Transportation Research Part E 181 (2024) 103359

Contents lists available at ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

Modelling the pedestrian's willingness to walk on the subway platform: A novel approach to analyze in-vehicle crowd congestion

Di Huang [a,b,1], Yuwei Yang [a,1], Xinyi Peng [a], Jiangyan Huang [c], Pengli Mo [a], Zhiyuan Liu [a,*], Shuaian Wang [d]

[a] *School of Transportation, Southeast University, China*
[b] *Key Laboratory of Transport Industry of Comprehensive Transportation Theory (Nanjing Modern Multimodal Transportation Laboratory), Ministry of Transport, China*
[c] *Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*
[d] *Department of Logistics & Maritime Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*

ARTICLE INFO

Keywords:
Urban subway system
Crowd congestion
Willingness to walk
Least square method
Pedestrian simulation

ABSTRACT

A common behavior pattern observed on subway platforms is that pedestrians walk downstairs from the escalator and choose a door to wait for a rail train. Interestingly, pedestrians often walk to farther doors rather than the nearest one to the escalator. This paper proposes a new concept, called willingness to walk (WTW), to describe pedestrians' behavioral characteristics, including their psychological tendency to stay in their original queue or to walk to farther queues. Two regression models are proposed to quantitatively measure WTW, both of which are calibrated using an inequality-based least square method. Observation data are collected through field investigation and simulation software. The calibration results confirm the existence of WTW. The proposed method is then applied to analyze the distribution of waiting passengers on the platform and the level of in-vehicle crowding. Simulation results demonstrate that the proposed WTW models can reliably approximate the actual passenger load in carriages.

1. Introduction

The urban subway system has become the dominating public transport mode that carries a large number of urban travel demands in most large cities worldwide. A major advantage of the subway system lies in the schedule adherence of on-time performance because it is free from complex traffic conditions (Huang and Wang, 2022). However, during peak hours, the dwell time at the station would be delayed due to overlong boarding and alighting times. One contributing factor to this problem is the concentrated boarding/alighting flows on the subway platform (Oliveira et al., 2019). This phenomenon often occurs in the areas where passengers easily congregate, such as the areas near staircases, indicating that a large proportion of passengers wait in the same area and board using the same door (Zhang et al., 2021).

The disadvantage of passenger aggregation on the platform is threefold: (1) It may lead to the unbalanced distribution of passengers

* Corresponding author.
E-mail address: zhiyuanl@seu.edu.cn (Z. Liu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.tre.2023.103359
Received 11 October 2022; Received in revised form 7 November 2023; Accepted 8 November 2023
Available online 8 December 2023
1366-5545/© 2023 Elsevier Ltd. All rights reserved.

Fig. 1. Illustration of the information signs of in-vehicle crowd congestion level displayed on the screens of platforms.

in carriages, i.e., some carriages being over-crowded while some remain relatively empty (Oliveira et al., 2019; Xu et al., 2023); (2) Highly crowded areas on the platform would block pedestrians from walking to less crowded areas (Fang et al., 2019); and (3) High-density crowds would increase conflicts between alighting and boarding passenger flows and cause safety issues (Sun et al., 2021). Hence, when the crowd on the subway platform requires to be dispersed, it is necessary to manage passenger distribution.

In practice, visual information signs have been widely applied to reduce crowding on the platform and to guide passengers to spread out along the platform. Fig. 1 shows an example of information signs displayed on the screens of platforms, which clearly shows the congestion level of each carriage of the coming train and directs pedestrians to walk to the doors where the corresponding carriage is less congested. The on-board passenger distribution, namely, congestion level in each carriage, is usually measured by weight load data collected by the airbag-based suspension system (Fang et al., 2019). However, for the urban subway systems where the weight load data is unavailable, indirect methods, such as analyzing the passenger distribution based on the pedestrian's behaviors are needed. This study intends to investigate the pedestrian's waiting area choice behavior by proposing a new concept, namely, willingness to walk (WTW), to describe the pedestrian's psychological tendency to walk or wait in a specific area. The proposed method presents a new method of analyzing in-vehicle crowd congestion without weight carriage load data. The distribution of waiting passengers along the platform can be estimated by incorporating WTW and automated passenger counts (APC) data.

1.1. Literature review

Pedestrian dynamics and crowd behavior are critical in managing crowd movement in urban subway systems (Yuan et al., 2022; Stanitsa et al., 2023). Subway station planning and management problems, such as station layout design and emergency evacuation management, are dependent on the pedestrian's internal psychological features and external physical forces simultaneously (Yu et al., 2020; Yang et al., 2022; Chen and Zhu, 2022; Cheng et al., 2024; Shi et al., 2023). A considerable number of models, describing and simulating the behavior involved in pedestrian activities at subway stations, have been developed from macro-, meso-, and micro-scopes. Interested readers may refer to Duives et al. (2013) for a comprehensive review.

A pedestrian macroscopically decides her/his destinations and directions, and moves along the chosen route to the destination (Mo et al., 2022; Zhang et al., 2023). The pedestrian's movement at a subway station is composed of several sub-procedures, such as staircase movement (Qu et al., 2014), waiting (including queuing, clustering, and scattering) on the platform (Zhou et al., 2019; Yuan et al., 2022), boarding-alighting process (Seriani and Fernandez, 2015; Yang et al., 2017; Qu et al., 2019; Qu et al., 2022). According to the internal psychological forces and external physical forces, pedestrians would update their motions and velocities to avoid collisions such as obstacles or other pedestrians, which are influenced by specific choice behaviors include vertical facilities choice (Qu et al., 2014), exit choice (Lovreglio et al., 2014), path choice (Kim et al., 2015), and waiting area choice (Krstanoski, 2014; Yang et al., 2017; Zhou et al., 2019).

From the operational level, the pedestrian motion decision is influenced by physical properties, such as the distance with respect to objects. However, it is widely recognized that psychological/sociological/physical forces from the strategic level play essential roles in pedestrian motion decisions (Duives et al., 2013). Kim et al. (2015) investigate the crowding stress in the logit-based path choice model. Results show that crowding itself has a significant effect on path choice behavior than delay. Yang et al. (2017) modify the pedestrian's impatience factor in the social force (SF) model to better reflect the change of psychological states for predicting the waiting area choice. It demonstrates that pedestrians will become impatient and speed up to walking to the waiting are when a train is coming. Pedestrian's psychological state would become unstable in case of emergencies (Yang et al., 2022). For instance, Zhou et al. (2021) find that panic pedestrians are less vulnerable from other pedestrians.

Given the increasing importance placed on the platform crowd safety, it becomes salient to investigate the pedestrian's choice behavior of waiting areas as well as which carriage to board. The pedestrian motion on the platform will influence the distribution of passengers along the rail train, which further affects the on-board congestion. In literature, the evaluation methods of pedestrian on-platform crowding can be divided into two categories (i.e., direct and indirect). Direct methods using manual and electronic counting systems, e.g., infrared or surveillance video, could provide the exact distribution of pedestrians. But such systems are usually expensive and the data is sometimes unavailable because of personal privacy issues. Hence, evaluating the in-vehicle crowding (i.e., knowing the exact number of passengers in each carriage) is a challenging problem. Though very few, some efforts have been devoted to semi-direct methods, such as using train's loadweight data (Nielsen et al., 2014; Fang et al., 2019; Peftitsi et al., 2020).

Indirect methods intend to investigate the fundamental features of pedestrian dynamics in choosing the desired waiting area on the platform given the fact that the automated passenger count system is lacking. Davidich et al. (2013) choose the cellular automata (CA) model to evaluate the impact of waiting zones on pedestrian dynamics. Results show that pedestrians do not always walk continuously on the platform, i.e., changing their routes or stopping occasionally. Johansson et al. (2015) extend the SF model the inclusion of waiting pedestrians on the platform. The density distribution of waiting passengers is highly dependent on the passing flow. For instance, pedestrians are pushed to the sides in the case with larger passing traffic volume. Yang et al. (2017) provide a SF-based modeling method for predicting pedestrian's waiting area choice. It finds that pedestrians intend to choose farther waiting areas

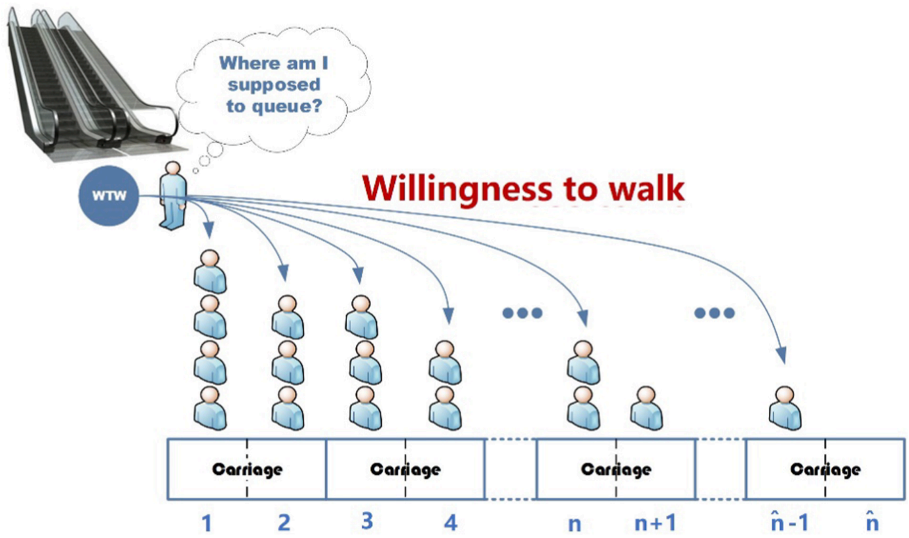


Fig. 2. Illustration of pedestrians' WTW.

when the nearby waiting areas are crowded. According to Ding et al. (2021), the choice of a waiting area by pedestrians is defined as a form of attraction, which is influenced by a combination of factors, including queue length, in-vehicle congestion level, and travel distance.

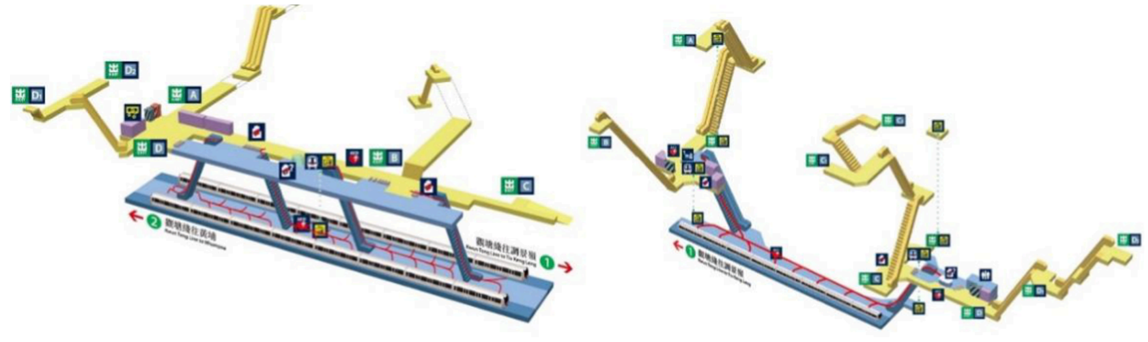
In any of the indirect methods mentioned above, pedestrian choice behavior is incorporated into classical simulation models, including CA and SF. Some efforts have also been devoted to modeling the choice behavior of waiting areas using discrete choice models. For instance, Kim et al. (2014) formulate a nested logit model to analyze the influencing factors affecting pedestrians' choice of waiting areas on the platform or a specific car of a train. Walking distance is considered to be the most decisive factor in the pedestrians' choice. Zhou et al. (2019) extend the fuzzy logic-based model by incorporating pedestrians' choice of waiting areas on the platform. Three waiting modes are investigated, namely, queuing mode, clustering mode, and scattering mode. The scattering mode is found to be the least efficient mode by causing waiting passengers to frequently rotate and reducing their speed. The study concludes that an even distribution of waiting passengers is helpful in enhancing the operational efficiency of subway platforms and mitigating congestion resulting from high passenger density. Peftitsi et al. (2021) develop a dynamic and stochastic transit assignment model to capture pedestrians' train carriage choice. The results indicate that the large traffic volume during peak hours would lead to a more even distribution of passengers across individual train cars.

1.2. Aims and contributions

According to the literature, understanding the features of pedestrian's the psychological effects is one of the most challenging tasks in modelling pedestrian dynamics. Complex pedestrian group behavior can be simulated using well-known microscopic models, such as the SF model and CA model. Pedestrians' psychological factors (e.g., force and desired velocity) are usually abstracted in model parameters. Though there have been some studies in extending classical models by incorporating more psychological factors, such as repulsive action (Guo, 2014), impatience (Yang et al., 2017), detour (Qu et al., 2018), physics-based models can hardly describe psychological effects on pedestrian's decision-making process (Wang, 2016). Alternatively, utility-based discrete choice models, including logit (Qu et al., 2018), cross-nested logit (Antonini et al., 2006), and mixed logit (Lovreglio et al., 2014), are applied to describe pedestrian's choice. However, due to the dependence of the logit-based model on the premise of independence of irrelevant alternatives (IIA), the correlation effects among pedestrians' psychological tendencies cannot be overcome, and it is difficult to calibrate the parameters of the nonlinear utility function.

In this study, a novel psychological factor, namely, willingness to walk (WTW), is proposed. It is assumed that after pedestrians come down the elevator to the platform level, they will first join the queue closest to the escalator exit, and then decide which queue to finally line up in. The proposed WTW describes the cognition of pedestrians in the above decision-making process regarding queue length and walking distance. A larger WTW indicates that pedestrians are more likely to leave the first queue and go to a queue with fewer people. During periods of peak demand, when the number of passengers queuing on the platform is high, concerns may arise among passengers about being unable to board the train. Hence, the WTW model can serve as an indicator of passengers' propensity to board the train. Specifically, a higher WTW value implies a stronger willingness to board and a greater tendency to opt for queues with fewer people.

To sum up, the contribution of this paper is threefold. First, it discusses the concept of WTW within a comprehensive conceptual framework and proposes a quantitative measurement of WTW using observation data. Second, it devises a novel method for evaluating in-vehicle crowd congestion by incorporating WTW and the passenger's in-vehicle movement. Third, a simulation framework is



(a) Lam Tim Station (b) Whampoa Station

Fig. 3. Two types of station layout in Hong Kong's MTR.

Table 1
List of notations.

Notation	Description
d_n	The perpendicular distance between the 1st queue and the n-th queue.
Δd	The distance between two adjacent screen doors.
i	ID of a collected record.
k	The sequence number of the newly arrived pedestrian in the 1st queue.
k^i	The sequence number of the newly arrived pedestrian in the 1st queue of the data record i.
n	The sequence number of a queue, $n = 1, 2, \ldots, \widehat{n}$.
$\widehat{n}$	The maximum number of queues.
N	The sequence number of the queue which the newly arrived passenger finally chooses, $N = 1, 2, \ldots, \widehat{n}$.
N^i	The sequence number of the queue which the newly arrived passenger finally chooses to walk to or the current 1st queue of the data record i.
q_n	The total number of waiting passengers in the n-th queue, $n = 1, 2, \ldots, \widehat{n}$.
q_n^i	The total number of waiting passengers in the n-th queue in data record i.
$\widehat{q}_1$	The maximum value of $W(k)$ in the collected data
r	The total number of collected data records.
R_1	Data set of collected data that the newly arrived passenger who will stay in the 1st queue and is not willing to go to other queues to wait for the subway.
R_N	Data set of collected data that the newly arrived passenger who will choose to walk to the N-th queue, $N = 2, \ldots, \widehat{n}$.
$W(k)$	Willingness to walk of the k-th position in the 1st queue.
λ	Parameter of unifying the units of the number of passengers in the queue to which the newly arrived passenger expects to walk and the distance between the target queue and the current queue.

developed in the Legion simulation environment, where a real-world subway line is built to design numerical experiments.

The remainder of this paper is organized as follows. The basic concept of WTW is defined in Section 2. Section 3 presents a general model and a parametric model of WTW. Section 4 formulates the simulation framework in Legion Studio. Section 5 presents the model of passenger's in-vehicle crowdedness and movement. Next, an application of WTW in a real-world subway line is presented in Section 6. Finally, in Section 7 some conclusions and recommendations are delivered.

