Dean&Francis
ISSN 2959-6157

Analysis of the Relationship between Zhengzhou Subway Passenger Flow and Weather

Tiantian Zhang

College of Software Engineering, Zhengzhou University of Light Industry, Zhengzhou, Henan, China
outlook_D97521C3FC5F37EC@outlook.com

Abstract:

Nowadays, subways have become the primary mode of transportation due to their various advantages, such as large carrying capacity, convenience, speed, and low cost. Zhengzhou, a new first-tier city known as a transportation hub, is primarily populated by young working professionals. The young group has a fast-paced commute and high demand for transportation efficiency, so they tend to prefer the subway as their daily commuting mode. Different weather conditions also affect people's choice of transportation mode. Weather, as an important factor of daily change, also influences people's choice of transportation mode. Based on this social reality, this paper employs a combination of correlation analysis, multiple linear regression analysis, and stepwise regression analysis to study the impact of weather on subway ridership. Through data analysis from April to September 2025, it was found that there is a negative correlation between subway ridership on the same day and both minimum temperature and total precipitation. When the minimum temperature or total precipitation increases, subway ridership on the same day decreases accordingly. The study confirms that minimum temperature and total precipitation have a tangible impact on daily ridership. This research can provide reference suggestions for subway dispatch management, ridership prediction, and emergency management.

Keywords: Urban rail transit; Weather factors; Zhengzhou.

1. Introduction

With the development of transportation, subways have become a crucial player in addressing urban traffic congestion issues. Their advantages, such as high efficiency, punctuality, limited boundaries, and large capacity, have contributed significantly to optimizing urban structure, protecting the environment, and enhancing residents' happiness. By the beginning of 2025, a total of 325 subway lines had been opened in 54 cities in China, with a total length of approximately 10,947.3 kilometers of subway lines built and put into use. Currently, the operational mileage of urban transportation in China has reached 11,000 kilometers, firmly ranking first in the world. However, as a complex public service system, the passenger flow of the subway system is jointly influenced by multiple factors, among which meteorological conditions, as an important external factor, cannot be ignored. Yueyan Lyu(2021) and Longzhu Xiao(2020), who studied and predicted the social activity of subway rail transit using Shenzhen and Suzhou as examples, inspired me to research the analysis of subway passenger flow and weather conditions centered around Zhengzhou City [1,2].

The data on subway ridership varies significantly across different regions and weather conditions. For example, Sun Yingbao(2024) mentioned that on the day after the heavy snowstorm in Shenyang in 2021, road traffic basically collapsed across the board, leading to a surge in subway ridership. The sudden increase in passenger flow caused a significant disaster for subway operations and also affected the normal travel of many residents [3]. Xu Bo and others(2023) used the nine-period moving average method to demonstrate that within a certain range of precipitation, the fluctuation rate of passenger flow decreases as rainfall increases [4]. Jiang et al.(2025) studied the impact of weather on coastal rail transit and concluded that when the peak rainfall occurs before or after the peak storm surge, surface water will gradually rise in multiple stages; when the peak rainfall coincides with the peak storm surge, surface water will rapidly rise in a very short period of time, providing reliable information reference for rescue and evacuation efforts [5]. From this, it can be seen that quantifying the impact of weather on subway ridership provides subway operating companies with an additional tool to predict potential changes in ridership under adverse weather conditions, allowing them to take corresponding measures.

It is important and indispensable to study and analyze the impact of weather on subway ridership. In terms of research methodology, a large number of researchers have analyzed the relationship between subway passenger flow and weather. For example, Nicole S. Ngo and others(2024) used least squares regression analysis to conclude that the number of people using public transportation decreases during extreme weather conditions [6]. İsmail Adalıoğlu(2025) utilized the capability of multiscale geographically weighted regression analysis to explain local variations in travel behavior, emphasizing the crucial role of terminal accessibility, socioeconomic characteristics, and mode integration in shaping transportation preferences [7]. Stephen Kome Fondzenyuy et al.(2025) analyzed the on-site speed data and road attribute data of 39,413 vehicles using multiple linear regression (MLR) and multilevel modeling techniques, and concluded that there are significant speed levels and differences between rural and urban areas [8]. Eva Brumercikova et al.(2020) determined through regression and correlation analysis that the number of residents, average monthly salary, minimum wage, and unemployment rate indeed have a certain impact on the number of public transportation passengers [9]. Yang Xiaohong(2024) verified the promoting effect of rail transit development on urban tourism by analyzing the correlation between urban rail transit and urban tourism [10]. These studies provide a theoretical foundation and methodological references for this article. This study will utilize correlation analysis and multiple linear regression methods to analyze and investigate the passenger flow volume of Zhengzhou Metro from April to October 2025 in conjunction with weather data, aiming to discover the correlation between them and provide some reference data for the planning of the metro operating company.

