

Using the unbiased grey Markov chain optimization model to forecast passenger flows through existing urban rail transit stations

Fuquan Pan^{1*}, Wenzheng Li¹, Lixia Zhang¹, Xiaoxia Yang¹, Hailiang Tang² and Yongkai Xia²

¹ School of Intelligent Transportation and Vehicle, Qingdao University of Technology, Qingdao 266520, China

² Qingdao Metro Operation Co., Ltd., Qingdao 266100, China

* Correspondence: fuquanpan@yeah.net (Pan F)

Abstract

To accurately forecast the passenger flows of existing stations in urban rail transit and provide a scientific basis for subway operation management departments, a passenger flow prediction model optimized by an unbiased grey Markov chain was constructed. Using 18 months of average daily passenger flow data from Shilaoren Bathing Beach Station on Qingdao Metro Line 2, a grey prediction model was established and tested. The data were classified into six states according to their relative errors for Markov state modeling, and the classification results were validated. A state transition probability matrix and the corresponding k-step transition probabilities were then established. The unbiased grey Markov chain prediction model was used to predict the passenger flow data of existing stations in the next 3 months. By comparing the prediction results with the actual passenger flow data, it was found that the prediction error after optimization by the unbiased grey Markov model was controlled within 10%, which corresponded well with the actual characteristics of urban rail transit passenger flows. This model is suitable for predicting the passenger flows of existing urban rail transit stations and can provide an effective reference for managing subway operations and ensuring safety.

Keywords: Passenger flow forecast, New additional stations, Grey theory, Unbiased test, Markov test

Citation: Pan F, Li W, Zhang L, Yang X, Tang H, et al. 2026. Using the unbiased grey Markov chain optimization model to forecast passenger flows through existing urban rail transit stations. *Digital Transportation and Safety* 5(1): 22–31 <https://doi.org/10.48130/dts-0026-0002>

Introduction

Urban rail transit, with its advantages of large capacity, high speed, and low pollution, has become a core aspect of modern urban public transportation systems. It can effectively alleviate ground traffic congestion and promote the transformation of the urban transportation structure towards low carbon and high efficiency. However, the growing passenger demand has posed greater challenges to rail transit operations and management, especially during holidays, when a surge in passenger flows in a short period of time can easily create operational stress. In this context, accurate passenger flow forecasting has become the key to ensuring operational safety and improving service quality. It can also provide a scientific basis for line operation management departments to optimize capacity allocation and formulate emergency plans^[1].

At present, passenger flow prediction methods are mainly divided into two categories: Traditional prediction methods and intelligent algorithms. Traditional prediction methods are based on linear assumptions and low-dimensional data processing, covering time series methods, gray model-based methods, linear regression methods, Kalman filter methods, etc. Zhou et al.^[2] combined singular spectrum analysis (SSA) with an AdaBoost-weighted extreme learning machine to establish an urban rail transit transfer passenger flow prediction model. Jiao^[3] successfully predicted passenger flows during regular periods by fitting the linear trend of historical passenger flows, based on the autoregressive integrated moving average (ARIMA) model. Wang et al.^[4] achieved short-term predictions of stable passenger flows through linear state estimation with the Kalman filter model. In addition, most models are based on clear mathematical assumptions, and the parameters' meanings are intuitive and easy to understand and explain. Carmona-Benítez et al.^[5]

optimized the damped trend grey prediction model (DTGM) using seasonal dynamic damping factors and constructed a seasonal ARIMA damped trend grey prediction model (SDTGM). The seasonal damping factor was used to directly quantify the seasonal fluctuations of passenger flows, and the model was highly interpretable. This type of method does not require large-scale sample support and can achieve low-cost and rapid predictions in scenarios where the amount of passenger flow data is small, the data fluctuations are gentle, and the linear pattern is significant.

In view of the limitations of linear modeling, existing improvements have mostly been made by adding correction factors to the traditional model and improving the Kalman filter on the basis of the maximum entropy principle. Peng et al.^[6] combined the traditional grey prediction model with the Markov chain state transfer matrix to construct a grey Markov prediction model of railway passenger flows. Li et al.^[7] constructed a grey system theory model and a grey Markov chain prediction model based on traffic accident data over several years to predict data for the next 2 years. Ye et al.^[8] proposed an adaptive grey Markov prediction model with fusion parameters, applied it to the passenger flow prediction of Chengdu Metro Line 1, and improved the accuracy of predicted passenger flows by correcting the absolute error. Ding et al.^[9] constructed a combination model with gray Grey Model (1,1) and Markov chain, and used the particle swarm algorithm to iteratively optimize it. The average error after optimization was reduced by 37%, and the prediction accuracy was significantly improved. Cai et al.^[10] improved the defect of the traditional Kalman filter, which exhibits large errors in the presence of non-Gaussian noise. However, these improvements remain local optimizations within the linear modeling framework, and they have not fundamentally overcome its structural constraints.

Intelligent algorithm methods are centered on nonlinear modeling and high-dimensional feature extraction, covering support vector regression (SVR), neural networks (Back Propagation [BP], recurrent neural networks [RNN]) and deep learning models (long short-term memory [LSTM], convolutional neural networks [CNN], graph convolutional network [GCN]). They do not require a preset data distribution and can automatically extract nonlinear features from large-scale, high-dimensional data. They are suitable for complex scenarios such as abnormal passenger flow during holidays and special events. Hu et al.^[11] used the typical support vector regression (SVR) algorithm to construct an urban rail transit passenger flow prediction model and optimized the prediction model with the improved particle swarm optimization (IPSO) algorithm. Shi et al.^[12] successfully captured the sudden fluctuations in holiday passenger flows by improving the Variational Mode Decomposition-Genetic Algorithm-Back Propagation Neural Network (VMD-GA-BP) model. Xue et al.^[13] proposed a hybrid deep neural network framework based on traditional smart card data and social media data to construct a station entry flow and social media interference model for predicting subway passenger flows prediction during special events. Mulerikkal et al.^[14] used an RNN to generate an intermediate feature space, integrated spatial features into the time series, and introduced an outlier detection and elimination algorithm based on support vector machines (SVMs) to improve the performance in predicting subway passenger flows. Bapaume et al.^[15] proposed a computer vision framework based on deep learning methods for predicting the real-time passenger flows and departure intervals of subway lines in urban transportation networks. Tu et al.^[16] integrated internet event data based on the DeepSPE model to achieve multistep passenger flow prediction during large-scale events. Yue et al.^[17] used an LSTM network to process historical transfer passenger flow data and combined it with the Transformer prediction model to predict short-term transfer passenger flows between integrated transportation hubs in urban agglomerations. In addition, some scholars have proposed a deep learning model for predicting rail transit passenger flows based on a bidirectional LSTM (BiLSTM) network that considers temporal characteristics^[18,19]. Some scholars have combined CNNs with LSTM for predicting short-term passenger flows in urban rail transit^[20,21]. Other scholars have proposed a multigraph convolutional recurrent neural network (MGC-RNN) and flow-similarity attention graph convolutional network (F-SAGCN) model that comprehensively considers various factors affecting passenger flows and is suitable for predicting short-term passenger flows in urban rail transit systems^[22,23]. Xiu et al.^[24] combined the correlation-based spatiotemporal feature selection (Cor-STFS) model for optimal input selection with the STA-PTCN-BiGRU model that can capture dynamic patterns to propose a new subway passenger flow prediction framework. This method improves the prediction efficiency and accuracy through parallel computing. Although deep learning models have certain advantages, their network structure requires a large number of labeled samples for training and requires high-performance computing support. In small-sample scenarios, overfitting or a sharp drop in prediction accuracy is prone to occur.

