GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

International Journal of Engineering Advances

Article

Study on Spring Festival Passenger Flow Forecasting of Rail Transit in South China Based on Optimized SARIMA Model--Taking Guilin City as an Example

Xiang Wang [1,2], Jianqiu Chen [1], Xueyu Zhuge [2], Guoyan Che [2] and Guobin Gu [1,*]

[1] Guangxi Key Laboratory of International Join for China-ASEAN Comprehensive Transportation, Nanning University, Nanning, 530200, China
[2] Guangxi Key Laboratory of Intelligent Transportation System (ITS), Guilin University of Electronic Technology, Guilin, 541004, China
* Correspondence: Guobin Gu, Guangxi Key Laboratory of International Join for China-ASEAN Comprehensive Transportation, Nanning University, Nanning, 530200, China

Abstract: In this study, the optimized SARIMA model is used to forecast the spring rail transit passenger flow in Guilin City in 2019-2025, and external factors such as GDP and mobile population are introduced to improve the forecasting accuracy. First, the data range is extended to 11 years to enhance the model learning ability and reduce the impact of epidemics. Subsequently, the SARIMA parameters are optimized using grid search and AIC criterion to ensure the optimal fitting effect, and the prediction stability is improved by error control strategy. The experimental results show that the optimized SARIMA model performs well in short-term prediction, and the prediction error decreases year by year. Based on the prediction results, strategies such as mobility management, ticket optimization and capacity allocation are proposed to alleviate the pressure of passenger flow during the Spring Festival. This study provides a scientific basis for rail transportation capacity planning, which is of great significance to improve the transportation efficiency during the Spring Festival.

Keywords: SARIMA model; fassenger flow prediction; Spring Festival travel rush; time series forecasting; transportation capacity optimization

Received: 8 March 2025
Revised: 16 March 2025
Accepted: 27 March 2025
Published: 31 March 2025

1. Introduction

Rail transit plays a crucial role in China's transportation industry, especially in South China where the economy is developing rapidly [1-3]. China's Spring Festival has been called "the world's largest cyclical population migration", and the peak passenger flow during the Spring Festival every year puts tremendous pressure on the transportation system [4-6]. As an important tourism and transportation hub in South China, Guilin's rail transportation demand is especially prominent during the Spring Festival. However, with the growth of the city's economy and the increase of the mobile population, the further rise of the passenger flow during the Spring Festival has become a problem that cannot be ignored. In the face of the growing demand for Spring Festival, rail transportation capacity still has certain limitations. Therefore, how to accurately predict the future Spring Festival passenger flow and provide a scientific basis for rail transportation capacity deployment has become an urgent problem to be solved.

Traditional time series forecasting methods include ARIMA, exponential smoothing, etc. ArunKumar and other scholars compared ARIMA and SARIMA with deep learning-based RNN (GRU and LSTM) models in the application of COVID-19 forecasting by optimizing the parameters (p, d, q and P, D, Q) [7]. The results showed that in most countries, the deep-learning-based model had higher prediction accuracy and a much lower RMSE than ARIMA. However, when it comes to rail transit passenger flow prediction, especially in the face of unexpected factors like the epidemic, deep-learning models have some limitations, such as poor prediction stability.

In this study, we chose the SARIMA model instead of deep - learning methods like LSTM for the following reasons:

1) Limited data volume, with SARIMA being more stable and interpretable

The data used in this study has a limited time span (2019-2025), which is medium-to-short - term time - series data with a relatively small data volume. Deep-learning methods like LSTM often rely on a large number of training samples to effectively capture long-term dependency features. When the number of samples is insufficient, problems such as overfitting or unstable prediction are likely to occur. In contrast, the SARIMA model is more robust to small-sample data and can effectively model periodic and seasonal fluctuations, making it more adaptable to the seasonal patterns of rail transit passenger flow during the Spring Festival.

2) Transparent model structure with strong interpretability

When forecasting traffic passenger flow, external interference factors such as policy adjustments, holidays, and the epidemic need to be considered. In practical applications, model interpretability is of great significance. The parameters of the SARIMA model have clear statistical meanings, which is convenient for subsequent analysis and adjustment. However, the LSTM belongs to a "black-box model". Although it has high prediction accuracy in some scenarios, it lacks a transparent interpretation mechanism, which is not conducive to using model results for policy reference and practical applications.