2. Method

2.1 Section Headings

The data used in this study include the public data of subway, as shown in Figure 1, and the public data of historical weather [11,12]. These data sets include the daily passenger flow, temperature, wind, weather, and other data of Zhengzhou from April to October 2025. This study integrates subway passenger flow and weather conditions through time data to ensure data consistency.

ISSN 2959-6157

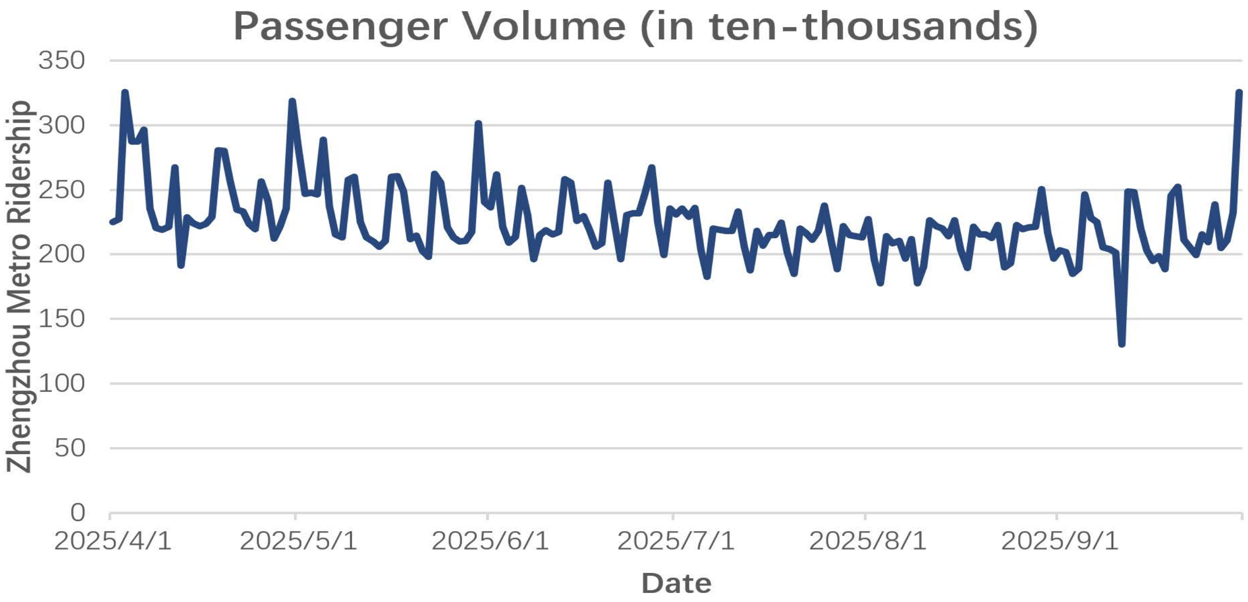

Fig. 1 Diagram of daily subway passenger flow data (Data from: [11, 12]

2.2 Index Selection and Description

Depending on varying weather conditions, people's tendency to take the subway differs. This study selected passenger volume, precipitation, wind speed, and temperature as indicators, as shown in Table 1.

Table 1. Indicator Selection

	Average	maximum	minimum	standard deviation
passenger volume	224.7	325.6	130.4	27.5
precipitation	4.6	79.3	0	1.12
wind speed	2.7	23.8	3.7	1.11
temperature	26.0	40.5	8.5	5.11

The selection of these indicators comprehensively considers the availability of data and reflects the impact of restrictions on subway passenger flow through all-around weather factors.

2.3 Method Introduction

Firstly, Pearson correlation analysis was used to explore the linear relationship between weather variables and passenger flow. On this basis, a multiple linear regression model was established.

