

Spatiotemporal Evolution Patterns and Intelligent Forecasting of Passenger Flow in Megacity High-Speed Rail Hubs: A Case Study of Guangzhou South Railway Station

Kangni Liu

Guangzhou Railway Polytechnic, Guangzhou, China

Email: liukangni@gtxy.edu.cn

How to cite this paper: Liu, K.N. (2026) Spatiotemporal Evolution Patterns and Intelligent Forecasting of Passenger Flow in Megacity High-Speed Rail Hubs: A Case Study of Guangzhou South Railway Station. *Journal of Transportation Technologies*, 16, 142-149.
<https://doi.org/10.4236/jtts.2026.161009>

Received: December 5, 2025

Accepted: January 6, 2026

Published: January 9, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Flow in high-speed rail (HSR) hubs serves as a “barometer” for factor mobility within urban agglomerations, and its accurate forecasting is crucial for capacity allocation and emergency management. This paper focuses on two core aspects: passenger flow characterization and intelligent forecasting methodology. Taking Guangzhou South Railway Station (GSRS) as a typical case, it utilizes multi-source big data to deeply excavate the refined spatiotemporal distribution patterns and structural characteristics of hub passenger flow. Furthermore, a hybrid VMD-CNN-BiLSTM-Attention-XGBoost forecasting model integrating time series decomposition, deep learning, and ensemble learning is constructed. The study finds that passenger flow exhibits a pattern of “dual peaks on weekdays for commuting and a single peak on weekends for leisure”, with the Shenzhen/Hong Kong SAR direction accounting for over 30%. The constructed hybrid model demonstrates significantly superior forecasting accuracy (MAPE = 3.76%) compared to benchmark models. This research provides methodological and decision-making support for the transition of mega HSR hubs from “experience-based operation” to “data-driven” precise governance.

Keywords

High-Speed Rail Hub, Passenger Flow Characteristics, Spatiotemporal Patterns, Variational Mode Decomposition (VMD), Hybrid Deep Learning Model, Passenger Flow Forecasting

1. Introduction

As a critical node in China’s “Eight Vertical and Eight Horizontal” HSR network

and the core gateway of the Guangdong-Hong Kong SAR-Macao Greater Bay Area (GBA), Guangzhou South Railway Station serves over 350,000 passengers daily [1]. Minor fluctuations in its passenger flow can trigger significant ripple effects on urban transportation. Traditional passenger flow analysis often remains at the aggregate level or simple temporal statistics, failing to reveal the underlying complex spatiotemporal heterogeneity, purpose-based structure, and multi-factor driving mechanisms [2]; in terms of forecasting, single time-series or regression models struggle to effectively respond to multiple external shocks such as holidays, weather, and major events, leading to forecast failures at critical junctures [3].

Therefore, answering the following two core scientific questions is of urgent importance for enhancing hub operational resilience: First, what refined and quantifiable regular patterns do passenger flows in mega HSR hubs exhibit across spatiotemporal dimensions? Second, how can a high-accuracy forecasting model be constructed that simultaneously captures the intrinsic temporal patterns of passenger flow and external complex factors? This paper aims to systematically address these questions through multi-dimensional, long-time-series data analysis and modeling of GSRS, forming a replicable and scalable analytical and forecasting framework.

2. Research Framework and Methodology

This research follows the logical sequence of “Characterization Analysis-Pattern Mining-Model Construction-Forecasting Application”. First, multi-source data covering the period from January 2021 to December 2023 are utilized, including HSR ticket data (daily granularity), metro AFC data (hourly granularity), and urban Points of Interest (POI) data [2]. These datasets underwent anonymization, temporal alignment, and fusion using a rule-based matching approach based on temporal and spatial keys. Secondly, methods such as spatiotemporal heatmaps, cluster analysis, and OD linkage strength models are employed to deconstruct passenger flow characteristics from three dimensions: time, space, and structure. Finally, to address the nonlinear and non-stationary nature of passenger flow series, an innovative VMD-CNN-BiLSTM-Attention-XGBoost hybrid forecasting model is constructed and compared against benchmark models like ARIMA, Prophet, and single LSTM/XGBoost [4]-[6].