3) Superior computational efficiency and implementation cost

The SARIMA model has high computational efficiency and is easy to tune parameters, making it suitable for prediction tasks with limited resources or those that require rapid deployment. In contrast, the LSTM model usually requires high-end computing resources, a long training period, and parameter tuning of the network structure, which increases the complexity of modeling [8].

4) SARIMA is more suitable for medium-to-short-term trend modeling considering the epidemic situation

In this study, the epidemic has significantly disturbed the passenger flow in some years. The SARIMA model has strong trend-removal and seasonal adjustment capabilities and can effectively model the trend changes before and after the epidemic through differencing and seasonal structures. However, LSTM's capture of non-linear mutations is highly dependent on the coverage of training data and is prone to prediction drift.

Based on the above reasons, this paper gives priority to using the SARIMA model for passenger flow prediction and analyzes and optimizes the model error. In subsequent work, we will also explore the hybrid modeling method of combining SARIMA and LSTM to improve the comprehensive performance of prediction.

2. Data Processing and Model Optimization

2.1. Data Collection and Pre-Processing

This study collects springtime rail traffic flows from all train stations in Guilin (including Guilin, Guilin North, Guilin West, Gongcheng, Yangshuo, Wutong, and Sanjiang South stations) from 2019-2025. Here, the data for 2025 is historical data, specifically the data from January 14th to February 22nd. As shown in Figure 1.

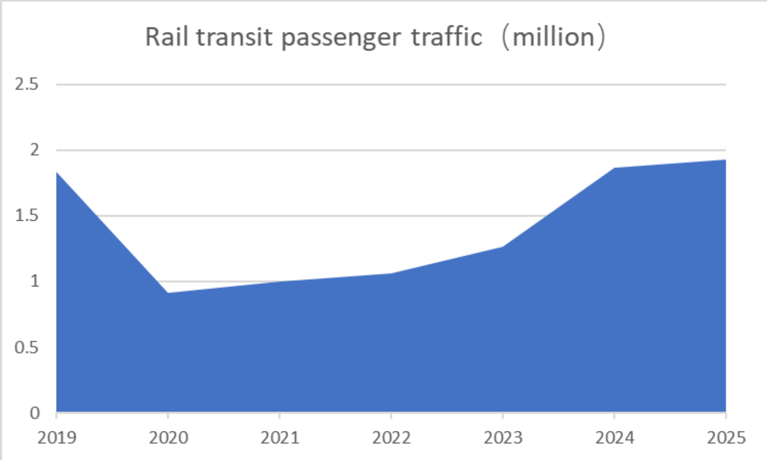

Figure 1. Spring traffic flow of Guilin rail transit, 2019-2025.

As can be seen from Figure 1, affected by the new crown epidemic, the spring passenger flow of Guilin rail transit has more ups and downs in 2019-2022, and shows a continuous growth in 2023-2025, which speculates that if it is not affected by the emergent situation, the spring passenger flow of Guilin rail transit will show a trend of continuous growth.

2.2. Model Optimization

In order to improve the accuracy of the SARIMA model in predicting the spring traffic flow of Guilin Railway, a series of optimization strategies are adopted in this paper. The optimization process mainly includes three aspects: data expansion, parameter adjustment and error control.

2.2.1. Data Extensions

The SARIMA model has a strong dependence on time series data, and a shorter data time span may lead to insufficient model training, which in turn affects the prediction accuracy. In order to solve this problem, this paper extends the original 2019-2025 data to 2015-2025, and adds the 2015-2018 Guilin Railway Spring Passenger Flow data, so that the data covers 11 years, which enhances the learning ability of the model.

It should be noted that there are certain deviations between the data from 2015 - 2018 and that from 2019 - 2022 due to the significant impact of the epidemic on passenger flow. In the actual data processing, no detrending or outlier treatment was performed on the data during the epidemic period, nor were there any special corrections. When expanding the data, we mainly introduced data from more years to balance the impact of abnormal years and enhance the model's ability to grasp the overall trend.

Advantages of extending the data:

1) Provide a more complete trend of passenger flow changes, allowing the model to better identify long-term trends and seasonality.
2) Reduces the interference of epidemics on model training, enabling the model to more reasonably predict the recovery of passenger flow after an epidemic.