Because of the high randomness and volatility of urban rail transit passenger flows, traditional prediction models, though suitable for small sample sizes and linear scenarios, struggle to cope with the periodic and sudden fluctuations in passenger flows. They can only fit long-term trends and rely heavily on empirical factors to correct for short-term fluctuations. Intelligent optimization algorithms,

though capable of handling nonlinear and complex passenger flows, are poorly adapted to small sample sizes and have high training costs. Their predictions for small sample sizes are suboptimal, capturing short-term fluctuations but lacking stability and interpretability for long-term trends. Furthermore, most model improvements rely on high-precision multisource data or are designed only for specific scenarios, making them challenging to implement in scenarios where basic data are missing or in cross-scenario applications. To address these shortcomings, this study proposes a prediction model based on the combined optimization of an unbiased grey model and a Markov algorithm to predict irregular fluctuations in urban rail transit passenger flows. The model first eliminates the systematic deviation of passenger flow data through unbiased processing, uses the unbiased grey model to accurately fit the long-term trend to make up for the trend prediction defects of the traditional model, and then introduces the Markov optimization model to divide the state through the data and quantify the short-term fluctuations, thereby achieving full coverage prediction of various samples and multiple scenarios. Ultimately, it provides a high-precision and easy-to-implement prediction plan for the average daily passenger flows entering the station on weekdays in the next 3 months, effectively making up for the shortcomings of the existing model in terms of scenario adaptability, modeling accuracy, and practical universality.

Materials and methods

The model's technology

The unbiased grey Markov model can be used in conjunction with passenger flow data (inbound, outbound, and section) generated for opened lines to accurately predict future passenger flows. The specific implementation, concept, and processing of the model are shown in Fig. 1.

Establishment of the unbiased grey prediction model

Unbiased processing of existing passenger flow data

The unbiased grey model is designed for existing stations, where the data's characteristics typically involve small samples, nonideal distributions, and random disturbances. It does not require a preset data distribution and only needs a small number of samples to capture the long-term trends. It can effectively adapt to these characteristics and does not require accumulation or subtraction operations for data processing. Its nonlinear fitting capability and small-sample adaptability are superior to those of other models, effectively ensuring the predictive accuracy of the combined model. To effectively apply the grey Markov chain prediction model, existing subway passenger flow data must be properly analyzed. The grey model is then used to process the existing subway passenger flow data in an unbiased manner as described below.

A sample sequence of the initial passenger flow data $X^{(0)}$ is established as shown below Eq. (1):

$$X^{(0)} = \{X_1^{(0)}, X_2^{(0)}, X_3^{(0)}, \dots, X_n^{(0)}\} \quad (1)$$

where, n is the number of sample data and $X^{(0)}$ is the initial passenger flow data.

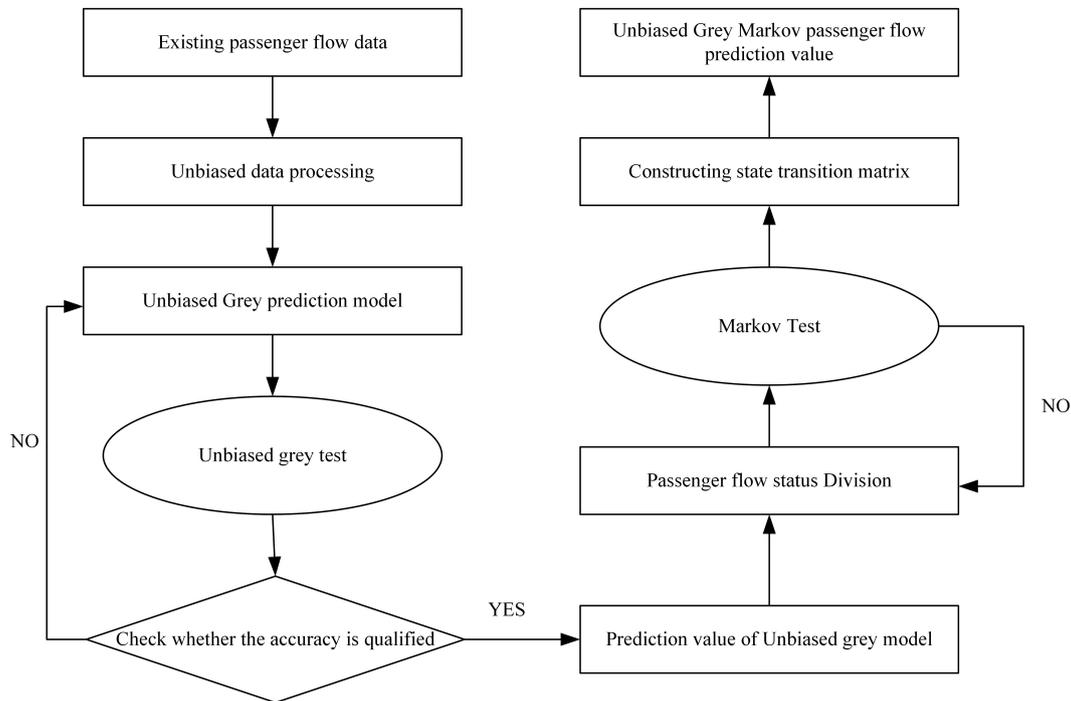


Fig. 1 Technological road map.

