



# Learning dynamic interaction for multimodal transportation systems: A joint passenger flow prediction approach based on spatiotemporal hypergraph convolution networks<sup>☆</sup>

Dongyu Luo<sup>a</sup>, Jiangfeng Wang<sup>b,\*</sup>, Wenqi Lu<sup>c,\*</sup>, Weidong Ding<sup>b</sup>, Xuedong Yan<sup>b,d</sup>, Hai Yang<sup>c</sup>

<sup>a</sup> School of Traffic Management, People's Public Security University of China, Beijing 100038, China

<sup>b</sup> School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

<sup>c</sup> Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong, China

<sup>d</sup> School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, China



## ARTICLE INFO

### Keywords:

Multimodal transportation systems  
Joint passenger flow prediction  
Spatiotemporal characteristic  
K-means clustering  
Temporal convolution  
Hypergraph convolution

## ABSTRACT

In metropolitan areas, multimodal transportation systems form a complex passenger flow network with intricate spatiotemporal dependencies across different transit modes. Existing approaches often treat metro and bus systems as independent, station-based modes, neglecting their direct flow transfer and mutual influence. To address this gap, this study proposes a joint passenger flow prediction framework that explicitly models inter-modal passenger flow transfer between metro and bus systems. Firstly, A hypergraph game K-means clustering algorithm is introduced to capture latent structural correlations between the two modes. Then, a multi-layer stacked Spatiotemporal Hypergraph Convolutional Network (STHCN) is proposed for the first time, integrating temporal gated convolutions with spatial hypergraph convolutional networks and systematically incorporating a hypergraph self-attention mechanism to effectively capture both intra- and inter-modal spatiotemporal dependencies of the two public transportation systems. Finally, real-world large-scale multimodal datasets covering metro and bus passenger flows in Beijing are employed to validate the effectiveness and robustness of the proposed approach. Extensive experiments indicate that the proposed STHCN outperforms several classic and state-of-the-art models across different forecasting horizons and spatial resolutions, achieving a 6.79% and 3.38% overall improvement over the best-performing baseline model at metro and bus passenger flow predictions, respectively. Ablation studies further confirm the effectiveness of each module in the proposed architecture and highlight the model's ability to learn dynamic interdependencies within multimodal passenger flow networks robustly.

## 1. Introduction

Urbanization is accelerating globally, presenting unprecedented challenges in urban development (Yu et al., 2024). The rapid increase in urban population and sustained economic growth have intensified travel demand, straining urban transportation systems.

<sup>☆</sup> VSI: This article belongs to the Virtual Special Issue on "Data Analytics and ML".

\* Corresponding authors.

E-mail addresses: [wangjiangfeng@bjtu.edu.cn](mailto:wangjiangfeng@bjtu.edu.cn) (J. Wang), [wenqilu@ust.hk](mailto:wenqilu@ust.hk) (W. Lu).

Public transportation is critical in mitigating urban traffic congestion and enhancing travel efficiency. Taking Beijing as an example, the bus system recorded a total of 2.09 billion passengers, and the metro system reached 3.45 billion in 2023. These data underscore the significance of public transportation in fulfilling travel demand and highlight the urgent requirement for effective management and optimization of public transportation systems.

To enhance the service quality of metro and bus systems, it is essential to monitor and predict passenger flow in real time and with high accuracy (Zhang et al., 2017). Passenger flow is defined as the number of individuals boarding at bus stations or entering metro stations, with its variation closely linked to both temporal and spatial factors. Analyzing passenger flow data provides insights into travel frequency and demand fluctuations in specific areas, but also aids in identifying traffic hotspots and assessing service levels. This information is crucial for operators, enabling them to allocate resources effectively, optimize travel lines, and improve user experience (Xu et al., 2023). Research in passenger flow prediction has garnered increasing interest, particularly concerning single traffic modes, where significant advancements have been made.

