Received 21 August 2024
Revised 23 September 2024
Accepted 23 September 2024

Research on passenger flow control at metro transfer stations based on real-time flow calculation of streamlines

Bin Lei and Zhuoxing Hou
School of Urban Development and Modern Transportation, Xi'an University of Architecture and Technology, Xi'an, China
Yifei Suo
School of Transportation Engineering, Chang'an University, Xi'an, China, and
Wei Liu, Linlin Luo and Dongbo Lei
Xi'an Metro Operations Weihe Depot, Xi'an Rail Transit Group Co., Ltd. Operations Branch, Xi'an, China

Abstract

Purpose – The volume of passenger traffic at metro transfer stations serves as a pivotal metric for the orchestration of crowd flow management. Given the intricacies of crowd dynamics within these stations and the recurrent instances of substantial passenger influxes, a methodology predicated on stochastic processes and the principle of user equilibrium is introduced to facilitate real-time traffic flow estimation within transfer station streamlines.

Design/methodology/approach – The synthesis of stochastic process theory with streamline analysis engenders a probabilistic model of intra-station pedestrian traffic dynamics. Leveraging real-time passenger flow data procured from monitoring systems within the transfer station, a gradient descent optimization technique is employed to minimize the cost function, thereby deducing the dynamic distribution of categorized passenger flows. Subsequently, adhering to the tenets of user equilibrium, the Frank–Wolfe algorithm is implemented to allocate the intra-station categorized passenger flows across various streamlines, ascertaining the traffic volume for each.

Findings – Utilizing the Xiaozhai Station of the Xi'an Metro as a case study, the Anylogic simulation software is engaged to emulate the intra-station crowd dynamics, thereby substantiating the efficacy of the proposed passenger flow estimation model. The derived solutions are instrumental in formulating a crowd control strategy for Xiaozhai Station during the peak interval from 17:30 to 18:00 on a designated day, yielding crowd management interventions that offer insights for the orchestration of passenger flow and operational governance within metro stations.

Originality/value – The construction of an estimation methodology for the real-time streamline traffic flow augments the model's dataset, supplanting estimated values derived from surveys or historical datasets with real-time computed traffic data, thereby enhancing the precision and immediacy of crowd flow management within metro stations.

Keywords Metro transfer station, Passenger flow control, Flow streamline, Stochastic process, User equilibrium

Paper type Research paper

Shaanxi Province Department of Science and Technology Social Development Field Project (2021SF-486); Shaanxi Province Transportation Science and Technology Project (20-05R).

1. Introduction

With the rapid development of cities and a significant increase in population size, the demand for travel by residents continues to rise. Subways have gradually become one of the important means to solve urban traffic congestion problems. An increasing number of cities are actively responding to the construction and operation of subways in the hope of improving urban traffic conditions. Compared to road traffic, subways have a stronger ability to resist interference. However, when the passenger flow at a station exceeds the emergency safety management range, passenger flow management becomes more complex, leading to a sharp increase in the risk of safety accidents. Transfer stations, as key intersection nodes in the subway network, bear more passenger flow than ordinary stations. Due to the passengers transferring different lines at transfer stations, there are large numbers of passenger flows during peak hours or special circumstances, such as holidays and large-scale events (China Association of Metros, 2024). The convergence and flow of large numbers of passengers during peak hours or in special circumstances (e.g. holidays, large-scale events) place tremendous pressure on passenger management at metro transfer stations. This research is critical in addressing these challenges by providing real-time solutions for managing passenger flow efficiently, ultimately reducing safety risks and improving the travel experience for passengers. Real-time and dynamic streamline passenger flow can ensure that on the basis of fully understanding the internal passenger flow status of the station, reasonable and efficient passenger flow control strategies are proposed to ensure passenger safety. Therefore, the study of the distribution of streamlines within metros stations and real-time streamline passenger flow is of great significance for the research on passenger flow control.

Due to the highly intricate internal streamlines within transfer stations, these streamlines interweave, creating a complex network structure that poses significant challenges for passenger flow monitoring. The use of video recognition technology for real-time statistical classification of passenger flow is particularly difficult to implement in transfer stations due to the complexity of the streamlines, resulting in higher costs. Based on data from the Automatic Fare Collection (AFC) system, real-time entry and exit passenger flow information in transfer stations can be obtained, while transfer passenger flow information can only be deduced after passengers exit the station based on entry and exit information, making it difficult to acquire in real time (Zhang, Tian, & Hou, 2023). Therefore, relying on a single method of passenger flow monitoring is insufficient to meet the demand for obtaining real-time, dynamic streamline passenger flow data.Existing research on the real-time estimation of passenger flow within transfer stations primarily relies on historical data regarding the proportion of passengers choosing various streamlines to allocate real-time entry passenger flow. However, this method also has certain limitations and cannot fully describe the dynamics of actual passenger flow characteristics, which implies that under certain specific conditions, such as emergencies or adverse weather, the accuracy of this method may be significantly affected. Therefore, in order to effectively address the passenger flow control issues in metro stations, there is an urgent need for a viable method for real-time estimation of streamline passenger flow in transfer stations, based on which the design of passenger flow control methods for transfer stations can be formulated.

2. Literature review

In the field of research related to the estimation of passenger flow within metro stations, the current sources of metro station passenger flow data can be primarily categorized into three directions: short-term subway passenger flow forecasting (time series, neural networks, evolutionary analysis, etc.), acquisition of real-time dynamic passenger flow data (video detection, image processing, WiFi probes, etc.), and the estimation of real-time dynamic passenger flow data based on monitoring equipment. Pan (2011) conducted an in-depth study on the application of time series in short-term passenger flow forecasting using SAS analytical software. Ma, Li, Zhu, and Tian (2020) found that the extreme selection of time granularity does not directly enhance the effectiveness of short-term passenger flow forecasting. Li, Ni, Sun, and Lv (2020) constructed a short-term passenger flow forecasting framework characterized by the temporal dependency, spatial correlation, and randomness of external factors of passenger flow. Zhang (2011) used image processing technology to accurately identify large passenger flows, achieving real-time dynamic monitoring and display through image technology. Shi, Che, Li, and He (2019) integrated Time-of-Flight (TOF) imaging technology, binocular stereo vision technology, and intelligent video analysis technology; Yu, Yang, and Li (2021) employed various passenger flow monitoring technologies such as infrared sensing, video recognition, and WiFi sniffing; Sun (2012) mined IC card data and established a real-time arrival number estimation model for the platform based on the distribution pattern of passengers' walking time within the station. Chen (2018) utilized real-time OD (Origin-Destination) estimation data, in conjunction with the theory of stochastic service systems, to estimate the passenger flow of stations and lines in real time. (Shi, Tu, & Zhao, 2022) optimized the train operation plan within a special network layout, specifically a city rail transit corridor with a dead-end terminal yard, by decomposing the problem into two sub-problems: train timetable optimization and locomotive circulation optimization. (Li, Xue, Shao, Zhu, & Liu, 2023) conducted a comprehensive analysis of the current state of global railway intelligent digital transformation, focusing on the characteristics and applications of smart digital transformation technologies. (Wen, Bai, Zhang, Chen, & Li, 2023) developed a mixed-integer nonlinear programming (MINLP) model for optimizing the last subway train schedule, considering constraints such as the maximum and minimum headway and the latest operational end time. (Wei, Yang, Xu, & Shi, 2023) addressing the autoregressive nature of daily passenger volume in high-speed railways, proposed a dual-layer parallel wavelet neural network (DLP-WNN) model for medium-term (around 120 days) forecasts of daily passenger volume.