The initial sample data sequence is accumulated to reduce the random volatility of the data by correlating the generated data. The sequence is generated from a first-order accumulation $X^{(1)}$ Eq. (2):

$$X_{(k)}^{(1)} = \sum_{i=1}^k X_{(i)}^{(0)}, k = 1, 2, 3, 4, \dots, n \quad (2)$$

where, $X^{(1)}$ is the first-order cumulative generation sequence.

The grey differential equation is established as follows Eq. (3):

$$\frac{dX^{(1)}}{dt} + \lambda X^{(1)} = v \quad (3)$$

where, λ is the sequence coefficient and v is the differential result.

The least squares method is used to solve for the parameters Eqs (4)–(6):

$$\hat{\alpha} = [\lambda \ v]^T = (B^T B)^{-1} B^T Y_n \quad (4)$$

$$B = \begin{bmatrix} -\frac{1}{2}(X_1^{(1)} + X_2^{(1)}) & 1 \\ -\frac{1}{2}(X_2^{(1)} + X_3^{(1)}) & 1 \\ -\frac{1}{2}(X_3^{(1)} + X_4^{(1)}) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(X_{n-1}^{(1)} + X_n^{(1)}) & 1 \end{bmatrix} \quad (5)$$

$$Y_n = [X_2^{(0)}, X_3^{(0)}, X_4^{(0)}, X_5^{(0)}, \dots, X_n^{(0)}]^T \quad (6)$$

where, $\hat{\alpha}$ is the quadratic parameter, B is the sequence of adjacent cumulative means, and Y_n is the initial data column matrix.

The solution of the initial differential equation of the grey model obtained by combining the formulae above is shown below Eq. (7):

$$\hat{X}_{k+1}^{(1)} = \left[X_1^{(0)} - \frac{v}{\lambda} \right] e^{-\lambda k} + \frac{v}{\lambda}, k = 1, 2, 3, \dots, n-1 \quad (7)$$

where, $\hat{X}_{k+1}^{(1)}$ is the initial prediction.

Substituting the parameters α and γ into Eq. (8), we can get the simulated value of the first-order sum sequence. After cumulative reduction, we can get the predicted passenger flow.

$$\begin{cases} X_1^{(0)} = X_1^{(1)} \\ X_{k+1}^{(0)} = X_{k+1}^{(1)} - X_k^{(1)} \end{cases} \quad (8)$$

where, $X_{k+1}^{(0)}$ is the initial prediction and $X_1^{(0)}$ is the original value.

Although the traditional grey prediction model is easy to calculate and requires fewer data, it has limitations, such as poor accuracy and large deviations. The unbiased grey prediction model proposed in the literature^[25] effectively reduces deviation but does not require cumulative reduction operations, simplifies the modeling process, and has a broader applicability than the traditional model. The unbiased grey prediction model is described as follows Eqs (9)–(11):

$$b = \ln \left(\frac{2-\lambda}{2+\lambda} \right) \quad (9)$$

$$A = \frac{2v}{2+\lambda} \quad (10)$$

$$\begin{cases} \hat{X}_1^{(0)} = X_1^{(0)} \\ \hat{X}_{k+1}^{(0)} = A e^{bk}, k = 1, 2, 3, \dots, n \end{cases} \quad (11)$$

where, $\hat{X}_{k+1}^{(0)}$ is the initial prediction, $X_1^{(0)}$ is the original value, A is the prediction model's coefficient, and b is the prediction model's index.

Unbiased grey model test

After the establishment of the unbiased grey model, it is necessary to test the accuracy of the prediction data. The mean and variance of the sample series are as follows Eq. (12):

$$\begin{cases} \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k^{(0)} \\ S_1^2 = \frac{1}{n} \sum_{k=1}^n (X_k^{(0)} - \bar{X})^2 \end{cases} \quad (12)$$

where, \bar{X} is the mean of the sample series and S_1^2 is the variance of the sample series.

The deviation between the sample data and prediction data reflects the predictive accuracy. The formulae for calculating the

mean and variance of the relative deviation are as follows Eqs (13) and (14):

$$\begin{cases} q^{(0)}(k) = x_k^{(0)} - \hat{x}_k^{(0)} \\ \bar{q} = \frac{1}{n} \sum_{k=1}^n |q^{(0)}(k)| \\ S_2^2 = \frac{1}{n} \sum_{k=1}^n (q^{(0)}(k) - \bar{q})^2 \end{cases} \quad (13)$$

$$D = \frac{S_2}{S_1} \quad (14)$$

where, $q^{(0)}(k)$ is the relative deviation, \bar{q} is the mean relative deviation, S_2^2 is the variance of the relative deviation, and D is the threshold of the deviation ratio.

We can use the predicted data to assess the validity of the prediction model. The reference level of inspection is shown in Table 1.

When the deviation threshold is $D < 0.35$, the inspection accuracy is of the first order and the inspection effect is ideal; when the deviation threshold is $0.35 \leq D < 0.5$, the inspection accuracy is of the second order, and the inspection effect is acceptable; when the deviation threshold is $0.5 \leq D < 0.75$, the inspection accuracy is of the third order and the inspection effect meets the qualification standard; and when the deviation threshold is $D \geq 0.75$, the inspection accuracy is of the fourth-order, and the inspection effect fails to meet the qualification standard. The grey model must be improved until the inspection effect meets the qualification standard.

Establishment of a modified Markov optimization model

The Markov model can effectively describe the random dynamic behavior of data. Establishing the relationships among different data states improves the accuracy of prediction. With increasing time and changes in the season, subway passenger flow data fluctuate irregularly, preventing the passenger flows from continuously increasing or decreasing. The Markov model offers the advantages of flexible forecasting and high accuracy, which compensate for the grey model's inability to predict random passenger flows effectively. The data are categorized into different states, each of which is independent and discrete, and different random processes are combined to form a Markov random state[26]. The states are categorized using the following formula Eq. (15):

$$X^1(n) = X(t)P^{(n-t)} \quad (15)$$

where, $X^1(n)$ is the prediction at time n , $X(t)$ is the initial data at time t , and $P^{(n-t)}$ is the probability matrix for the $n - t$ step transition.