Early studies predominantly employed time series and machine learning models for passenger flow prediction (Bratsas et al., 2020; Filipovska and Mahmassani, 2020). However, these traditional approaches often struggle to capture complex spatiotemporal dependencies and nonlinear characteristics, leading to limited predictive accuracy. The advent of deep learning techniques, particularly Graph Convolutional Network (GCN), has prompted researchers to explore new methods capable of extracting intricate spatiotemporal insights from extensive traffic datasets, owing to their robust feature learning capabilities (Li et al., 2018). Despite advancements in various methodologies to enhance passenger flow prediction performance, traffic modes remain isolated, with metro and bus passenger flows predicted separately. In megacities like Beijing, particularly in central urban areas, the transfer rate at bus stations within 50 m of metro stations has reached 86 %. This high-intensity transfer demand implies an implicit relationship between metro and bus passenger flows, a factor that has long been underestimated in passenger flow prediction. For instance, demand at a metro station can influence the utilization of nearby bus lines and vice versa. Passenger transfer behavior between different traffic modes and various activities throughout the travel chain contributes to the observed spatiotemporal correlations among these modes (Liang et al., 2022). Therefore, predicting metro or bus passenger flow in isolation often fails to capture the comprehensive characteristics. Effectively integrating passenger flow information from different traffic modes and conducting joint passenger flow predictions presents a key challenge. However, while this issue may appear to be a straightforward multimodal transportation problem, the solution is complicated by data limitations. Despite various proposed prediction methods, metro and bus passenger flow predictions remain insufficiently explored.

The challenges associated with joint passenger flow prediction in multimodal transportation systems arise from several current circumstances:

(1) Although metro and bus systems are station-based, their data sources, such as bus card swiping data and metro entry records, are highly heterogeneous. These datasets are vast, structurally complex, and challenging to synchronize due to differences in regulatory agencies and operational practices. This lack of temporal and spatial alignment, particularly at the spatial granularity level, complicates the integration and alignment of multimodal transportation systems and poses a critical issue requiring urgent attention.

(2) While integrating multimodal transportation for passenger flow prediction is not novel, most existing studies treat station-based traffic modes (e.g., metro and bus systems) as a single category, often combining them with region-based traffic modes like shared bicycles and taxis, and examining the interactions between these two modes. This oversimplified approach fails to capture the intrinsic relationship involved in the direct passenger flow transfer between metro and bus systems, which is the primary focus of this study.

To address these challenges, historical full-scale metro and bus data from Beijing propose a joint passenger flow prediction method based on multimodal dynamic hypergraphs. This method emphasizes the advantages of data integration, direct passenger flow transfer modeling, and enhanced model applicability, paving the way for more accurate and effective predictions in multimodal transportation systems. The main contributions of this work are as follows:

(1) A novel Hypergraph Game K-Means (HGK-Means) clustering algorithm is built for the dynamic hypergraph representation of multimodal transportation systems. It encodes intra- and inter-modal high-order relationship among stations by using hyperedges to connect stations with similar characteristics spatially. The HGK-Means algorithm effectively identifies potential correlations between stations with identical flow patterns and functions.

(2) A Spatiotemporal Hypergraph Convolutional Network (STHCN) for joint passenger flow prediction is proposed. This model combines a temporal gated convolutional network with a spatial hypergraph convolutional network, incorporating a hypergraph self-attention mechanism to effectively capture complex spatiotemporal dependencies and seamlessly integrate data from different traffic modes.

(3) Extensive experiments use historical full-scale multimodal transportation datasets encompassing full-time and peak-hour records. Ablation studies, dynamic interactions analysis, and spatiotemporal aggregation experiments are performed to evaluate the model comprehensively. Comparative results with existing baseline methods demonstrate the effectiveness and practicality of the proposed framework.

The rest of this paper is organized as follows. Section 2 overviews related work in passenger flow prediction and hypergraph theory. Section 3 details the dynamic hypergraph representation, and Section 4 outlines the joint passenger flow prediction approach. Section 5 discusses data sources, feature analysis, and baseline methods, while Section 6 presents model performance evaluations, ablation experiments, and a discussion of the results. Section 7 concludes the article and suggests directions for future research.