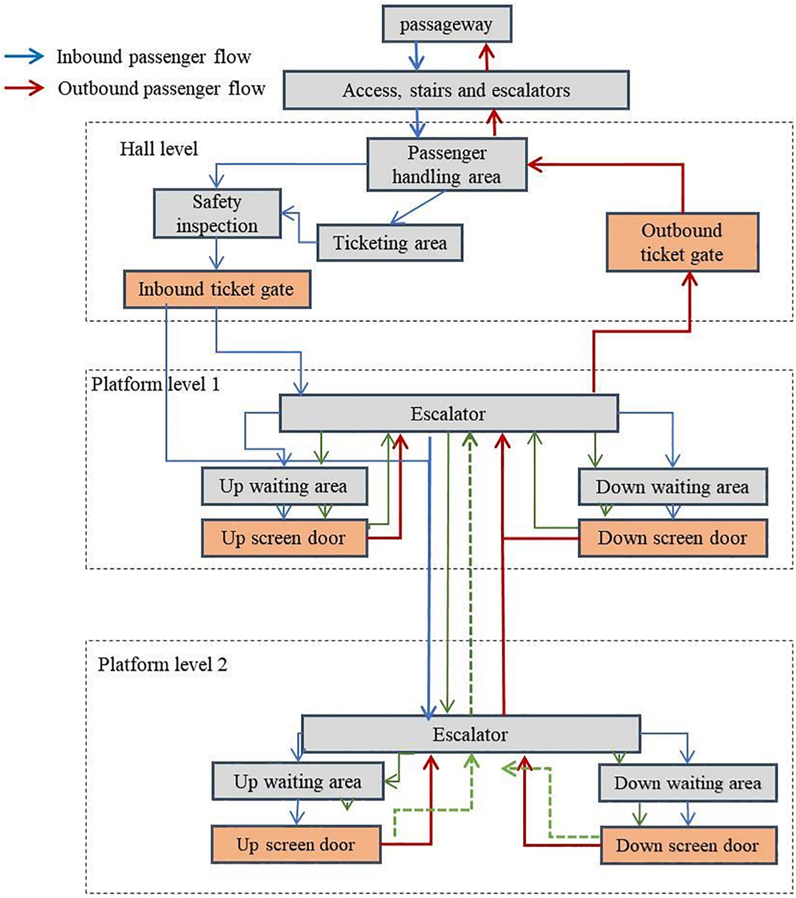
The synthesis of the aforementioned research findings indicates that current studies predominantly rely on historical data or field survey data to estimate the proportion of passengers choosing various streamlines within the station, thereby calculating the current types of passenger flow data (Jiang, Fan, Liu, Zhu, & Gu, 2018). However, the characteristics of actual passenger flow do not entirely conform to the patterns of historical data, with the presence of some random and unexpected circumstances. Besides, traditional estimation methods based on historical data may have limitations when dealing with the dynamic changes of passenger flow data. Therefore, there is a need for in-depth research into a new method capable of accurately and in real-time capturing and analyzing the characteristics of passenger flow data. Consequently, this paper utilizes real-time passenger flow monitoring data to propose a real-time flow method for transfer station streamlines based on stochastic processes and user equilibrium theory. Firstly, under real-time passenger flow monitoring conditions, by streamline analysis and stochastic process theory, the paper employs the gradient descent method to deduce passenger destination data that are difficult to directly monitor, using readily available real-time passenger flow data. Secondly, based on the user equilibrium theory, dynamic allocation of passenger flow for the same origin and destination is conducted, combining the BPR (Bayesian Personalized Ranking) function to describe passengers' path selection behavior, and employing the Frank-Wolfe algorithm to predict the passenger flow through various streamlines and facilities within the station. Finally, the method is validated with a case study at the Xiaozhai Station of Xi'an Metro.

3. Real-time passenger flow classification and estimation model

3.1 Selection of data sources

At present, the primary methods for acquiring real-time dynamic passenger flow data within metro stations include AFC system gate data, video detection, image processing, and WiFi probes. The AFC data has a broad scope of application, comprehensive data coverage, and strong reliability, but the data is only obtained after passengers complete their entire journey, resulting in poor timeliness (Li, 2019). The application range of video monitoring for measuring passenger flow is limited, but it offers better immediacy. Therefore, these two types of data can be integrated to solve the problem together, avoiding the shortcomings of using a single data source. Passenger flow monitoring data during the exit process is more detailed and easily obtained. Consequently, a passenger flow model for the exit process can be established, using AFC data obtained from exit gates combined with disembarking passenger data to estimate the transfer ratio and exit ratio of disembarking passengers.

Based on the aforementioned analysis, the locations of the passenger flow monitoring points utilized in this study are depicted in Figure 1, and the data sources and types are presented in Table 1.



Source(s): Authors' own work

Figure 1. Schematic diagram of passenger flow monitoring point selection

Monitoring point location	Detection methods	Monitored passenger flow data	Monitored data details
Entry ticket barriers	AFC system	Inbound passenger flow	The real-time entry passenger count is monitored, but it is not possible to determine immediately which line or direction the entering passengers are heading toward
Egress ticket barriers	AFC system	Outbound passenger flow Passenger flow origin	By combining exit data with entry information, the system can track which line and direction passengers used before exiting
Platform screen doors	Video surveillance	Boarding passenger flow Alighting passenger flow	Boarding passenger flow consists of both new entries and transferring passengers, while alighting passenger flow is divided into passengers exiting the station and those transferring to other lines

Source(s): Authors' own work

Table 1.
Summary table of passenger flow data monitoring points

3.2 Stochastic processes theory

In the field of traffic engineering, the translation with an academic tone could be: stochastic theory is primarily utilized to delineate data in real life and the stochastic phenomena observed in the objective world. It encompasses the collection, organization, and depiction of data, along with an in-depth analysis, focusing on characterizing the probabilities of specific events occurring. Subsequently, it proposes optimization recommendations based on these probabilities (Che, 2020).

In the routine operations of metro stations, the spatial distribution of passengers is characterized by uncertainty, thereby constituting a stochastic variable in and of themselves. Furthermore, the flow state of passengers within the station exhibits distinct stochastic characteristics, most notably the time passengers spend walking within the station, and this parameter is also considered as a stochastic variable. As these parameters evolve over time, from a statistical perspective, they can be classified as stochastic variables.

To thoroughly analyze passenger behavior patterns, the theory of stochastic processes is adopted as an investigative instrument. It is postulated that T is an infinite set of real numbers, and the collection of infinitely many random variables dependent on the parameter $t \in T$ is termed a stochastic process, denoted as $\{X(t), t \in T\}$.

The walking time of alighting passengers reaching the exit ticket gates is governed by an exponential distribution with a parameter of λ, where $\lambda > 0$ is a constant, denoted as $X \sim E(\lambda)$. The exponential distribution represents the time intervals between independent random events. The probability density function of the exponential distribution is Formula (1).

