

Civil Engineering Journal

Vol. 6, No. 2, February, 2020



Weather Impact on Passenger Flow of Rail Transit Lines

Yongqing Guo^{a, c}, Xiaoyuan Wang^{b, c*}, Qing Xu^d, Shanliang Liu^b, Shijie Liu^a, Junyan Han^a

^a School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255000, China.

^b College of Electromechanical Engineering, Qingdao University of Science & Technology, Qingdao 266000, China.

^c Joint Laboratory for Internet of Vehicles, Ministry of Education - China Mobile Communications Corporation, Tsinghua University, Beijing 100048, China.

^d School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China.

Received 17 September 2019; Accepted 06 January 2020

Abstract

Passenger flow prediction is important for the planning, design and decision-making of urban rail transit lines. Weather is an important factor that affects the passenger flow of rail transit line by changing the travel mode choice of urban residents. A number of previous researches focused on analyzing the effects of weather (e.g. rain, snow, and temperature) on public transport ridership, but the effects on rail transit line yet remain largely unexplored This study aims to explore the influence of weather on ridership of urban rail transit lines, taking Chengdu rail transit line 1 and line 2 as examples. Linear regression method was used to develop models for estimating the daily passenger flow of different rail transit lines under different weather conditions. The results show that for Chengdu rail transit line 1, the daily ridership rate of rail transit ecreases with increasing temperature. While, for Chengdu rail transit line 2, the daily ridership rate of rail transit decreases with increasing wind power. The research findings can provide effective strategies to rail transit operators to deal with the fluctuation in daily passenger flow.

Keywords: Weather Effect; Rail Transit Line; Passenger Flow; Estimation Model.

1. Introduction

The passenger flow estimation of urban rail transit is widely used as the foundation for the planning, design, and daily operations of rail transit. Weather can influence people's travel behavior and traffic safety, and then have an impact on passenger flow of rail transit line. For example, rain is considered one of the most common adverse weather that may lead to change or cancellation of trips. But, weather factors are not usually presented in the existing models for estimating rail transit line ridership, which results in an insufficient or excessive estimation in the design stage, and unexpected large fluctuations in operation stage. It is essential to identify the impacts of weather factors on passenger flow of rail transit line. The relevant research mainly includes three aspects: data preprocessing of passenger flow [1, 2] quantitative analysis of impact factors [3-7] and development of estimation models [8, 9].

Several studies have explored the effects of rain and snow on public transit ridership. Inclement weather has an impact on people's travel modes and travel routes, and further effects on passenger flow in public transport [10, 11]. Changnon [12] found that summer rain days have a reduced number of passengers using public buses compared to summer sunny days. Cravo et al. [13] found that rain and snow have negative impacts on passenger flow of bus and subway. Guo et al.

* Corresponding author: wangxiaoyuan@qust.edu.cn

doi) http://dx.doi.org/10.28991/cej-2020-03091470



© 2019 by the authors. Licensee C.E.J, Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).

Civil Engineering Journal

[14] investigated the impact of weather elements, and revealed that rain has a negative impact on bus and rail ridership. Zhou et al. [15] found that the negative impact of rain on bus ridership appears obvious during off-peak time, and no significant effect shows during peak hours.

Arana et al. [16] showed increasing temperature leads to an increase in bus ridership on weekends. However, other studies show inconsistent results regarding the impact of temperature on passenger flow. For example, Kashfi et al. [17] exhibited no obvious relationship between temperature and bus ridership. Stover et al. [18] found that snow temperature has no obvious impact on bus passengers.

Some studies focused on developing the estimation models for daily passenger flow. Cravo et al. [13] used a crosssectional regression model to determine the impact of weather on New York City Transit's daily ridership. Stover et al. [18] used the least square methods to analyze the impact of weather on bus ridership. Zhao et al. [19] identified the impact of weather factors on passenger flow rate using multiple linear regression.

Most existing research has focused on the effects of weather (e.g. rain, snow, and temperature) on public transport ridership, but the effects on rail transit remain largely unexplored. On the other hand, little evidence is available to examine the weather-transit ridership in different lines of urban rail transit. Hence, there is a need for transport scholars to begin to determine the effects of weather on rail transit ridership in multiple lines. This study will identify the impacts of weather on ridership of rail transit line, in the two rail transit lines in Chengdu. Six months of data will be used to model the relationship between weather and ridership with linear regression method.

