subsequent model validation and calibration. $\beta_2$ is a sensitive positive parameter for scaling. $\alpha_2$ is an inertia positive parameter which determines the attractive ability of the left space of a waiting area.

In this paper, $L^w_{after}(t)$ and $L^w_{before}(t)$ are given according to [42]:

$$L^w_{after}(t) = 0.694 n^w(t)^{0.510}, \quad L^w_{before}(t) = 0.685 n^w(t)^{0.546}. \tag{6}$$

Here, $n^w(t)$ denotes at the waiting area $w$ at time $t$.

The passenger density $\rho^w(t, x, y)$ in the area $S_w$ shown in Figure 3 is another factor that needs to be considered. Let $C^{'iw}(t, x, y)$ denote this influence factor; we define this factor as

$$C^{'iw}(t, x, y) = \exp \left( \frac{\rho^w(t, x, y)}{\beta_3} \right). \tag{7}$$

$\beta_3$ is a sensitive positive parameter for scaling the expected cost $C^{'iw}$. Basically, the alighting passengers can leave the platform within a short time, and therefore they affect the waiting area choice behaviors mostly concentrated in the start stage of the separation time.

Therefore, the optimal waiting area $w^*$ for the passenger $i$ is given by

$$w^* = \arg \min \ C^{'iw}, \quad w = 1, 2, 3, \ldots, n - 1, n. \tag{8}$$

We define $n$ as the total number of waiting areas that is related to the physical structure of a platform.

In this paper, we assume that passenger determines an optimal waiting area from time to time until his or her distance to the optimal waiting area is less than a detection threshold. Behavior, like changing to another waiting area during boarding, is not considered in this paper.

### 2.2. Modeling Passenger Movement

In this section, we will give a brief description of the passenger driven model based on the SFM. The SFM is proposed by Helbing et al. [13, 23], where pedestrians are driven by three types of forces: the desired force, $\overrightarrow{f^i}$; the interaction force between pedestrians $i$ and $j$, $\overrightarrow{f_{ij}}$; the interaction force between the pedestrian $i$ and walls $w$, $\overrightarrow{f_{iw}}$. The SFM has been a prevalent microscopic simulation model in pedestrian dynamics and is still being investigated and embedded into the numerical simulation software such as Anylogic [46] and FDS+Evac [29]. Some self-organization phenomena are also represented through the application of the SFM [47], which further reveals the usability of the model.

The mathematical formula of the SFM is expressed by

$$m_i \frac{d\overrightarrow{v_i}}{dt} = \overrightarrow{f^i} + \sum_{j \neq i} \overrightarrow{f_{ij}} + \sum_{w} \overrightarrow{f_{iw}}, \tag{9}$$

where $m_i$ is the mass of pedestrian $i$ and $\overrightarrow{v_i}(t)$ is his/her walking velocity at time $t$. $\overrightarrow{f^i}$ indicates the pedestrian's willingness to achieve the desired speed.

At the subway station, we can always observe the bonded groups such as families, friends, colleagues, and couples, especially on the weekends. This paper also considers the effects of bonded groups based on the SFM, and we directly adopt the bonding force proposed in [1] which has already been calibrated and validated. As bonded groups could bear the shorter distance between each other because of their special relationships, the bonding force $k_{ij}^{bond}$ has the opposite direction of the force $\overrightarrow{f_{ij}}$. The force-driven equation for passengers in the bonded group is given by

$$m_i \frac{d\overrightarrow{v_i}}{dt} = \overrightarrow{f^i} + \sum_{j \neq i, j \in B(i)} \overrightarrow{f_{ij}} + \sum_{w} \overrightarrow{f_{iw}} + \sum_{j \in B(i)} (k_{ij}^{bond} \overrightarrow{f_{ij}}). \tag{10}$$

$f_{ij}^{bond}$ is the interaction force between passengers $i$ and $j$ who belong to the set of bonded groups $B(i)$. For passengers in the same bonded group, we assume that they would choose the same waiting area.

It is easy for a pedestrian to vibrate continuously in a high density crowd, especially when he or she is in the bottleneck area [48]. Pelechano et al. introduced a "stopping rule" to avoid this behavior, where his/her own personality, the walking directions of others, and pedestrian's current situation were all taken into account [48]. Besides, a "respect" mechanism as a self-stopping mechanism was introduced by Parisi et al., which reproduced the experimental data and also avoided the vibration [49]. In this paper, we adopt the same "respect" mechanism in [49]. The respect distance $D_R$ for the passenger $i$ is $D_R = R_F \cdot r_i$, where $R_F$ is the respect factor. Once any other pedestrian touches the respect area of the pedestrian $i$ which is $\pi \cdot D_R^2$, the desired walking speed $v_i^0$ will be set to 0 until the respect area is free. In this paper, it is also assumed that $R_F = 0.7$, and we refer the readers to [49] for more details.