2. Problem description

The concept of WTW is developed based on a common pedestrian activity on the subway platform. When a pedestrian descends the escalator, she/he will continue to walk on the platform and wait in a designated area for the train to arrive. Typically, potential boarding waiting areas are specified in advance with respect to the locations of platform screen doors, where pedestrians will line up to form queues along the platform marking lines. One interesting phenomenon that can be observed is that pedestrians have the different willingness (or desire) to walk on the platform and to choose a particular waiting area (see Fig. 2). Consider a special situation where a train has just left the station and the platform is empty. In this case, pedestrians are more willing to wait in front of the nearest screen door to the escalator rather than walk farther. When pedestrians accumulate in the waiting area of the nearest door, newly arriving pedestrians are more willing to walk to farther waiting areas with fewer waiting pedestrians, as well as to avoid on-board crowding. Peftitsi et al. (2021) also observe this phenomenon through the agent-based simulation and conclude that experienced pedestrians may make trade-offs between on-platform walking and on-board crowding.

Obviously, the layout of the platform, especially the location and direction of escalators, has a dominant impact on the distribution of pedestrians on the platform, thereby affecting the on-board crowding level in each carriage (see Fig. 3). Hence, it is of considerable importance to understand the pedestrian's decision-making process on the platform and quantitatively estimate the WTW values of different pedestrians. The on-board crowding level in each carriage can then be estimated by combining WTW and the APC system.

2.1. Notations

The notations used to formulate the model in this paper are introduced in Table 1.

2.2. Willingness to walk

As shown in Fig. 2, assume that there is only one queue of passengers waiting in front of the nearest screen door to the escalator, which is numbered as the 1st screen door on the platform. Each door corresponds to a waiting area (or queue). The total number of waiting areas that a pedestrian could walk to is denoted by $\widehat{n}$. The pedestrian's decision-making process can be described as follows: when a newly arrived pedestrian exits the escalator, she/he will decide whether to join the 1st queue or continue to walk to another queue with fewer waiting passengers. The 1st queue here refers to the queue formed in front of the screen door closest to the escalator exit.

The passenger's choice of the waiting area is the result of her/his psychological process. Liu et al. (2016) find that passengers standing in the front of the queue are also more ambitious to board, while those standing at the back are less willing to board considering the on-board crowding level. Similarly, if the queue is long enough, the newly arrived passenger is less likely to take the risk of not boarding than walking to farther waiting areas with fewer passengers. In this regard, when a pedestrian joins the tail of the 1st queue, the number of waiting passengers already in the queue determines her/his level of WTW to other waiting areas.

Let q_1 denote the current number of passengers waiting in the 1st queue. The sequence number of each passenger in the queue is numbered from one. Thus, the newly arrived passenger would wait at the (q_1+1)-th position of the 1st queue. We use a nonnegative $W(k)$ to denote the WTW of the k-th passenger in the 1st queue, $k=1,2,\ldots,q_1+1$. According to the above descriptive analysis, passengers waiting at the back of the queue have larger WTW than those in the front, that is,

Assumption 1. $W(k)\leqslant W(k+1)$, $k=1,2,\ldots,q_1+1$.

2.3. Decision-making process

According to **Assumption 1**, a pedestrian standing at the tail of the 1st queue is always more willing to walk to farther waiting areas. Let n denote the sequence number of queues along the platform. The decisions on whether to stay in the 1st queue and which queue/waiting area to walk to are jointly determined by three aspects: (1) $W(k)$; (2) walking distance to the n-th queue, denoted by d_n; and (3) the current number of passengers waiting in the n-th queue, denoted by q_n. Note that the last aspect is considered because pedestrians would avoid long queues and put catching the train as a priority (Krstanoski, 2014; Oliveira et al., 2019) To jointly consider the impacts of the last two aspects, the parameter λ is introduced to unify d_n and q_n. In sum, the pedestrian's final decision depends on her/his position in the 1st queue as well as potential queues. Thus, the following assumption can be made:

Assumption 2. If $W(k)\geqslant d_n+\lambda\cdot q_n$, the passenger will walk to the n-th queue; otherwise, the passenger will stay in the tail of the 1st queue.

Remark. When the platform is over-crowded, one natural queue cannot accommodate all the waiting passengers. However, the premise of the analysis in this paper is that passengers are queuing up in an orderly manner, regardless of how crowded the platform is. This is an assumption for the analysis. Therefore, when the number of passengers is small, we only have one queue, and when the number of passengers is large, there will be multiple queues. However, regardless of the number of queues, the assumption of this analysis is the same, that is, passengers at the back of the queue will have a relatively large WTW (waiting time to walk), and thus they will choose to walk to a farther door. Therefore, the analytical framework proposed in our paper still holds.

Assumption 2 reflects the heterogeneous forces on the choice between staying in the 1st queue and walking to other queues, which can also be interpreted as the tradeoff between on-platform and on-board crowding (Peftitsi et al., 2021). Let $\widehat{n}$ denote the total number of potential queues on the platform and N the sequence number of the queue a pedestrian finally chooses. According to **Assumptions 1** and **2**, the value of N should satisfy the following relationship:

$$N=\begin{cases}1, & \text{if } W(k)\leqslant d_1+\lambda\cdot q_1\\ \max\{n=2,\ldots,\widehat{n}\,|\,W(k)>d_n+\lambda\cdot q_n\}, & \text{otherwise}\end{cases} \tag{1}$$

where $d_n=\Delta d\cdot(n-1)$, and Δd is the distance between two adjacent screen doors.

For each newly arrived pedestrian who just walks downstairs from the escalator, two observations can be obtained until she/he chooses a queue: (1) the sequence number of the queue that she/he finally chooses, N; and (2) the number of passengers waiting in each queue, i.e., q_n, $n=1,\ldots,N$. Suppose that the above observations for an individual pedestrian is represented by a data record, and that the total number of data records is R. Data records can be further categorized into the following cases.

Case 1. $N=1$.

This case implies that $W(k)\leqslant\lambda\cdot q_1$, and the newly arrived passenger will stay in the (q_1+1)-th position of the 1st queue. Data records of Case 1 are grouped in set R_1.

Case 2. $1<N\leqslant\widehat{n}$.

This case indicates that the pedestrian is willing to walk to the N-th queue. Similarly, data records of Case 2 are grouped in set R_N. According to **Assumption 2**, we have

$$W(k)>\Delta d\cdot N+\lambda\cdot q_N\,. \tag{2}$$

Table 2
Errors between the observed information and the estimated for different cases.

Case	Comparison between observed information and the estimated WTW	Error
$N = 1$	$W(k) \leqslant \lambda \cdot q_1$	0
	$W(k) > \lambda \cdot q_1$	$W(k) - \lambda \cdot q_1$
		$error = \max(W(k) - \lambda \cdot q_0, 0)$
$N \neq 1$	$W(k) > \Delta d \cdot N + \lambda \cdot q_N$	0
	$W(k) \leqslant \Delta d \cdot N + \lambda \cdot q_N$	$W(k) - (\Delta d \cdot N + \lambda \cdot q_N)$
		$error = \min(W(k) - (\Delta d \cdot N + \lambda \cdot q_N), 0)$
	$W(k) \leqslant \Delta d \cdot n + \lambda \cdot q_n$	0
	$W(k) > \Delta d \cdot n + \lambda \cdot q_n$	$W(k) - (\Delta d \cdot n + \lambda \cdot q_n)$
		$error = \max(W(k) - (\Delta d \cdot N + \lambda \cdot q_N), 0)$

It also indicates that the pedestrian skips the 2nd to $(N-1)$-th queues, which results in the following two constraints,

$$W(k) \leqslant \Delta d \cdot n + \lambda \cdot q_n, n = 2, \ldots, N-1, \tag{3}$$

$$W(k) > \lambda \cdot q_1 . \tag{4}$$

Generally, each data record $i = 1, 2, \ldots, r$ contains three types of data: (1) q_1^i: the number of passengers in the 1st queue; N^i: the sequence number of the queue that a pedestrian finally chooses; and (3) q_n^i , $n = 2, \ldots, \widehat{n}$: the numbers of waiting passengers of the n-th queue.

3. Models of WTW

As discussed above, the pedestrian's WTW mainly depends on the current distribution of waiting passengers on the platform. If the specific value of $W(k)$ is obtained, the pedestrian's WTW can be predicted. Moreover, the number of waiting passengers in each queue can subsequently be predicted given the total number of passengers entering the subway station during a certain period of time. This section intends to quantitively describe $W(k)$ by proposing two regression models based on the observed data.

3.1. General model

According to **Assumption 1**, the passenger standing at the back of the queue always has a larger WTW than front passengers. Hence, it is reasonable to assume that $W(k)$ is a monotonically non-decreasing function with respect to the passenger's position, $W(k)$. In each data record $W(k)$, let k denote the maximum value of $W(k)$ in the collected data, i.e., $\widehat{q}_1 := \max\{q_1^i, \quad i = 1, 2, \ldots, r\}$. The general model of $W(k)$ can then be represented as follows

$$W(k), k = 1, 2, \ldots, \widehat{q}_1 + 1 . \tag{5}$$

Because the WTW of any passenger, which is considered as a psychological factor in this study, cannot be observed as an input, it is exceedingly difficult to calibrate the aforementioned model using traditional regression methods such as the least square method (LSM). However, in this problem we can only observe an inequality relationship, that is, whether a passenger's WTW is larger or smaller than a particular value according to **Assumption 2.**

In this regard, an inequality-based LSM is employed to calibrate $W(k)$ due the lack of errors between dependent and independent data points. Specifically, the error only occurs when the observed data violates the inequality presented in **Assumption 2.** For instance, in ***Case 1*** if the newly arrived pedestrian remains in the 1st queue, we have q_0^i, and no error occurs; otherwise, an error occurs because this record is not in accordance with **Assumption 2**. The calculation process of errors in different cases is summarized in Table 2.

As aforementioned, each data record i contains three components: q_1^i, N^i, and q_n^i , $n = 1, 2, \ldots, \widehat{n}$. The sum of square errors can be defined as

$$\begin{aligned} Z_1 = {} & \sum\nolimits_{i \in R_1} \left[\max\left\{W\left(q_1^i + 1\right) - \lambda \cdot q_1^i, 0\right\}\right]^2 + \\ & \sum\nolimits_{i \in R_N} \left\{ \left[min\left\{W\left(q_1^i + 1\right) - \left(\Delta d \cdot N^i + \lambda \cdot q_N^i\right), 0\right\}\right]^2 + \sum\nolimits_{n=2}^{N-1} \left[max\left\{W\left(q_1^i + 1\right) - \left(\Delta d \cdot n + \lambda \cdot q_n^i\right), 0\right\}\right]^2 \right\}. \end{aligned} \tag{6}$$

Note that $(q_0 + 1)$ -th represents the position of the newly arrived pedestrian in the 1st queue, which can be replaced by k. In sum, the least square method, which aims to minimize the sum of square errors $W(k)$, can be solved by the following model **[M1]**.

[M1]

$$\min \ Z_1 \tag{7}$$

s.t.

$$W(k) \geqslant 0, \ 1 \leqslant k \leqslant \widehat{q}_1 + 1, \tag{8}$$

$$W(k) \leqslant W(k+1),\ 1 \leqslant k \leqslant \widehat{q}_1. \tag{9}$$

It can be observed that **[M1]** is a non-differential due to the maximize and minimize operators. With a little abuse of notation, we introduce two auxiliary variables u^i, $i \in R_1 \cup R_N$; v^i, $i \in R_N$. Then, **[M1]** can be reformulated as the following convex quadratic program.

[M1′]

$$\min\ Z_2 = \sum\nolimits_{i \in R_1} \left(u^i\right)^2 + \sum\nolimits_{i \in R_N} \left[\left(u^i\right)^2 + \left(v^i\right)^2\right] \tag{10}$$

s.t.

$$W(k) \leqslant W(k+1),\ 1 \leqslant k \leqslant \widehat{q}_0\ , \tag{11}$$

$$u^i \geqslant \max\left(W\left(q_1^i + 1\right) - \lambda \cdot q_1^i, 0\right),\ u^i \geqslant 0, i \in R_0\ , \tag{12}$$

$$v^i \leqslant \min\left(W\left(q_1^i + 1\right) - \left(\Delta d \cdot N^i + \lambda \cdot q_N^i\right),\ 0\right),\ v^i \leqslant 0, i \in R_N\ , \tag{13}$$

$$u^i \geqslant \max\left(W\left(q_1^i + 1\right) - \left(\Delta d \cdot n + \lambda \cdot q_n^i\right), 0\right),\ u^i \geqslant 0, n = 2, \ldots, N-1, i \in R_N\ . \tag{14}$$

3.2. Parametric model

In Section 3.1, we simply assume that $W(k)$ is a monotonically non-decreasing function with no parameters but an independent variable k. However, this model cannot be extended to the cases where no data record was observed. For instance, the total number of passengers waiting in queue is 20 due to the limited space available in the waiting area. In this regard, **[M1]** can only obtain the WTW of these pedestrians, and it cannot be extended to the case with the queue of 30 passengers because the choice results of these additional 10 passengers cannot be observed. Because **Assumption 1** only assumes that $W(k)$ is a monotone function of k, but the form of $W(k)$ is unknown, we define $W(k)$ by the parametric models without exact forms but only contains a set of parameters which can be calibrated by observed data. Let W denote the collection of $W(k)$, which is indexed by the parameter set Θ. For each $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_s) \in \Theta$, $W_{\boldsymbol{\theta}}(k)$ denote the corresponding member of the collection W. For ease of calibration, all parameters are assumed to be positive. In sum, the parametric models of $W(k)$ can be defined as follows

$$W = \{W_{\boldsymbol{\theta}}(k) | \boldsymbol{\theta} \in \Theta\}. \tag{15}$$

In this regard, for each $\boldsymbol{\theta} \in \Theta$, the sum of square errors in Eq. (6) can be written as

$$\begin{aligned} Z_3 = & \sum\nolimits_{i \in R_1} \left[\max\left\{W_{\boldsymbol{\theta}}(k) - \lambda \cdot q_1^i, 0\right\}\right]^2 + \\ & \sum\nolimits_{i \in R_N} \left\{\left[\min\left\{W_{\boldsymbol{\theta}}(k) - \left(\Delta d \cdot N^i + \lambda \cdot q_N^i\right), 0\right\}\right]^2 + \right. \\ & \left. \sum\nolimits_{n=2}^{N-1} \left[\max\left\{W_{\boldsymbol{\theta}}(k) - \left(\Delta d \cdot n + \lambda \cdot q_n^i\right), 0\right\}\right]^2 \right\}. \end{aligned} \tag{16}$$

[M1] can then be modified to the following model **[M2]**:

[M2]

$$\min\ Z_3 \tag{17}$$

s.t.

$$\theta_1, \theta_2, \ldots, \theta_s > 0. \tag{18}$$

3.3. Weighted least square method

In general, the regression models **[M1′]** and **[M2]** can be calibrated by LSM using observation data. The calibration process of LSM is usually dominated by a certain category of observations while the number of observations in this category is large, thus giving rise to inaccurate results. This key deficiency is called the sample selection bias problem (Qu et al., 2015). Hence, the weighted least square method (WLSM) is applied to remedy this deficiency. The basic idea of WLSM is that if a category of observations is overrepresented, lower weights should be imposed on the corresponding observation to ensure that this observation category would not dominate the calibration process.