2. Research Methodology

2.1. Data Collection

2.1.1. Passenger Flow Data and Weather Data

The datasets used in this study consist of two types: passenger flow data and weather data. The passenger flow data covering a period from Jan 01, 2017 to Jun 01, 2017 was obtained from China Railway Corporation official website. It should be noted that Chengdu opened a new rail transit line on Jun 02, 2017, which led to a significant change of passenger flow. Therefore, the 2017 data after Jun 02 was excluded. The average daily passenger flow was calculated by weekday, weekend and holiday, shown in Table 1. The data were collected from two Chengdu rail transit lines, lines 1 and 2 (see Figure 1). It was observed that the daily average passenger flow is higher on weekday than on weekend, and weekend is higher than holiday in the two lines.



Figure 1. Chengdu rail transit lines 1 and 2

I inc	Average of	laily passenger flow	(thousand)
Line	Weekdays	Weekends	Holidays a
Line 1	722	517	391
Line 2	663	628	543

Table 1. Passenger flow of rail transit in the two lines

Note: a-Include seven national holidays in China, see Table 2 for details.

The meteorological data form Jan 01, 2017 to Jun 01, 2017 were gained from the World Weather Online [20], including daily temperatures (°C) (the lowest and highest), weather conditions (sunny, cloudy, rainy and snowy), wind speed (m/s), cloud fraction (%), precipitation (mm), air pressure (bar) and humidity (%). The flowchart of this research method is available in Figure 2.



Figure 2. The flowchart of the research method

2.2. Data Processing

2.2.1. Passenger Flow of Rail Transit Line

Before analyzing the relationship between passenger flow and weather, it is necessary to clean raw data. Daily passenger flow of rail transit line is affected by many factors such as holidays, large-scale events, and emergencies. The purpose of data cleaning is not only to detect and correct incomplete or inaccurate records, but also to reduce or eliminate the effects of other factors except weather.

First, it is important to identify the outliers of passenger flow data. Outliers are observations that lie abnormal distances from other values. Fig. 3shows the daily passenger flow of Chengdu rail transit line 1 between 01/01/2017 and 01/06/2017. It was observed that several obvious outliers are presented at the end of January and the beginning of February. This time period is exactly in the longest national holiday in China (Lunar New Year). China has seven national holidays (shown in Table 2), in which the passenger flow varies significantly. Therefore, the holiday effect needs to be removed from passenger flow data.



Figure 3. Daily passenger flow of Chengdu rail transit line 1 in 2017 (Jan 01-Jun 01)

Holidays in China	Date	Total days
New Year's Day	2016.12.31~2017.01.02	3
Lunar New Year	2017.01.27~2017.02.02	7
Qingming Festival	2017.04.02~2017.04.04	3
May Day	2017.04.29~2017.05.01	3
Dragon Boat Festival	2017.05.28~2017.05.30	3
Mid-autumn Festival and National Day	$2017.10.01\!\sim\!2017.10.08$	8

 Table 2. Holiday schedule in 2017

T I I I I	1	1 4	•	1 1 0	64 1 1* 1
Table 4 Standard	deviations of	data secilen	res removing	davs before	or after holiday
rubic 5. Standard	uc mations of	uata seguen	cos i emoving	uays belore	or arter nonuay

T in a	Denne alimine te d	0	1	2	3	4
Line	Days eliminated	SD	SD	SD	SD	SD
	Before holiday	14.82	14.61	14.67	14.82	14.97
Line 1	After holiday	14.82	14.96	15.13	15.19	15.33
	Before and after holiday	14.82	14.74	14.98	15.21	15.55
	Before holiday	9.38	8.85	8.85	8.96	9.08
Line 2	After holiday	9.38	9.46	9.57	9.63	9.75
	Before and after holiday	9.38	8.92	9.03	9.20	9.47

Note: SD- Standard Deviation

It was also noticed that significant changes in passenger flow usually present in several days before or after a holiday. To determine the days, the standard deviations of data sequences were compared. Standard deviation is the measure of dispersion of a set of data from its mean. The smaller the standard deviation, the lower the dispersion. Results show that the standard deviation of the dataset decreases, excluding the first day before holiday, and the first day before and after holiday, compared to the original dataset. When two or more days before and after holiday are removed, the standard deviations do not change much. For example, the standard deviation of line 1 reduces from 14.82 to 14.61 after removing the first day before holiday. And then, the standard deviation begins to slightly increase after elimination of two or more days before holiday. Therefore, to keep the original data as much as possible, only data of the first day before holiday were selected to eliminate (shown in Table 3).