3. Multi-Dimensional Characterization of Passenger Flow at Guangzhou South Railway Station

3.1. Temporal Distribution: Multi-Level Periodicity and Shock Effects

As shown in **Table 1**, analysis reveals a stable “three-level periodicity” structure in the temporal distribution of passenger flow at GSRS [1]:

Daily Cycle: Exhibits a distinct pattern of “dual peaks on weekdays, single peak on weekends”. The weekday morning peak (8:00-10:00) is dominated by commut-

ers on the Guangzhou-Shenzhen and Guangzhou-Zhuhai corridors, while the evening peak (18:00-20:00) combines arriving and departing flows. Weekend peaks are more evenly distributed between 13:00 and 18:00.

Weekly Cycle: Passenger volume climbs from Monday, peaks on Thursday-Friday for weekdays, and reaches the weekly maximum on Saturday, approximately 25% higher than the average weekday volume.

Annual Cycle: Characterized by four major peak periods: the “Spring Festival Extreme Peak”, “Summer Transport Sub-peak”, “Short Holiday Pulse Peaks”, and the “Canton Fair Peak” [1].

Table 1. Passenger flow impact indices for key holidays at GSRS (2023).

| Holiday/Event | Peak Daily Passenger Volume (10,000 persons) | Increase vs. Normal Weekday | Passenger Flow Impact Index* |
|---------------------------------------------------------------------------|----------------------------------------------|-----------------------------|------------------------------|
| National Day (Oct 1) | 55.3 | 70.2% | 1.70 |
| Spring Festival Travel (28 th of Lunar 12 th Month) | 53.8 | 65.5% | 1.66 |
| Canton Fair (Phase I Opening Day) | 46.2 | 42.1% | 1.42 |
| Labor Day (May 1) | 49.1 | 51.0% | 1.51 |

Note: Impact Index = Peak Daily Volume/Monthly Average Volume.

3.2. Spatial Distribution: Directional Agglomeration and Transfer Choice

Spatial distribution shows strong “directional agglomeration” and “transfer dependency”, as shown in **Figure 1**.

OD Agglomeration: The Shenzhen/Hong Kong SAR direction accounts for the highest share (31.3%), forming, together with Zhuhai/Macao (19.0%) and Changsha/Wuhan (18.2%), the top three dominant flows, constituting nearly 70% of the

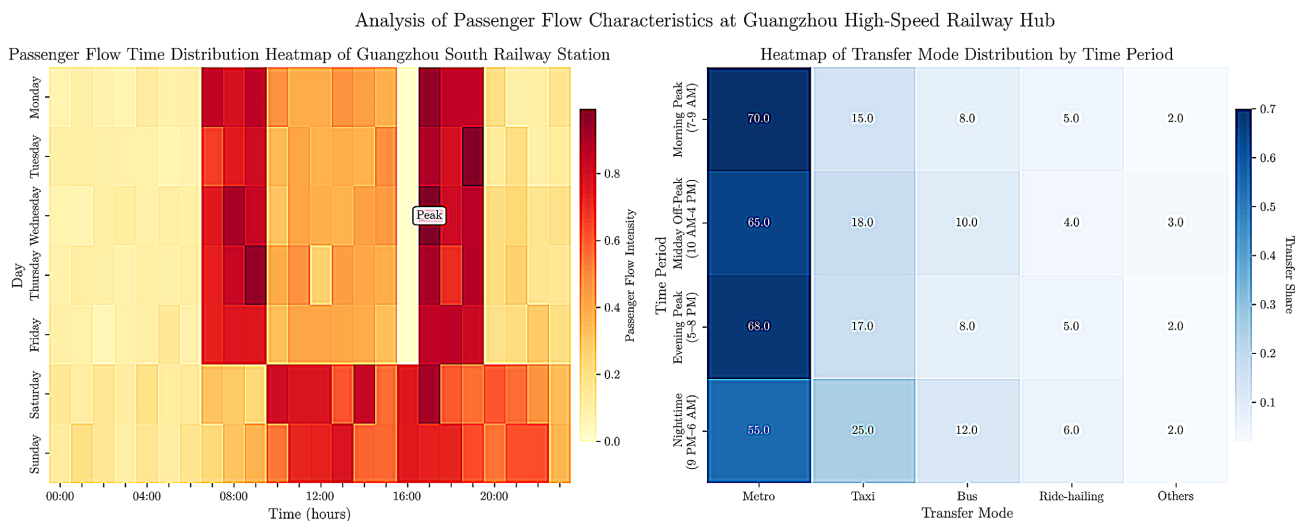


Figure 1. Spatiotemporal heatmap of passenger flow (Left) and major OD flow diagram (Right).