When passengers arrive at the target waiting area, we assume they will queue up to two columns at the mark
insertions of the waiting area, and the desired positions will relate to \( L_{\text{after}}^w(t) \) or \( L_{\text{before}}^w(t) \). Moreover, this paper mainly focuses on the waiting area choice behavior of passengers at the tactical level, and the alighting and boarding behaviors are not investigated.

2.3. Modeling of Passenger Distribution at the Platform. Passenger distribution at the subway platform could be predicted by the combination of waiting area choice model and passenger driven model. The target waiting area \( w^* \) determined by (8) affects a passenger’s desired walking direction \( \bar{e}_i^0 \) in the SFM. In particular, the flow diagram of the movement process of passengers at the platform is shown in Figure 6, and the detailed description is given as follows:

1. Build the platform according to the CAD diagram. Generate passengers and populate them at the platform near the stairs/escalators with random positions. Their initial speeds are set to be 1 m/s, and the desired walking directions point to the front waiting area directly for simplicity. The number of passengers generated is evenly distributed over time, while both the total number of passengers and the ratio of the passenger quantity from the left stairs/escalators to that from the right are set according to the actual demands.

2. Calculate the simulation time. If the time length exceeds the separation time, end the simulation.

3. Compute a target waiting area \( w^* \) according to the choice model proposed in this paper. Determine whether or not changing the target waiting area. If the passenger keeps the previous choice of the waiting area, keep the previous desired walking direction and update the position according to the SFM. Else, change the desired walking direction according to...
the new target waiting area. Then, update the new position according to the SFM.

(4) For each passenger, repeat step (3) until all passengers finish their updating.

(5) Repeat steps (2), (3), and (4) until reaching the required simulation time.

3. Case Study

3.1. Passengers’ Basic Attributes. According to the statistics, Beijing metro network shown in Figure 7 now has 18 lines in operation with a total length of over 550 km. In accordance with the current plan, the mileage of Beijing metro will reach 997 km by 2020. In addition, the carrying capacity of urban rail transit increases year by year, the daily passenger volume of Beijing subway reaches over 10,000,000. The platform of line 4 of Beijing Xuanwumen subway station shown in Figure 8 is chosen as an investigation platform in this paper for the observation and data collection. Xuanwumen station is an interchange station of line 2 and line 4. Ridership at Xuanwumen station is very large, especially during the morning and evening peak hours. According to the transportation experience in Beijing, 7:00-9:00 am and 17:00-19:00 pm are rush hours, while the other hours are all considered as off-peak hours.

In this paper, we choose $P_0$ in Figure 8 as the observation place to collect the data of passengers’ basic attributes from 18:00 pm to 20:00 pm. The collected attributes mainly consist of male-to-female ratio, age structure, bonding rate, and luggage statistics. Some statistics data are listed in Table 2. This table reflects that there is not any significant difference in the gender ratio, and most passengers are young and middle-aged for about 95% because of the complex structure of the station and the existence of stairs/escalators. In addition, the bonding rate is around 10%, and the ratio of passengers carrying large pieces of luggage is around 5%. Passengers who take the large pieces of luggage always have the relatively low travel efficiency, and they mainly transfer to Beijing South Railway Station. The masses of passengers are set according to the statistics data from National Health and Family Planning Commission of the People’s Republic of China. Moreover, other data in Table 2 are consistent with that in [50].

There are stairs/escalators on both sides of the platform, through which passengers enter or leave the platform. The field data of the inflow and outflow from 18:30 pm to 20:00 pm are collected at the observation places $P_l$ and $P_r$. The software SPSS is applied to test the collected data for the
of broadcasting the coming of a train. Once broadcasting starts, \( N \) before the open of train doors in each cycle time \( O \)

statistically significant correlations. The testing results show that \( I_1 \sim N(85, 36), O_1 \sim N(78, 26), I_2 \sim N(76, 28), \) and \( O_2 \sim N(78, 27) \) with a 5% significance level. \( I_1 \) and \( O_1 \), respectively, denote the entering and leaving numbers of passengers from the observation place \( P_1 \) during a cycle. \( I_2 \) and \( O_2 \) are corresponding values from \( P_2 \) in a cycle, respectively. The mean value of \( I_1 \) is obviously larger than that of \( I_2 \), which could directly result in the difference in passenger distribution at the platform. During our simulation, the ratio of inflow from \( P_1 \) to that from \( P_2 \) also keeps the same value with our field data.