2.2.2. Weather Data

In this study, weather data was cleaned in the following three aspects:

- (1) Eliminate the data whose corresponding passenger flow data were removed.
- (2) Eliminate the data with missing values.
- (3) Eliminate the data of extreme weather events, such as hailstorm.

2.3. The Estimation Model for Daily Passenger Flow

2.3.1. Correlation Analysis between Weather and Passenger Flow of Rail Transit Lines

Because the passenger flow distribution presents a significant difference on weekend and on weekday, the flow data of weekend and weekday were separated. Scatterplots were used to analyze if there are relationships between weather factors and passenger flow of rail transit line. In Figure 4, it was found that temperature (the average of the highest and lowest temperature) has an effect on passenger flow of rail transit line 1, as well as wind speed has an influence on passenger flow of rail transit line 2. It was also observed that the weekend flow is more likely to be affected by weather than the weekday flow.

2.3.2. Models of Daily Passenger Flow Estimation

The daily passenger flow presents a periodic fluctuation between weekday and weekend, and obvious changes in holidays (shown in Figure 3). Thus, as developing the daily passenger flow estimation model, the day factor (DF, weekday and weekend) and holiday factor (HF) should be included. Otherwise,



Figure 4. Effects of temperature and wind speed on passenger flow of Chengdu rail transit lines

Temperature Factor (TF) and Wind Power Factor (WPF); it was converted by wind speed) should be also included due to their impacts on passenger flow.

Based on Singhal's study [1], a multiple linear regression model was established for estimating daily passenger flow in Chengdu rail transit lines. In the model, the dependent variable is daily passenger flow, and the independent variables include the four influencing factors, as shown in Equation 1.

$$DR = \alpha + \beta_1 DF_{c,b} + \beta_2 HF_{c,d} + \beta_3 TF_{c,t} + \beta_4 WPF_{c,wp}$$
⁽¹⁾

Where DR estimated daily passenger flow of rail transit is lines; α is model constant. β_1 , β_2 , β_3 , β_4 is coefficient for estimation. DF_{c,b} is day factor for rail transit line $\in \{\text{line 1}, \text{line 2}\}$ for day type $b \in \{\text{Mon, Tues, Wed, Thu, Fri, Sat, Sun}\}$. HF_{c,d} is holiday factor for rail transit line c during holiday d. TF_{c,t} is temperature factor for rail transit line c during temperature interval $t \in [0, 10) \cup [10, 20) \cup [20, 30)$. WPF_{c,wp} is wind power factor for rail transit line c having a wind power wp $\in \{\text{light air and lower, light air to light breeze, gentile to moderate breeze}\}$.

Civil Engineering Journal

Based on Kashfi's study [21], day factor ($DF_{c,b}$) and holiday factor ($HF_{c,d}$) for two rail transit lines were calculated by Equations 2 and 3 respectively, and the calculated values are shown in Table 4.

$$DF_{c,b} = \frac{\sum_{i=1}^{N_b} DR_{c,b,i}}{N_b DR_{c,av}}$$
(2)

Where $DR_{c,b,i}$ is original daily passenger flow of rail transit line c on day i of day type b. N_b is the number of relevant days of day type b. $DR_{c,av}$ is original average daily passenger flow of rail transit line c.

$$HF_{c,d} = \frac{\sum_{i=1}^{N_d} DR_{c,d,i}}{N_d DR_{c,av}}$$
(3)

Where $DR_{c,d,i}$ is original daily passenger flow of rail transit line c on day i of holiday d. N_d is the number of relevant days of holiday d.