total and forming a stable “GBA Core Corridor” [2].

Transfer Structure: Metro is the overwhelmingly dominant transfer mode, with a share of 65.2%. Notably, the combined share of taxis and ride-hailing services can surge to over 35% during late-night hours (after 22:00) and under adverse weather conditions, demonstrating significant “spatiotemporal elasticity”.

3.3. Passenger Composition: Purpose Segmentation and Passenger Profiling

Passenger categorization was performed using a combination of K-means clustering (applied to features such as booking lead time, travel frequency, ticket class, and temporal patterns) and rule-based classification (e.g., same-day return trips on weekdays classified as business travel).

Through analysis of ticket class, booking lead time, and POI correlation, passengers are profiled in detail [2]:

Business Travelers (42.3%): High-frequency travelers between Guangzhou-Shenzhen/Zhuhai, concentrated on weekdays, highly sensitive to departure/arrival times, and with the lowest tolerance for transfer time.

Tourists (28.7%): Concentrated on weekends, holidays, and summer travel periods, often carrying luggage, showing greater concern for wayfinding signage and rest facilities within the hub.

Commuters (8.9%): Exhibit stable “tidal” characteristics and are a primary source of pressure on metro systems during peak hours.

4. A Hybrid Deep Learning Forecasting Model Based on VMD-CNN-BiLSTM-Attention and XGBoost

4.1. Overall Model Architecture and Core Innovations

Addressing the highly nonlinear, non-stationary nature of HSR hub passenger flow series and their complex influence by multiple external factors [3], this study proposes a combined forecasting framework integrating hybrid deep learning and ensemble learning. The model architecture was designed to sequentially address different aspects of the forecasting challenge: VMD handles non-stationarity and multi-scale patterns; CNN extracts local spatial-temporal features; BiLSTM captures long-term bidirectional dependencies; Attention focuses on relevant historical periods; and XGBoost integrates deep features with external variables for robust ensemble learning. The core innovation lies in its three-stage architecture of Decomposition-Reconstruction-Fusion.

1) Signal Decomposition Layer: Employs Variational Mode Decomposition (VMD) to adaptively decompose the original passenger flow series into a set of quasi-stationary sub-series [7].

2) Deep Forecasting Layer: For each decomposed sub-series, a CNN-BiLSTM-Attention neural network is designed for deep feature extraction and forecasting [6] [8].

3) Ensemble Output Layer: The deep forecasting outputs are combined with

external features and fed into an XGBoost model for nonlinear ensemble and residual correction [4].

The mathematical representation of the model is:

$$\hat{Y}_t = F_{XGBoost} \left(\left[IMF^1, \dots, IMF^K \right], E_t \right) + \epsilon_t$$

where IMF^K is the k-th modal component forecasted by the deep sub-network, E_t is the vector of external features at time t .

4.2. Stage 1: Passenger Flow Series Decomposition via VMD

VMD decomposes the original signal $f(t)$ into K band-limited Intrinsic Mode Functions (IMFs) $u_k(t)$ by solving a constrained variational problem [8]:

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \quad \text{s.t.} \quad \sum_{k=1}^K u_k = f(t)$$

The decomposition number K was determined through a combination of spectral analysis and trial evaluation on a validation set. $K=6$ was selected as it clearly separated the series into interpretable components: ultra-high-frequency noise, daily periodicity, weekly periodicity, holiday/monthly periodicity, seasonal trend, and long-term trend.