3.2. Model Calibration. This paper focuses on investigating passengers’ waiting area choice behaviors, and field data at the platform with time is collected. In each cycle time, the collected data mainly contain the number of alighting passengers, \( N_{\text{alight}} \); the number of passengers who could not board the train for some reason in the previous cycle time, \( N_{\text{wag}} \); an increase in the number of waiting passengers during the time between the initial of a new cycle time and being informed of an arrival of a train, \( N_{\text{before}} \), and an increase in the number of passengers during the time between being informed of the coming of a train and the open of train doors, \( N_{\text{increase}} \). Therefore, the total number of passengers before the open of train doors in each cycle time \( N_{\text{read}} \) is \( N_{\text{wag}} + N_{\text{before}} + N_{\text{increase}} \). Note that the station staff always broadcast the coming of a train. Once broadcasting starts, we will record the required \( N_{\text{before}} \) thereby. According to the observation and statistics, one reason for not boarding may be that the space in the train is not enough for the waiting passengers; another reason may be that the train does not pass passengers’ destination station because of the operation mode of the long-short routing. In this paper, we do not consider the strategic level of their destinations but regard the results of these passengers’ choices as input data.

As mentioned above, large difference in the passenger traffic for two different driving directions at the platform of line 4 of Xuanwumen subway station exists. In addition, the traffic of boarding passengers with Anheqiao North direction is not very large during the evening rush hours, while the traffic of alighting passengers is relatively large. We choose to use the field data of \( N_{\text{before}}, N_{\text{increase}}, N_{\text{wag}}, \) and \( N_{\text{alight}} \) in each cycle during the time from 18:30 pm to 19:00 pm for 24 waiting areas with Anheqiao North direction at Xuanwumen subway station, and the mean values of the field data and their corresponding approximate integer values marked by “estimated mean value” are shown in Figures 9, 10, and 11 which also indicate the position of stairs. Note that there is no passenger who could not board in the dwell time. For our statistic data in each cycle time, we can find the significant difference between the total number of waiting passengers \( \sum_{w=1}^{24} N_{\text{read}} \) and the alighting passengers \( \sum_{w=1}^{24} N_{\text{alight}} \). The statistic results indicate that the mean value of \( \sum_{w=1}^{24} N_{\text{before}} \) during a cycle time is 36 with a standard deviation 9 and the mean value of \( \sum_{w=1}^{24} N_{\text{increase}} \) is 20 with a standard deviation 3, while the mean value of \( \sum_{w=1}^{24} N_{\text{alight}} \) is 153 with a standard deviation 29. These numerical fluctuations of \( \sum_{w=1}^{24} N_{\text{before}} \) and \( \sum_{w=1}^{24} N_{\text{increase}} \) are not very great, which provide us the possibility of calibrating the model based on these data. Though the statistic data of the number of passengers at each waiting area during each cycle time always vary randomly within a certain range, the overall distribution is similar with more passengers on both ends of the platform.

According to statistics and timetable of trains, train departure interval is 180 s during our investigation time from 18:30 pm to 19:00 pm with Anheqiao North direction. Generally, the dwell time for each train ranges from 30 s to 45 s, and passengers are usually informed of the coming of a train in advance through broadcasts and displays. We assume the

---

### Table 2: Passengers’ basic attributes at the platform of line 4 of Xuanwumen subway station.

<table>
<thead>
<tr>
<th>Passenger category</th>
<th>Young and middle-aged (Male)</th>
<th>Young and middle-aged (Female)</th>
<th>Child</th>
<th>Elderly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>18≤ age &lt; 60</td>
<td>18≤ age &lt; 60</td>
<td>age &lt; 18</td>
<td>age ≥ 60</td>
</tr>
<tr>
<td>Proportion (%)</td>
<td>47.5</td>
<td>48</td>
<td>3.1</td>
<td>1.4</td>
</tr>
<tr>
<td>Mass (kg)</td>
<td>66 ± 15</td>
<td>57 ± 15</td>
<td>30 ± 15</td>
<td>65 ± 15</td>
</tr>
<tr>
<td>Radius (m)</td>
<td>0.270 ± 0.020</td>
<td>0.240 ± 0.020</td>
<td>0.210 ± 0.015</td>
<td>0.250 ± 0.020</td>
</tr>
<tr>
<td>Desired speed (m/s)</td>
<td>1.35 ± 0.20</td>
<td>1.15 ± 0.20</td>
<td>0.90 ± 0.30</td>
<td>0.80 ± 0.30</td>
</tr>
<tr>
<td>Reaction time (s)</td>
<td>1 ± 0.2</td>
<td>1 ± 0.2</td>
<td>1 ± 0.2</td>
<td>1 ± 0.2</td>
</tr>
</tbody>
</table>