Categorization of passenger flow status

The randomness of the passenger flow produces different states, where the data in the same state vary according to certain rules. The passenger flow state is categorized according to the characteristics of the relative error between the initial data and the unbiased grey model's predicted data. The passenger flow data can be evenly distributed across states to ensure the rationality of the categorization scheme. The formula for calculating the relative error $R(k)$ is as follows Eq. (16):

$$R(k) = \frac{X_k^{(0)} - \hat{X}_k^{(0)}}{X_k^{(0)}} \times 100\% \quad (16)$$

where, $R(k)$ is the relative error, $X_k^{(0)}$ is the actual passenger flow, and $\hat{X}_k^{(0)}$ is the predicted passenger flow.

The volatility of the passenger flow data results in positive and negative relative errors. To ensure that the optimized data are

Table 1. Inspection-level parameters.

Inspection-level accuracy	First-order accuracy	Second-order accuracy	Third-order accuracy	Fourth-order accuracy
D	$D < 0.35$	$0.35 \leq D < 0.5$	$0.5 \leq D < 0.75$	$D \geq 0.75$

highly accurate, the positive and negative relative errors are separated. The relative error is used to categorize the data into s states. These states are expressed as M_i ($i = 1, 2, 3, \dots, s$), and the state interval is expressed as $M_i = [T_i, H_i]$, where T_i and H_i are the lower and upper limits of the state i , respectively.

Markov test

Since the passenger flows at different stations and lines are affected by many factors, the degree of fluctuation in the passenger flows varies and it is impossible to accurately divide the states. Therefore, in order to ensure that the divided passenger flow state meets the Markov characteristics, the Markov model verification method based on the χ^2 -test is used to determine the reasonable number of states and state intervals[27]. The calculation formula is as follows Eq. (17):

$$\begin{cases} P_{oj} = \frac{\sum_{i=1}^s E_{ij}}{\sum_{i=1}^s \sum_{j=1}^s E_{ij}} \\ P_{ij} = \frac{E_{ij}}{\sum_{i=1}^s E_{ij}} \\ \chi^2 = 2 \sum_{i=1}^s \sum_{j=1}^s E_{ij} \left| \ln \left(\frac{P_{ij}}{P_{oj}} \right) \right| \sim \chi_\alpha^2((s-1)^2) \end{cases} \quad (17)$$

where, P_{oj} is the transition rate from the initial state i to state j , and P_{ij} and E_{ij} are the transition probability and the number of elements transferred from state i to state j , respectively.

The original data are divided into s states. If $\chi^2 > \chi_\alpha^2((s-1)^2)$ at a given significance level α , the Markov test is considered valid and conforms to the Markov property. Otherwise, it is necessary to redivide the passenger flow states and conduct the Markov test again until the test requirements are met.

Construction of the state transition matrix

Calculating the state transition matrix is the key step in building the Markov model. Different states have different rules for a change in passenger flows, and there is an inevitable connection between the states. With the help of this matrix, the probability distribution of transitions between states can be intuitively presented, and then probabilistic methods can be used to predict the change trend of passenger flows in the next time period, significantly improving the predictive accuracy[28]. If the total amount of data corresponding to State W_i is E_i , and the amount of data entering State E_j after k steps of transition is $E_{ij}(k)$, then the state transition probability is as follows Eq. (18):

$$P_{ij}(k) = \frac{E_{ij}(k)}{E_i} \quad (18)$$

where, $P_{ij}(k)$ is the k -step transition probability from State i to State j , E_i is the amount of data in State i , and $E_{ij}(k)$ is the amount of data entering State E_j after k steps of transition.

Arranging all the one-step transition probabilities p_{ij} in state order forms the state transition probability matrix $P(r)$, which is shown below Eq. (19):

$$P(r) = \begin{bmatrix} p_{11}(r) & p_{12}(r) & \dots & p_{1m}(r) \\ p_{21}(r) & p_{22}(r) & \dots & p_{2m}(r) \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1}(r) & p_{m2}(r) & \dots & p_{mm}(r) \end{bmatrix} \quad (19)$$

where, $P(r)$ is the state transition matrix.

Taking the passenger flow data of the latest existing station as the initial state vector V , the k -step prediction is determined using a k -step state transition matrix formula:

$$P(k) = V \times P(r)^k \quad (20)$$

where, $P(k)$ is the result of prediction after k steps of state transition, V is the initial state vector, and $P(r)^k$ is the state transition probability matrix $P(r)$ raised to the power of k .

Forecast of passenger flows through existing stations

The unbiased grey prediction model can only predict the long-term trend in the passenger flow data, whereas the actual passenger flow data exhibit fluctuations under various conditions. Therefore, the Markov optimization model is used to partition the data state interval, define the upper and lower limits of variations in the passenger flow data, and refine the prediction data of the unbiased grey model. If we assume that the unbiased grey predicted passenger flow at time t is $\hat{X}_n^{(0)}$, the optimization formula of the forecast passenger flow is as follows Eq. (21):

$$\hat{y}(n) = \frac{\hat{X}_n^{(0)}}{1 \pm \frac{|T_i + H_i|}{2}} \quad (21)$$

where, H_i is the upper limit of State i and T_i is the lower limit of State i ; the "+" or "-" symbol is selected according to how the state interval has

been partitioned. If the upper and lower limits of the state are negative, then "+" is used in the formula. If the upper and lower limits of the state are positive, "-" is used in the formula. Ultimately, the formula above is used to obtain the Markov-optimized forecast of passenger flows.

Case study and results

Qingdao is the largest city in Shandong Province, China. Rail transit has been in operation Qingdao since 2015 and is called the Qingdao Metro. Qingdao Metro Line 2 was used as a case study. The first phase of operation of the first section of the Qingdao Metro Line 2 began on December 10, 2017, when 18 stations from Licun Park Station to Zhiquan Road Station were opened. The second phase of operation of the first section consisted of putting Taishan Road Station, Lijin Road Station, and Taidong Station into operation on December 16, 2019. Lines 2 and 3 cross at Licun Station and Wu Si Square Station. The opening of these stations along Metro Line 2 marked a new era for the Qingdao Metro. The routes of the M2 and M3 Lines are shown in Fig. 2.

The second phase of operation of the first section of Metro Line M2 provides public transport to a densely populated and commercial area. In addition to considering the operation status of the existing line, it is very important to study the influence of access during this second phase on the passenger flows at the stations along Lines 2 and 3, as well as the transfer passenger flows. This is crucial for vehicle scheduling and organizing passenger transport for Lines 2 and 3.