Assuming that the dependent variable, i.e., the passenger flow, is y, the independent variable, i.e., the influencing factors are x1, x2, x3..., xk respectively, and the independent variable and the dependent variable are linear, the multiple linear regression model is:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + e \quad (1)$$

Where: b0 represents constant coefficient, b1, b2, ..., bk represents regression coefficient. The regression coefficient b1 represents the change of y when x1 increases or decreases when other independent variables are fixed, that is, the partial regression coefficient of x1 to y. Similarly, b2, b3, ..., bk can be obtained.

3. Results and Discussion

3.1 Data Visualization

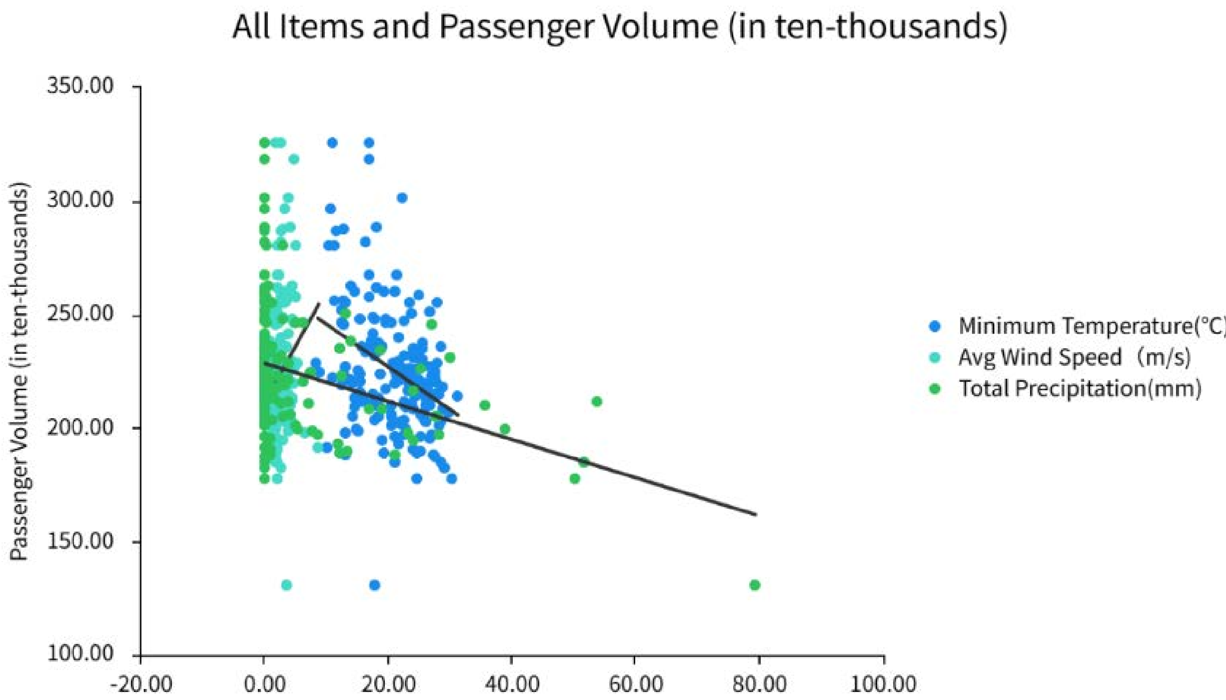

Fig. 2 Relationship between all items and passenger volume (Picture credit: Original)