---

Figure 8: The simplified 2D diagram of the platform of line 4 at Xuanwumen subway station.

Table 2: Passengers’ basic attributes at the platform of line 4 of Xuanwumen subway station.
The mean number of passengers before the arrival of a train $N_{\text{before}}^w$ with Anheqiao North direction.

The identification number of the waiting area $w$

Mean value of field data

Estimated mean value of field data

Figure 9: The field number of passengers at each waiting area before being informed of the arrival of a train $N_{\text{before}}^w$ with Anheqiao North direction.

An increase in the number of passengers

The number of alighting passengers

Mean value of field data

Estimated mean value of field data

Figure 10: An increase in the field number of passengers at each waiting area $N_{\text{increase}}^w$ with Anheqiao North direction.

The longest time for passengers knowing the coming of a train is 55 s. For the feasibility of simulations, the total number of passengers with Anheqiao North direction in a cycle time is 56 and $\sum_{w=1}^{24} N_{\text{alight}}^w = 153$ during our simulation, which keep the same with the mean field values, among which the number of passengers coming from the left stair/escalator is 30, and 26 passengers are from the right stair/escalator. Assume that passengers’ waiting area choice behaviors are not affected by passengers with the other train driving direction in this paper.

Basically, parameter calibration of a model is very critical to simulations [1]. Parameters in the passenger driven model of this paper have already been adapted in [1, 23], while parameter calibration in the waiting area choice model still requires further investigations. As shown in Section 2, $d_0$, $\alpha_2$, $\beta_1$, $\beta_2$, and $\beta_3$ are the sensitivity parameters to be calibrated. The values of these parameters are related to the probability of choosing a waiting area. The method of setting parameters in this paper refers to [1]: experiments with different values of above parameters are run for the investigation of the influence of these sensitivity parameters, associated with the perception of the simulation dynamics and actual observations at the platform. Meanwhile, we propose to determine the above parameters based on the field data, and the magnitudes of $C_{1w}^w$, $C_{2w}^w$, and $C_{3w}^w$ are recorded with the repeated numerical simulations in order to regulate the influence degree of different factors. Furthermore, through minimizing the mean error $E = (\sum_{w=1}^{24} |N_{\text{real}}^w - N_{\text{simulation}}^w|)/24$, the parameters could be finally determined. Among which, $N_{\text{simulation}}^w$ is the simulation result of the number of passengers at each waiting area before the opening of train doors.

During the parameter calibration, the mean values of the numbers of passengers from the left and right stairs/escalators in the simulation runs are set according to those in Figures 9, 10, and 11. Considering all of the above criteria, parameters in this paper are set as $d_0 = 10$, $\alpha_2 = 2.9$, $\beta_1 = 110$, $\beta_2 = 0.8$, and $\beta_3 = 100$.

After using the above parameters, the dynamic characteristics for passengers when searching for the waiting areas could be found in the simulation snapshots shown in Figure 12. During the first few seconds of the separation time, alighting passengers occupy the main position at the platform as shown in Figure 12(a). After that, there are passengers entering the platform continuously and choosing an appropriate waiting area as shown in Figures 12(b) and 12(c). During our field observation, stairs on both sides of the platform mainly serve outbound passengers during the initial stage of the separation time, so does the simulation. In Figure 13, the box-plot shows the field number of passengers at each waiting area before the opening of doors during each cycle time through statistics, and also the simulation results of a random experiment marked with magenta asterisks. Note that the central red mark in Figure 13 is the median value of the field number of passengers at each waiting area, and the bottom and top edges of the blue box are the 25th and 75th percentiles of all collected field data, respectively.
Table 3: Scenario setting and experiment results.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Passenger number (Total)</th>
<th>Passenger number (Left stair/escalator)</th>
<th>Passenger number (Right stair/escalator)</th>
<th>Proportion (In blue boxes)</th>
<th>Proportion (Between maximum and minimum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>44</td>
<td>23</td>
<td>21</td>
<td>97%</td>
<td>100%</td>
</tr>
<tr>
<td>S2</td>
<td>56</td>
<td>30</td>
<td>26</td>
<td>82.3%</td>
<td>100%</td>
</tr>
<tr>
<td>S3</td>
<td>68</td>
<td>36</td>
<td>32</td>
<td>76.3%</td>
<td>95.8%</td>
</tr>
</tbody>
</table>