4.3. Stage 2: CNN-BiLSTM-Attention Forecasting Sub-Network for Each IMF Component

Each IMF component is input into an independent, identically structured deep neural network consisting of three layers:

- 1) 1D Convolutional Layer (1D-CNN)

Extracts local fluctuation patterns:

$$h_{conv} = \text{ReLU}(W_{conv} * x_{window} + b_{conv})$$

- 2) Bidirectional Long Short-Term Memory Layer (BiLSTM)

Comprehensively learns the dynamic evolution of each IMF within its full context:

$$H_t = \left[\vec{h}_t; \overleftarrow{h}_t \right]$$

- 3) Attention Mechanism Layer:

Dynamically weights historical time steps, allowing the model to focus on key historical periods relevant to the current forecast [8]: $C = \sum_t \alpha_t H_t$. The context vector C serves as the final deep feature representation for forecasting that IMF component.

4.4. Stage 3: Multi-Source Feature Fusion and XGBoost Ensemble

This stage aims to fuse the intrinsic temporal patterns extracted by the deep network with rich external influencing factors.

- 1) External Feature Engineering: A feature pool containing over 30 features

across 5 categories is constructed (temporal, historical, economic/event, weather, competing transport). Examples include: binary indicators for public holidays and Canton Fair periods; temperature, precipitation, and visibility as weather variables; and average ticket prices for competing transport modes (e.g., flights, coaches).

2) XGBoost Nonlinear Ensemble: The forecasted values of all IMF components are concatenated with the external feature vector as input to XGBoost. XGBoost learns the complex nonlinear mapping between these features and the final passenger flow target, performing residual correction [4]. Its objective function is:

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

4.5. Model Performance Evaluation and Comparative Analysis

The dataset was split into training (Jan 2021-Jun 2023), validation (Jul 2023-Sep 2023), and test (Oct 2023-Dec 2023) sets. Missing values were forward-filled, and outliers beyond three standard deviations were winsorized, as shown in **Table 2**. Evaluation on an independent test set shows the proposed model significantly outperforms benchmark models across multiple metrics [5], as shown in **Table 3**.

Table 2. Model hyperparameters and tuning strategy.

| Component | Hyperparameter | Value/Tuning Method |
|-----------|----------------|------------------------------|
| VMD | K (modes) | 6 (spectral validation) |
| CNN | Filters | 64 (grid search) |
| BiLSTM | Units | 128 (grid search) |
| Attention | Mechanism | Bahdanau (fixed) |
| XGBoost | Learning rate | 0.05 (Bayesian optimization) |
| | Max depth | 8 |
| | N estimators | 300 |

Table 3. Comprehensive comparison of model forecasting performance (test set results).

| Model | MAPE (%) | RMSE (10 k) | MAE (10 k) | R ² | Peak Forecast Error (%) |
|-----------------------------------------|----------|-------------|------------|----------------|-------------------------|
| ARIMA | 8.72 | 2.89 | 2.21 | 0.891 | 18.5 |
| Prophet | 7.15 | 2.45 | 1.92 | 0.922 | 15.2 |
| LSTM | 6.83 | 2.31 | 1.85 | 0.930 | 13.8 |
| XGBoost | 5.94 | 2.05 | 1.67 | 0.945 | 11.3 |
| VMD-LSTM | 5.12 | 1.78 | 1.45 | 0.958 | 9.7 |
| VMD-CNN-BiLSTM-Attention-XGBoost (Ours) | 3.76 | 1.32 | 1.08 | 0.977 | 6.4 |

Accuracy Improvement: Compared to the second-best model (VMD-LSTM), our model reduces MAPE by a further 26.6%.

Peak Forecasting Capability: For the National Day holiday peak contained in the test set, our model's forecast error for the single-day maximum passenger volume is only 6.4%, demonstrating excellent fitting ability for extreme values.

5. Forecasting Application: Insights for Passenger Flow Trends and Operational Implications

Based on the model, forecasts for GSRS passenger flow in 2024 are generated, yielding the following management insights:

Trend Forecast: The average daily passenger volume in 2024 is projected to reach 382,000, representing a year-on-year growth of approximately 9% [1]. Growth drivers primarily stem from the deepening integration of the Shenzhen-Hong Kong SAR innovation corridor and the full recovery of the summer tourism market.