Instead, the objective functions Z_2 and Z_3 in **[M1′]** and **[M2]** can be replaced by the following $\overline{Z}_2$ and $\overline{Z}_3$, respectively:

$$\overline{Z}_2 = \sum\nolimits_{i \in R_1} \varpi^i \left(u^i\right)^2 + \sum\nolimits_{i \in R_N} \varpi^i \left[\left(u^i\right)^2 + \left(v^i\right)^2\right], \tag{19}$$

Table 3
Instance of weight calculation.

i	k^i	N^i
1	1	1
2	2	3
3	2	1
4	3	3
5	4	3
6	4	2
7	4	3
8	5	2

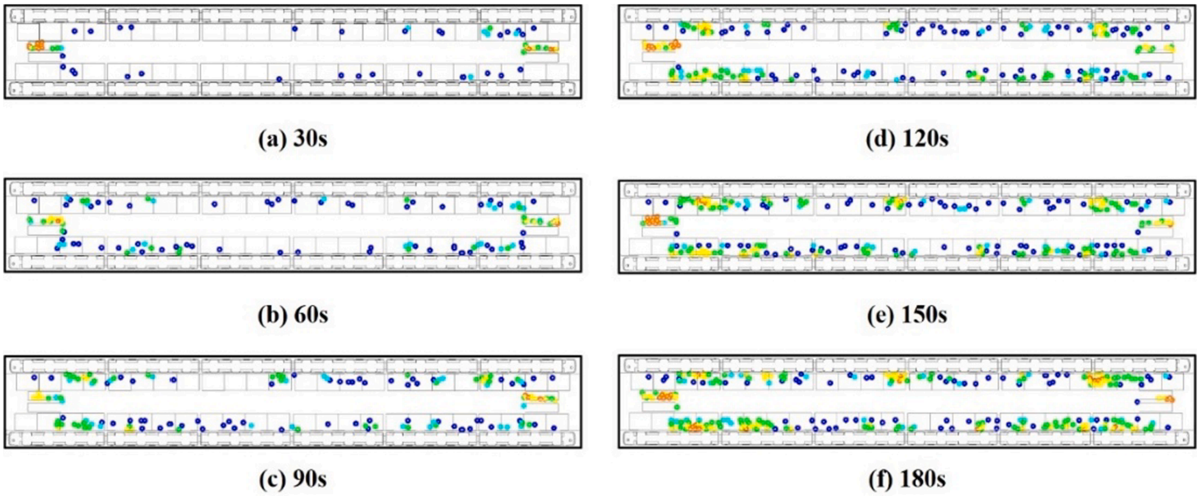


Fig. 4. Snapshots of waiting pedestrians at different time steps after the start of simulation.

$$\overline{Z}_3 = \sum\nolimits_{i\in R_1} \varpi^i \left[\max\left(W_\theta(k) - \lambda \cdot q_1^i, 0\right)\right]^2 + \sum\nolimits_{i\in R_N} \varpi^i \left\{\left[\min\left(W_\theta(k) - \left(\Delta d \cdot N^i + \lambda \cdot q_N^i\right), 0\right)\right]^2 + \sum\nolimits_{n=2}^{N-1} \left[\max\left(W_\theta(k) - \left(\Delta d \cdot n + \lambda \cdot q_n^i\right), 0\right)\right]^2\right\}. \tag{20}$$

where ϖ_i is the weight of observation i. A weight determination method is proposed following the basic idea of WLSM aiming to eliminate the effects of overrepresentation. Define m_j as the number of data records satisfying $k^i = j$, $j = 1, 2, 3, \cdots, \overline{k}$, and $\overline{k} := \max\left\{k^i, i \in R_1 \cup R_N\right\}$.

If $m_j \neq 0$, the weight of data record i can be easily obtained by the inverse of m_j, that is,

$$\varpi^i = \frac{1}{m_j}, i \in R_1 \cup R_N . \tag{21}$$

Eq. (21) indicates that the if there are many data records with the same queue choice result, then the weight of these data records should be small. However, Eq. (21) is invalid if $m_j = 0$, which means that there are no data records of a certain queue choice.

To develop a general form of Eq. (21), we further introduce two new sets $R_i^+ := \left\{i \middle| (k^i)' = j-1, i \in R_1 \cup R_N\right\}$ and $R_i^- := \left\{i \middle| (k^i)' = j+1, i \in R_1 \cup R_N\right\}$, and

$$(j-1)^+ = \max\left\{N^i \middle| i \in R_i^+\right\}, j = 2, 3, 4, ..., \overline{k} - 1, \tag{22}$$

$$(j+1)^- = \min\left\{N^i \middle| i \in R_i^-\right\}, j = 2, 3, 4, ..., \overline{k} - 1. \tag{23}$$

For a certain data record i, k^i is the sequence number of the newly arrived pedestrian in the 1st queue. Hence, R_i^+ represents the set of data records and their sequence number in the 1st queue is $k^i - 1$; and R_i^- contains the data records with the sequence number $k^i + 1$. $(j-1)^+$ and $(j+1)^-$ represent the maximum and minimum N^i in sets R_i^+ and R_i^-, respectively. As aforementioned, the weight of a data

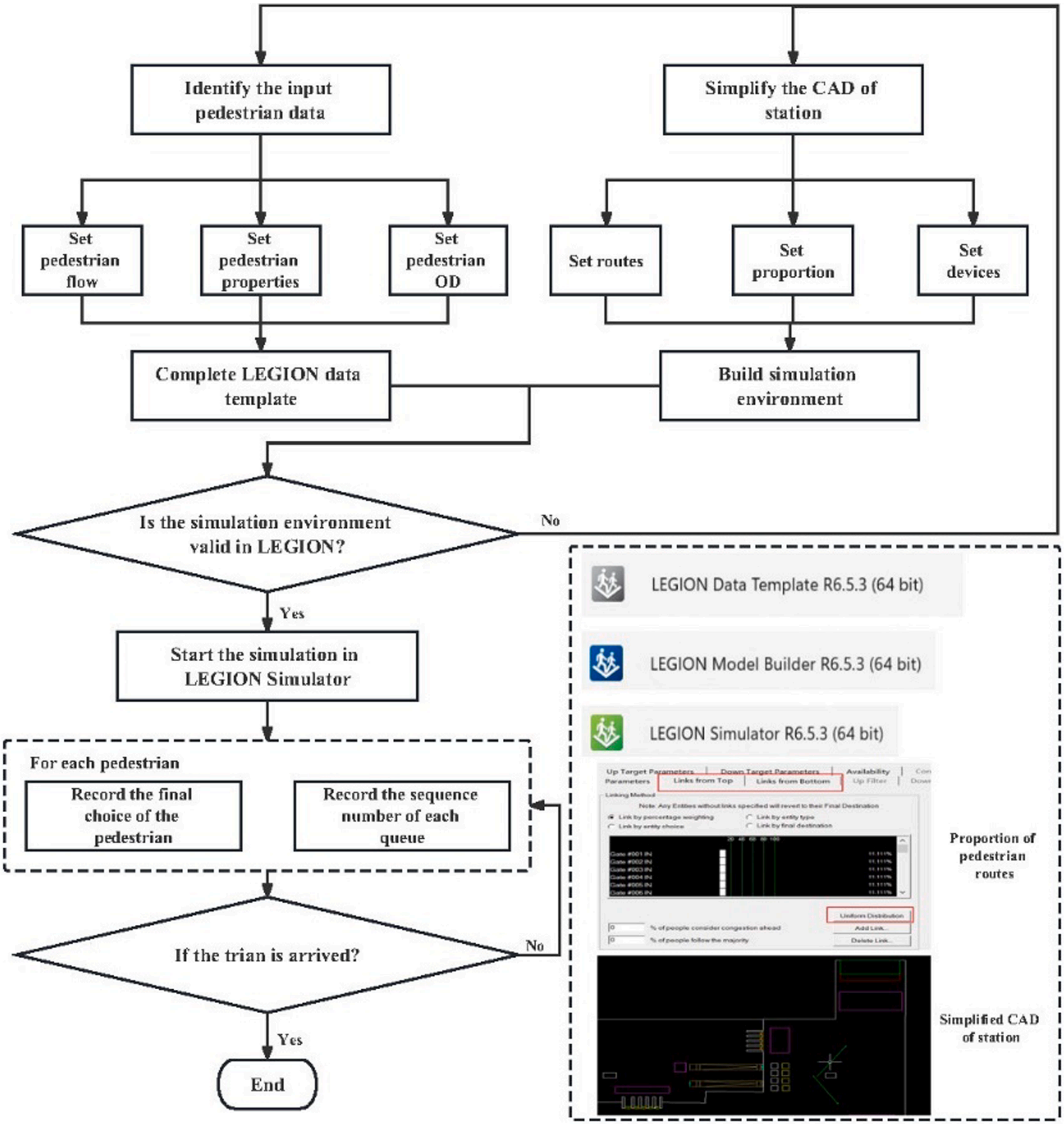


Fig. 5. The simulation framework based on Legion.

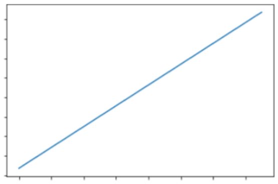
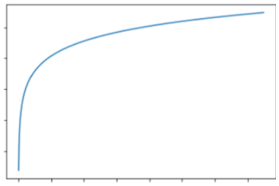
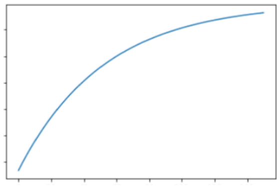
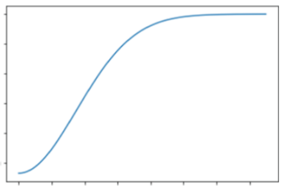
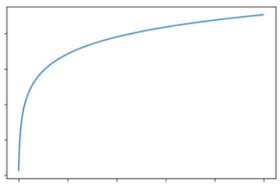
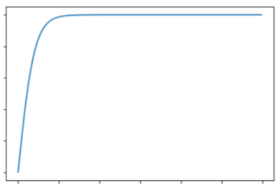
record is in inverse proportion to the number of the same kind of records.

Following the idea that the weight of a data record is in inverse proportion to the number of the same kind of records, Eq. (21) can then be rewritten as follows

$$\varpi^i = \begin{cases} \dfrac{1}{m_j}, & j = 1 \text{ or } \overline{k} \\ \dfrac{1}{m_j} \cdot \dfrac{|(j-1)^+ - (j+1)^-| + 1}{2}, & j = 2, 3 \ldots \overline{k} - 1 \end{cases} . \tag{24}$$

For instance, suppose that we have collected eight data records (see Table 3). The weights of data records $i = 1$ and $i = 8$ equals one because $m_1 = 1$ and $m_5 = 1$. For data records $i = 2$, we have $k^2 = 2$, $N^2 = 3$, $(1)^+ = 1$, $(3)^- = 3$, $m_2 = 2$. Hence, $\varpi^2 = (1/2) \cdot (|1 - 3| + 1)/2 = 0.75$.

Table 4
Six parametric models.

Parametric models	Function	Shape
Greenshields et al. (1935)	$W(k) = \theta_1 + \theta_2\left(1 - \frac{k}{\theta_3}\right)$	
Greenberg (1959)	$W(k) = \theta_1 + \theta_2 \ln\left(\frac{\theta_3}{k}\right)$	
Underwood (1961)	$W(k) = \theta_1 + \theta_2 exp\left(-\frac{k}{\theta_3}\right)$	
Northwestern (Drake et al., 1967)	$W(k) = \theta_1 + \theta_2 exp\left(-\frac{1}{2}\left(\frac{k}{\theta_3}\right)^2\right)$	
Liu et al. (2016)	$W(k) = \theta_1 + \frac{\theta_2}{k^{\theta_3}}$	
Sigmoid	$W(k) = \theta_1 + \frac{\theta_2}{1 + \theta_3 \exp(-\theta_4 k)}$	

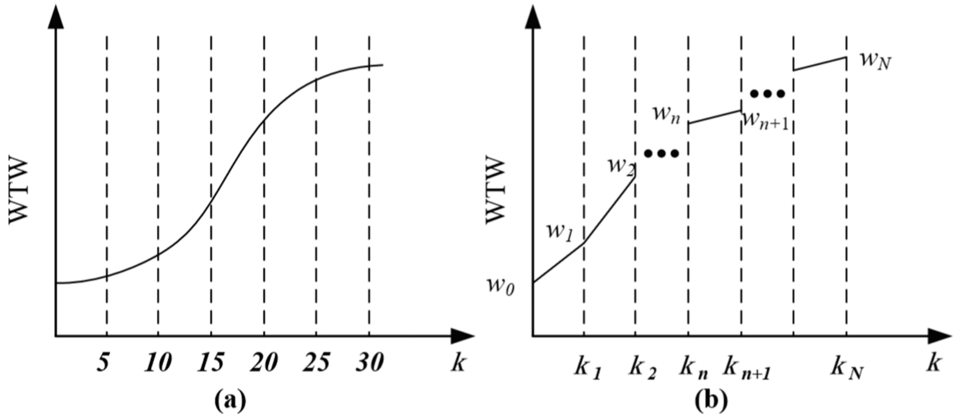


Fig. 6. A comparison of (a) a general form $\phi(k)$ and (b) a piecewise linear form $\phi_N(k)$.

4. Simulation and calibration

4.1. Experimental settings

The simulation model is built in the pedestrian microsimulation software Legion Studio, which provides an effective tool for the analysis of general mass behavior and interaction between individuals. The basic scenario of the model is a subway platform with two rail trains on each side (see Fig. 4). A train has six carriages, each of which has four doors. The length and width of the waiting area in front of a door are 4.7 m and 3.55 m, respectively. Two escalators are situated at opposite ends of the platform. To simplify the scenario model, the alighting pedestrian is not considered here. The simulation mainly concerns the analysis of the pedestrian's walking behavior on the platform such as which door will be chosen when a pedestrian walks downstairs from the escalator. We intend to investigate how the distribution of waiting passengers in front of each door affects the pedestrian's decision. Hence, we focus on the collection of data in the simulation including: (1) pedestrian door choice, and (2) the number of waiting passengers in front of each door.

The simulation framework in Legion Studio is presented in Fig. 5. The necessary input data for the simulation includes pedestrian features, traffic flow, and origin–destination (OD) information. The built-in module "Model Builder" is utilized to automatically identify and construct the simulation environment based on the station's CAD drawing, which includes the layout of turnstiles, elevators, platforms, and trains. During the simulation, the final choice of the queue for each pedestrian to wait in, as well as the total number of people queuing in each queue when a train arrives, are collected as inputs for the proposed model.

4.2. Collection of parametric models

4.2.1. Continuous model

In Section 3.2, the parametric model of WTW is defined as a collection of monotone non-decreasing functions of k, and the exact form of $W(k)$ is unknown. Following the traffic flow theory, we adopt four classic fundamental diagrams of freeway traffic that have been widely adopted to calibrate the speed-density relationship, namely, Greenshields et al. (1935), Greenberg (1959), Underwood (1961), and Northwestern (Drake et al., 1967). These models cannot be directly used to calibrate $W(k)$ because the dependent variable (i.e., speed) is a decreasing function of the independent variable (i.e., density). Hence, the four classic models are modified by adding non-negative parameters θ_1 and θ_2. The parametric model proposed by Liu et al. (2016) is also considered, which is a power function plus a constant. In addition, the sigmoid function which has been widely used in machine learning is also applied and modified by adding four non-negative parameters accordingly. Table 4 lists six parametric models, each with three or four parameters.

4.2.2. Piecewise linear model

The preceding section outlines six function forms with unknown parameters to calibrate. In the literature, functions without a predetermined form are commonly approximated using a piecewise linear function, which divides the independent variable into equal intervals (Huang et al., 2016). Fig. 6 illustrates a typical nonlinear curve and its corresponding piecewise linear form.

Let $\phi(k)$ denote general form of WTW. k is the independent variable which represents the sequence number of the arrived pedestrian in the 1st queue, $k = 1,2,...,q$, where q is total number of passengers in the 1st queue. $\phi(k)$ can then be divided into N equal intervals, each of which can be accurately approximated by a linear function. This linear function can be defined by its vertexes $\boldsymbol{w} = (w_0, w_1, ..., w_n, ...w_N)$. This piecewise linear approximation function can be defined as follows:

$$\overline{\phi}_n(k) = w_i + \frac{w_{i+1} - w_i}{k_{i+1} - k_i}(k - k_i), k_i \leqslant k < k_{i+1}, i = 1, 2, ..., n, \tag{25}$$

where k_1 is the length of the interval and k_{n+1}=q; $w_0, w_1, \cdots, w_n, \cdots w_N$ are the parameters which need to be calibrated. According to the

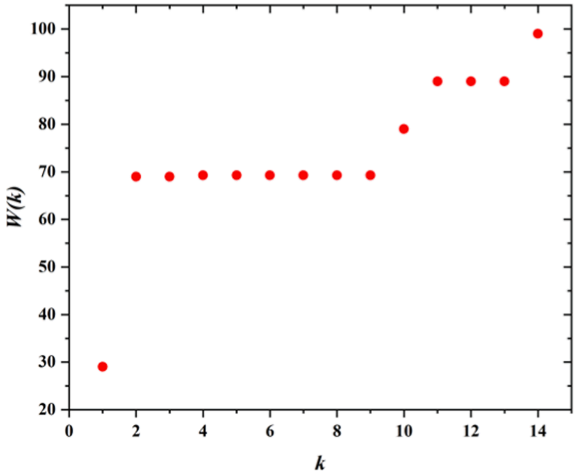


Fig. 7. Calibration results for the general model.