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x > 0 \\ 0, x \leq 0 \end{cases} \tag{1}$$

The probability that alighting passengers arrive within the expected time follows a Poisson distribution with a parameter of λ. This distribution's core parameter represents the average frequency of random events per unit of time. The mathematical expectation and variance of this distribution are consistent, both equal to this key parameter. The Probability Mass Function (PMF) of the Poisson distribution is given by Formula (2):

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \tag{2}$$

3.3 Model building

By examining the relationships among various types of passenger flows during the entry and exit processes, a real-time passenger flow estimation model is established. Based on real-time entry and exit AFC data and real-time alighting passenger flow data, the model estimates the inflow, outflow, and transfer passenger flows. As analyzed previously, the passenger flow data that cannot be directly obtained includes the transfer passenger flow in each direction and the flow of passengers boarding trains in various line directions. These can be transformed into the calculation of two unknowns: the transfer ratio and the entry destination ratio.

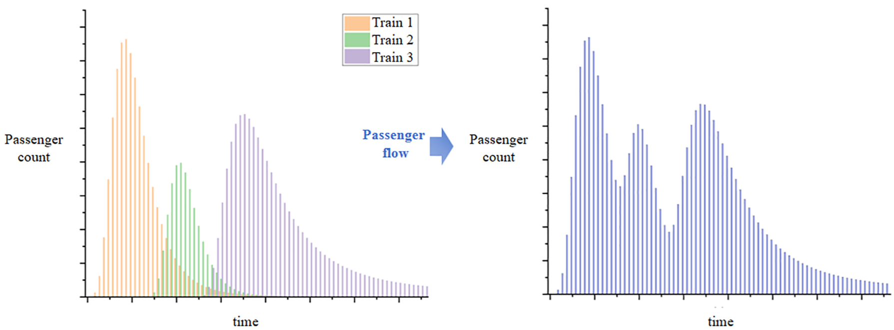
The alighting moments of passengers disembarking from the same train are approximated as the moments when the platform screen doors open. This paper takes the transfer station where two lines intersect as an example. Each line can be further divided into the upstream direction and the downstream direction, denoted as $R = \{1, 2, 3, 4\}$, specifically the upstream direction of Line 1, the downstream direction of Line 1, the upstream direction of Line 2, and the downstream direction of Line 2.

3.3.1 Outbound passenger flow estimation. Taking the upstream direction of Line 1 as an example, let the proportion of passengers who choose to exit the station after alighting be denoted as $B_1(t)$, $B_1(t) \in [0, 1]$ and those who choose to transfer be denoted as $1 - B_1(t)$.

During a certain metro station operating period, the actual number of passengers exiting the station at the exit gates during time interval $[a_j, b_j]$ is recorded as $K_1(j)$ and the estimated number of exiting passengers is denoted as $G_1(a_j, b_j), j = 1, 2, \cdots, m$, where there are n train arrivals, thus $b = a + n\Delta t$.

At the exit gates during time interval $[a_j, b_j]$, the exiting passengers may come from different trains, and the arrival times of these trains are recorded as t_{ij}, $i = 1, 2, \cdots, n$. The passenger count is the sum of the exiting passengers from each train, as shown in Figure 2.

When the train's arrival time at the station is later than the end of the estimated time interval b_j recorded by the exit gate statistics, passengers alighting from this train cannot reach the exit gate during the estimated time interval $[a_j, b_j]$. Therefore, the subscript i of t_{ij} needs to be traced back from time b, denoted sequentially from the end to the beginning as $t_{1j}, t_{2j}, \cdots, t_{nj}$, as shown in the Figure 3.



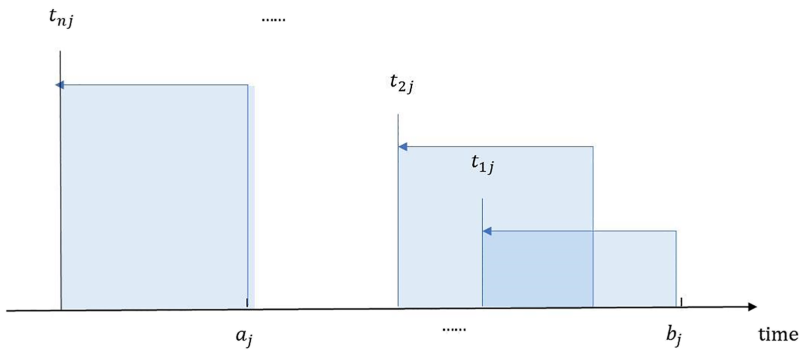
Source(s): Authors' own work

Figure 2. Schematic diagram of passenger flow superposition

The value of n should be ensured such that the probability of passengers alighting before time t_n, after walking within the station and exiting, falls within the time interval $[a_j, b_j]$ is sufficiently small to be negligible.

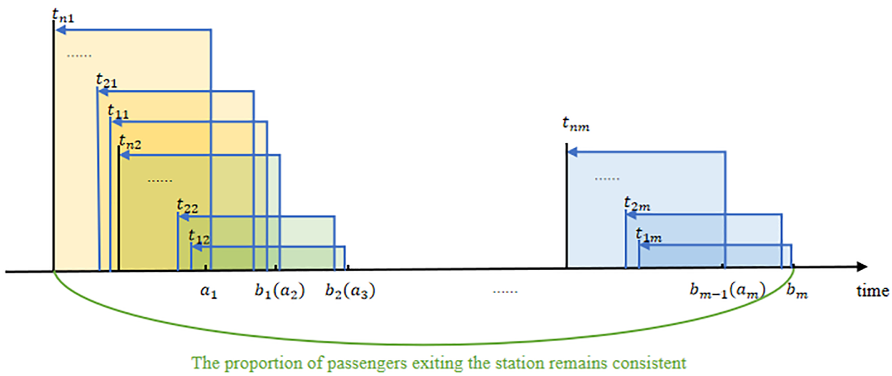
To enhance computational efficiency and ensure the timeliness of the estimation, it is assumed that the proportion of alighting passengers who exit the station within a certain period, denoted as period $[t_{1n}, b_m]$, remains constant, represented by the variable $B_1(t_i)$, as illustrated in Figure 4. The determination of the value m requires empirical setting to ensure that there is a sufficient amount of data while also maintaining an insignificant variation in the proportion of passengers exiting the station during this time frame.

During the operational period of a certain metro station, within a specific time frame, the estimated number of passengers exiting for a particular line, denoted as $G_1(a_j, b_j)$, is approximately equal to the sum of the product of the alighting passenger counts, the proportion of alighting passengers, and the probability of passengers reaching the fare gates for trains 1 to n that have arrived during that period. The time at which passengers alight should be greater than time point $a - t_i$ and not exceed time point $b - t_i$; otherwise, passengers who alighted would not have exited the station within the time interval $[a, b]$. The probability within this interval can be calculated using $\sum_{t=a-t_i}^{b-t_i} P(T = t)$.