This paper uses the average daily passenger flow data of Qingdao Metro Line 2 Shilaoren Bathing Beach Station from January 2018 to June 2019 as a sample for analysis, compares the predictive accuracy of the unbiased grey prediction model and the grey

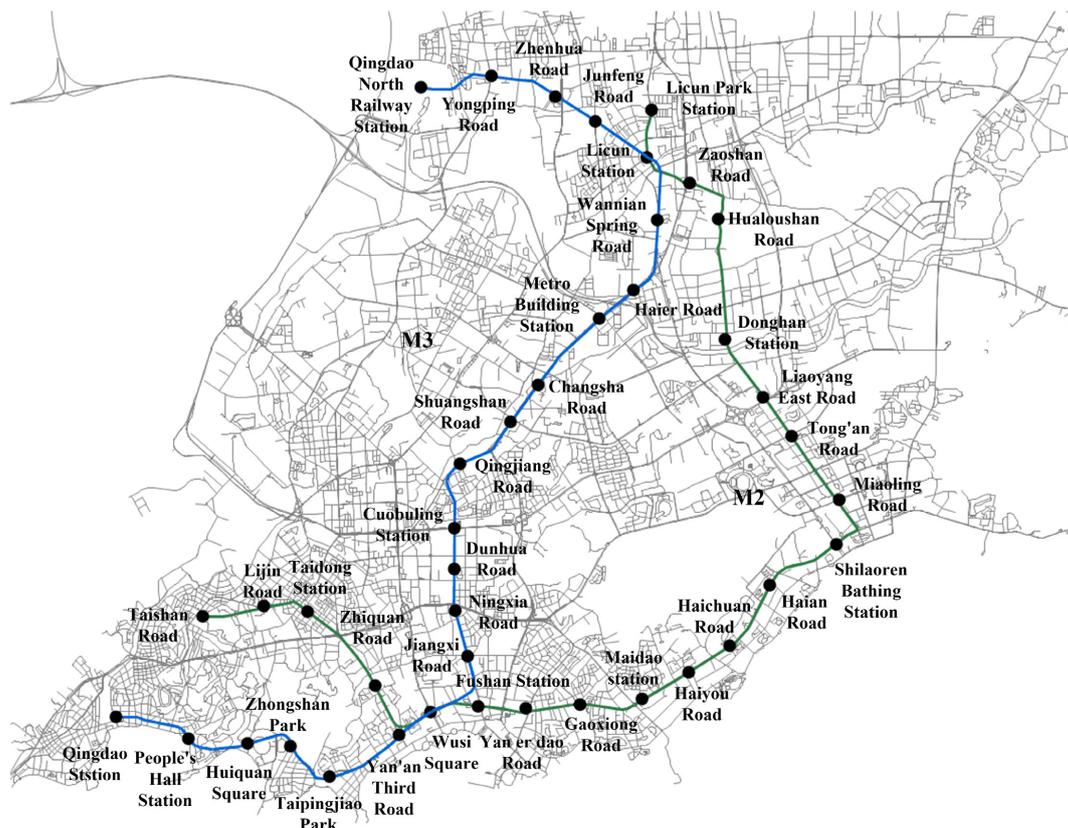


Fig. 2 Route map of Lines 2 and 3. Source: the authors.

Markov model, and predicts the average daily passenger flows for weekdays in the next 3 months, providing a theoretical basis for the operations management department. The average daily passenger flows of Shilaoren Bathing Beach Station on weekdays is shown in Table 2.

Shilaoren Bathing Beach Station is a famous scenic spot in Qingdao. It can be seen from the data in Table 2 that the overall passenger flow of Line 2 shows a steady upward trend during operating hours. During the peak tourist season (June, July, and August), the passenger flow through the station increases significantly. Metro travel has become a key part of Qingdao's transportation network and is the main choice for many passengers. Therefore, accurate predictions of passenger travel data can provide an effective reference for the subway's operations management department, and contribute significantly to ensuring operational safety and improving service quality.

After sorting the average daily passenger flow data in Table 2 and performing a series of cumulative calculations, the following model parameters can be obtained: $b = -0.0059$, $A = 8,308.8$. Substituting the results into the formula (Eq. 11), the grey prediction model can be obtained. According to the grey prediction model, the simulated prediction value and the relative error corresponding to the original passenger flow data can be obtained, as shown in Table 3.

Both the actual and grey forecast passenger flows through Shilaoren Bathing Beach Station from January 2018 to September 2019 are fitted. The fitting results are shown in Fig. 3.

Figure 3 shows that the grey prediction of passenger flows can capture the overall development trend of the passenger flows but does not effectively reflect the short-term fluctuations at specific nodes in the passenger flow data. Therefore, the predicted passenger flows must be further optimized.

The unbiased grey test model is used to obtain the test parameters shown in Table 4.

Table 2. Average daily passenger flows into Shilaoren Bathing Beach Station on weekdays

Date	Average daily passenger flow (persons)	Date	Average daily passenger flow (persons)
2018.1	6,138	2018.10	8,049
2018.2	6,284	2018.11	8,100
2018.3	7,155	2018.12	8,098
2018.4	8,064	2019.1	7,923
2018.5	8,266	2019.2	7,836
2018.6	9,406	2019.3	8,699
2018.7	11,484	2019.4	8,789
2018.8	13,036	2019.5	9,065
2018.9	8,976	2019.6	10,051

Table 3. The grey prediction model's predictions and relative error

Date	Predicted value (persons)	Relative error (%)	Date	Predicted value (persons)	Relative error (%)
2018.1	6,138	0.00	2018.10	8,792	-9.23
2018.2	8,383	-33.41	2018.11	8,844	-9.19
2018.3	8,433	-17.87	2018.12	8,897	-9.86
2018.4	8,484	-5.20	2019.1	8,950	-12.96
2018.5	8,534	-3.24	2019.2	9,003	-14.90
2018.6	8,585	8.73	2019.3	9,057	-4.11
2018.7	8,636	24.80	2019.4	9,111	-3.66
2018.8	8,688	33.36	2019.5	9,165	-1.11
2018.9	8,740	2.63	2019.6	9,220	8.27

According to the test parameter table, $D = 0.658 < 0.75$. As shown in Table 1, the data's fitting accuracy is at Level 3, which is marginally qualified. Therefore, it is necessary to use the Markov model to further refine the predicted passenger flow data.

The relative error (Eq. 16) is shown in Table 3. The state interval is divided into six intervals, as shown in Table 5.