## 2. Literature review

This section reviews existing research on single-mode passenger flow prediction for metro and bus systems and provides an

overview of the theories and methods relevant to multimodal passenger flow prediction. Furthermore, recent advancements in hypergraph embedding are briefly summarized, providing a foundation for the proposed model.

## 2.1. Metro passenger flow prediction

Early metro prediction methods primarily relied on statistical and machine learning models, which typically treated historical passenger data as time series with periodic fluctuations. Common models included the Autoregressive Integrated Moving Average (ARIMA), Kalman filter, probability trees, and logic-based models. [Leng et al. \(2013\)](#) proposed a prediction model using probability trees and global networks, enhancing accuracy through historical origin-destination data. [Ling et al. \(2018\)](#) employed Smart Card Data (SCD) from the Shenzhen metro in conjunction with density-based clustering and regression models to identify regular and anomalous phenomena at metro systems. However, these models, though adequate for linear relationships, often struggled to address the complexities of large-scale nonlinear data, limiting their predictive power.

With the rise of deep learning, methods such as Recurrent Neural Network (RNN), particularly Long Short-Term Memory (LSTM), have gained attention for their ability to capture complex, nonlinear spatiotemporal relationships ([Zhao et al., 2017; Hao et al., 2019](#)). Gated Recurrent Unit (GRU), known for its simpler structure and ability to address the vanishing gradient problem, has also become famous for passenger flow prediction. [Sha et al. \(2020\)](#) used GRU to predict rolling flow, incorporating weather data and time-based clustering. [Mulerikkal et al. \(2022\)](#) combined spatial and temporal information from Automatic Fare Collection (AFC) systems with spatiotemporal LSTM to improve prediction accuracy. Traditionally used in image classification, Convolutional Neural Network (CNN) has also been applied. Despite CNN's effectiveness in segmenting metro systems into 2D grids, the subjective choice of grid size and irregular spatial units poses challenges. Recent advances in GCN offer a solution for graph-based data analysis. [Lu et al. \(2021\)](#) introduced a dual-attention GCN to simultaneously capture spatial and temporal influences using weighted metro maps. [Hu et al. \(2024\)](#) highlighted the effectiveness of GCN in modeling metro systems.

Research on metro passenger flow prediction primarily focuses on comparing the performance of various models, such as LSTM, CNN, and GCN, in capturing temporal and spatial dependencies. However, different models serve distinct roles in predicting metro passenger flow, with their applicability often influenced by data characteristics and task requirements. A significant advantage of hybrid models is their capacity to leverage the strengths of each model for specific tasks, enhancing feature extraction capabilities comprehensively ([Liu and Chen, 2017; Yu et al. 2018](#)). Combining these models allows for a holistic consideration of spatiotemporal features, improving adaptability to various data types and noise ([Zhao et al. 2020b](#)). Consequently, integrating multiple network architectures or models has emerged as a prevailing research trend, and [Table 1](#) summarizes the hybrid models that apply metro passenger flow datasets.

In [Table 1](#), research on metro passenger flow prediction has increasingly favored the integration of GCN with other modeling approaches, demonstrating significant advantages. Combining GCN with models such as LSTM or GRU allows for integrating complex spatial relationships with dynamically changing temporal responses, resulting in a more comprehensive and robust prediction tool. Moreover, existing studies primarily focus on short-term predictions, with a 15-minute time granularity emerging as a common standard, as confirmed by the works of [Zhong et al. \(2016\)](#) and [Zhao et al. \(2020a\)](#). However, while the current hybrid models adequately address the needs of real-time dispatching for single-mode passenger flow prediction, they encounter limitations when applied to metro and bus prediction. Traditional GCN often struggles to capture the intricate interactions between metro and bus systems when representing their higher-order relationships. For instance, during peak hours, the direct interactions between metro and