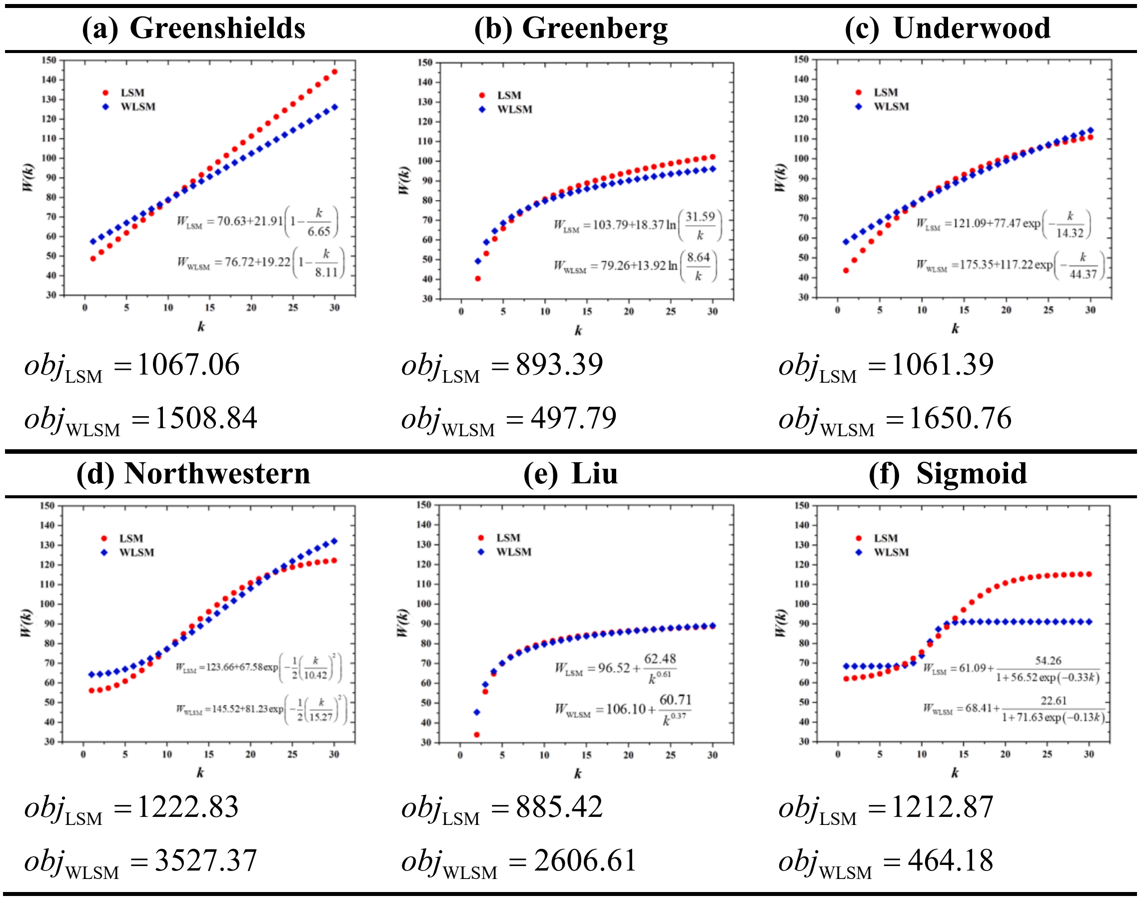


Fig. 8. Calibration results of parametric models.

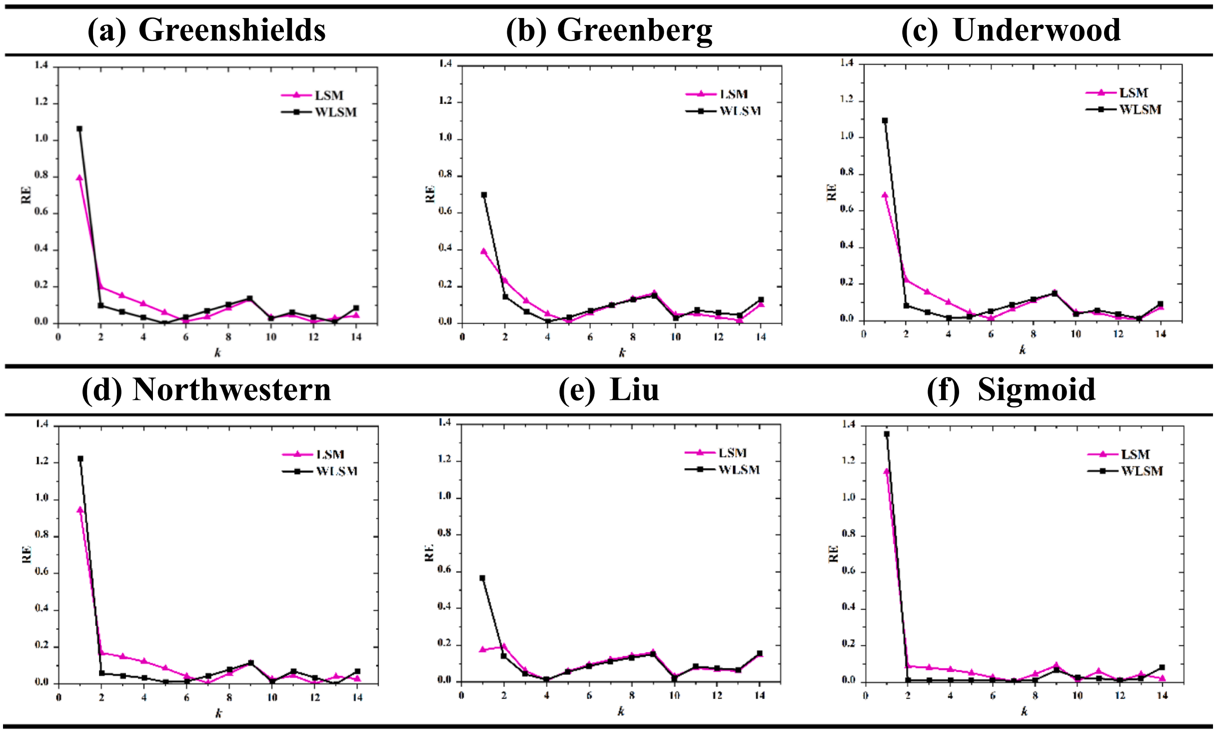


Fig. 9. Relative errors of parametric models.

least square method introduced in Section 3, the sum of square errors in Eq. (16) can be revised as

$$Z_3 = \sum_{i\in R_1}\left[\max\left\{\phi_w(k) - \lambda\cdot q_1^i, 0\right\}\right]^2 + \sum_{i\in R_N}\left\{\left[\min\left\{\phi_w(k) - \left(\Delta d\cdot N^i + \lambda\cdot q_N^i\right), 0\right\}\right]^2 + \sum_{n=2}^{N-1}\left[\max\left\{\phi_w(k) - \left(\Delta d\cdot n + \lambda\cdot q_n^i\right), 0\right\}\right]^2\right\}. \quad (26)$$

Furthermore, the regression model of the piecewise linear function can be formulated as

[M2′]

$$\min\ Z_3 \quad (27)$$

s.t.

$$w_0, w_1, ..., w_n, ...w_N > 0. \quad (28)$$

4.3. Calibration results

4.3.1. Calibration results for general and parametric models

The collected data are applied to calibrate **[M1′]** and **[M2]**, respectively. Two general observations can be obtained from the calibration results in Figs. 7 and 8: (1) Both general and parametric models show the existence of WTW; (2) In the general model, it can be identified that there is an increasing trend of $W(k)$ with regard to k. This is consistent with the phenomenon that the passengers at the back of the queue have a greater willingness to move to other queues with fewer waiting passengers.

There is an obvious gap in the value of $W(k)$ in the general model with the change of the sequence number of the newly arrived pedestrian k. For instance, the $W(k)$ of the first pedestrian is 30, while the $W(k)$ of the second to the ninth pedestrian is 70, which is about two times larger than the first pedestrian. This gap reduces with the increase of k. Fig. 9 presents the calibration results of six different types of parametric models. A clear difference between LSM and WLSM can be observed from the following aspects: (1) The Liu model performs well using LSM, while the Sigmoid performs well using WLSM; (2) In addition to the Liu model (Liu et al., 2016), there exists a clear gap between LSM and WLSM curves; (3) The WLSM method not always gives better calibration results (i.e., smaller values of errors defined in Eq. (16)). Only two models have lower values of errors, that is, the Greenberg model and the Sigmoid model.

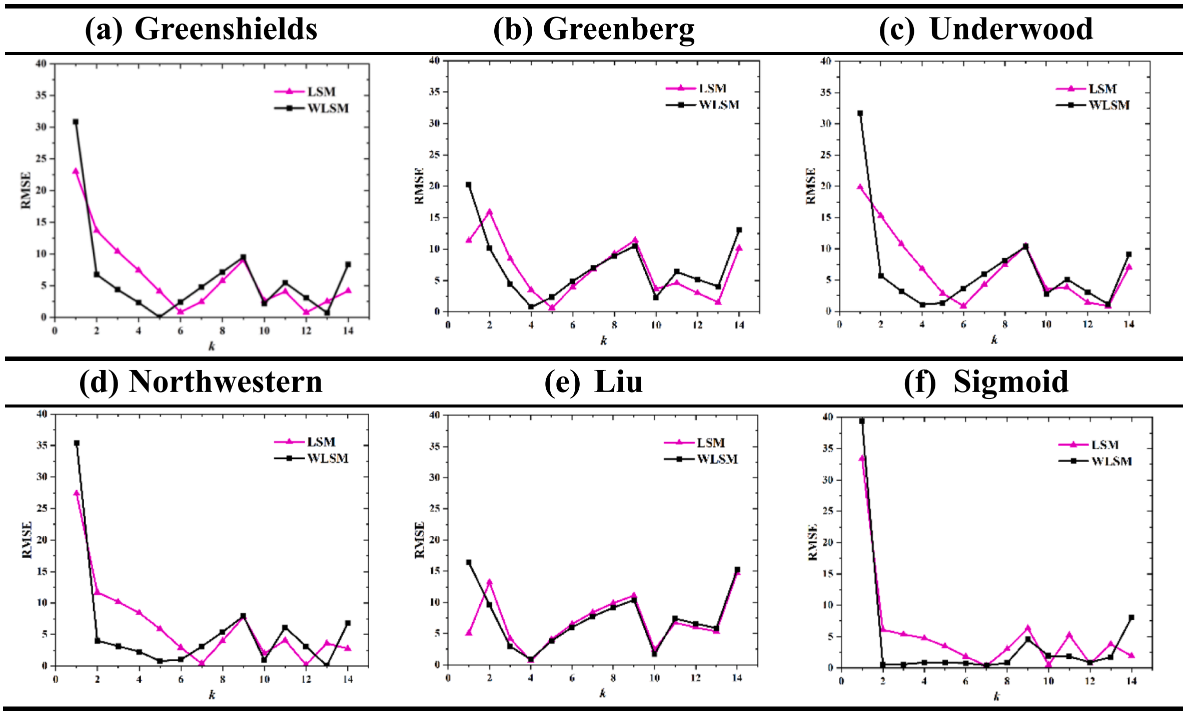


Fig. 10. Root-mean-square errors of parametric models.

Table 5
Comparison of average RE and RMSE.

Parametric models	LSM		WLSM	
	RE	RMSE	RE	RMSE
Greenshields	0.12	6.49	0.13	6.28
Greenberg	0.10	6.72	0.12	7.15
Underwood	0.12	6.81	0.13	6.60
Norethwestern	0.13	6.52	0.12	5.73
Liu	0.09	7.02	0.12	7.41
Sigmoid	0.12	5.48	0.11	4.52

4.3.2. Model validation

This section quantitatively analyzes the performances of the six models. Two error test estimators are applied which have been widely used in transportation data analysis. The relative error (RE) aims to compare the approximation of numbers of widely differing sizes. It is calculated by the ratio of absolute error and the actual value

$$\mathrm{RE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|W_i - \widehat{W}_i|}{W_i}, \tag{29}$$

where $\widehat{W_i}$ is the pedestrian's WTW predicted by the model; W_i is the actual value. Considering the fact that the pedestrian's actual WTW cannot be directly obtained, the WTW calculated by the general model is used instead.

The root-mean-square error (RMSE) represents the sample standard deviation of the differences between actual data and predicted values

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(W_i - \widehat{W}_i)^2}. \tag{30}$$

As can be seen in Figs. 9 and 10, both LSM and WLSM have higher values of RE and RMSE when k is small. This is because of the limited quantity of data records regarding the pedestrian standing in front of the nearest screen door. The average values of RE and RMSE of considered parametric models are summarized in Table 5. It shows that the Liu model has the lowest value of RE but the largest value of WLSM in LSM; while the Sigmoid performs best using WLSM, which is consistent with the observation in the above subsection.

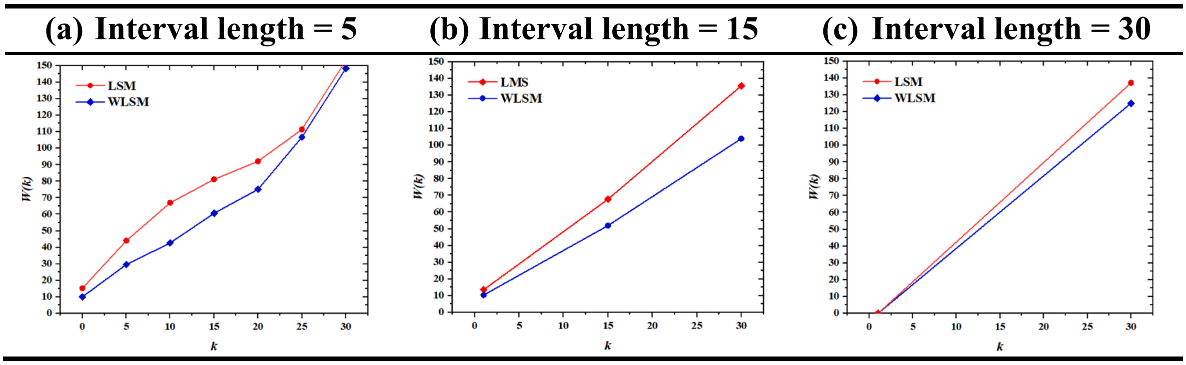


Fig. 11. Calibration results of piecewise linear models.

Table 6
Comparison of average RE and RMSE.

Piecewise line models	LSM		WLSM	
	RE	RMSE	RE	RMSE
Interval length = 5	1.70	128.78	1.16	94.46
Interval length = 15	0.55	41.92	0.47	26.94
Interval length = 30	0.32	22.01	0.29	19.64

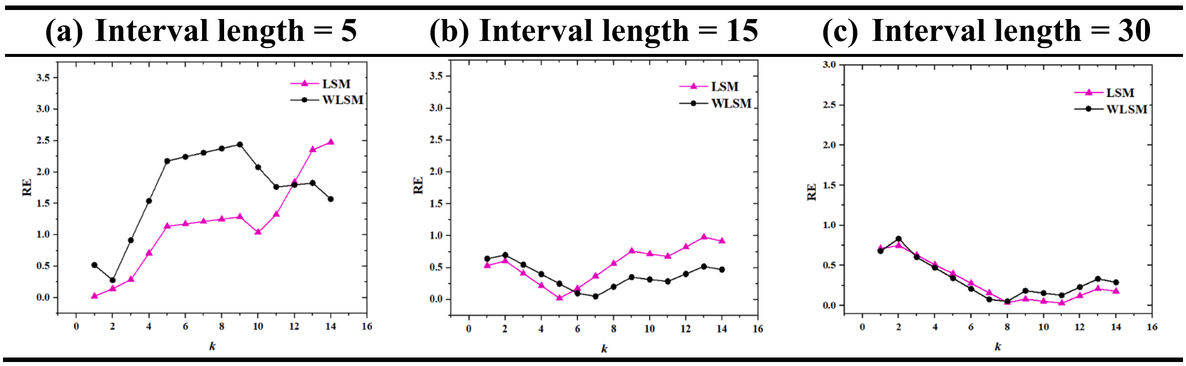


Fig. 12. Relative errors of piecewise linear models.