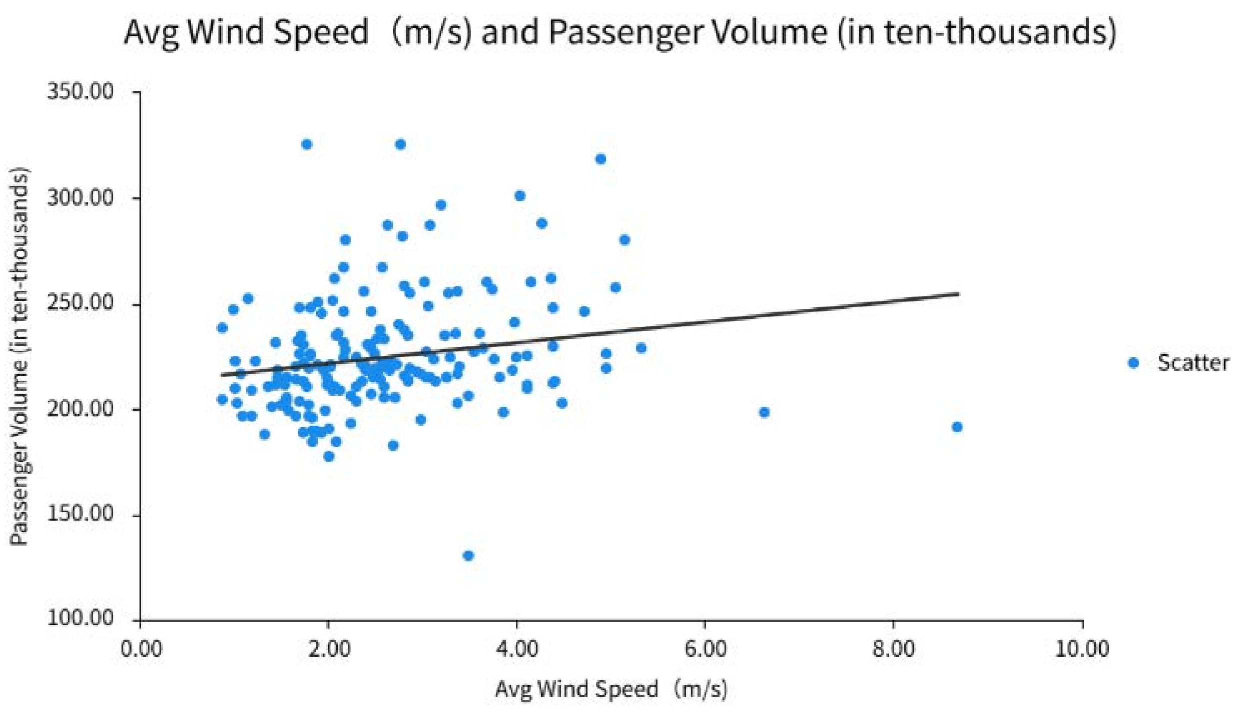

Fig. 3 Relationship between Avg wind speed and passenger volume (Picture credit: Original)

ISSN 2959-6157

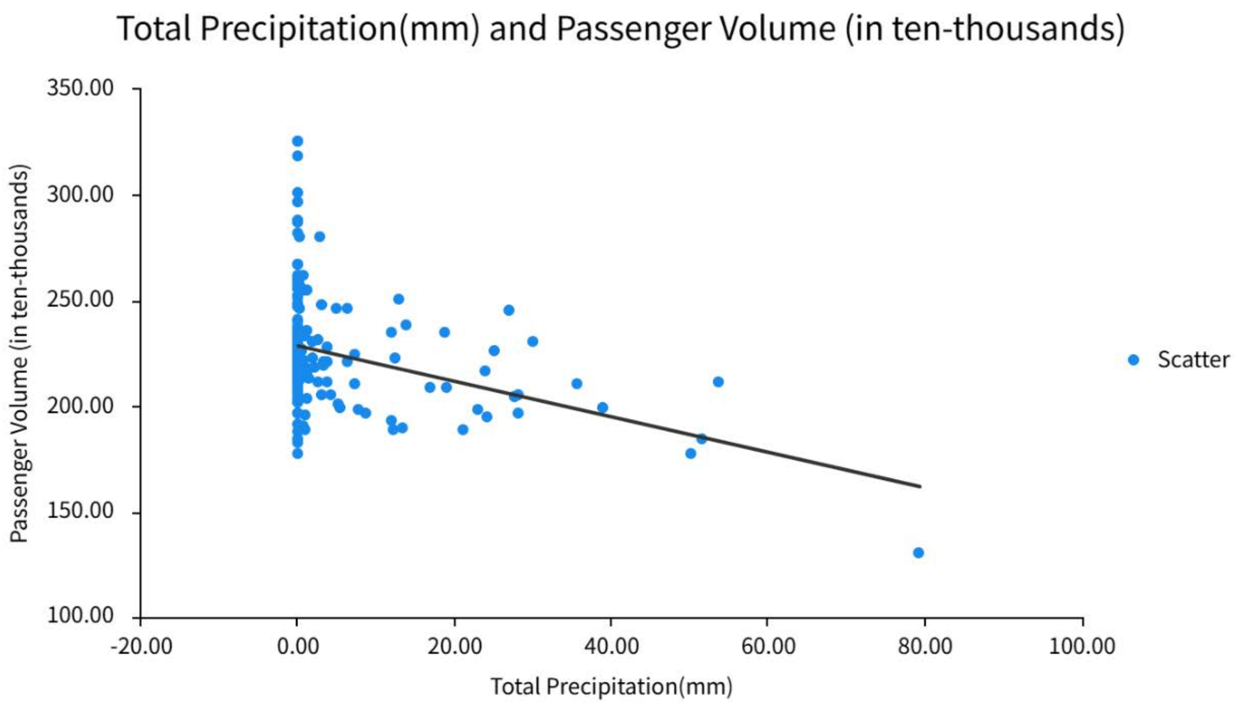

Fig. 4 Relationship between total precipitation and passenger volume (Picture credit: Original)