Source(s): Authors' own work

Figure 3.
Schematic diagram of outbound passengers' alighting times recorded during the time interval



Source(s): Authors' own work

Figure 4.
A schematic diagram of the data required for a single estimation process

Therefore, the estimated number of passengers exiting for this line within the time interval $[a_j, b_j]$, denoted as $G_1(a_j, b_j)$, is expressed as Formula (3):

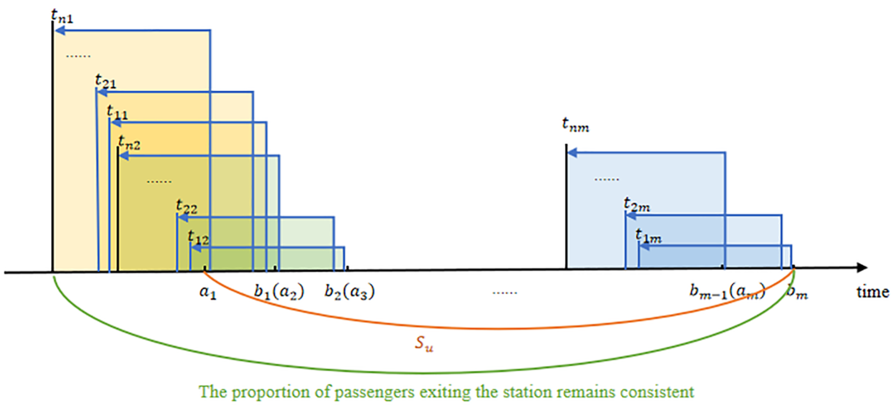
$$G_1(a_j, b_j) \approx \sum_{i=1}^{n} V_1(t_i, t_i + \Delta t) \cdot B_1(t_j) \cdot \sum_{t=a-t_i}^{b-t_i} P(T = t) \tag{3}$$

In the formula, $V_1(t_i, t_i + \Delta t)$ represents the volume of passengers alighting within the time interval $[t_i, t_i + \Delta t]$, and $B_1(t_j)$ denotes the proportion of passengers alighting in the upstream direction of Line 1 who choose to exit the station.

Considering the proportion of exiting passengers $B_1(t_j)$ as the independent variable, the dependent variable $G_1(a_j, b_j)$ exhibits a linear relationship with the independent variable $P(t_j)$. The value of $G_1(a_j, b_j)$ is solely related to the parameter j, and can be simplified to Formula (4):

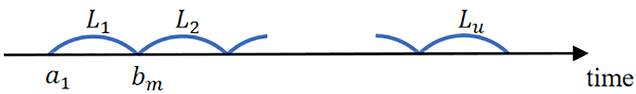
$$G_1(a_j, b_j) \approx A_j \cdot B_1(t_j) \tag{4}$$

To enhance the accuracy of the estimated proportion of exiting passengers denoted as $B_1(t_j)$, optimization is performed based on the actual number of exiting passengers $K_1(j)$. To ensure the uniqueness of the estimation results, it is necessary to guarantee that there is only one result within the same time interval. The estimated number of passengers exiting in each time interval $[a_1, b_m]$ is denoted as L_u, $u = 1, 2, \cdots, r$, with L_u corresponding to time intervals that are contiguous, as shown in Figures 5 and 6. According to the method of least squares, the loss function between the estimated number of exiting passengers and the actual number of exiting passengers within the time interval $[a_j, b_j]$ is obtained as Formula (5):



Source(s): Authors' own work

Figure 5. The temporal granularity determination of the model



Source(s): Authors' own work

Figure 6. Uniqueness determination of the estimation results

$$L_u = \sum_{i=1}^{n} [G(a_j, b_j) - K(j)]^2 \tag{5}$$

The loss function is reducible to a quadratic function of the independent variable $B_1(t_j)$, with the coefficient of the quadratic term being strictly positive, thus ensuring the existence of a minimum value for this loss function. To guarantee that the error L_u is minimized within the u time interval, as shown in the Formula (6):

$$\min L_u = \sum_{i=1}^{n} [G_1(a_j, b_j) - K_1(j)]^2 \ \min L_u = \sum_{i=1}^{n} [G_1(a_j, b_j) - K_1(j)]^2 \tag{6}$$

3.3.2 Inference of inbound passenger flow destinations. By the same token, the inbound passenger flow will be diverted into the following directions: Line 1 upstream, Line 1 downstream, Line 2 upstream, and Line 2 downstream. Taking the inbound flow heading towards the upstream direction of Line 1 as an example, the proportion of passengers heading towards the upstream of Line 1 is $H_1(t_j)$. The sum of the inbound passenger flow heading towards the upstream of Line 1 and the passenger flow transferring from the upstream and downstream of Line 2 to Line 1 can be approximately equal to the boarding passenger flow during this period. It can be expressed as Formula (7):

$$S(a_j, b_j) \approx \sum_{e=1}^{s} E(t_e, t_e + \Delta t) \cdot H(t_j) \cdot \sum_{t=a-t_i}^{b-t_i} P(T = t) + \sum G(a_j, b_j) \cdot B(t_j) \tag{7}$$

From this, the proportion $H_1(t_j)$ of passengers entering the station heading towards the upstream direction of Line 1 can be calculated.

3.4 Algorithm design

The determination of the proportion of passengers exiting the station $B_1(t_j)$ can be formulated as an optimization problem that minimizes a loss function. The gradient descent method is employed to iteratively refine the solution. During each iteration, the algorithm updates the parameters based on the direction of the gradient, progressively honing in on the local minimum of the function. Through successive iterations, the parameters are adjusted to drive the loss function towards its minimal value, thus facilitating the precise computation of the exiting passenger ratio.

During the determination of the exiting passenger ratio, the issue can be reformulated as an unconstrained optimization problem, focusing on identifying the parameter values that minimize the loss function. The gradient descent algorithm operates by computing the gradient of the loss function relative to the parameters and iteratively updating in the direction antithetical to the gradient, thereby progressively converging on the optimal solution. This methodology offers not only enhanced computational efficiency but is also adept at managing extensive datasets, rendering it appropriate for addressing real-world problems, including the calculation of the ratio of passengers exiting the station.

In the application of the gradient descent method to determine the ratio of passengers exiting the station, it is imperative to choose a suitable loss function and an appropriate step size for iteration to guarantee the convergence and stability of the algorithmic process.

4. Passenger flow distribution methods within metro stations

4.1 User equilibrium theory

Passenger flow within a metro station can primarily be categorized into three major classes: in-station flow, out-station flow, and transfer flow. These distinct categories of passengers utilize various facilities within the station in a specific sequence, with each passenger group interacting with and using different equipment and facilities. To more clearly delineate the activity patterns of passengers within the station, a network diagram is employed for description. In this network diagram, each node represents the various facilities within the station, while directed arcs illustrate the sequence of passengers' movement between these node facilities.

Taking a specific transfer station as an example, the equipment-related network diagram is depicted as shown in Figure 7.

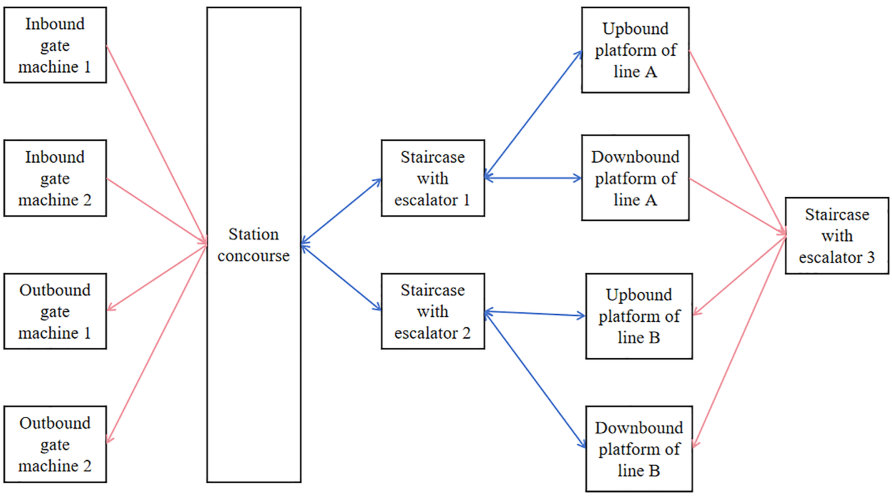
The User Equilibrium theory posits that each traveler has complete information and rational decision-making capabilities, enabling them to independently choose the shortest travel path based on their needs and traffic conditions. The core tenet of this theory is that when individuals within a transportation network act to maximize their own benefits, seeking to minimize travel time and making independent route choices, the entire network system evolves into an equilibrium state of traffic flow distribution (Zhang, Li, Liu, & Chen, 2019).