Although the IPAT equation is generally used for environmental impact analysis, in this study, we use it to indicate the influence of factors such as population and affluence on passenger flow. However, it may be more appropriate to use more common transportation demand prediction models or introduce elasticity coefficient analysis in future research. For now, this study introduces the GDP as well as the number of mobile population in Guilin City from 2019-2025, as shown in Figure 2 and Figure 3.

$$I=a\times P^b\times A^c\times T^d\times e \quad (1)$$

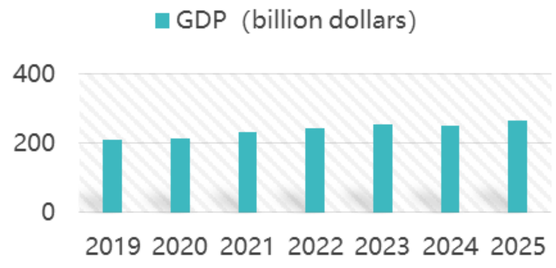

Figure 2. 2019-2025 Guilin City GDP.

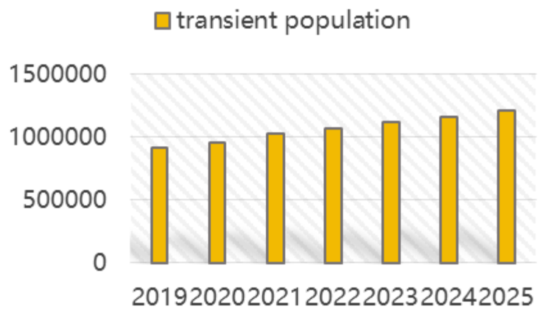

Figure 3. 2019-2025 Guilin City Mobile Population.

From Figure 2 and Figure 3, it can be seen that the GDP and mobile population of Guilin City from 2019 to 2025 both show a continuous upward trend. Among them, the mobile population rises year by year, with an average growth rate of about 4.2%. The overall program of GDP is on an upward trend, and the GDP in 2025 takes the planning value of Guilin.

2.2.2. Parameter Adjustment

After the data expansion, this paper adopts Grid Search + AIC (Akai Information Criterion) method to screen out the optimal SARIMA model parameters. The finalized model parameters are:

1) Non-seasonal part (p, d, q):(1,1,1), i.e., the first-order difference is used to eliminate the trend, and an autoregressive term and a moving average term are used to model the residuals.
2) Seasonal component (P, D, Q, s):(1,1,1,2), i.e., considering seasonal variations with a period of 2 years, first-order differencing is used to eliminate the seasonal trend and an autoregressive and a moving average term are added to model the seasonal effects.

Key points for parameter optimization:

1) AIC minimization: select the model with the smallest AIC through multiple sets of parameter tests to ensure the model has the best fitting effect.
2) Stability test: check the unit root test results of the model to ensure that the residuals are smooth to improve the predictive stability.

2.2.3. Error Control

Since the key objective of forecasting is to ensure a small error, this study adopts an iterative optimization strategy to continuously adjust the SARIMA parameters and optimize the model effect: firstly, the SARIMA model is initially trained to obtain the 2019-2025 forecasting results, and the errors (MSE, MAPE) are calculated. Then check the prediction error of 2023-2025, the error is considered large if the error is more than 10%, and adjust the parameters if the error is more than 10%.

3. Passenger Flow Forecast

3.1. Experimental Environment Configuration

The experimental environment uses Windows 11 as the operating system, PyTorch 1.11 as the deep learning framework, and Nvidia GeForce RTX 2080TI graphics card as the GPU to accelerate the training simulation process, which is shown in the following Table 1.

Table 1. Experimental environment configuration.

Item	Environment configuration and version	Note
operating system	Windows11	
CPU	Intel(R) Core(TM) i7-14650HX	Main frequency: 2.20 GHz
RAM	16GB	
ROM	512GB	
GPU	Nvidia GeForce RTX 4060	Memory Capacity: 7957MB
PyTorch	1.11	
CUDA	11.3	
Python	3.8.10	

3.2. Model Accuracy Validation

In order to verify the predictive accuracy of the optimized SARIMA model, this study used the model to predict the predicted values for 2019-2025 and compared them with the true values to derive the predictive accuracy of the model.

The prediction error is larger in 2019-2021 due to the epidemic, and the prediction error in 2022-2025 decreases year by year, which proves that the optimized SARIMA model has high accuracy in short-term prediction. The comparison of real and predicted values and the prediction accuracy are shown by Figure 4 and Figure 5.