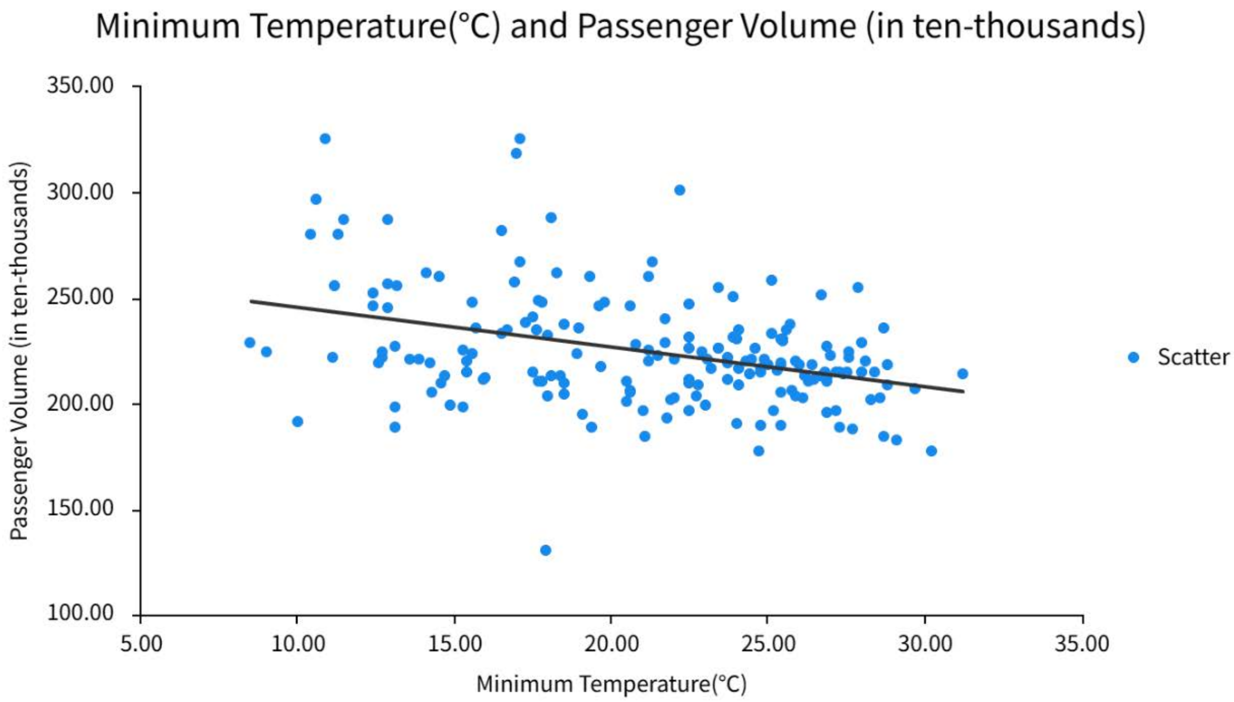

Fig. 5 Relationship between minimum temperature and passenger volume (Picture credit: Original)

Figure 2, Figure 3, Figure 4 and Figure 5 respectively show the relationship between average wind speed (m/s), total precipitation (mm), minimum temperature (°C) and passenger flow. The correlation between each variable and passenger flow can be seen simply through a linear trend.