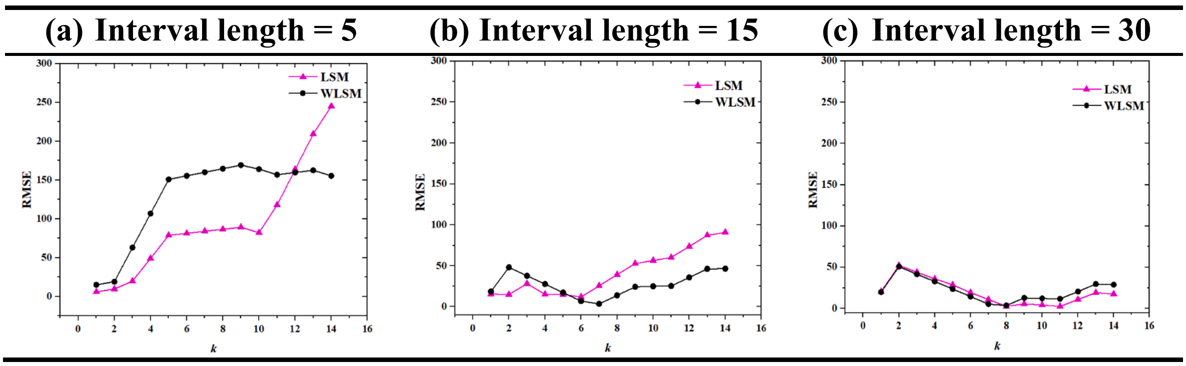


Fig. 13. Root-mean-square errors of piecewise linear models.

Index of pedestrian	Number of pedestrians in the first queue	Final choice	Number of pedestrians in the final choice queue
1	0	3	0
2	0	7	0
3	0	2	0
4	0	1	1
5	1	3	2
6	2	4	0
7	2	2	0
8	2	3	0
9	2	2	2

Fig. 14. On-site survey in Wangzhuanglu Station, Wuxi, China.

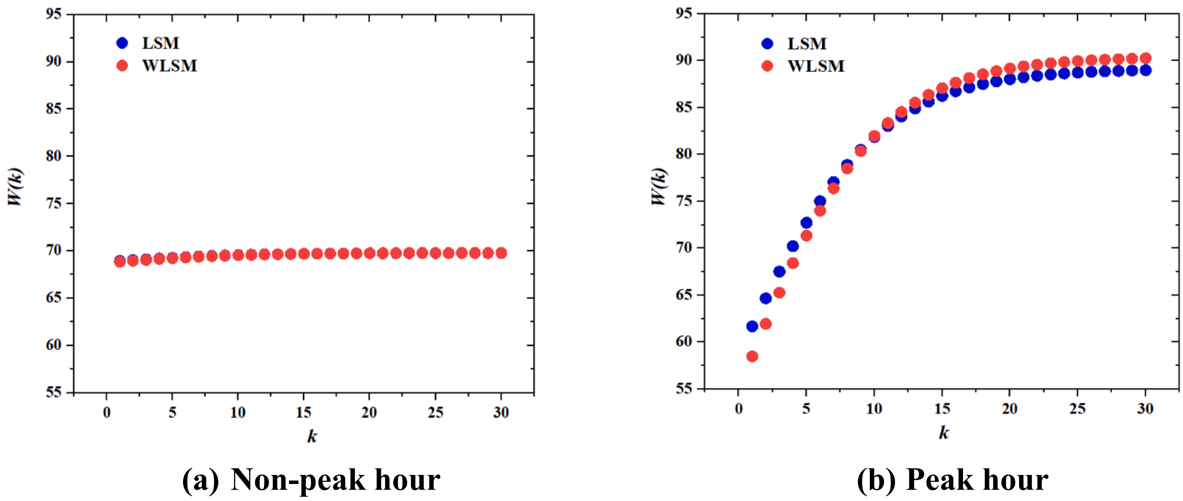


Fig. 15. Calibration results of parametric models using real data.

4.3.3. Validation of the piecewise linear model

This section provides a detailed analysis of the performance of the piecewise linear model proposed in Section 4.2.2. The overall accuracy of the approximation model is obviously dependent on the length of each interval. Hence, the section analyzes intervals of three different lengths separately. As shown in Fig. 11, the pedestrian's WTW described by the piecewise linear model shows an increasing trend with respect to k. The results obtained from the WLSM for fitting are more conservative than those obtained from the LSM. Moreover, for the same length of interval, the values of RE and RMSE obtained from WLSM are lower than those obtained from LSM (see Table 6). Once again, this proves that in the case of non-uniform samples, the fitting effect of WLSM is better than that of traditional LSM. The segmented linear model does not exhibit the situation of mutual intersection seen in continuous models. This is because there is an increasing relationship between the decision variables $w_0, w_1, \ldots, w_n, \ldots w_N$ when calibrating **[M2′]**.

As aforementioned, the length of the interval has a significant impact on the accuracy of the approximation model. In general, the smaller the interval, the closer the linear function within each interval will approximate the curve. However, as shown in Table 6, the values of RE and RMSE decrease with an increase in the length of the interval. This is because after dividing the independent variable k into intervals, the sample size in each segment becomes smaller, which leads to an increase in fitting error and makes the model less accurate. Therefore, when the interval is 30, it contains the most samples and has an error that is closer to the continuous model mentioned in the previous section. Additionally, in Figs. 12 and 13, the change in RE and RMSE with respect to k also supports this point. Among the samples we collected, the sample ratio for the intervals [1,5], [6,10], [11, 15] is 1:3:1. When the length of interval takes 15, the values of RE and RMSE for [6, 10] are significantly lower than those for the other two intervals, and the performance of WLSM is still better than that of LSM.

Table 7
Comparison of average RE and RMSE for non-peak and peak hours.

Hours	LSM		WLSM	
	RE	RMSE	RE	RMSE
Non-peak	0.001	0.093	0.001	0.101
Peak	0.056	4.188	0.051	4.255

Table 8
Classification of carriage congestion level.

Count of passengers	Congestion level	Service level
≤180 per/carriage	comfortable	A
180 < per/carriage < 300	general	B
≥300 per/carriage	crowded	C

Table 9
The scenarios of passenger movement.

Scenarios	Service level of the current carriage,i	Service level of the adjacent carriage,j	Moving willingness,$x_{i,j}$
Scenario 1	C	A	2
Scenario 2	C	B	1
Scenario 3	B	A	1

4.4. *Calibration results by real data*

In addition to calibrating the proposed model using simulation data, we also collected real-world data to calibrate the model. The Wangzhuanglu Station, Wuxi, China, is chosen as the investigation site. As shown in Fig. 14, the distinctive feature of this station is that its elevators are positioned at two separate sections of the platform, with no intervening walls obstructing the view. As a result, this station is particularly well-suited for the observation of passenger choices. We deployed three cameras on the platform, situated at both ends and the midpoint. The recorded information encompasses the initial and ultimate queue designations opted for by pedestrians, as well as the queue sizes at each door upon vehicle arrival. Based on the collected data, the parametric model is calibrated according to the proposed method in Section 4.2.

We collected the queuing data from more than 200 individuals during both the morning and evening peak hours. The calibration results based on the Liu model are shown in Fig. 15. It can be observed that during non-peak hours, due to the limited occurrence of queuing, the value of WTW is a constant regarding the location of passengers. It can be contributed to the fact that pedestrians tend to choose queues more casually for waiting. It also raises an essential discussion on the scope of application of the proposed WTW with respect to the congestion level on the platform. The analysis on the validity of WTW will be further discussed in Section 6.5. The average values of RE and RMSE of the model for non-peak hours is presented in Table 7, which demonstrates good fitting performance of the proposed model when applied to real-world data.

5. Model of passenger's in-vehicle crowdedness and movement

The train equipped with the airbag-based suspension system can weigh the passengers in each carriage, such as the Hammersmith & City line of the London Underground (Fang et al., 2019). In this regard, the number of passengers in each carriage can be directly obtained through loadweight measurements. For trains without pressure sensors, it is possible to estimate the number of passengers entering each car based on the distribution of pedestrians on the platform. In addition to the location of the waiting area, the degree of crowding inside the carriage is also an important factor that influences passengers' door choice behavior (Yang et al., 2017; Peftitsi et al., 2021). However, as passengers waiting on the platform cannot observe the number of people inside each car, the WTW model proposed in this paper does not take into account the congestion levels of the cars from the passengers' perspective.

In the actual investigation, passengers may encounter two scenarios during the boarding process: (1) upon train arrival, passengers may switch to board another carriage if they observe the current carriage to be crowded; (2) after boarding a carriage, passengers may relocate to a less crowded carriage if the initial carriage is too crowded. Both situations mentioned above involve passengers making a decision to switch carriages or doors after perceiving overcrowding in the current queue or carriage. Hence, it is reasonable to use a unified model to describe the impact of in-vehicle congestion on passengers' carriage or door selection behavior.

In order to accurately quantify the level of service inside carriages, we have categorized in-vehicle congestion into three levels according to the pedestrian's level-of-service (Fruin, 1970) (see Table 8). As previously mentioned, some passengers may choose to switch their target carriage and move to a nearby one when the train is approaching or after entering a carriage due to in-vehicle congestion. Based on the above analysis, the passenger's movement is categorized into three scenarios, and passengers have a stronger inclination to move from crowded carriages to less crowded ones (see Table 9). A binary logistics model is formulated to

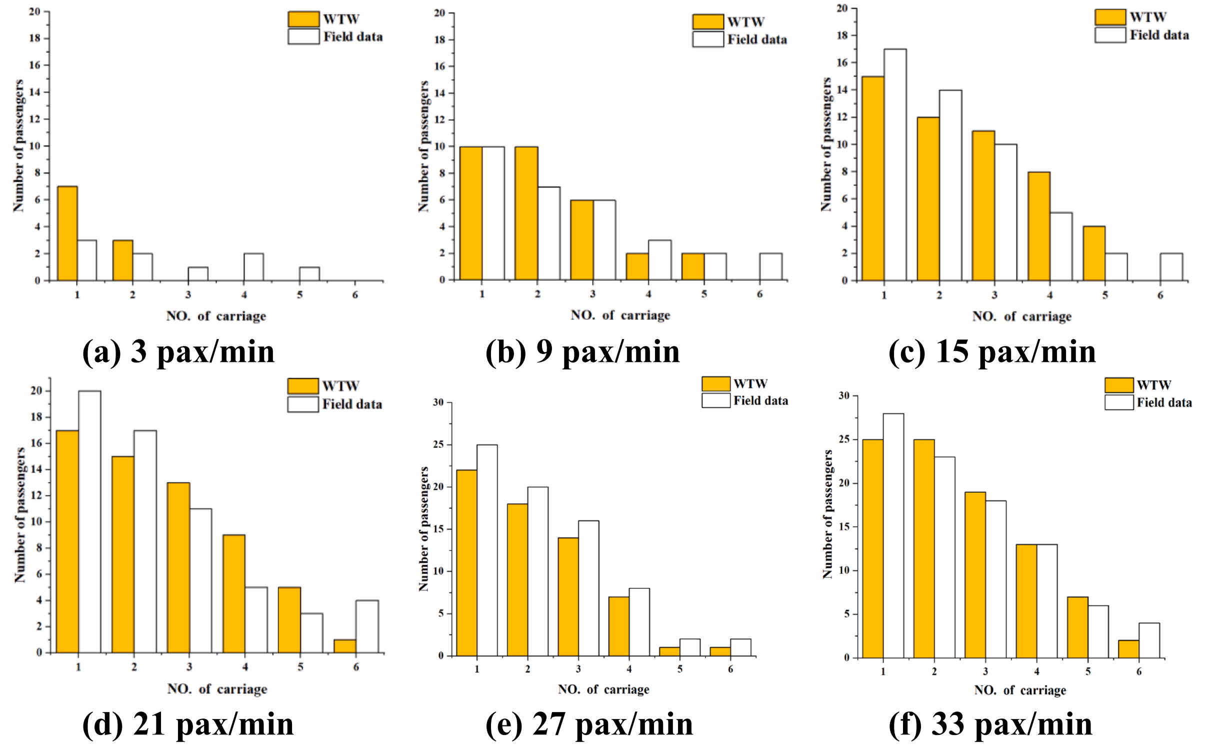


Fig. 16. The comparison between the results obtained by WTW and field data with respect to the passengers volume.

predict the proportion of passengers who are willing to go to nearby carriage. y is defined as the response variable. When a passenger is willing to move to nearby carriage, $y = 1$; otherwise, $y = 0$. The independent variable $x_{i,j}$ represents the passenger's moving willingness. The conditional probability that passengers move to adjacent carriages according to the in-vehicle congestion level is

$$p_{i,j} = p\left(y_{i,j} = 1 \middle| x_{i,j}\right) = \frac{\exp\left(b + ax_{i,j}\right)}{1 + \exp\left(b + ax_{i,j}\right)}. \tag{31}$$

where $p_{i,j}$ is the probability that passengers is willing to move from carriage i to j, which is a function of independent variable $x_{i,j}$; a is the corresponding coefficient of $x_{i,j}$ and b is the intercept of the equation.

6. Case study

The concept of WTW is proposed to estimate the distribution of waiting passengers along the subway platform. In this section, an experiment for Metro Line 3 in Wuxi, China, is presented. Five stations of Metro Line 3 with different platform layouts are built in Legion Studio. Each train has six carriages, and each carriage has four doors (as well as waiting areas).

The Liu model (Liu et al., 2016) and the Sigmoid model which have been validated to have better performance using LSM and WLSM are applied to model the pedestrian's queue choice process.

6.1. Comparison between the predicted queue length and field data

To visually exemplify the efficacy of the proposed WTW model, this section entails a comparative analysis between the results obtained from the WTW model and real data. As mentioned in Section 4.4, when the passenger volume is low, the disparities in pedestrian's WTW are not significant. Hence, in consideration of the applicability of WTW concerning congestion levels, this section refrains from a simplistic categorization of the collected data into peak and non-peak periods. Instead, the data is classified based on the volume of commuters entering the station. It employs the passenger volume walking downstairs from the escalator and then waiting on the platform per minute as the principal criterion for classification. Accordingly, the data is categorized into six distinct scenarios, as illustrated in Fig. 16.

On the whole, the outcomes derived from WTW closely align with the real data, exhibiting a nearly identical distribution trend in queuing distribution. When the passenger volume reaches 9 pax/min, the results obtained through WTW closely mirror the real data. However, for passenger volumes less than 9 pax/min, in which cases the platform is less congested, the results of WTW exhibit an overestimation in proximity to the elevator, compared to the real data. The real data demonstrates a certain degree of long-tail effect.