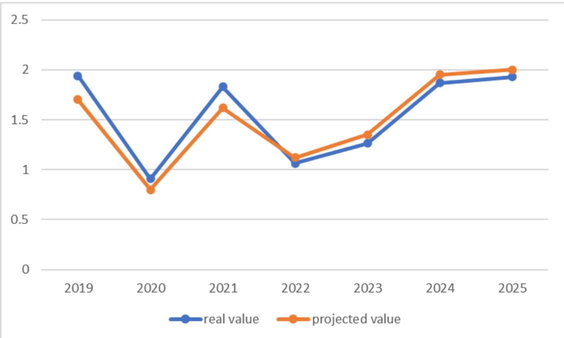

Figure 4. Comparison of real and predicted values.

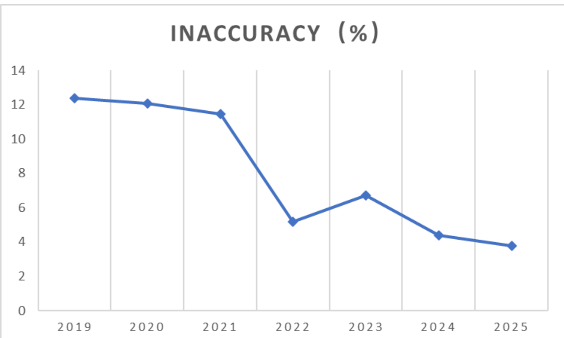

Figure 5. Prediction accuracy.

3.3. Forecast of Spring Passenger Traffic of Guilin Rail Transit in the Next Three Years

The passenger traffic volume during the Spring Festival of Guilin Railway Transportation in 2026-2028 predicted based on the optimized SARIMA model is shown in Figure 6.

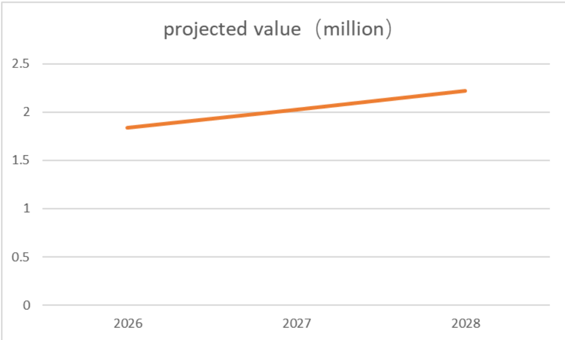

Figure 6. 2026-2028 Forecast value of passenger traffic during spring transportation of Guilin rail transit.

According to the development plan and development situation of Guilin, the GDP of Guilin will grow steadily in the next three years, and the mobile population will continue to grow, so the passenger volume of rail transportation during the Spring Festival will continue to rise in the next three years. From Figure 6, it can be seen that the predicted value of the model is in line with the rising trend.

3.3. Suggestions for Strategies to Reduce Passenger Pressure on Rail Transit

The passenger flow of Guilin rail transit will continue to grow in the future, especially during the Spring Festival, and there is no expansion plan regarding the Guilin Railway Station, therefore, this study proposes strategies to reduce the pressure of rail transit passenger flow in terms of future mobile population control and other aspects.

Mobile population management: Strengthen the research on urban population distribution, accurately predict the travel demand of the mobile population, and formulate the corresponding capacity deployment plan; promote staggered vacation of enterprises, and reduce the concentrated travel pressure during the Spring Festival peak period by encouraging enterprises to arrange vacations in batches; optimize the intercity passenger transport connection, and strengthen the linkage with highways and airlines, so as to provide more travel options for the mobile population.

Ticket optimization: implement time-sharing fares, and appropriately increase fares during peak periods to encourage passengers to travel in a staggered manner; expand advance ticket purchase preferential policies, such as discounts for tickets purchased more than 30 days in advance, to reduce pressure on ticket purchases during the approaching travel period; and optimize refund and reissuance policies to encourage passengers to adjust their itineraries ahead of time and reduce the need for temporary ticket purchases.