3.2 The Results of the Correlation Analysis

Table 2. Pearson correlation

	Passenger volume(in ten-thousands)
Minimum temperature (°C)	-0.368**
Avg wind speed (m/s)	0.199**
Total precipitation (mm)	-0.342**

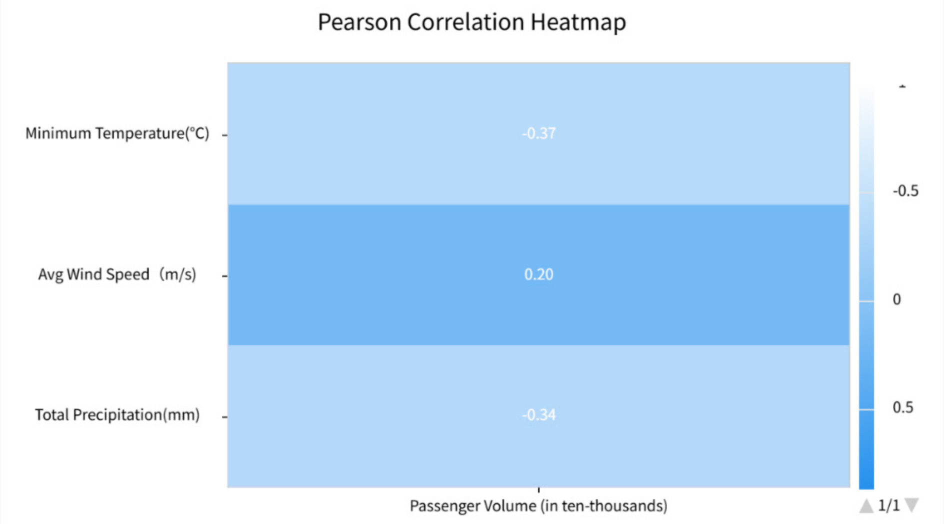

Fig. 6 Pearson related thermodynamic diagram (Picture credit: Original)

In order to explore the correlation between related variables, this study uses the Pearson correlation coefficient for correlation analysis, and the value range is [-1,1]. The closer the absolute value of the Pearson coefficient is to 1, the greater the correlation is, and the closer it is to 0, the smaller the correlation is. A positive value indicates a positive correlation. When one variable increases (decreases), the other variable also increases (decreases); Negative values indicate negative correlation. When one variable increases (decreases), the other variable decreases (increases).

It is observed from Table 2 and Figure 6 that the Pearson coefficient of the minimum temperature is a complex number, indicating that with the increase of the minimum temperature, the daily passenger volume shows a decreasing trend; The coefficient of average wind speed is positive, indicating that with the increase of average wind speed, the daily passenger volume shows an increasing trend; The coefficient of total precipitation is also complex, indicating that with the increase of daily total precipitation, the daily passenger flow shows a decreasing trend.

3.3 Results of Multiple Linear Regression Analysis

Table 3. Results of linear regression analysis (n=183)

	Denormalization coefficient		Standardization coefficient	t	p	Collinearity diagnosis	
	B	Standard error	Beta			VIF	Tolerance
constant	264.489	9.832	-	26.901	0.000**	-	-
Minimum temperature (°C)	-1.834	0.345	-0.358	-5.311	0.000**	1.095	0.913
Average wind speed (m/s)	1.038	1.685	0.042	0.616	0.539	1.123	0.890

ISSN 2959-6157

Total precipitation (mm)	-0.831	0.161	-0.338	-5.166	0.000**	1.029	0.972
R 2	0.256						
AdjustR 2	0.243						
F	F (3,179)=20.506, p=0.000						
D-W value	1.375						

Note: dependent variable=passenger volume (in ten-thousands)

* p<0.05 ** p<0.01

As shown in Table 3, it can be easily seen from the table that the p-value of the average wind speed is 0.539, which is significantly greater than 0.05, so the average wind speed has no significant impact on the daily passenger flow. Next, this study will use a stepwise regression analysis model for further analysis.

3.4 Results of Stepwise Regression Analysis

Table 4. Results of stepwise regression analysis (n=183)

	Denormalization coefficient		Standardization coefficient	t	p	Collinearity diagnosis	
	B	Standard error	Beta			VIF	Tolerance
constant	268.596	7.214	-	37.232	0.000**	-	-
Minimum temperature (℃)	-1.897	0.329	-0.371	-5.758	0.000**	1.000	1.000
Total precipitation (mm)	-0.848	0.158	-0.345	-5.354	0.000**	1.000	1.000
R2	0.254						
Adjust R2	0.246						
F	F (2,180)=30.675, p=0.000						
D-W value	1.369						

Note: dependent variable=passenger volume (in ten-thousands)

* p<0.05 ** p<0.01

As shown in Table 4 above, this study used stepwise regression analysis to exclude the non-significant influence factor of average wind speed, and used passenger flow as the dependent variable, and the minimum temperature and total precipitation as the independent variables for regression analysis. The p-value of total precipitation and minimum temperature is zero, indicating that the two independent variables have a significant impact on passenger flow. Then can accurately know from the positive and negative of the regression coefficient b that the total precipitation and the minimum air temperature are negatively correlated with the passenger flow, and when the total precipitation and the average temperature increase, the passenger flow will be relatively reduced. From the collinearity VIF values in the table are less than 5, it can be concluded that the independent variables selected in this study do not have the problem of collinearity. Finally, from R2 of 0.254, it can be concluded that the weather indicators selected in this study can only explain 25.4% of the changes in passenger volume, and nearly three-quarters of the changes are caused by other reasons.