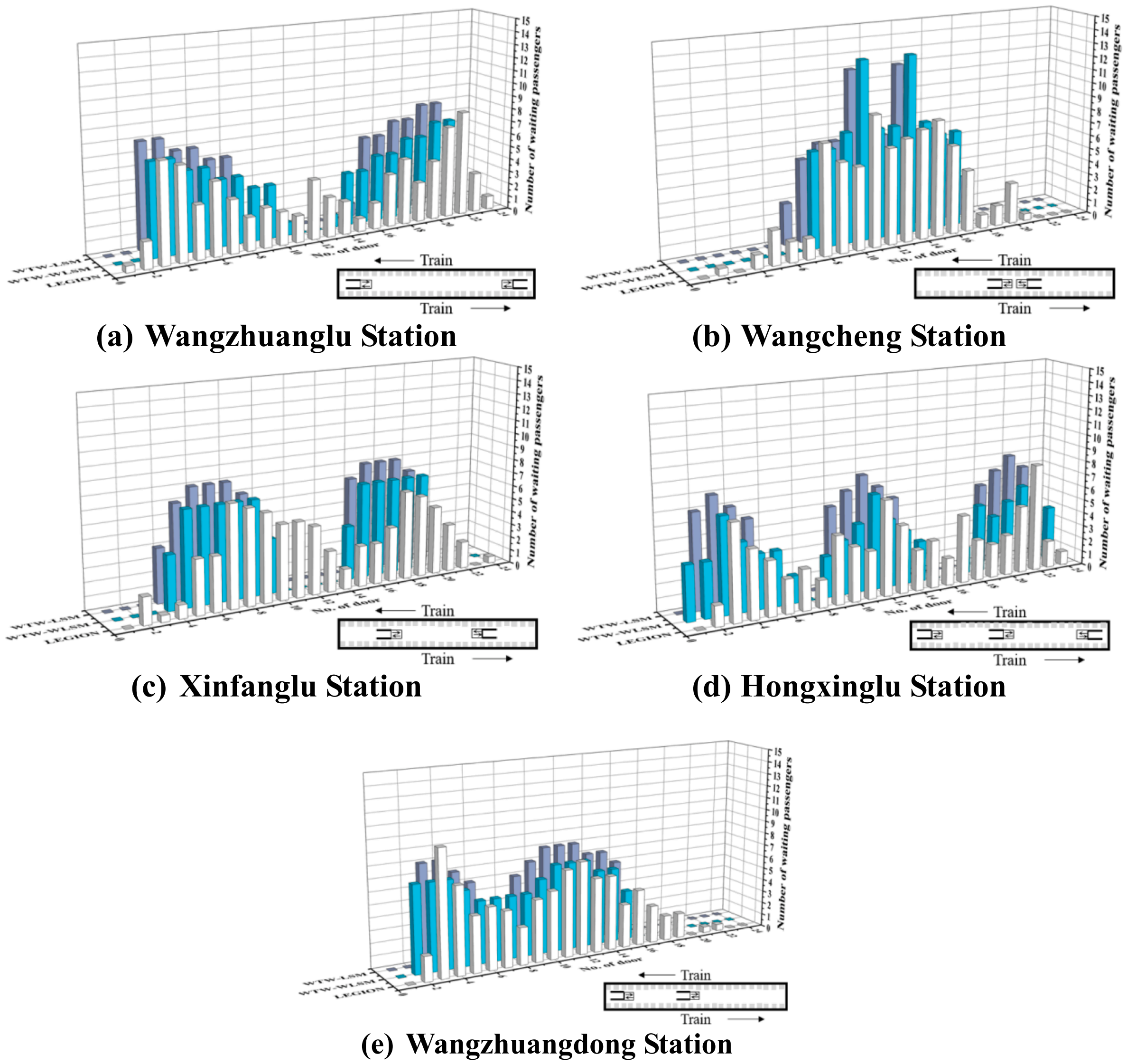


Fig. 17. Distribution of waiting passengers along the platform using simulation and WTW.

This phenomenon can be attributed to the increased variability in passenger queue choice behavior in non-congestion scenarios, where passengers exhibit a higher degree of arbitrariness in their queuing choices. Beyond the threshold of 9, as the platform gradually becomes more congested, the overall outcomes yielded by WTW tend to be conservative. Nevertheless, the disparities between each queue and the field data remain within a range of three individuals.

6.2. Analysis of on-platform pedestrian distribution

Fig. 17 depicts the distribution results of waiting passengers along the platform within a given headway between two adjacent rail trains in five stations. The layout of each platform (especially the location of escalators) is presented as well. Three experiments are conducted using Legion simulation and two pedestrian queue choice models developed based on WTW, which are called WTW-LSM and WTW-WLSM. The results show that the phenomenon of passenger aggregation is significant in all experiments, that is, passengers are more likely to choose the nearest door when they walk downstairs from the escalator. The increase in walking distance leads to a decrease in passengers. Another observation is that the passenger choice models based on WTW can obtain an approximated result compared with the simulation tool. This similarity lies in the fact the Legion Studio is developed based on the classic CA model which could describe the interaction between individuals and the surrounding environment. The pedestrian's choice is measured by three types of physical and psychological factors, namely, inconvenience, discomfort, and frustration (Seriani and Fernandez, 2015). These factors could be comprehensively considered by the unique factor WTW proposed in this paper.

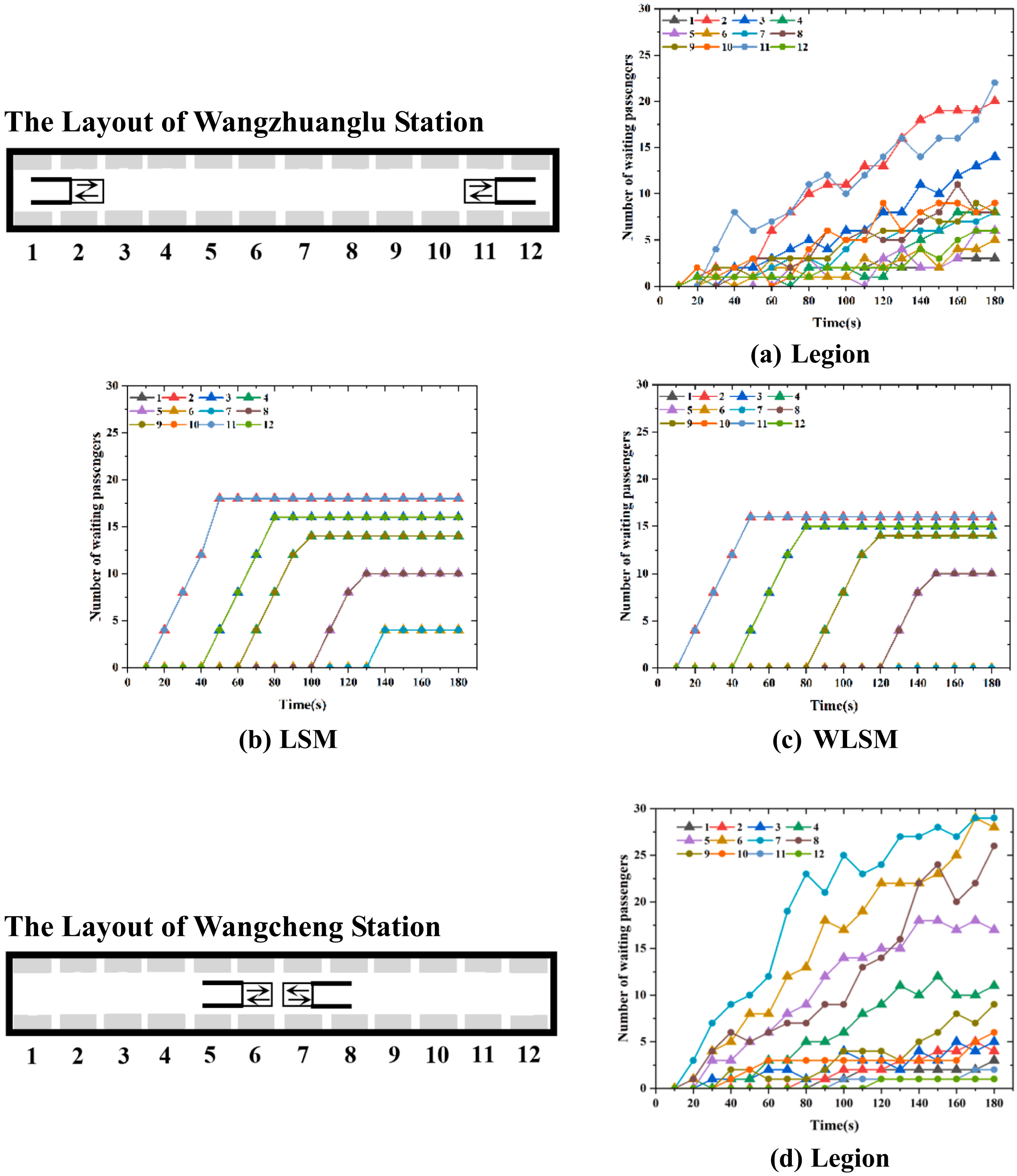


Fig. 18. Dynamics of pedestrian's queuing process.

Fig. 18 captures the dynamics of the pedestrian's queue choice process. To clarify the graphic representation, we combine the results of adjacent queues by reducing the number of queues from 24 to 12. The simulation result of Legion shows that pedestrians accumulate at the nearest waiting area to the escalator from the beginning of the simulation, and the increasing rate of waiting passengers decrease over time. This phenomenon is significant in the Wangcheng Station where two escalators are located adjacent in the middle of the platform. The pedestrian's queue choice process using WTW is simpler because the pedestrian can easily determine which queue to wait depending on her/his location of the nearest queue to the escalator and the number of waiting passengers in other queues.

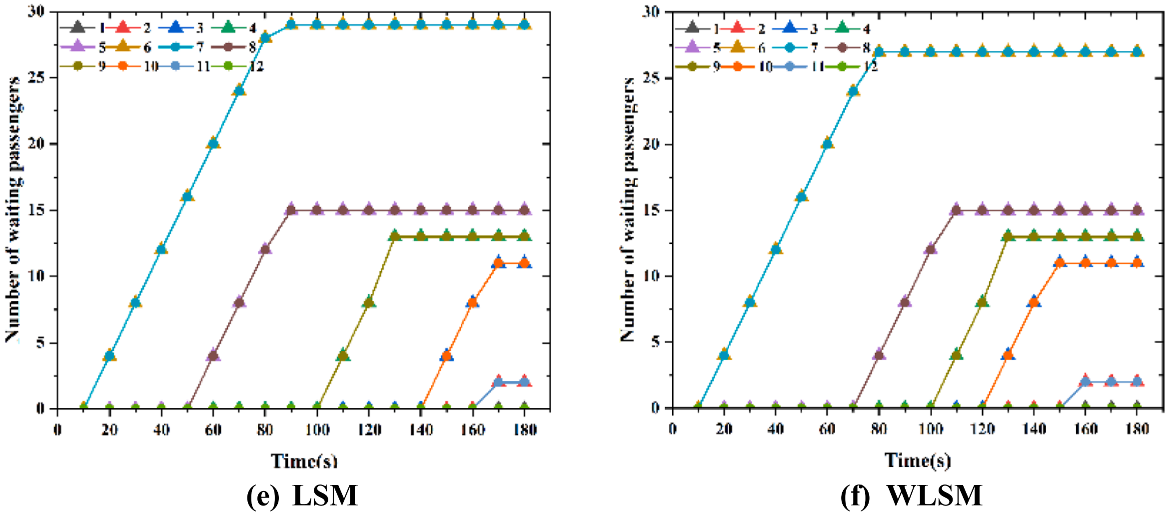


Fig. 18. (*continued*).

6.3. Analysis of in-vehicle crowd congestion

The on-platform passenger distribution has an immediate impact on the on-board crowd congestion level because the effect of pedestrian flow through carriages is insignificant (Xu et al., 2016; Luangboriboon et al., 2021). Fig. 19 depicts the carriage passenger load when a train departs from stations. It can be seen that the distribution of passengers in each carriage changes during the operation of a train, and the distribution result is highly dependent on the location of the escalators. For instance, at the beginning station, i.e., the Wangzhuanglu Station, carriages 3 and 4 are less congested because escalators of this station are located near carriages 1 and 6. However, carriages 3 and 4 become increasingly congested during the operation because in the following stations there is at least one escalator located near carriages 3 or 4. The average difference with respect to the number of passengers in each carriage between Legion simulation and WTW-LSM and WTW-WLSM are 5 and 4, respectively, which shows that the proposed WTW model can give a reliable approximation to the actual situation.

Fig. 20 presents the spatial and temporal dynamics of in-vehicle crowd congestion derived from the Legion and the WTW-based model. The analysis focuses on five rail trains operating on Metro Line 3 from 8:00 a.m. to 12:00 a.m. Dynamic OD information between stations was extracted from Automated Fare Collection (AFC) data and used as input for the proposed Legion and WTW models. These models were utilized to simulate vehicle operations, as well as passenger boarding and alighting processes. The Legion simulation software allows direct output of passenger counts in each carriage. In the WTW-based model, the initial step involves calculating the number of queues in front of each door, followed by determining the movement of queued passengers entering the carriages using the model introduced in Section 5. This process ultimately yields the passenger count for each carriage. Fig. 20 illustrates that the proposed method provides a rough estimation of the congestion level inside the carriages, offering valuable reference information for passengers at the next station.

6.4. Analysis of the impact of information signs on pedestrian distribution

In the previous section, it was observed that the WTW-based model is capable of obtaining the level of crowding inside each carriage when it departs from the platform. As mentioned in the Introduction, the purpose of this paper is to propose a WTW model with the aim of estimating in-vehicle crowd congestion solely based on AFC data, without the need for additional data such as weight carriage load. Therefore, in this section, we further analyze the impact of information signs on passengers based on the calculation of queue size using the WTW-based model.

The information sign shown in Fig. 1 records the level of congestion inside the approaching train, and such signs are part of the subway's internal guidance system. Their purpose is to guide pedestrians' movement and avoid congestion. In literature, signs are typically considered in the tactical level behavior of pedestrians, i.e., path choice (Johansson et al., 2015; Huo et al., 2023; Huang and Liao, 2023). However, due to the heterogeneity of pedestrians, it is challenging to characterize the impact of information signs on passengers accurately. For instance, Johansson et al. (2015) divide the crowd into two groups, namely, those with interest and those with no interest on the information sign.

In this section, we divide the process of passenger queue selection at the platform into two stages. The first stage involves using the WTW-based model to select a queue. Then, we employ the model described in Section 5 to make the final queue selection based on the information sign. As an example, we consider a specific train in Section 6.2, and Fig. 21 illustrates the display of the information sign for this train before its arrival at each station.

Fig. 22 demonstrates the impact of considering the in-vehicle congestion level on passenger selection. For any given station, the left

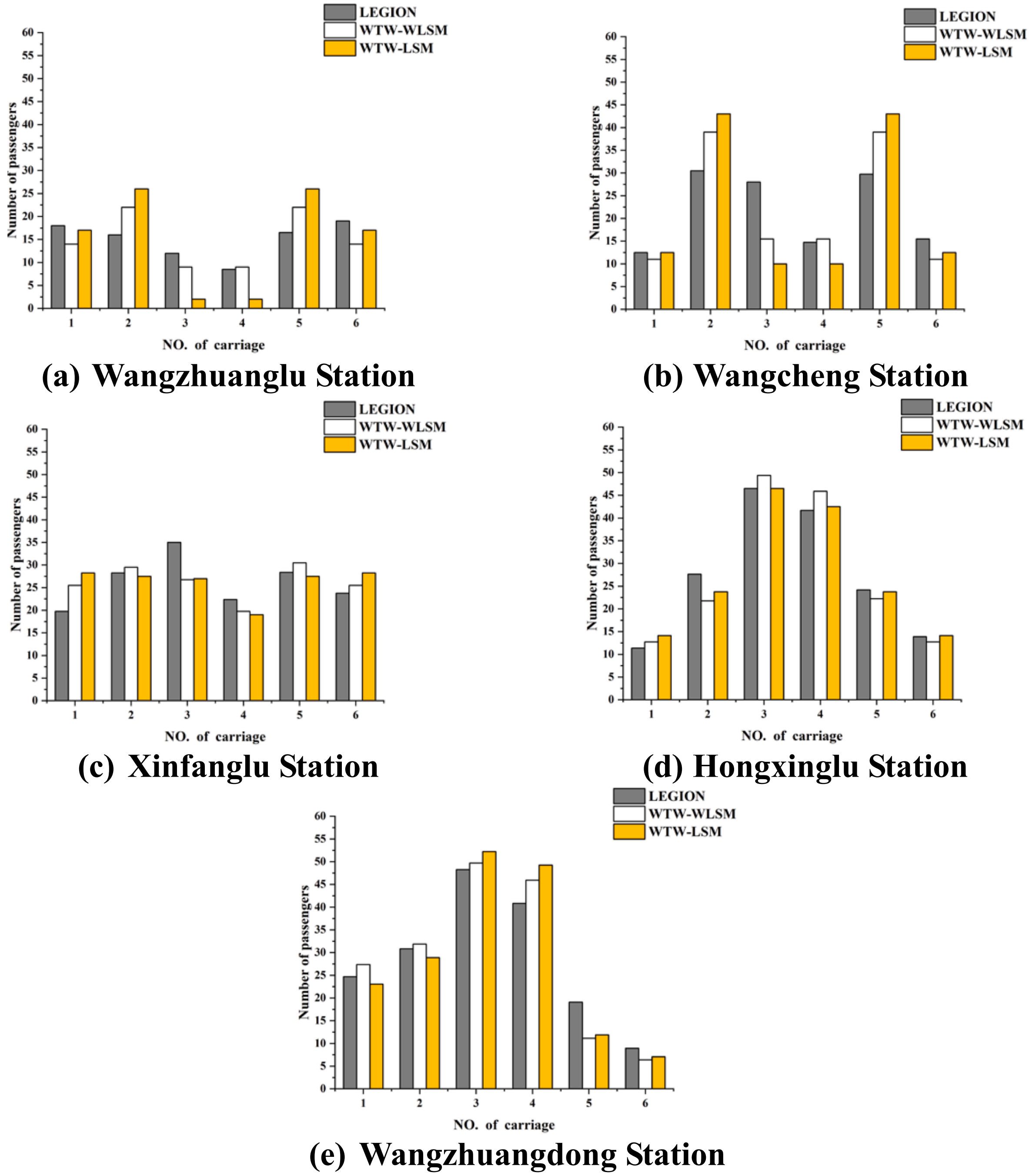


Fig. 19. Carriage passenger load when the train departing from a station.

plot illustrates the scenario where only the in-vehicle congestion level is considered, without taking the information sign into account. In this case, pedestrians move inside the carriage based on the level of congestion observed within the carriage. The right plot illustrates the scenario where both the information sign (i.e., on-platform movement) and the in-vehicle congestion level (i.e., in-vehicle movement) are considered. After selecting a queue using WTW, pedestrians reevaluate their choice based on the in-vehicle congestion level displayed on the information sign for the approaching train. Once inside the carriage, they continue to adjust their positions according to the level of congestion observed within the carriage. It can be observed that passenger choices are influenced by the information sign. Taking Xinfanglu Station as an example, the information sign displays that the 2nd and 5th carriages are more crowded. Pedestrians, based on this information, choose to move towards the less congested carriages on either side to avoid crowding.