Temperature factor $(TF_{c,t})$ and wind power factor $(WPF_{c,tw})$ for the two rail transit lines were calculated by Equations 4 and 5 respectively, and the calculated values are demonstrated in Table 5.

$$TF_{c,t} = \frac{\sum_{i=1}^{N_t} DR_{c,t,i}}{N_t DR_{c,av}}$$
(4)

Where $DR_{c,t,i}$ is original daily passenger flow of rail transit line c on day i for a given temperature interval t. N_t is the number of relevant days occurring in a temperature interval t.

$$WPF_{c,wp} = \frac{\sum_{i=1}^{Nwp} DR_{c,wp,i}}{N_{wp} DR_{c,av}}$$
(5)

Where $DR_{c,wp,i}$ = original daily passenger flow of rail transit line c on day i for a given wind power wp; N_{wp} = the number of relevant days having a wind power wp.

Table 4. Calculated values of day factor $(DF_{c,b})$ and holiday factor $(HF_{c,d})$ for two lines

Lino		V	Weekday	7		Wee	kend		Holiday				
Line	Mon	Tues	Wed	Thu	Fri	 Sat	Sun	-	New Year's	Spring	Qingming	May	Dragon Boat
Line 1	1.15	1.14	1.15	1.14	1.19	0.84	0.76		0.75	0.37	0.79	0.85	0.74
Line 2	1.03	1.02	1.03	1.02	1.13	0.99	0.94		1.00	0.51	1.12	1.14	1.00

Table 5. Calculated values of temperature factor ($TF_{c,t}$) and wind power factor ($WPF_{c,wD}$) for two lines

Line -	ŋ	Femperature (°C)		Wind power	
	0~10	10~20	20~30	1	2	3
Line 1	1.04	1.06	1.07	1.03	1.06	1.05
Line 2	0.99	1.03	1.03	1.02	1.02	1.00

3. Results and Discussion

SPSS software was used to perform multiple linear regression, and stepwise regression method was used to eliminate non-significant variables. Table 6 shows the parameter estimates for the regression model, including the correlation coefficient (B), standard error (S. E), t-statistics, P value and R² value. R² is a goodness-of-fit measure for the model: the higher the R² value, the better the estimation model fits the passenger flow data. If R² value is more than 0.8, this value is generally considered strong effect size. If |t| is greater than 1.96 at a 5% significance level, we are 95% confident that the variable has a significant impact on the daily passenger flow in rail transit lines, otherwise the variable will be eliminated.

It was seen that the day and holiday variables have high coefficients, which indicates that they play a dominant role in daily passenger flow (P < 0.001). The temperature has a significant impact on passenger flow for rail transit line 1 (P < 0.05), and the wind power has a significant impact on passenger flow for rail transit line 2 (P < 0.05). According to the |t| value, the variables with a |t| value less than 1.96 were eliminated (see Table 6).

Line 1	В	S. E	t	Р	Line 2	В	S. E	t	Р
Constant	-41.86	6.33	-5.98	0.000	 Constant	-17.92	4.90	-4.48	0.000
DF _{c,b}	47.68	2.83	18.96	0.000	DF _{c,b}	31.15	5.02	6.01	0.000
HF _{c,d}	54.66	3.70	16.38	0.000	HF _{c,d}	51.55	4.84	10.45	0.000
TF _{c,t}	2.97	6.74	2.07	0.042	TF _{c,t}	-6.03	6.58	- 0.99	0.324
WPF _{c,wp}	7.12	10.29	1.36	0.177	WPF _{c,wp}	-1.20	2.87	-1.99	0.047
	R ²		R ²	=0.852					

Table 6. Fitting results of regression model

Note: Significant level $\alpha = 0.05$

From the results of regression model fitting, R^2 values with more than 0.8 indicate that the model produced a high goodness of fit. In order to improve the accuracy of passenger flow estimation, the daily passenger flow estimation models for Chengdu rail transit line 1 and line 2 were established respectively, based on the results in Table 6, as shown in Equations 6 and 7.