The User Equilibrium theory can be represented by the following formulas:

$$\min z(x) = \sum_{a \in A} \int_0^{x_a} t_a(x)dx \tag{8}$$

$$q^\omega = \sum_{k \in K^\omega} f_k^\omega, \forall \omega \in W \tag{9}$$

$$x_a = \sum_{\omega} \sum_{k \in K^\omega} \delta_{ak}^\omega f_k^\omega \tag{10}$$



Source(s): Authors' own work

Figure 7.
Equipment association network diagram

$$f_k^\omega \geq 0, \forall k \in K^\omega, \omega \in W \tag{11}$$

In the Formula (8), x_a represents the traffic flow on road segment a, and f_k^ω represents the traffic flow on path k of OD pair ω; the travel cost function for the links is denoted as $t_a(x)$, the set of all road segments is represented by A, and K^ω is the set of all paths for the OD pair ω; the distributed traffic volume for OD pair ω is represented by q^ω, and W is the set of all OD pairs. δ_{ak}^ω is the connectivity variable, where if path k includes road segment a, then $\delta_{ak}^\omega = 1$, otherwise $\delta_{ak}^\omega = 0$.

4.2 Quantification of passenger node selection behavior

Passengers within a station tend to prefer service nodes that are closer in distance and have a higher service quality to meet their service needs. In order to maximize the satisfaction of their individual service requirements. This behavior reflects the dual considerations of efficiency and quality in the selection process by passengers. Accurately capturing and reflecting the characteristics of this selection behavior is particularly important, as it is crucial for enhancing the service quality of the station and optimizing the passenger's travel experience.

The BPR function, a Bayesian personalized ranking function, relies on the extensive field traffic survey data from the Federal Highway Administration in the United States. It aims to accurately quantify the traffic impedance encountered by vehicles during their travel on road sections, thereby objectively assessing the smoothness of vehicle travel on those sections.

It can be expressed by Formula (12):

$$T_i(x) = T_{i0} \cdot \left[1 + \alpha\left(\frac{x}{C}\right)^\beta\right] \tag{12}$$

In the formula, $T_i(x)$ represents the travel time on road section i; T_{i0} represents the free-flow travel time on road section i; x represents the total traffic flow on the road section; C represents the actual capacity of the road section; α, β is the parameter within the function. The Federal Highway Administration of the United States provides reference values based on the traffic conditions in the United States, which are $\alpha = 0.15$ and $\beta = 4$, respectively.

This study refers to the calibration parameter values of the BPR function for various spaces within the metro station as presented in reference (Yuan, 2022), with the values listed in Table 2.

4.3 Algorithm design

This paper employs the Frank-Wolfe algorithm to solve the user equilibrium assignment model. By computation, it obtains the specific distribution of the total passenger flow at each node and between flow lines on the equipment-related network, providing strong support for passenger flow management and optimization at the station. This algorithm can simplify complex nonlinear programming problems into a series of linear programming problems, allocating flow to the most cost-effective paths with the lowest travel costs (Bliemer & Raadsen, 2020). Finally, the cost is calculated based on the overall traffic volume of the road

Table 2. Parameter values of the BPR function for various spaces within the subway station

Type of space	α	β
Passageway	0.079	1.79
Platform	0.17	1.42

Source(s): Authors' own work

sections. The determination of the feasible descent direction can be based on the difference between the results of the sub-process and the current process solutions.

5. Case study validation

5.1 Field investigation at Xiaozhai station

Taking the Xiaozhai Transfer Station of Xi'an Metro as an example, this paper conducts a case study analysis. The transfer node stairs at Xiaozhai Station are organized in a downward direction from the node, forming a "T"-shaped transfer between Line 2 and Line 3.

The passenger flow data for a certain period at Xiaozhai Station is shown in Table 3.

By analyzing the aforementioned data, it can be observed that, with the exception of special reasons on certain dates, the passenger flow at Xiaozhai Station far exceeds the designed capacity of the station. For most periods, it operates under a state of high passenger flow.

Using the passenger flow data from a specific day in 2019 at Xiaozhai Station as a reference, the simulation process's passenger flow is set, and a passenger flow control plan is formulated based on the peak flow data. In the Xiaozhai Station case study, simulation parameters were selected based on real-world data, such as the number of passengers entering and exiting the station. The peak flow data from the AFC system for the highest traffic period (17:30–18:00) was used to set simulation conditions, and AnyLogic simulation software was employed to model crowd behavior based on this empirical data. The selection of these parameters ensures the accuracy and applicability of the model. AFC data is collected in 30-min intervals, revealing that the highest passenger flow of the day occurred between 17:30 and 18:00, with 3,162 people entering on Line 2 and 2,105 people entering on Line 3, 3,103 people exiting on Line 2, and 971 people exiting on Line 3. Through on-site surveys at Xiaozhai Station, it was found that the ratio of male to female passengers is roughly equal, showing a relatively balanced distribution. Detailed statistics of passengers of different age groups were conducted to facilitate further analysis of the characteristics of the passenger composition, with specific data shown in Table 4.

5.2 Model establishment

This study employs simulation modeling with AnyLogic to generate a dataset for validation. This approach effectively substantiates the scientific validity of the calculation method. The simulation steps are illustrated in Figure 8.

Date	Entry	Exit	Daily transfer volume	Passenger volume
1	6,941	7,863	15,213	30,017
2	9,292	10,082	28,788	48,162
3	10,478	11,611	34,667	58,758
4	11,761	12,808	39,197	63,766
. . .	. . .	. . .	. . .	. . .
30	31,283	33,705	102,217	167,205
31	32,636	35,808	102,666	170,900

Source(s): Authors' own work

Table 3. Passenger flow data for a specific period at the Xiaozhai transfer station

Passenger types	Children	Middle-aged youth	Elderly
Proportions (%)	7.3	78.2	14.6

Source(s): Authors' own work

Table 4. Proportions of passenger types by age group at Xiaozhai station

5.3 Data set acquisition

To facilitate data acquisition and subsequent model analysis, data can be programmatically written into an Excel spreadsheet to create a simulation data set. After running the simulation model, a total of 4,038 data entries were obtained.

Statistical analysis of the data set can reveal the central tendency, variability, and distribution patterns of the data. The statistical results are presented in Table 5.

Visualizing the numerical variables of the data set through scatter plots, as shown in Figure 9, can assist in identifying the relationships between variables. It is observed that the constructed data set meets the requirements for the randomness of passenger walking behavior as stipulated in the study.