4. Conclusion

This paper focuses on the impact of weather on urban subway passenger flow. Taking Zhengzhou as an example, it deeply studies the correlation between a variety of weather factors (such as temperature, precipitation, and wind speed) and the fluctuation of subway daily passenger flow. This study collected the daily weather data of Zhengzhou from April to September 2025 and the subway passenger flow data of the same period, and integrated them through time. First, Pearson correlation analysis was used to screen variables; Secondly, multiple linear regression analysis was used to separate the variables and screen them one by one; Finally, stepwise regression analysis was used to remove the factors without significant correlation.

Through correlation analysis, multiple linear regression analysis and stepwise regression analysis, this study concluded that among the selected weather factors, precipitation and temperature were the biggest factors affecting the subway passenger flow. With the increase in precipitation and minimum temperature, the number of people taking the subway decreased. However, the impact of the data selected in this study on the subway passenger flow is only 25.4%, only about a quarter.

This study has great limitations, the data is not carefully differentiated, there is no time period, only the number of days; There is no distinction between holiday and weekday passenger flow. Future research can expand the number of samples and take into account factors such as large-scale social activities and holidays. In terms of research methods, machine learning models can be further used to obtain more accurate data.

The research results can be used as a reference for subway operation scheduling and refined management, and help to improve and optimize the urban public rail transit system.

References

[1] Lyu Y. Short-term passenger flow forecast of urban subway transportation based on deep learning methods. Journal of Physics: Conference Series, 2021, 1915(2): 022064.

[2] Xiao L, Lo S, Zhou J, Liu J, Yang L. Predicting vibrancy of metro station areas considering spatial relationships through graph convolutional neural networks: The case of Shenzhen, China. Urban Analytics and City Science, 2020, 0(0): 1-22.

[3] Sun Yingbao. Research on passenger flow forecasting of urban rail transit under snowfall. Shijiazhuang Railway University, 2024.

[4] Xu Bo, Zhang Nan, Wei Juntao, Du Mengmeng, Zhang Gaojian. Research on the impact of rainfall on the fluctuation of rail transit passenger flow. Shaanxi Meteorology, 2023, (06): 59-63.

[5] Jiang Y, Gao Y, Yuan Q, et al. The combined influence of extreme weather and sea-level rise on water damage around the entrance of rail transit in coastal areas. Journal of Ocean Engineering and Marine Energy, 2025, 11: 327-338.

[6] Ngo N S, Bashar S. The impacts of extreme weather events on U.S. public transit ridership. Transportation Research Part D: Transport and Environment, 2024, 113: 104185.

[7] Adduōğlu I, Ergün M. Evaluating multimodal travel choices through MGWR: Insights from the Ankara-Istanbul corridor. Travel Behaviour and Society, 2025, 32: 100789.

[8] Fondzenyuy S K, Dahdah S, Neki K, Usami D S, Burlacu A F, Feudjio S L T, Persia L. Developing operating speed prediction models in mixed traffic environments using multiple linear regression and multilevel modeling approaches: Evidence from five low- and middle-income countries. Results in Engineering, 2025, 27: 107799.

[9] Brumercikova E, Bukova B. The regression and correlation analysis of carried persons by means of public passenger transport of the Slovak Republic. Transportation Research Procedia, 2020, 44: 61-68.

[10] Yang Xiaohong. Correlation analysis between urban rail transit and urban tourism. Urban Rail Transit Research, 2024, 27(1): 21-22.

[11] Metro Passenger Flow Data. Daily passenger flow data of Zhengzhou metro. 2025. Retrieved from https://www.metrolinehub.com/zh

[12] Soil and Water Conservation Archive. Historical weather data of Zhongyuan District in October 2025. 2025. Retrieved from https://www.yanshou100.com/