Pressure Warning: The model forecasts that single-day passenger volume during the 2024 National Day holiday may exceed 580,000. It is recommended to activate the "Large Passenger Flow Contingency Plan" in advance.

Precision Scheduling Suggestions: Based on time-sliced forecast results, the timetable for metro shuttle trains can be dynamically adjusted. For instance, morning peak shuttle trains should precisely correspond to Guangzhou-Shenzhen intercity train arrivals between 7:45 and 8:30, achieving "train arrival-departure synchronization" and reducing average transfer waiting time from 8 minutes to under 5 minutes [9].

Practical Implementation Considerations: Deploying such a forecasting system in practice would require reliable real-time data feeds (e.g., from ticket systems, weather APIs), adequate computational resources for model retraining, and staff trained in data science and operational analytics. Collaboration between transport authorities and technical teams would be essential for sustainable implementation.

6. Conclusions and Future Work

Through an in-depth analysis of big data on passenger flow at GSRS, this study clearly delineates the multi-dimensional patterns of passenger flow in mega HSR hubs across time, space, and structure, and constructs a high-performance intelligent forecasting model. The main conclusions are as follows:

- 1) Hub passenger flow exhibits stable multi-level periodicity and significant directional agglomeration, with business and tourist flows forming the main body and displaying distinct behavioral patterns [1] [2].
- 2) The proposed VMD-CNN-BiLSTM-Attention-XGBoost hybrid forecasting model effectively handles the complex characteristics of passenger flow series, with forecasting accuracy significantly improved compared to traditional methods [7], demonstrating practical application value.
- 3) Proactive operational scheduling based on high-accuracy forecasts is key to

alleviating hub congestion and improving service quality [9].

Future research can expand in two directions: First, incorporating real-time GPS, mobile phone signaling, and other finer-grained data to track and forecast the microscopic movement trajectories of passengers within the hub [2]. Second, integrating the forecasting model with a digital twin platform to develop a visual, interactive passenger flow simulation and decision support system, advancing HSR hub operation towards genuine “smart” management.

Funding

1) National Natural Science Foundation of China: 2022 Guangdong Provincial Key Platform and Research Project “Innovation Platform for Integration of Industry and Education in Guangdong-Hong Kong SAR-Macao Rail Transit” (Grant No. 2022CJPT016); 2) Teaching and Research Cultivation Project of Guangzhou Railway Polytechnic: Research and Practice of Modular Teaching in the Context of Sino-Foreign Cooperative Education—A Case Study of Railway Transportation Operation Management Major” (Project No.: GTXYY2203).

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Guangzhou Municipal Transportation Bureau (2024) Guangzhou Transportation Operation Annual Report (2023).
- [2] Wang, X.D. and Li, S.M. (2023) Urban Rail Transit Passenger Flow Characteristics Analysis Based on Multi-Source Data Fusion. *Journal of Transportation Systems Engineering and Information Technology*, **23**, 45-52.
- [3] Wang, X.D. and Li, S.M. (2023) Identification and Governance of the “Last Mile” Transfer Efficiency Bottleneck in Comprehensive Transportation Hubs. *Journal of Transportation Systems Engineering and Information Technology*, **2**, 1-10.
- [4] Chen, T. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 785-794. <https://doi.org/10.1145/2939672.2939785>
- [5] Liu, J.H. and Wang, Z.Q. (2022) A Review of Deep Learning-Based Models for Traffic Passenger Flow Forecasting. *Journal of the China Railway Society*, **44**, 1-12.
- [6] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [7] Dragomiretskiy, K. and Zosso, D. (2014) Variational Mode Decomposition. *IEEE Transactions on Signal Processing*, **62**, 531-544. <https://doi.org/10.1109/tsp.2013.2288675>
- [8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Polosukhin, I., et al. (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems*, **30**, 5998-6008.
- [9] Chen, L. and Zhang, D. (2022) Forecasting High-Speed Rail Passenger Demand with Hybrid ARIMA and Machine Learning Models. *Transportation Research Part A: Policy and Practice*, **156**, 78-92.