According to the formula (Eq. 18) and Table 5, the one-step state transition probability matrix is calculated as follows:

$$P(1) = \begin{bmatrix} 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/4 & 1/4 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

According to the state intervals divided in Table 5, the χ^2 statistic is used to test whether the sequence has Markov properties. The calculated statistic $\chi^2 = 2 \sum_{i=1}^s \sum_{j=1}^s E_{ij} \left| \ln \left(\frac{P_{ij}}{P_{oj}} \right) \right| = 13.52667$. The significant level $\alpha = 0.05$ is selected, the number of states is 6, and the table gives $\chi_{0.05}^2[(6-1)^2] = 11.0705$. The verification analysis shows that $11.0705 < 13.52667$. Therefore, the Markov test is valid at a confidence level of 95% and Markov correction can be performed.

The Markov optimization model is applied to the predicted data of the grey model. The optimization results are shown in Table 6.

To evaluate the effectiveness of the model, we selected the BP neural network model (a commonly used machine learning model) as a comparison. A three-layer feedforward neural network was constructed as the baseline prediction model, using a sliding window mechanism. The historical passenger flows from the previous

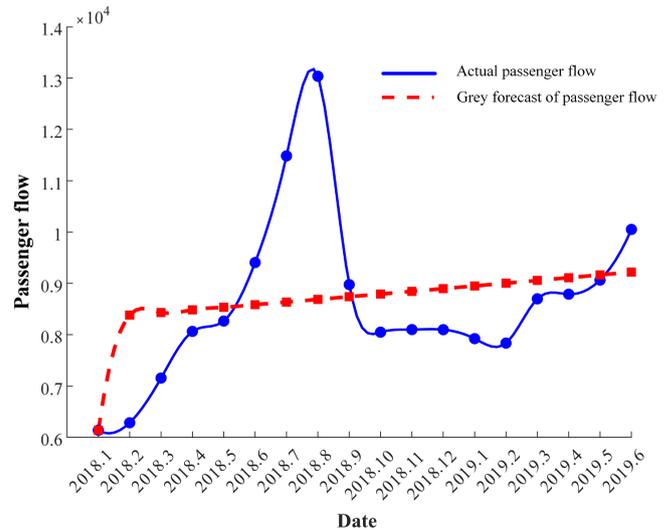


Fig. 3 Fitting diagram of the actual passenger flow and the grey model's forecasted passenger flow.

Table 4. The grey model's inspection parameters

Parameter	\bar{X}	S_1	\bar{q}	S_2	D
Numerical value	8,634	1,661	1,023	1,093	0.658

Table 5. Division of status intervals

Status range	E_1	E_2	E_3	E_4	E_5	E_6
Interval division	[-34, -12]	[-10, -9]	[-6, -3]	[-2, 0]	[2, 9]	[24, 34]

There are many transfer stations along Qingdao Metro Line 2, including Wu Si Square Station, Miaoling Road Station, Licun Station, Liaoyang East Road Station, Taidong Station, and Taishan Road Station. Accurate forecasting of transfer passenger flows is crucial for predicting the overall passenger flows for the line. Figures 6 and 7 show the forecast and actual direction of the flows through Wu Si Square Station, respectively, for the morning and evening peak hours during the first half of 2020.

An analysis of the predicted data in Fig. 6 and the actual data in Fig. 7 reveals that the relative error of the Markov-optimized passenger flows is less than 10%, indicating that the predicted results fall within the acceptable error range. During the morning peak period, the primary direction of the passenger flow from Line 2 through Wu Si Square Station is toward Qingdao Station. This direction can be explained in terms of the economic development, mature land use, abundant job opportunities, and numerous tourist attractions, all of which attract substantial transfer passenger flows. The transfer passenger flow from Line 2 to Line M3 via Wu Si Square Station during the evening peak hours is the direction of the greatest

passenger flow from Licun Park Station. The passenger flows through stations between Licun Park Station and Wu Si Square Station account for a large proportion of the passenger flows along Line 2. The line from Wu Si Square Station to Qingdao North Station is long and mostly crosses commercial land, which is a high-employment area. Many passengers travel home from work during the evening peak hours, mostly transferring at Wu Si Square Station. The transfer passenger flows from Line M3 to Line 2 mainly go from Qingdao North Station to Licun Park Station during the morning peak hours. Many passengers travel the line from Wu Si Square Station to Qingdao North Station. The most convenient way for residents in this area to access Line 2 is to transfer at either Wu Si Square Station or Licun Station.

Discussion

This study aims to predict passenger flows at existing urban rail transit stations. It conducts a multidimensional analysis of historical passenger flow data from typical stations on Qingdao Metro Line 2.

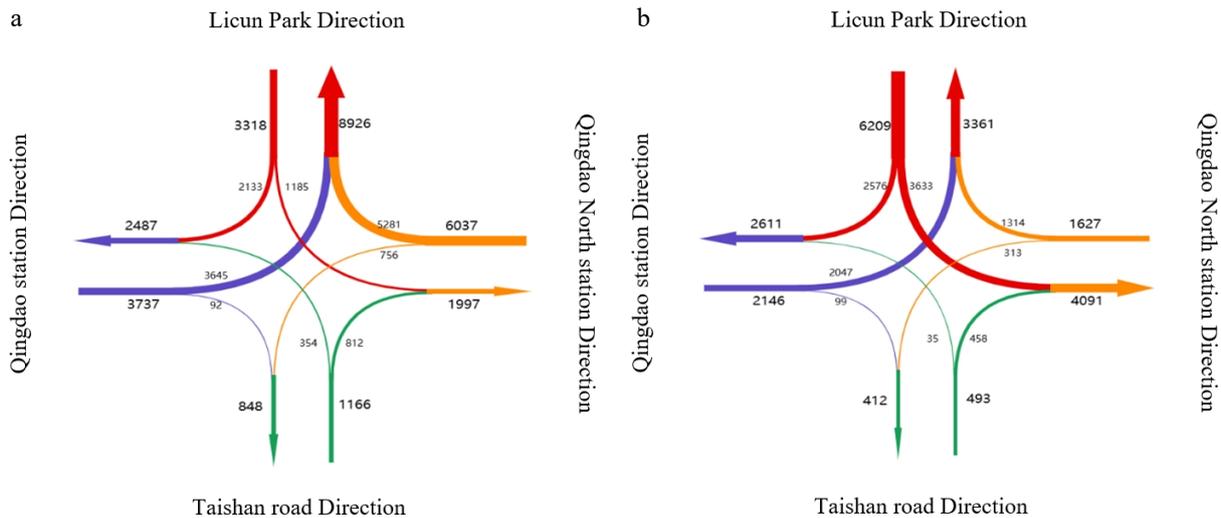


Fig. 6 Predicted direction of passenger flows through Wu Si Square Station in the morning and evening peak times. (a) Morning rush hour. (b) Evening rush hour.