Line 1: DR =
$$\alpha_1 + \beta_{11} DF_{c,b} + \beta_{12} HF_{c,d} + \beta_{13} TF_{c,t}$$
 (6)

Line 2:
$$DR = \alpha_2 + \beta_{21}DF_{c,b} + \beta_{22}HF_{c,d} + \beta_{23}WPF_{c,wp}$$
 (7)

For Chengdu's rail transit line 1 and line 2, the estimates using the newly-developed models were compared with the original ridership rates. The estimated daily passenger flow, calculated using Eq. (6) for line 1 and Eq. (7) for line 2 respectively, were compared with the original daily passenger flow (shown in Figs. 5and 6). It was observed that in general, the estimates for the two lines are in close proximity to the actual values. Line 1 primarily provides commuter service to people, including businessmen, white-collar, IT elite, and business travellers. The target passenger flows. Weekly patterns are repeated within each month block, causing some peaks and troughs from the actual data were not fully captured. However, Line 2 shows subtle differences between weekday and weekend passenger flows. Similar to Line 1, some peaks and troughs from the actual data were not fully captured.



Figure 5. Time series graph of actual daily passenger flow trend & fitting of Model for rail transit line 1



Figure 6. Time series graph of actual daily passenger flow trend & fitting of Model for rail transit line 2

This study analyzed the effects of temperature and wind power on daily passenger flow in rail transit lines. It was found that temperature has a significant impact on passenger flow for Chengdu rail transit line 1, and wind power has a significant impact on passenger flow for Chengdu rail transit line 2.

3.1. Temperature

In the temperature ranging from 20 to 30°C, temperature has a significant impact on daily passenger flow of Chengdu rail transit line 1, and the correlation coefficient is positive. The results indicate that the daily passenger flow of rail transit line 1 increases with an increase of temperature.

3.2. Wind Power

Within the wind power ranging from level 1 to level 2, wind power has a significant impact on daily passenger flow of Chengdu rail transit line 2, and the correlation coefficient is negative. The results indicate that the daily passenger flow of rail transit line 2 reduces as wind power increases.

4. Conclusion

This study performed large-scale data analysis on the data of daily passenger flow and weather elements to explore the impacts of weather factors on usage of rail transit line. The daily ridership estimation models were established under different weather conditions for different rail transit lines. The results show that in Chengdu, the increase in temperature is associated with increasing ridership in rail transit line 1, and the increase in wind power is associated with decreasing ridership in rail transit line 2. These findings provide rail transit operators with valuable information to deal with daily passenger flow fluctuation related to varying weather conditions.

5. Funding

This study was supported by the Joint Laboratory for the Internet of Vehicles, Ministry of Education – China Mobile Communications Corporation under Project (Grant No. ICV-KF2018-03), Qingdao Top Talent Program of Entrepreneurship and Innovation (Grant No. 19-3-2-8-zhc), the National Natural Science Foundation of China (Grant Nos. 71901134, 61074140, 61573009, 51508315), the Natural Science Foundation of Shandong Province (Grant No. ZR2017LF015).

6. Conflicts of Interest

The authors declare no conflict of interest.

7. References

 Singhal, Abhishek, Camille Kamga, and Anil Yazici. "Impact of Weather on Urban Transit Ridership." Transportation Research Part A: Policy and Practice 69 (November 2014): 379–391. doi:10.1016/j.tra.2014.09.008.