Figure 12: The snapshots of the 2D passenger movement corresponding to a simulation during the model calibration, t=5 s, t=50 s, and t=100 s. Blue dot markers represent alighting passengers and red dot markers represent passengers coming from the left stair/escalator, while magenta dot markers represent passengers coming from the right stair/escalator.

Figure 13: Box-plot for the field number of passengers at each waiting area and the simulation results of a random experiment.

Moreover, the dashed lines extend to the maximum and minimum values not considering the red outliers which are separately plotted. From Figure 13, we can observe the simulation data are all within the blue boxes, which indicates that the waiting area choice model proposed in this paper can reflect the distribution of passengers in the waiting areas to a certain extent. Considering some random factors of passenger movement, another repeated 20 simulations are run for each different scenario set in Table 3. In this table, the total numbers of passengers coming from the stairs/escalators on both sides of the platform in the scenarios S1, S2, and S3 are the minimum, mean, and maximum values of the field data, respectively. Results indicate that the majority of the simulation data can fall in the blue boxes of the field data and outliers only exist in very few cases. Taking into account some random characteristics, such errors are acceptable, which further reflect the ability and effectiveness of this model to capture passengers’ characteristics of the waiting area choice behaviors.

3.3. Model Validation. We start from the observations of passenger behaviors at the platform; we want to achieve these goals by the proposed modeling method, and so we take the following steps in order to ensure that our simulation results are indeed close to observations. Simulation experiments in the case of the platform with Tiangongyuan direction which is opposed to the mentioned Anheqiao North direction are
passenger density, the area of each waiting area is different from each other. Note that during the computing of the platform during the simulation experiment keeps the same settings, and the simulation results do not have significant differences from the field data. During the simulation, the number of entering passengers from \( P \) is set to 110, while 99 passengers enter the platform from \( P_r \). Besides, \( \sum_{w=1}^{24} N_{\text{wait}}^w = 71 \), and the initial distribution of these passengers at the platform during the simulation experiment keeps the same with the field data. Figure 15 shows the pseudo-color map of the variation of the passenger density with time, from which we can get the information of real-time density at each waiting area. Note that during the computing of the passenger density, the area of each waiting area is different from the field data. Figure 16 shows the pseudo-color map of the variation of passenger density with time at each waiting area.

Another 15 simulation experiments with different settings which are corresponding to the field data in 15 different cycle time between 18:30 pm and 20:00 pm are carried out. This further indicates that inflows from \( P_i \) and \( P_r \) are set differently in each simulation experiment according to different field data. As shown in Figure 17, the mean value \( E \) and the standard deviation \( \delta \) of \( |N_{\text{real}}^w - N_{\text{simulation}}^w| \) at 24 different waiting areas in 15 serial simulation experiments are applied to measure the difference, among which each simulation experiment is done repeatedly for 20 times. The \( t \)-test is used to guarantee that the model can predict the general passenger distribution at the platform. The result of \( t \)-test validates the hypothesis that \( E=2.5 \), as the observation value of statistics 0.7317 is less than the test statistic value 1.7613 when the confidence level is 95%. In addition, subfigure in Figure 17, that is, \( \sigma = (E \cdot 24)/\sum_{w=1}^{24} N_{\text{real}}^w \), is applied to measure the total deviation which is around 15%. Furthermore, another 15 simulation experiments at the platform with Tiangongyuan direction, using the field data in 15 different cycle times between 9:30 am and 11:00 am, are carried out. Note that this time period is among the off-peak hours. The corresponding comparison results are given in Figure 18. The result of \( t \)-test validates the hypothesis that \( E=0.5 \) when the confidence level is 95%. Besides, the total deviation \( \sigma \) is about 20%. Inevitably, the difference in the number of pedestrians at each waiting area between the field data and the experiment result exists. There are some reasons for this difference. One reason is the randomness characteristic of the passengers’ choice behaviors. Another reason is that passenger distribution at the platform has the relationship with the entering time into the platform. During our simulation, passengers enter the platform uniformly with time, which can further result in the existence of the distribution difference. Furthermore, manual collection error may also exist.