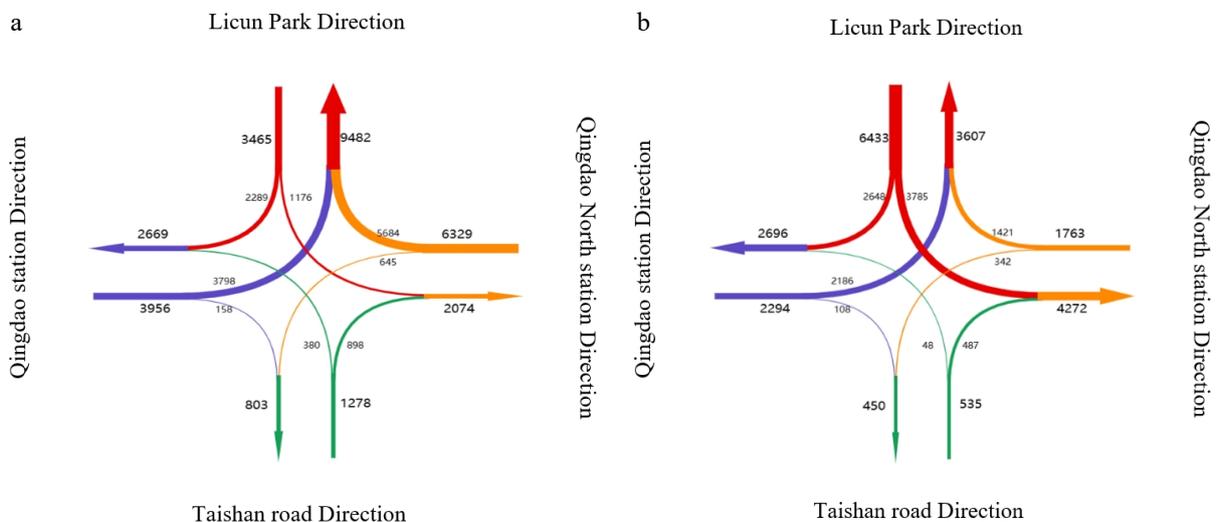


Fig. 7 Actual direction of passenger flows through Wu Si Square Station in the morning and evening peak times. (a) Morning rush hour. (b) Evening rush hour.

The model is systematically applied for construction, accuracy verification, passenger flow state classification, and case validation of an unbiased grey Markov chain combined prediction model. By avoiding the accumulation and subtraction operations of the traditional GM (1,1) model, the unbiased grey model effectively captures the long-term evolution of passenger flows. For Shilaoren Bathing Beach Station, the model's deviation threshold $D = 0.658$ meets the Level 3 standard of qualified accuracy. The Markov chain accurately corrects for random fluctuations in passenger flows by classifying passenger flows into six discrete states. Forecasts for passenger flows over the subsequent 3 months show that the relative error is within 10%. The model also effectively captures the passenger flow characteristics of stations in scenic areas during the peak tourist season, verifying its predictive effectiveness and providing a quantitative basis for optimizing line capacity allocation and organizing train operations. The model achieves stable predictions with only 18 months of sample data. It is particularly suitable for existing stations with a short operating life and for complex passenger flow scenarios characterized by both long-term trends and short-term fluctuations. The model demonstrates superior adaptability. Compared with existing research, this study replaces the traditional GM (1,1) model with an unbiased grey model, avoiding the bias caused by accumulation and subtraction operations. The model eliminates the need for complex data preprocessing, significantly simplifies the modeling process, and improves the engineering application efficiency. The model also significantly reduces the sample size requirement and eliminates the need for high-performance computing equipment, reducing the model's complexity and computational cost. This makes it easier for operations managers to understand, master, and apply it, resulting in greater practicality and operability.

This study has certain limitations. First, the spatiotemporal distribution of passenger flows varies across stations and lines, resulting in the uneven predictive accuracy of the Markov chain optimization model across lines. Furthermore, the current classification of passenger flow states relies on researchers' empirical judgment, making it difficult to completely eliminate the impact of subjective factors on the model's accuracy. The study only considered passenger flow scenarios under good weather conditions and normal operating conditions, and did not incorporate external disturbances such as inclement weather, major holidays, and large-scale sport or entertainment events, which can cause sudden changes in passenger flows. This results in insufficient adaptability to extreme scenarios.

Therefore, future improvements to the model should focus on the following directions:

1. Reduce subjectivity: Classify passenger flow states more objectively by incorporating clustering techniques to minimize human bias.
2. Enhance adaptability: Improve the model's robustness to highly volatile data by integrating adaptive algorithms or hybrid models capable of dynamically responding to sudden changes in passenger flows.
3. Incorporate external factors: Extend the model to account for external influences such as severe weather, major holidays, and large-scale events.

These improvements will significantly enhance the practical value and reliability of the model in diverse and dynamic environments.

Conclusions

Addressing the limitations of traditional prediction models in simultaneously capturing long-term trends and short-term fluctuations in urban rail transit passenger flows, this study constructs an

unbiased grey Markov chain combined prediction model. The effectiveness and applicability of the model are verified using Qingdao Metro Line 2 as a case study. This model integrates the core advantages of unbiased grey models and Markov chains, utilizing the state transition characteristics of Markov chains to dynamically correct the components of fluctuation that are not captured in the initial predictions of grey models, effectively compensating for the shortcomings of single models in characterizing complex passenger flows. Comparative experiments with BP neural network models demonstrate that the model constructed in this study fits the actual passenger flow data better, especially in capturing the surge during peak tourist seasons, showing adaptability to complex passenger flow scenarios. To verify the effectiveness and stability of the model, two representative existing stations—Shilaoren Bathing Beach Station and Wu Si Square Transfer Station were selected for study. The results show that the relative error between the model's predicted values and the measured values is controlled within 10%, indicating that the model can accurately reproduce both the overall passenger flow trends and short-term fluctuation patterns. The passenger flow data obtained from the existing station passenger flow prediction method proposed in this study can provide reliable data support for station capacity allocation, peak-hour passenger flow organization, emergency response scheduling, and other operational management work, thus helping to improve operational efficiency and service quality of urban rail transit systems.