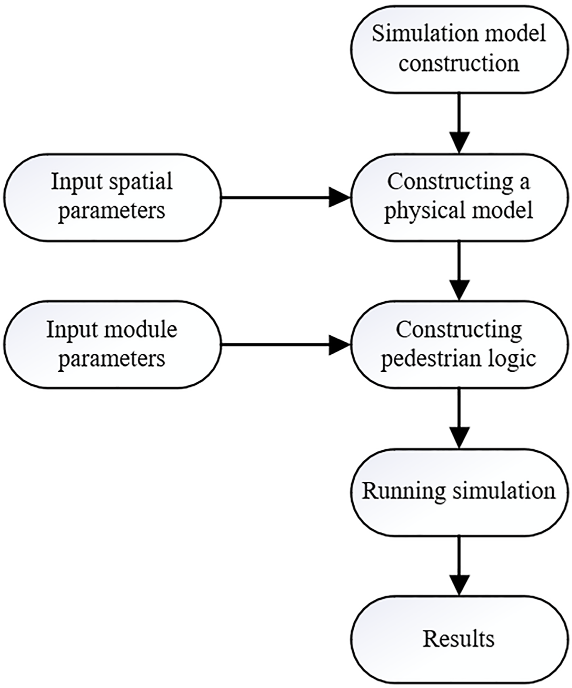
Based on the findings from the survey, the headway of subway trains is set at four minutes. The data set has been statistically analyzed with a time interval of four minutes, and the results are presented in Table 6.

5.4 Model calculation results

5.4.1 Estimation results of real-time passenger flow.

(1) Parameter settings

The model employs gradient descent to iteratively calculate key parameters such as passenger flow rate, transfer ratios, and time intervals. Based on field research at Xiaozhai Station, it is assumed that all passengers can exit the station within 12 minutes after disembarking.



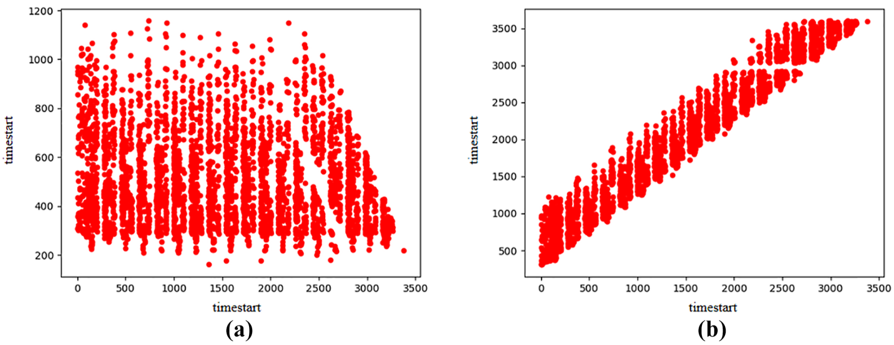
Source(s): Authors' own work

Figure 8. Simulation steps generated by AnyLogic

	Time start	Time end	Time stay
Count	4037.000000	4037.000000	4037.000000
Mean	1489.622530	1999.161283	509.538753
Std	915.878959	912.558515	187.618594
Min	0.282158	310.800000	160.914719
25%	687.397509	1224.000000	360.347565
50%	1448.084752	1965.900000	458.555483
75%	2265.861354	2740.200000	632.825990
Max	3380.590083	3599.700000	1157.314246

Source(s): Authors' own work

Table 5. Statistical results of the dataset



Source(s): Authors' own work

Figure 9. Variable distribution scatter plot

Time period	Number of disembarking passengers	Number of passengers exiting the station	Number of passengers transferring
[0,4]	396	195	201
(4,8]	352	165	187
(8,12]	287	149	138
(12,16]	283	146	137
(16,20]	409	214	195
(20,24]	279	129	150
(24,28]	310	145	165
(28,32]	352	176	176
(32,36]	261	140	121
(36,40]	273	142	131
(40,44]	257	110	147
(44,48]	284	149	135
(48,52]	198	134	64
(52,56]	92	66	26
(56,60]	4	3	1

Source(s): Authors' own work

Table 6. Passenger flow statistics for various time periods

The obtained data set spans a period of 60 minutes, and historical data from half an hour is used to estimate passenger flow data for the next five minutes, resulting in a total of eight sets of outcomes.

(2) Model evaluation metrics

To evaluate the performance of the passenger flow classification estimation model proposed in this study, the following evaluation metrics are employed to quantify the model's performance:

Mean Absolute Error (MAE): It measures the average of the absolute differences between predicted and actual values. The smaller the MAE, the higher the predictive accuracy of the model. The calculation formula is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i'| \tag{13}$$

where, n represents the number of samples; y_i denotes the actual observed values; y_i' indicates the predicted values; $|y_i - y_i'|$ is the absolute value of the difference between the actual and predicted values.

The R-squared (R^2) score: it reflects the extent to which the model explains the variation in the actual data. The calculation formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i')^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{14}$$

where, n denotes the number of samples; y_i represents the actual observed values; y_i' signifies the predicted values; $\sum_{i=1}^{n}(y_i - y_i')^2$ is the Residual Sum of Squares (RSS); and $\sum_{i=1}^{n}(y_i - \bar{y})^2$ is the Total Sum of Squares (TSS). The real-time passenger flow is estimated using the gradient descent method implemented on the Python platform. Some of the computational results and the evaluation metrics are illustrated in Table 7 and Figure 10, with MAE = 23.44 and $R^2 = 0.72$. The comparison between the estimated and actual values is depicted, where the solid line indicates the actual values and the dashed line represents the estimated values.

The MAE effectively assesses the average absolute discrepancy between predicted and actual values. It is calculated to be approximately 23.44 persons, indicating the overall average deviation of the predicted values from the actual values. This metric provides a quantitative basis for the accuracy of predictions and offers a more comprehensive assessment of the model's performance.

The R^2 score is an indicator that measures the model's ability to explain the variability in the data. With an R^2 score of 0.72, it suggests that the model accounts for about 72% of the data variability, capturing the fluctuations in the data to a considerable extent.

5.4.2 Calculation results of passenger flow allocation model.

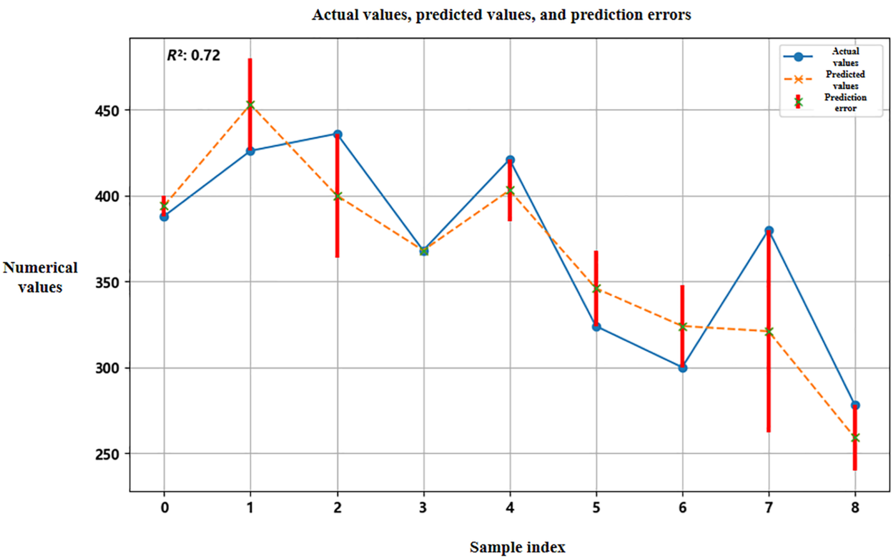
(1) Data input

Based on the results obtained from the passenger flow classification estimation, the user equilibrium theory is employed to allocate the passenger flow of the same Origin-Destination (OD) pairs to various flow lines. Due to the absence of transfer passenger flow