Another station, Shanghai natural history museum station in China, is chosen to have a further test of the validity of the proposed model. As shown in Figure 19, this station has 4 entrances into the platform which are a pair of stairs/escalators on both sides of the platform and another pair of stairs at the middle of the platform, respectively. The field data of passenger distribution at the platform is collected during the time period from 14:00 pm to 17:00 pm, which indicates most passengers entering the platform from the left stair/escalator because its location is near the park. We further do simulation experiments at the platform of Shanghai natural history museum station with Jinjun Road direction, and the corresponding comparison results.
Figure 16: Illustration of 2D passenger distribution corresponding to a simulation during the model verification, t=20 s and t=60 s. Black circles stand for passengers left in the last cycle time.

Figure 17: The values $E$ and $\delta$ of $|N_{\text{real}} - N_{\text{simulation}}|$ at 24 waiting areas.

Figure 18: The values $E$ and $\delta$ of $|N_{\text{real}} - N_{\text{simulation}}|$ at 24 waiting areas during the off-peak time period.

are given in Figure 20. The results of $t$-test validates the hypothesis that $E=0.47$ when the confidence level is 95%, which hence reflects the validity of the proposed model.

The prediction result $\sigma$ from the macroscopic level that only considers the distance factor in [38] is 17%, which is just the result of an experiment that is hardly representative. Besides, [39] models the passenger distribution at the subway platform using the ant colony optimization method, in which the mean prediction result $\sigma$ from multiple experiments is slightly above or below 17% within the acceptable range. It is worth noting that the result $\sigma$ obtained by the proposed method in this paper could also have the similar prediction accuracy compared with that in [39]. Moreover, this cost function approach could reflect more behavior dynamics in a way of considering more influence factors.

4. Conclusion

In this paper, we propose a cost function method to predict passenger distribution at the subway platform, which can be further for the passenger organization and the design of the layout of the platform. Through the field observation and video recording, a survey was done at Beijing Xuanwumen subway station for the statistics of passenger attributes and distribution at the platform. Based on the collected historical data and video, a waiting area choice model is established, considering many influencing factors such as the distance to the waiting area, passenger density in the visual field, and the length of waiting area occupied by passengers. Detailed individual characteristics such as gender, age, and luggage that affect the choice determination and walking dynamics are taken into account in the waiting area choice model and the SFM.

The model calibrated and validated by the field data from the platform exhibits a series of stochastic and complex dynamic phenomena. It captures the individual behaviors and also clusters characteristics during the process of choosing a waiting area, which was once very difficult to be modeled. Under 95% confidence level, the absolute deviation
of the number of passengers at each waiting area between the
field data and the experiment data is in an acceptable range,
which shows the validity of this model to mimic the waiting
area choice behaviors of passengers. Though Beijing subway
has currently 334 stations and on average almost 10 million
trips per day, most stations are new and many new stations
have the exactly same designs across the People’s Republic of
China. The analysis of Beijing Xuanwumen subway station
and Shanghai natural history museum station can provide
related insights into the design and the evacuation efficiency
that are relevant for the daily transportation of several hun-
dred million people across China. However, subway systems
in US, Europe, and Russia look very different; the method
proposed in this paper only provides a modeling idea of the
passenger distribution prediction which is also applicable to
other subway stations around the world, and the calibration
and validation of this model still require a research in the
future.

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 there are no conflicts of interest
regarding the publication of this paper.

Acknowledgments

This work is supported by Shandong Provincial Natural Sci-
ence Foundation of China under Grant ZR2018PF008, China
Postdoctoral Science Foundation under Grant 2018M632625,
and the Scientific Research Fee of Qingdao University under
Grant 4117010260. The authors would also like to thank Qianling Wang, Min Zhou, Jing Chen, Hong Lu, Shihang Lv, Chengjie Wei, Zhaquan Tang, Lei Zhang, Yubing Wang,
Xiaoyu Wang, Zhuopu Hou, Xiaowei Zhang, Qi Meng, ShiYu
Ning, et al. in Beijing Jiaotong University, as well as Yanjun
Zhang and Huai Zhan in Beijing MTR Corporation Limited
for the field data collection and video recording at the subway
station.

References

their impacts on pedestrian dynamics,” IEEE Transactions on
by public transport: The role of traveller information and
approach to modeling and analysis of a crowd dynamics,” IEEE