Author contributions

The authors confirm their contributions to the paper as follows: study conception and design: Pan F; literature collection and summarization: Pan F, Li W; analysis and interpretation of the results: Zhang L, Tang H; draft manuscript preparation: Yang X, Xia Y. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

This work was supported by the Innovation Institute for Sustainable Maritime Architecture Research and Technology, Qingdao University of Technology (CK-2024-0068); the Natural Science Foundation of China (62373209); and the Innovation and Development Joint Fund Project of Shandong Provincial Natural Science Foundation (ZR2024LZN012).

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 27 July 2025; Revised 20 November 2025; Accepted 12 January 2026; Published online 31 March 2026

References

- [1] Wang Y, Ma J, Zhang J. 2019. Metro passenger flow forecast with a novel Markov-grey model. *Periodica Polytechnica Transportation Engineering* 48(1):70–75

- [2] Zhou W, Wang W, Zhao D. 2020. Passenger flow forecasting in metro transfer station based on the combination of singular spectrum analysis and AdaBoost-weighted extreme learning machine. *Sensors* 20(12):3555
- [3] Jiao B. 2025. Analysis and prediction of railway passenger flow patterns based on the ARIMA model. *Journal of Engineering Innovations & Technology* 7(2):25–61
- [4] Wang H, Teng J, Ye L, Chen Y. 2024. Short-term prediction method for passenger density in urban rail transit based on nonlinear kalman filter. *Urban Mass Transit* 27(6):33–38,43
- [5] Carmona-Benítez RB, Nieto MR. 2020. SARIMA damp trend grey forecasting model for airline industry. *Journal of Air Transport Management* 82:101736
- [6] Peng L, Shao X, Huang W. 2022. Prediction of Yantai railway passenger volume based on gray Markov model. *Journal of Ludong University (Natural Science Edition)* 38(1):50–56 (in Chinese)
- [7] Li H, Wen C. 2024. Prediction of road traffic accidents based on the grey-Markov model. *Communications Science and Technology Heilongjiang* 47(5):170–174 (in Chinese)
- [8] Ye J, Xu Z, Gou X. 2022. An adaptive Grey-Markov model based on parameters self-optimization with application to passenger flow volume prediction. *Expert Systems With Applications* 202:117302
- [9] Ding T, Pan N, Du B, Ai W. 2023. Forecast of cargo throughput in port based on improved grey Markov. *Journal of Chongqing Jiaotong University (Natural Science)* 42(09):130–136 (in Chinese)
- [10] Cai L, Zhang Z, Yang J, Yu Y, Zhou T, et al. 2019. A noise-immune Kalman filter for short-term traffic flow forecasting. *Physica A: Statistical Mechanics and Its Applications* 536:122601
- [11] Hu S, Ji M, Chang Z, Wang H, Kong X. 2025. An improved particle swarm optimization algorithm based urban rail passenger flow prediction model: a case study in Beijing, China. *Digital Transportation and Safety* 4(2):101–107
- [12] Shi F, Yang X, Hu X, Xu G, Wu R. 2019. A VMD-GA-BP method for predicting non-holiday passenger flow of high speed railway based on data replacement correction. *China Railway Science* 40(3):129–136
- [13] Xue G, Liu S, Ren L, Ma Y, Gong D. 2022. Forecasting the subway passenger flow under event occurrences with multivariate disturbances. *Expert Systems with Applications* 188:116057
- [14] Mulerikkal J, Thandassery S, Rejathalal V, Kunnamkody DMD. 2022. Performance improvement for metro passenger flow forecast using spatio-temporal deep neural network. *Neural Computing and Applications* 34(2):983–994
- [15] Bapaume T, Côme E, Ameli M, Roos J, Oukhellou L. 2023. Forecasting passenger flows and headway at train level for a public transport line: focus on atypical situations. *Transportation Research Part C-Emerging Technologies* 153:104195
- [16] Tu Q, Geng G, Zhang Q. 2023. Multi-step subway passenger flow prediction under large events using website data. *Tehnički Vjesnik* 30(5):1585–1593
- [17] Yue M, Ma S. 2023. LSTM-based transformer for transfer passenger flow forecasting between transportation integrated hubs in urban agglomeration. *Applied Sciences* 13(1):637
- [18] Halyal S, Mulangi RH, Harsha MM. 2022. Forecasting public transit passenger demand: With neural networks using APC data. *Case Studies on Transport Policy* 10(2):965–975
- [19] Qi Q, Cheng R, Ge H. 2023. Short-term inbound rail transit passenger flow prediction based on BILSTM model and influence factor analysis. *Digital Transportation and Safety* 2(1):12–22
- [20] Cao B, Li Y, Chen Y, Yang A. 2024. A CNN-LSTM model for short-term passenger flow forecast considering the built environment in urban rail transit stations. *Journal of Transportation Engineering, Part A: Systems* 150(11):04024072
- [21] Bogaerts T, Masegosa AD, Angarita-Zapata JS, Onieva E, Hellinckx P. 2020. A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. *Transportation Research Part C: Emerging Technologies* 112:62–77
- [22] He Y, Li L, Zhu X, Tsui KL. 2022. Multi-graph convolutional-recurrent neural network (MGC-RNN) for short-term forecasting of transit passenger flow. *IEEE Transactions on Intelligent Transportation Systems* 23(10):18155–18174
- [23] Yu S, Luo A, Wang X. 2023. Railway passenger flow forecasting by integrating passenger flow relationship and spatiotemporal similarity. *Intelligent Automation and Soft Computing* 37(2):1877–1893
- [24] Xiu C, Zhan S, Pan J, Peng Q, Lin Z, et al. 2026. Correlation-based feature selection and parallel spatiotemporal networks for efficient passenger flow forecasting in metro systems. *Transportmetrica A: Transport Science* 22(1):1
- [25] Pan J, Ma C. 2018. Passenger flow forecast based on improved grey Markov model. *Technology & Economy in Areas of Communications* 20(6):52–56,69
- [26] Guo X, Chen X, Yu Q, Sun Y. 2019. Traffic state prediction based on cluster analysis and Markov model. *Highway* 64(8):304–309
- [27] Bao L. 2017. Real-time forecast of passenger flow volume in urban rail transit. *Urban Mass Transit* 20(5):104–106,111
- [28] Yu Q, Yao Z. 2019. Research on Markov particle filter for traffic flow prediction. *Journal of Transportation Systems Engineering and Information Technology* 19(2):209–215



Copyright: © 2026 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.