Sample index	Time period	Actual values	Estimated values (rounded)
0	(15,20]	388	394
1	(20,25]	426	453
2	(25,30]	436	400
3	(30,35]	368	368
4	(35,40]	421	403
5	(40,45]	324	346
6	(45,50]	300	324
7	(50,55]	380	321
8	(55,60]	278	259

Source(s): Authors' own work

Table 7. Real-time passenger flow calculation results



Source(s): Authors' own work

Figure 10. Comparison chart of actual values, predicted values, and prediction errors

data in the data from a certain day in 2019, the transfer ratio is set according to the calculation results of 4.1.

(2) Correspondence between station facilities and flow lines

Multiple passenger flow lines may converge on a single piece of equipment. By calculating the number of passengers passing through the converging equipment, the passenger flow on each flow line can be summarized. This allows for a clear expression of the interrelationships between various station facilities. This section will combine the inflow and outflow, as well as transfer passenger flow lines, with the connection to station facilities to study the passenger flow through each facility within the station (Khalid, Baten, Nawawi, & Ishak, 2015). The network diagram of the relationship between facilities will be abstracted into a passenger flow diagram to support the research approach based on flow line passenger flow control, as shown in Figure 11.

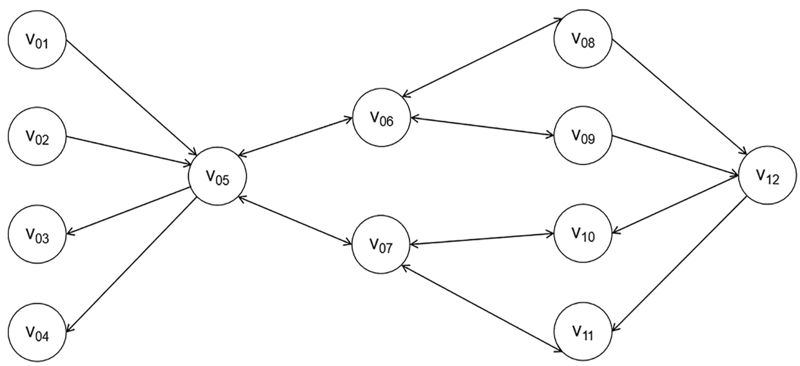
Based on the analysis of passenger flow lines and the associated facility network within the transfer station, the correspondence between equipment and flow lines can be determined, as shown in Table 8.

(3) Calculation results and analysis

The passenger flow on each flow line is calculated according to the passenger flow allocation model, with the results shown in Table 9.

5.4.3 Passenger flow control scheme. Through an in-depth analysis of the time-varying data of passenger flow into and out of Xiaozhai Station, a 30-min period is selected as a cycle for passenger flow control. Subsequently, the necessary data is input into the established passenger flow control model, which is solved using a genetic algorithm implemented in Python. Based on the solution, a passenger flow control plan for Xiaozhai Station during the period from 17:30 to 18:00 is formulated, as shown in Table 10.

Xiaozhai Station is equipped with dedicated stairs for transfers, which, although capable of meeting passengers' transfer needs during peak hours, faces significant passenger flow pressure and harbors safety hazards that cannot be ignored. Therefore, to ensure safety and order at the transfer stairs, it is imperative to arrange for station staff to be in place in a timely



Source(s): Authors' own work

Figure 11. Transfer station passenger flow diagram

Facility equipment	Inbound flow line set	Outbound flow line set	Transfer flow line set
v_{01}	{1, 2, 3, 4}	∅	∅
v_{02}	{5, 6, 7, 8}	∅	∅
v_{03}	∅	{9, 11, 13, 15}	∅
v_{04}	∅	{10, 12, 14, 16}	∅
v_{05}	{1, 2, 3, 4, 5, 6, 7, 8}	{9, 10, 11, 12, 13, 14, 15, 16}	{17, 18, 19, 20}
v_{06}	{1, 2, 5, 6}	{9, 10, 11, 12}	{17, 18, 19, 20}
v_{07}	{3, 4, 7, 8}	{13, 14, 15, 16}	{17, 18, 19, 20}
v_{08}	{1, 5}	{9, 10}	{17, 18, 21, 23}
v_{09}	{2, 6}	{11, 12}	{19, 20, 22, 24}
v_{10}	{3, 7}	{13, 14}	{17, 19, 21, 22}
v_{11}	{4, 8}	{15, 16}	{18, 20, 23, 24}
v_{12}	∅	∅	{21, 22, 23, 24}

Source(s): Authors' own work

Table 8. Station equipment interconnection table

Streamline number	Streamline type	Streamline passenger flow	Streamline equipment route
1	Inbound flow line	948	$v_{01}, v_{05}, v_{06}, v_{08}$
2		791	$v_{01}, v_{05}, v_{06}, v_{09}$
3		779	$v_{01}, v_{05}, v_{07}, v_{10}$
4		273	$v_{01}, v_{05}, v_{07}, v_{11}$
5		664	$v_{02}, v_{05}, v_{06}, v_{08}$
6		759	$v_{02}, v_{05}, v_{06}, v_{09}$
7		653	$v_{02}, v_{05}, v_{07}, v_{10}$
8		400	$v_{02}, v_{05}, v_{07}, v_{11}$
9	Outbound flow line	807	$v_{08}, v_{06}, v_{05}, v_{03}$
10		496	$v_{08}, v_{06}, v_{05}, v_{04}$
11		931	$v_{09}, v_{06}, v_{05}, v_{03}$
12		869	$v_{09}, v_{06}, v_{05}, v_{04}$
13		291	$v_{10}, v_{07}, v_{05}, v_{03}$
14		311	$v_{10}, v_{07}, v_{05}, v_{04}$
15		175	$v_{11}, v_{07}, v_{05}, v_{03}$
16		194	$v_{11}, v_{07}, v_{05}, v_{04}$
17	Transfer flow line	203	$v_{08}, v_{06}, v_{05}, v_{07}, v_{10}$
18		190	$v_{08}, v_{06}, v_{05}, v_{07}, v_{11}$
19		158	$v_{09}, v_{06}, v_{05}, v_{07}, v_{10}$
20		200	$v_{09}, v_{06}, v_{05}, v_{07}, v_{11}$
21		748	v_{10}, v_{12}, v_{08}
22		437	v_{10}, v_{12}, v_{09}
23		311	v_{11}, v_{12}, v_{08}
24		582	v_{11}, v_{12}, v_{09}

Source(s): Authors' own work

Table 9. Calculation results of Xiaozhai station passenger flow allocation

Passenger flow control node	Number of people passing through per unit time (people/min)	Passenger flow control method
Inbound Turnstile 1, Inbound Turnstile 2	82	Use crowd control barriers to set up a detour path in front of the security check equipment connected to Inbound Turnstile 1 and Inbound Turnstile 2, extending the walking time for passengers and controlling the speed of the security check
Station Concourse	53	Use crowd control barriers to set up a detour path in the station hall, extending the walking time for passengers, and take measures to release passengers in batches when necessary

Source(s): Authors' own work

Table 10. Passenger flow control scheme

manner to guide and maintain order among transferring passengers, and to strictly guard against potential safety threats such as congestion and stampede incidents.