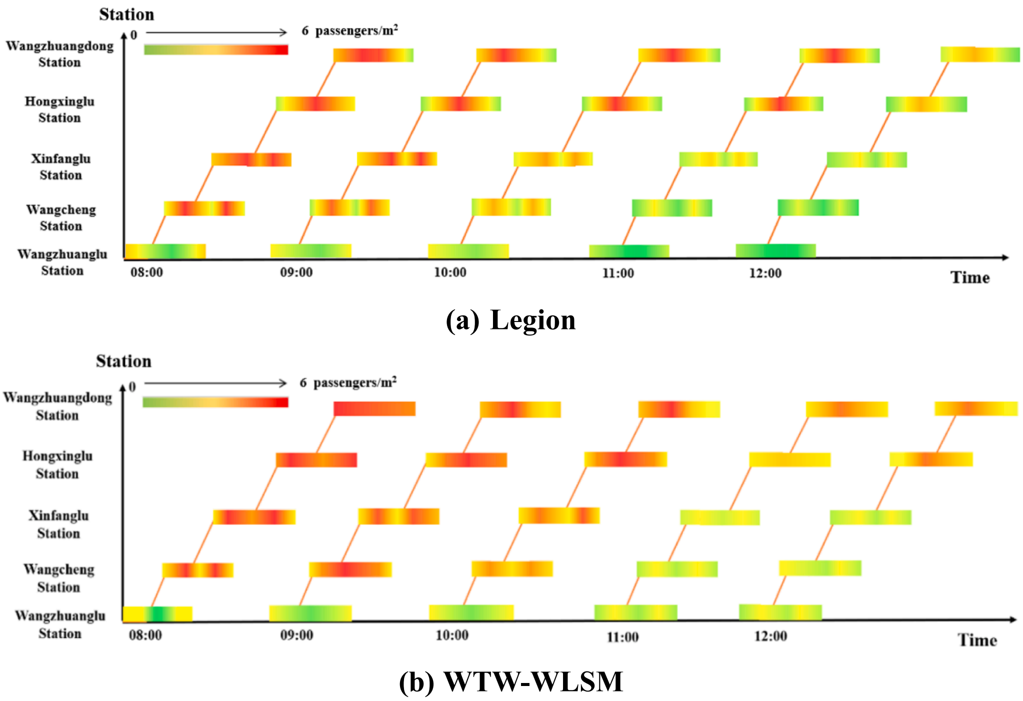


Fig. 20. In-vehicle crowd congestion level of five rail trains.

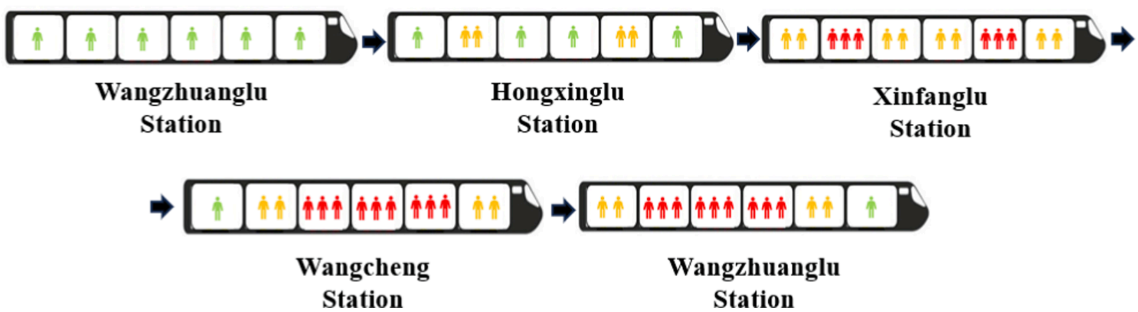


Fig. 21. The information sign for a train before its arrival at each station.

6.5. Validity of WTW with respect to the passenger volume

As evident from Fig. 15 in Section 4.4, the significance of pedestrian WTW is less pronounced when the passenger volume is low. This raises a pivotal inquiry concerning the applicability scope of the WTW model proposed in this study, particularly in relation to congestion levels. Hence, in alignment with Section 6.1, this section quantifies congestion levels by measuring the passenger volume descending from the escalator per minute.

From Fig. 23, it is evident that when the passenger volume is relatively low, specifically when the number of individuals descending from the elevator per minute is less than 9, the variations in pedestrian WTW are not sufficiently significant. Two main factors contribute to this phenomenon: Firstly, in scenarios with a low number of individuals, the formation of queues becomes challenging. Passengers often opt to take a few extra steps to ensure a position near the front of the queue, thereby resulting in the primary cause of the long-tail effect observed in situations of low passenger volume. Secondly, instances of passengers making secondary queue selections are comparatively infrequent. When the passenger volume is around 9 to 15, the WTW exhibits a notable upward trend. In this context, the waiting area has gradually established queues, prompting passengers to develop a certain sensitivity toward their queuing positions, thus giving rise to the emergence of WTW. Upon reaching a passenger volume of 15, the WTW for passengers further increases. This observation underscores that as the number of individuals queuing on the platform grows, the prominence of passenger WTW becomes increasingly evident.

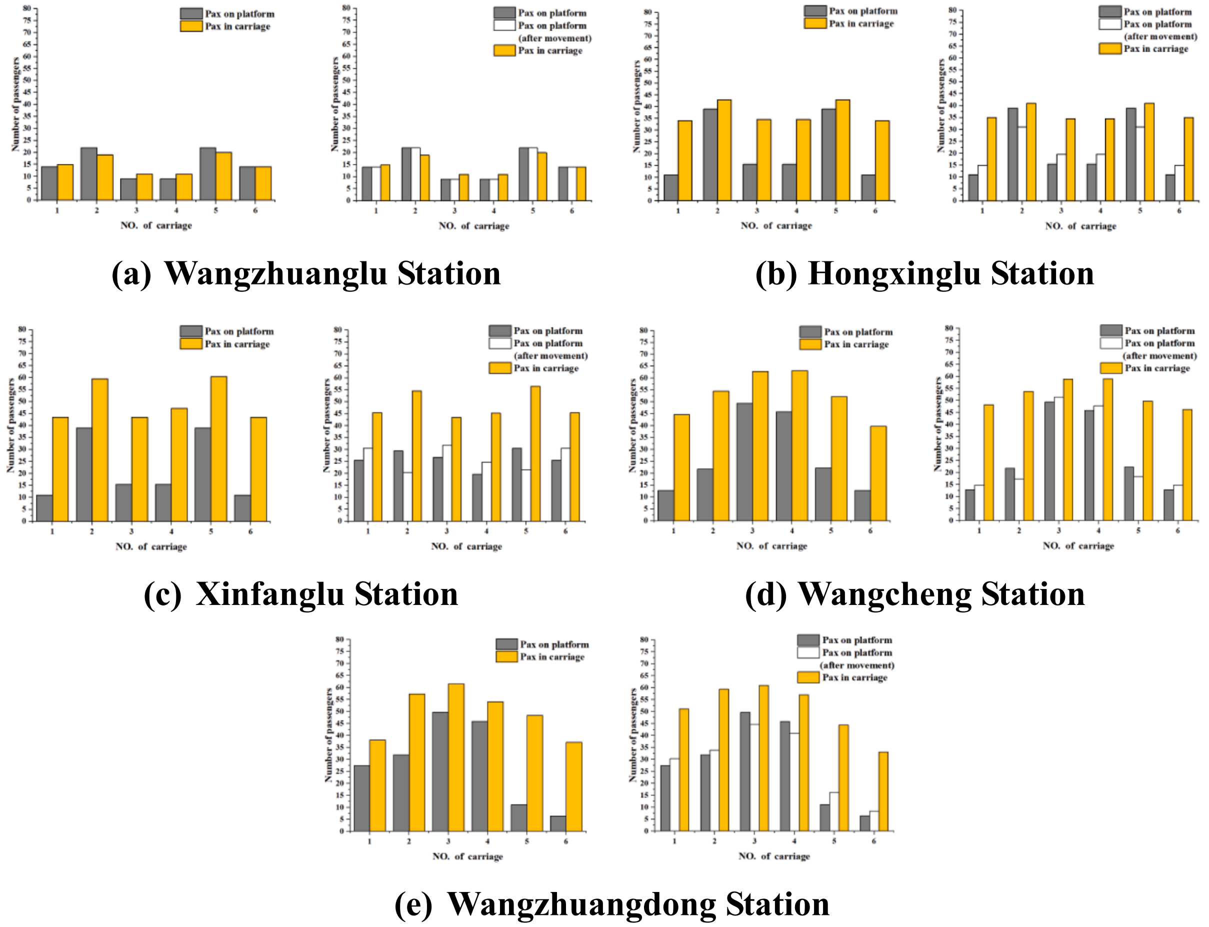


Fig. 22. Carriage passenger load when the train departing from a station considering movements.

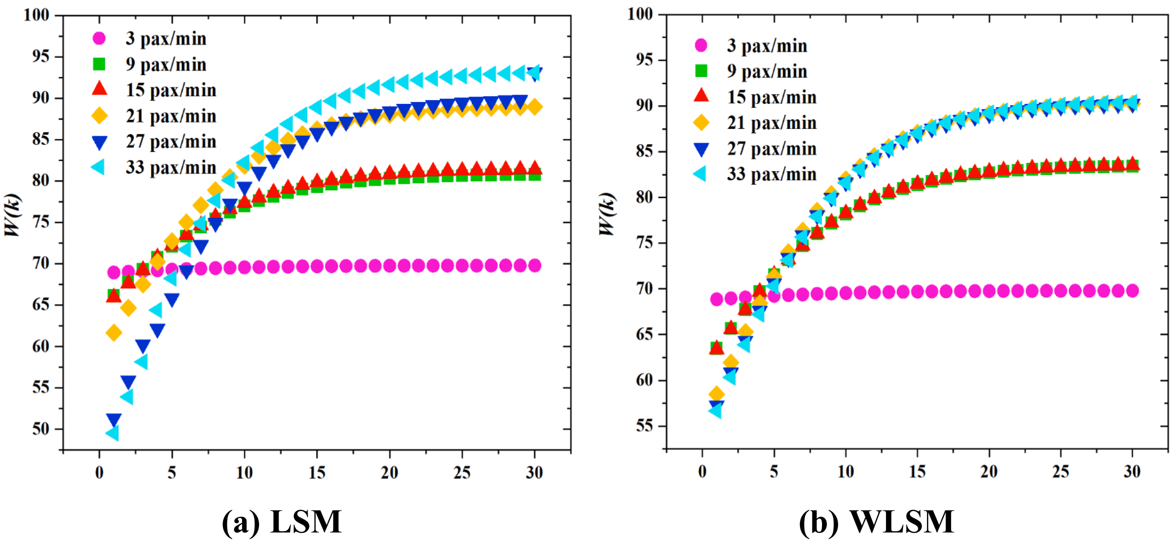


Fig. 23. Existence of WTW with respect to passenger volume.

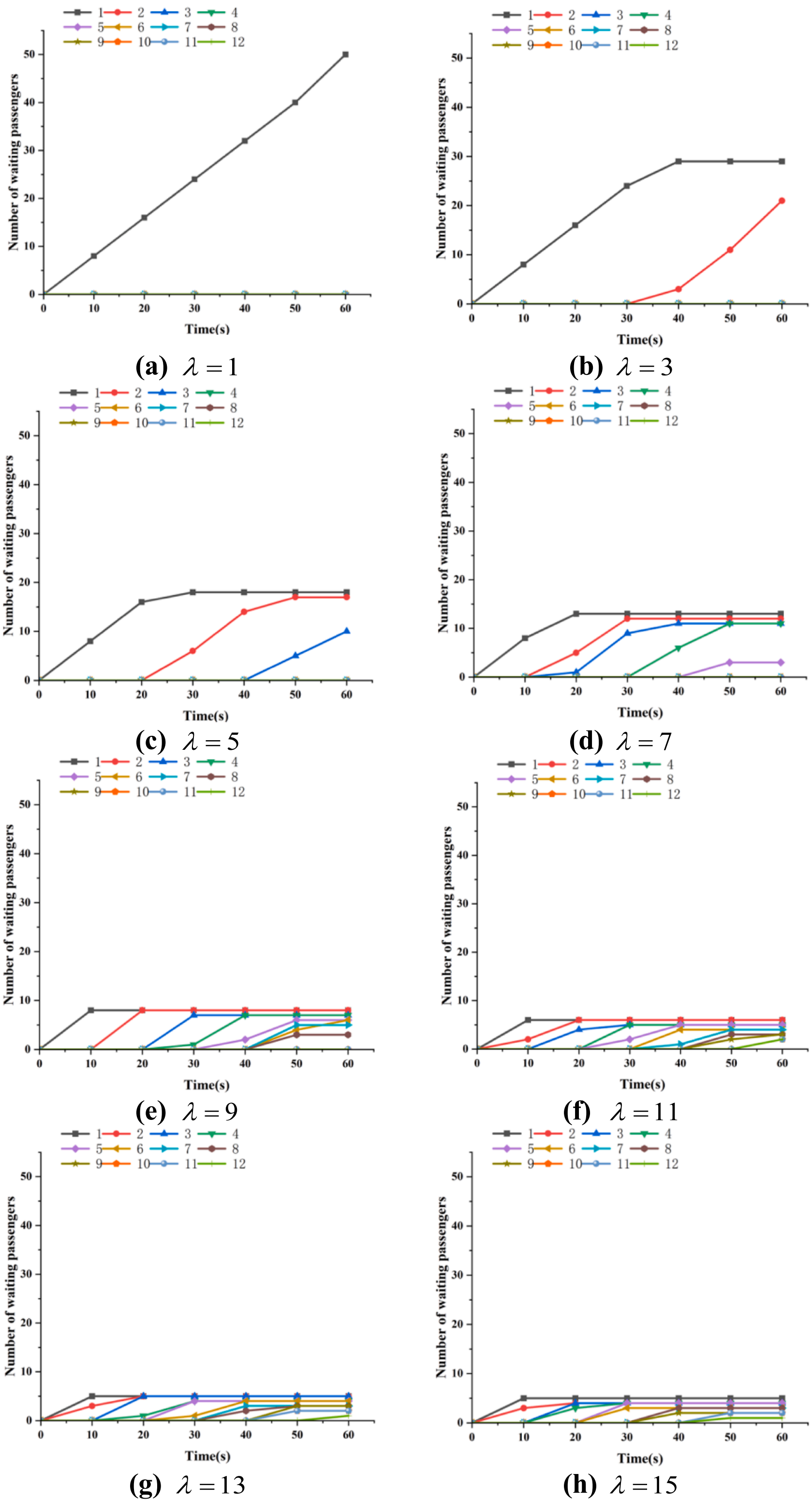


Fig. 24. Dynamics of pedestrian's queuing process with respect to λ.