- [2] Kalkstein, Adam J, Michael Kuby, Daniel Gerrity, and James J Clancy. "An Analysis of Air Mass Effects on Rail Ridership in Three US Cities." Journal of Transport Geography 17, no. 3 (May 2009): 198–207. doi:10.1016/j.jtrangeo.2008.07.003.
- [3] Keay, Kevin, and Ian Simmonds. "The Association of Rainfall and Other Weather Variables with Road Traffic Volume in Melbourne, Australia." Accident Analysis & Prevention 37, no. 1 (January 2005): 109–124. doi:10.1016/j.aap.2004.07.005.
- [4] Brijs, Tom, Dimitris Karlis, and Geert Wets. "Studying the Effect of Weather Conditions on Daily Crash Counts Using a Discrete Time-Series Model." Accident Analysis & Prevention 40, no. 3 (May 2008): 1180–1190. doi:10.1016/j.aap.2008.01.001.
- [5] Park, Byungkyu, Theresa K. Jones, and Stephen O. Griffin. Traffic analysis toolbox volume XI: weather and traffic analysis, modeling and simulation. No. FHWA-JPO-11-019. United States. Federal Highway Administration, 2010.
- [6] Liu, Chengxi, Yusak O. Susilo, and Anders Karlström. "Investigating the Impacts of Weather Variability on Individual's Daily Activity-travel Patterns: A Comparison Between Commuters and Non-Commuters in Sweden." Transportation Research Part A: Policy and Practice 82 (December 2015): 47–64. doi:10.1016/j.tra.2015.09.005.
- [7] Sohn, Keemin, and Hyunjin Shim. "Factors Generating Boardings at Metro Stations in the Seoul Metropolitan Area." Cities 27, no. 5 (October 2010): 358–368. doi:10.1016/j.cities.2010.05.001.
- [8] Taylor, Brian D., Douglas Miller, Hiroyuki Iseki, and Camille Fink. "Nature And/or Nurture? Analyzing the Determinants of Transit Ridership Across US Urbanized Areas." Transportation Research Part A: Policy and Practice 43, no. 1 (January 2009): 60–77. doi:10.1016/j.tra.2008.06.007.
- [9] Kusakabe, Takahiko, and Yasuo Asakura. "Behavioural Data Mining of Transit Smart Card Data: A Data Fusion Approach." Transportation Research Part C: Emerging Technologies 46 (September 2014): 179–191. doi:10.1016/j.trc.2014.05.012.
- [10] Gaudry, Marc, and André Arbic. "On the demand for and value of monthly transit passes within the framework of the demtec transit ridership demand model." Research paper, Agora Jules Dupuit, Publication AJD-21, University of Montreal, Canada (1996).
- [11] Outwater, Maren L., Greg Spitz, John Lobb, Margaret Campbell, Bhargava Sana, Ram Pendyala, and William Woodford. "Characteristics of Premium Transit Services That Affect Mode Choice." Transportation 38, no. 4 (April 22, 2011): 605–623. doi:10.1007/s11116-011-9334-0.
- [12] Changnon, Stanley A. "Effects of Summer Precipitation on Urban Transportation." Climatic Change 32, no. 4 (April 1996): 481–494. doi:10.1007/bf00140357.
- [13] Cravo, V. S., Cohen, J., Williams, A., "Impact of Weather on Transit Revenue in New York City". Transportation Research Board Meeting, 2009.
- [14] Guo, Zhan, Nigel H. M. Wilson, and Adam Rahbee. "Impact of Weather on Transit Ridership in Chicago, Illinois." Transportation Research Record: Journal of the Transportation Research Board 2034, no. 1 (January 2007): 3–10. doi:10.3141/2034-01.
- [15] Zhou, Meng, Donggen Wang, Qingquan Li, Yang Yue, Wei Tu, and Rui Cao. "Impacts of Weather on Public Transport Ridership: Results from Mining Data from Different Sources." Transportation Research Part C: Emerging Technologies 75 (February 2017): 17–29. doi:10.1016/j.trc.2016.12.001.
- [16] Arana, P., S. Cabezudo, and M. Peñalba. "Influence of Weather Conditions on Transit Ridership: A Statistical Study Using Data from Smartcards." Transportation Research Part A: Policy and Practice 59 (January 2014): 1–12. doi:10.1016/j.tra.2013.10.019.
- [17] Kashfi, Syeed Anta, Jonathan M. Bunker, and Tan Yigitcanlar. "Understanding the Effects of Complex Seasonality on Suburban Daily Transit Ridership." Journal of Transport Geography 46 (June 2015): 67–80. doi:10.1016/j.jtrangeo.2015.05.008.
- [18] Stover, Victor, and Edward McCormack. "The Impact of Weather on Bus Ridership in Pierce County, Washington." Journal of Public Transportation 15, no. 1 (March 2012): 95–110. doi:10.5038/2375-0901.15.1.6.
- [19] Zhao, Pengjun, Bin Lü, and Gert de Roo. "Impact of the Jobs-Housing Balance on Urban Commuting in Beijing in the Transformation Era." Journal of Transport Geography 19, no. 1 (January 2011): 59–69. doi:10.1016/j.jtrangeo.2009.09.008.
- [20] Available on: www.worldweatheronline.com/map/
- [21] Kashfi, Syeed Anta, Jonathan M. Bunker, and Tan Yigitcanlar. "Modelling and analysing effects of complex seasonality and weather on an area's daily transit ridership rate." Journal of transport geography 54 (2016): 310-324. doi:10.1016/j.jtrangeo.2016.06.018.