Rail transportation capacity deployment: increase temporary trains during the Spring Festival, and add temporary trains during peak periods, especially in the major migration directions from Guilin to Guangzhou, Nanning, Changsha, etc. Increase capacity; optimize the setting of departure stations, avoiding all trains being concentrated in Guilin Station, and diverting some long-distance passenger traffic through Guilin North Station and Guilin West Station, etc.; dynamically adjust the train configuration, and increase the number of moving train configurations and increase the number of single-ticket units appropriately during the Spring Festival. Dynamically adjust the train formation, appropriately increase the number of train sets during the Spring Festival, and improve the capacity of a single train.

5. Conclusion

The optimized SARIMA model performs well in the prediction of passenger flow of rail transit in Spring Festival, and the stability and prediction accuracy of the model are improved through data expansion, parameter optimization and error control, and the prediction accuracy is within 10%. Especially, it has high reliability in the prediction of passenger flow recovery trend after the epidemic.

External factors have a significant impact on the passenger flow of rail transportation and should be included in the prediction model. GDP growth and the increase of mobile population are important factors affecting the spring passenger flow, and the introduction of these variables in this study improves the accuracy of the prediction, and in the future, the weather, policy regulation and other factors can be further considered to optimize the planning and management of rail transit.

The passenger flow of Guilin City will continue to grow in the future Spring Festival, and the capacity deployment needs to be optimized in advance. Combined with the forecast results and urban development planning, rail transit will face greater pressure during the Spring Festival, and should be used to alleviate congestion by increasing temporary trains, optimizing ticketing policies, and strengthening the linkage of multiple modes of transportation.

Funding: This research was funded by Guangxi Science and Technology Base and Talent Project for Guangxi Science and Technology Plan Project: Construction of Guangxi Transportation New Technology Transfer Center Platform (No. Guike AD23026029); Construction of Scientific and Technological Innovation Bases for Science Research and Technology Development Plan Project of Nanning City: Intelligent Operation and Maintenance of Transportation Infrastructure and Health Monitoring Platform (No. 20223230).

References

1. M. Song, G. Zhang, W. Zeng, J. H. Liu, and K. Fang, "Railway transportation and environmental efficiency in China," *Transp. Res. Part D: Transp. Environ.*, vol. 48, pp. 488-498, 2016, doi: 10.1016/j.trd.2015.07.003.
2. X. Ding, K.-S. Choi, and X. Ji, "The impact of the China Railway Express on FDI inflows in Chinese prefecture-level cities," *Transp. J.*, vol. 62, no. 3, pp. 331-354, 2023, doi: 10.5325/transportationj.62.3.0331.
3. L. Zhang, Q. Cheng, and S. Qu, "Evaluation of railway transportation performance based on CRITIC-relative entropy method in China," *J. Adv. Transp.*, vol. 2023, pp. 1-11, 2023, doi: 10.1155/2023/5257482.
4. Y. Zou, W. Yang, J. Lai, J. Hou, and W. Lin, "Vaccination and quarantine effect on COVID-19 transmission dynamics incorporating Chinese-Spring-Festival travel rush: Modeling and simulations," *Bull. Math. Biol.*, vol. 84, no. 2, pp. 0-0, 2022, doi: 10.1007/s11538-021-00958-5.
5. J. Lin and K. Wu, "Intercity asymmetrical linkages influenced by Spring Festival migration and its multivariate distance determinants: A case study of the Yangtze River Delta region in China," *Humanit. Soc. Sci. Commun.*, vol. 10, no. 1, pp. 0-0, 2023, doi: 10.1057/s41599-023-02456-6.
6. Y. Zhang, N. Yu, M.-Y. Zhang, and Q. Yu, "Particulate matter exposures under five different transportation modes during Spring Festival travel rush in China," *Processes*, vol. 9, no. 7, p. 1133, 2021, doi: 10.3390/pr9071133.
7. K. E. ArunKumar, D. V. Kalaga, C. M. Sai Kumar, M. Kawaji, T. M. Brenza, "Comparative analysis of Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM) cells, Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA) for forecasting COVID-19 trends," *Alexandria Eng. J.*, vol. 61, no. 10, pp. 7585-7603, 2022, doi: 10.1016/j.aej.2022.01.011
8. M. A. Raza, A. Amin, M. Aslam, T. Nawaz, M. Irfan, and F. Tariq, "Nonparametric mixed exponentially weighted moving average-moving average control chart," *Sci. Rep.*, vol. 14, no. 1, p. 0, 2024, doi: 10.1038/s41598-024-57407-1.