6. Conclusion

This paper, starting from the perspective of ensuring reliability, first discusses the methods of passenger flow data collection and the selection of collection points. Due to the complexity of passenger flow in transfer stations, the proportion of passengers going to various lines after entering the station is unknown, and the proportion of passengers choosing to exit or transfer after disembarking is also unknown. Therefore, it is necessary to establish a real-time passenger flow estimation model for solution. Secondly, based on the analysis of passenger flow lines and combined with the theory of stochastic processes, this paper establishes a model for calculating the time passengers spend alighting, the service time of facilities, and the travel time of passengers, as well as the total time for exiting the station in metro stations. The model of the passenger exiting process time and the total time model are established, and various types of passenger flow estimation models are established based on the theory of stochastic processes, providing a reference for passenger flow control in urban rail transit stations. Based on the user equilibrium theory, passenger flow is allocated to various flow lines to achieve fine-grained acquisition of passenger flow data. Finally, taking the typical transfer station of Xi'an Metro - Xiaozhai Station as the research object, a case analysis of simulation is carried out. On the basis of the on-site investigation, basic data and passenger flow of the transfer station are obtained. Through AnyLogic simulation, the actual types of passenger flow that are difficult to obtain are obtained, a data set is created, and the effectiveness of the real-time passenger flow estimation model is verified. Relying on the passenger flow situation during the peak period of a certain day at Xiaozhai Station, a passenger flow control plan is formulated with a time interval of 30 minutes.

In this study, to reduce the computational complexity and ensure that the loss function has a unique minimum solution, the proportion of passengers exiting the station within a certain period is considered as a fixed value. However, in order to further refine this research outcome, efforts could be directed towards shortening the response time of the model, thereby enhancing its timeliness. Concurrently, the probability distributions utilized in the stochastic process theory within the real-time passenger flow estimation model are determined based on literature and empirical knowledge. Future research could obtain more accurate probability distributions through empirical measurements, which would help reduce errors and improve the model's precision.To deepen the findings of this paper, in the study of large passenger flow identification and control, integrating data from multiple sources could be beneficial to mitigate the potential inaccuracies arising from reliance on a single data source. This approach would contribute to a more robust and comprehensive understanding of passenger flow dynamics, ultimately enhancing the effectiveness of traffic engineering strategies in urban rail transit systems.

Reference

Bliemer, M. C. J., & Raadsen, M. P. H. (2020). Static traffic assignment with residual queues and spillback. *Transportation Research Part B: Methodological*, *132*, 303–319. doi: 10.1016/j.trb.2019.02.010.

Che, C. (2020). Metro station based on random process study on optimization of passenger flow organization. MA thesis. Chengdu: Sichuan: Southwest Jiaotong University.

Chen, Z. J. (2018). Real-time estimation model and control method for urban rail transit line load. PhD thesis. Beijing: Beijing Jiaotong University.

China Association of Metros (2024). Urban rail transit annual statistical and analytical report 2023. China Association of Metros.

Jiang, Z., Fan, W., Liu, W., Zhu, B., & Gu, J. (2018). Reinforcement learning approach for coordinated passenger inflow control of urban rail transit in peak hours. *Transportation Research Part C: Emerging Technologies*, *88*, 1–16. doi: 10.1016/j.trc.2018.01.008.

Khalid, R., Baten, M. A., Nawawi, M. K. M., & Ishak, N. (2015). Analyzing and optimizing pedestrian flow through a topological network based on M/G/C/C and network flow approaches. *Journal of Advanced Transportation*, *50*(1), 96–119. doi: 10.1002/atr.1330.

Li, Z. F., Ni, S. Q., Sun, K. Y., & Lv, M. M. (2020). Short-term passenger flow prediction of an urban rail transit based on multi-feature fusion. *Journal of Transportation Engineering and Information*, *18*(4), 10.

Li, Z. H., Tian, X. L., Li, Z. W., Zhou, W., Zhou, Z. J., & Zhong, M. H. (2019). Risk analysis of metro station passenger flow based on passenger flow patterns. *Journal of Tsinghua University* (Science and Technology), *10*, 7.

Li, P., Xue, R., Shao, S., Zhu, Y., & Liu, Y. (2023). Current state and predicted technological trends in global railway intelligent digital transformation. *Railway Sciences*, *2*(4), 397–412. doi: 10.1108/rs-10-2023-0036.

Ma, C. Q., Li, P. K., Zhu, C. H., & Tian, T. (2020). Short-term passenger flow forecast of urban rail transit based on different time granularities. *Journal of Chang'an University (Natural Science Edition)*, *40*(3), 9.

Pan, L. (2011). Research on short-term passenger flow prediction and early warning system for subway. MA thesis. Beijing: Capital University of Economics and Business.

Shi, C. L., Che, H. L., Li, Y., & He, J. Z. (2019). Study on passenger flow monitoring system of urban rail transit based on multi-data fusion. *Journal of Safety Science and Technology*, Supplement 1, 5.

Shi, F., Tu, X., & Zhao, S. (2022). Optimization method of urban rail train operational plan based on O-D time-varying demand. *Railway Sciences*, *1*(1), 148–166. doi: 10.1108/rs-04-2022-0008.

Sun, H. R. (2012). Research on real-time calculation of passenger flow on urban rail transit line. MA thesis. Beijing: Beijing Jiaotong University.

Wei, T., Yang, X., Xu, G., & Shi, F. (2023). Medium-term forecast of daily passenger volume of high speed railway based on DLP-WNN. *Railway Sciences*, *2*(1), 121–139. doi: 10.1108/rs-01-2023-0003.

Wen, F., Bai, Y., Zhang, X., Chen, Y., & Li, N. (2023). Last train timetable optimization for metro network to maximize the passenger accessibility over the end-of-service period. *Railway Sciences*, *2*(2), 273–288. doi: 10.1108/rs-03-2023-0012.

Yu, Q. H., Yang, H. H., & Li, A. S. (2021). Research on urban rail transit passenger flow monitoring system based on multi-source monitoring data fusion. *Communication and Shipping*, *8*(3), 68–72.

Yuan, Z. Z. (2022). Passengers' dynamic travel route choice modeling based on congestion index in urban rail transit networks. MA Thesis. Beijing: Beijing Jiaotong University.

Zhang, L. (2011). Technique and system about identification of large passenger flow and assessment of security status of Beijing mass transit railway. MA thesis, Beijing: Beijing Jiaotong University.

Zhang, R., Li, J. J., Liu, G. H., & Chen, S. K. (2019). Identification and optimization of resolving bottleneck for passenger flow in metro stations. *Journal of Transport Information and Safety*, *37*(2), 7.

Zhang, J., Tian, H., & Hou, X. (2023). *Frontier technology development and scientific issues in transportation and vehicle engineering (rail transit and maglev transportation systems)*. Beijing: China Communications Press.

Corresponding author
Bin Lei can be contacted at: 371503770@qq.com