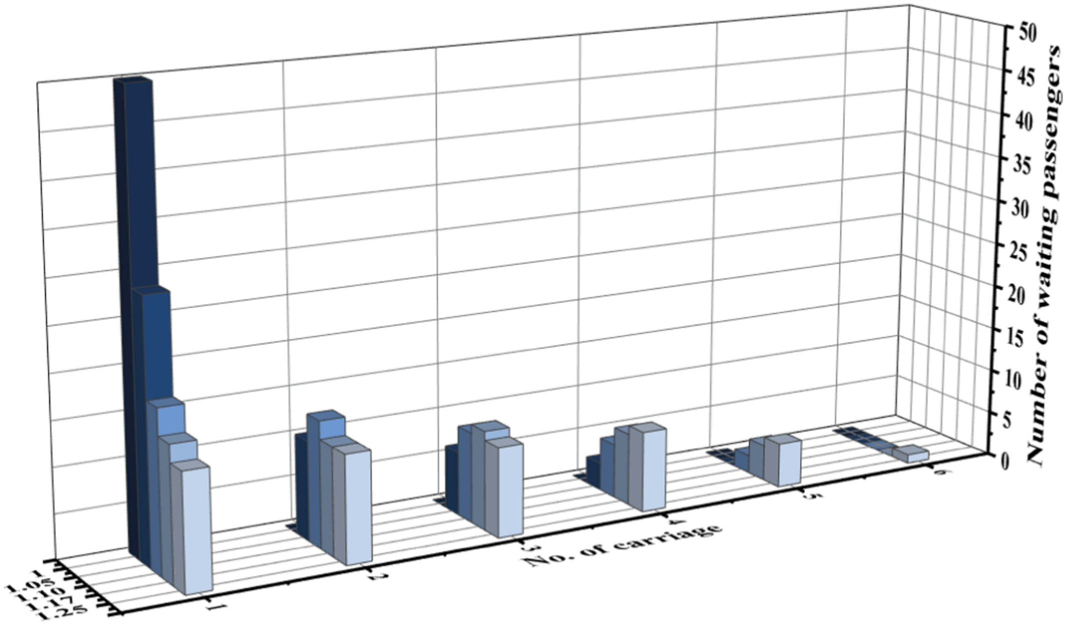


Fig. 25. The result of pedestrian distribution with respect to λ.

6.6. *Sensitivity analysis on* λ

As discussed in Section 2.3, the pedestrian's decision process is jointly dependent on the walking distance to a queue and the current number of passengers waiting in this queue which are unified by the parameter λ. A sensitivity analysis is conducted to examine the effect of the pedestrian's queue choice behavior. To simplify the testing scenario, this analysis only considers one escalator. The nearest door to the escalator is considered as the first door. Fig. 24 presents the pedestrian's queuing process under different values of λ. It clearly shows that with the increase of λ, the number of queues chosen by pedestrians will increase. But the total number of passengers waiting in each queue decreases. It indicates that λ reflects the pedestrian's tolerance of her/his position in a queue. Specifically, if λ is large, a pedestrian is more willing to walk farther doors to wait in the front of the queue. As a result, pedestrians are evenly distributed on the platform.

In the WTW model proposed in this paper, the hyperparameter λ is applied to unified two decision variables when measuring WTW, namely, d_n and q_n, where d_n represents the walking distance to the n-th queue and q_n the current number of waiting passengers in the n-th queue. According to **Assumption 2**, the passenger will walk to the n-th queue if her/his WTW is larger than $d_n + \lambda \cdot q_n$; otherwise, the passenger will stay in the tail of the 1st queue. The value of λ decides the weight of q_n when the passenger makes her/his decision. For instance, smaller λ indicates that passengers are more concerned about the walking distance rather than their position in the queue. As shown in Fig. 24, the value of λ also reflects the maximum queue length that pedestrians are willing to tolerate. The width of the platform determines the maximum queue length that pedestrians can tolerate, so the hyperparameter λ can also serve as a factor to measure the influence of platform width on WTW. Fig. 25 depicts the final distribution of waiting passengers as a function of the variation of λ. It illustrates that as λ decreases, the aggregation effect of passengers in individual queues decreases, indicating that the maximum queue length that passengers are willing to tolerate is decreasing. This also suggests that if the platform width becomes smaller, passengers will choose to queue up in a slightly farther queue.

7. Conclusions

This paper proposes the concept of WTW to describe the desire of waiting passengers (particularly newly arrived ones) to walk away from the queue where they are initially standing to wait for the incoming subway. This idea is derived from the commonly observed phenomenon that passengers who want to take the subway tend to choose to queue up in front of the first door after taking the escalator, but the relatively smaller number of waiting passengers in other queues may attract them to move away.

Hence, WTW describes the psychological tendency of pedestrians to walk on the platform. To quantitatively measure WTW, two regression models are formulated based on observation data. A simulation framework is developed in the Legion simulation environment to collect pedestrian's choice data. The calibration results of both general and parametric models confirm the existence of WTW. The proposed WTW-based passenger choice model contributes to the analysis of pedestrian behavior, resulting in the distribution of passengers along the subway platform and in-vehicle crowd congestion level.

However, there are also some limitations of this paper. First of all, the proposed concept of WTW lacks consideration of pedestrian heterogeneity and individual behaviors. Although socio-psychological crowding stress has been considered in WTW, the relationship

Fig. 26. Snapshots of passenger volumn 33 pax/min.

between crowding stress and pedestrian heterogeneity needs further investigation. Furthermore, WTW can also be incorporated into SF as a supplement to the self-driving force. Secondly, this study assumes that WTW only happens when a passenger walks downstairs from the escalator. However, passengers would make choices on which carriage to board when a train arrives, according to the in-vehicle congestion level. Hence, the perceived information that a passenger obtains about the in-vehicle congestion should be considered. Moreover, the in-vehicle crowd congestion level of the coming train is usually shown on the information signs on the platform. Hence, the passenger's judgement on the congestion state should be incorporated into WTW.

In Section 6.5, the validity of passenger WTW is examined under various passenger volumes. The analysis indicates that the influence of WTW becomes more prominent when there is a higher number of passengers waiting in the queue. However, the validity of WTW needs further investigation, particularly in specific scenarios. For example, as shown in Fig. 26, the passenger queuing situation is depicted at a passenger volume of 33 pax/min. Nevertheless, at certain transfer stations during peak hours, the platforms rapidly fill with passengers. In such congested situations, passengers may find it challenging to switch queues due to the high platform congestion. Therefore, in future research, it is imperative to consider the impact of the surrounding environment on passengers' decision-making processes. Models like the social force model (Johansson et al., 2015) can be employed to capture passengers' aversion to nearby pedestrians and obstacles.

CRediT authorship contribution statement

Di Huang: Conceptualization, Methodology. **Yuwei Yang:** Conceptualization, Methodology. **Xinyi Peng:** . **Jiangyan Huang:** Methodology. **Pengli Mo:** . **Zhiyuan Liu:** Supervision, Validation. **Shuaian Wang:** Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by the Fundamental Research Funds for the Central Universities (No. RF1028623241), the Open Project Program of the Key Laboratory of Transport Industry of Comprehensive Transportation Theory (Nanjing Modern Multimodal

Transportation Laboratory) (No. MTF2023001), the Natural Science Foundation of Jiangsu Province (No. BK20230852), the National Natural Science Foundation of China (No. 72301065).

References

Antonini, G., Bierlaire, M., Weber, M., 2006. Discrete choice models of pedestrian walking behavior. Transp. Res. B 40 (8), 667–687.
Chen, W., Zhu, F., 2022. Discussions on pedestrian delay models and applications at signalized crosswalks. Multimodal Transportation 1 (4), 100039.
Cheng, Q., Lin, Y., Zhou, X., Liu, Z., 2024. Analytical formulation for explaining the variations in traffic states: A fundamental diagram modeling perspective with stochastic parameters. Eur. J. Oper. Res. 312 (1), 182–197.
Davidich, M., Geiss, F., Mayer, H.G., Pfaffinger, A., Royer, C., 2013. Waiting zones for realistic modelling of pedestrian dynamics: A case study using two major German railway stations as examples. Transportation Research Part C 37, 210–222.
Ding, H., Di, Y., Zheng, X., Liu, K., Zhang, W., Zheng, L., 2021. Passenger arrival distribution model and riding guidance on an urban rail transit platform. Phys. A 571, 125847.
Drake, J.S., Schofer, J.L., May, A.D., 1967. A statistical analysis of speed–density hypotheses. Highw. Res. Rec. 154, 112–117.
Duives, D.C., Daamen, W., Hoogendoorn, S.P., 2013. State-of-the-art crowd motion simulation models. Transp. Res. C 37, 193–209.
Fang, J., Fujiyama, T., Wong, H., 2019. Modelling passenger distribution on metro platforms based on passengers' choices for boarding cars. Transp. Plan. Technol. 42 (5), 442–458.
Fruin, J.J., 1970. Designing for pedestrians a level of service concept. Polytechnic University.
Greenberg, H., 1959. An analysis of traffic flow. Oper. Res. 7 (1), 79–85.
Greenshields, B.D., Bibbins, J.R., Channing, W.S., Miller, H.H., 1935. A study of traffic capacity. Highway Research Board 14, 448–477.
Guo, R.Y., 2014. Simulation of spatial and temporal separation of pedestrian counter flow through a bottleneck. Phys. A 415, 428–439.
Huang, K., Liao, F., 2023. A novel two-stage approach for energy-efficient timetabling for an urban rail transit network. Transp. Res. E 176, 103212.
Huang, D., Liu, Z., Liu, P., Chen, J., 2016. Optimal transit fare and service frequency of a nonlinear origin-destination based fare structure. Transp. Res. E 96, 1–19.
Huang, D., Wang, S., 2022. A two-stage stochastic programming model of coordinated electric bus charging scheduling for a hybrid charging scheme. Multimodal Transportation 1 (1), 100006.
Huo, J., Liu, Z., Chen, J., Cheng, Q., Meng, Q., 2023. Bayesian optimization for congestion pricing problems: A general framework and its instability. Transp. Res. B 169, 1–8.
Johansson, F., Peterson, A., Tapani, A., 2015. Waiting pedestrians in the social force model. Phys. A 419, 95–107.
Kim, K.M., Hong, S.P., Ko, S.J., Kim, D., 2015. Does crowding affect the path choice of metro passengers? Transp. Res. A 77, 292–304.
Kim, H., Kwon, S., Wu, S.K., Sohn, K., 2014. Why do passengers choose a specific car of a metro train during the morning peak hours? Transp. Res. A 61, 249–258.
Krstanoski, N., 2014. Modelling passenger distribution on metro station platform. International Journal for Traffic & Transport Engineering 4 (4), 456–465.
Liu, Z., Wang, S., Chen, W., Zheng, Y., 2016. Willingness to board: a novel concept for modeling queuing up passengers. Transp. Res. B 90, 70–82.
Lovreglio, R., Borri, D., dell'Olio, L., Ibeas, A., 2014. A discrete choice model based on random utilities for exit choice in emergency evacuations. Saf. Sci. 62, 418–426.
Luangboriboon, N., Seriani, S., Fujiyama, T., 2021. The influence of the density inside a train carriage on passenger boarding rate. International Journal of Rail Transportation 9 (5), 445–460.
Mo, B., Koutsopoulos, H.N., Zhao, J., 2022. Inferring passenger responses to urban rail disruptions using smart card data: A probabilistic framework. Transp. Res. E 159, 102628.
Nielsen, B.F., Frølich, L., Nielsen, O.A., Filges, D., 2014. Estimating passenger numbers in trains using existing weighing capabilities. Transportmetrica A 10 (6), 502–517.
Oliveira, L.C., Fox, C., Birrell, S., Cain, R., 2019. Analysing passengers' behaviours when boarding trains to improve rail infrastructure and technology. Rob. Comput. Integr. Manuf. 57, 282–291.
Peftitsi, S., Jenelius, E., Cats, O., 2020. Determinants of passengers' metro car choice revealed through automated data sources: A Stockholm case study. Transportmetrica A 16 (3), 529–549.
Peftitsi, S., Jenelius, E., Cats, O., 2021. Evaluating crowding in individual train cars using a dynamic transit assignment model. Transportmetrica B 9 (1), 693–711.
Qu, Y., Gao, Z., Xiao, Y., Li, X., 2014. Modeling the pedestrian's movement and simulating evacuation dynamics on stairs. Saf. Sci. 70, 189–201.
Qu, X., Wang, S., Zhang, J., 2015. On the fundamental diagram for freeway traffic: A novel calibration approach for single-regime models. Transp. Res. B 73, 91–102.
Qu, Y., Xiao, Y., Wu, J., Tang, T., Gao, Z., 2018. Modeling detour behavior of pedestrian dynamics under different conditions. Phys. A 492, 1153–1167.
Qu, Y., Xiao, Y., Liu, H., Yin, H., Wu, J., Qu, Q., Li, D., Tang, T., 2019. Analyzing crowd dynamic characteristics of boarding and alighting process in urban metro stations. Phys. A 526, 121075.
Qu, Y., Liu, X., Wu, J., Wei, Y., 2022. Modeling pedestrian behaviors of boarding and alighting dynamics in urban railway stations. Transportmetrica A 1–24.
Seriani, S., Fernandez, R., 2015. Pedestrian traffic management of boarding and alighting in metro stations. Transp. Res. C 53, 76–92.
Shi, J., Yang, J., Yang, L., Tao, L., Qiang, S., Di, Z., Guo, J., 2023. Safety-oriented train timetabling and stop planning with time-varying and elastic demand on overcrowded commuter metro lines. Transp. Res. E 175, 103136.
Stanitsa, A., Hallett, S.H., Jude, S., 2023. Investigating pedestrian behaviour in urban environments: a Wi-Fi tracking and machine learning approach. Multimodal Transportation 2 (1), 100049.
Sun, L., Yuan, G., Yao, L., Cui, L.i., Kong, D., 2021. Study on strategies for alighting and boarding in subway stations. Phys. A 583, 126302.
Underwood, R.T., 1961. Speed, volume, and density relationship: quality and theory of traffic flow. Retrieved from *Yale Bureau of Highway Traffic* 141–188.
Wang, P. (2016). Understanding social-force model in psychological principles of collective behavior. *arXiv preprint arXiv:1605.05146*.
Xu, X.Y., Liu, J., Li, H.Y., Jiang, M., 2016. Capacity-oriented passenger flow control under uncertain demand: Algorithm development and real-world case study. Transp. Res. E 87, 130–148.
Xu, G., Liu, Y., Gao, Y., Liu, W., 2023. Integrated optimization of train stopping plan and seat allocation scheme for railway systems under equilibrium travel choice and elastic demand. Transp. Res. E 177, 103231.
Yang, X., Dong, H., Yao, X., 2017. Passenger distribution modelling at the subway platform based on ant colony optimization algorithm. Simul. Model. Pract. Theory 77, 228–244.
Yang, X., Zhang, R., Pan, F., Yang, Y.i., Li, Y., Yang, X., 2022. Stochastic user equilibrium path planning for crowd evacuation at subway station based on social force model. Phys. A 594, 127033.
Yu, C., Li, H., Xu, X., Liu, J., 2020. Data-driven approach for solving the route choice problem with traveling backward behavior in congested metro systems. Transp. Res. E 142, 102037.
Yuan, Y., Li, S., Yang, L., Gao, Z., 2022. Real-time optimization of train regulation and passenger flow control for urban rail transit network under frequent disturbances. Transp. Res. E 168, 102942.
Zhang, P., Sun, H., Qu, Y., Yin, H., Jin, J.G., Wu, J., 2021. Model and algorithm of coordinated flow controlling with station-based constraints in a metro system. Transp. Res. E 148, 102274.
Zhang, W., Xu, M., Wang, S., 2023. Joint location and pricing optimization of self-service in urban logistics considering customers' choice behavior. Transp. Res. E 174, 103128.
Zhou, R.u., Cui, Y., Wang, Y., Jiang, J., 2021. A modified social force model with different categories of pedestrians for subway station evacuation. Tunn. Undergr. Space Technol. 110, 103837.
Zhou, M., Dong, H., Wang, F.Y., Zhao, Y., Gao, S., Ning, B., 2019. Field observations and modeling of waiting pedestrian at subway platform. Inf. Sci. 504, 136–160.