**Table 1**  
The hybrid models for metro passenger flow prediction.

Authors	Hybrid models	Metro datasets	Time granularity	
			Collection	Prediction
Wei and Chen (2012)	EMD and BPN	Taipei rapid transit	15 min	15 ~ 60 min
Sun et al. (2015)	Wavelet and SVM	Beijing	15 min	15 min
Jia et al. (2019)	SAE and LSTM	Guangzhou (AFC)	15 min	15 min
Zhang et al. (2019)	K-Means and LSTM	Beijing (APC)	60 min	60 min
Ma et al. (2019)	CNN and Bi-LSTM	Beijing (SCD)	2 min	10 ~ 20 min
Du et al. (2020)	ResNet and LSTM	Beijing (SCD)	15 min	15 min
Ye et al. (2020)	GCN and LSTM	Shenzhen (AFC)	30 min	30 min
Zhang et al. (2020)	GCN and 3D CNN	Beijing (SCD)	10 ~ 30 min	10 ~ 30 min
Zhang et al. (2021)	GCN, ResNet, and LSTM	Beijing (APC)	10 ~ 30 min	10 ~ 30 min
Zhang et al. (2022)	GCN, Bi-LSTM, and Transformer	Shanghai (SCD)	10 ~ 30 min	10 ~ 30 min
He et al. (2022)	GCN and LSTM	Shenzhen (AFC)	15 min	15 ~ 60 min
Xiu et al. (2022)	GCN and Bi-GRU	Chengdu (APC)	5 min	5 ~ 30 min
Bao et al. (2022)	Multiple GCN and Seq2Seq	Nanjing (SCD)	10 min	10 min
Li et al. (2023a)	Multiple GCN and MTL	Suzhou (AFC)	15 min	15 min
Zeng and Tang (2023)	Relational GCN and LSTM	Shenzhen, Hangzhou	10 ~ 15 min	1 ~ 15 min
Yin et al. (2023)	GCN, GAT and Gated CNN	Hangzhou	10 min	10 min
Li et al. (2023b)	CNN and LSTM	Suzhou (AFC)	15 min	15 min
Huang et al. (2024)	Multiple EMD, DPG, and LSTM	Chengdu (AFC)	10 min	10 ~ 40 min
Zhan et al. (2024)	Multiple GCN and GRU	Hangzhou, Shanghai	15 min	15 min

bus passengers are frequent, yet the underlying relationship behind these interactions remain insufficiently explored.

Hypergraph Convolutional Network (HGCN) has emerged as a promising alternative. HGCN can effectively manage high-order relationships among nodes and simultaneously account for multiple traffic modes, providing a more comprehensive representation of the interaction characteristics between metro and bus systems. By incorporating hypergraph structures, HGCN enhances the modeling of joint predictions, enabling the capture of multimodal passenger flows and ultimately improving the accuracy and practicality.

## 2.2. Bus passenger flow prediction

Most spatiotemporal modeling approaches primarily analyze the metro system. However, notable differences exist between the metro and bus passenger flows, rendering the direct application of metro prediction methodologies to the bus system inadequate (Baghbani et al., 2025).

A key distinction lies in temporal dependence. The bus system is susceptible to external factors such as traffic congestion, weather conditions, and special events, requiring models to account for multi-scale temporal variations. For instance, Sun et al. (2020) used clustering and genetic algorithms to model the Qingdao bus system, focusing on weekly patterns. Wang et al. (2022) developed a hybrid model combining multiple linear regression, K-nearest neighbors, extreme gradient boosting, and GRU to address the inherent multi-scale variability. Spatial dependence in the bus system also presents challenges. The complexity of bus lines and the large number of stations necessitate considering multi-level spatial dependencies, including line, station, and service levels, alongside internal and external factors. Liu et al. (2021) utilized decision tree-based models and modular CNN to predict macro and micro passenger flow. Luo et al. (2022) introduced a spatiotemporal hash GCN that constructs two types of subgraphs to capture temporal and spatial dependencies at both the station and line levels. More recently, Mei et al. (2023) developed a model using an attention mechanism to encode station-level spatial dependencies and parallel computing for evaluating station correlations.

Despite these advancements, significant challenges remain in accurately modeling uncertainties and external factors, mainly when predicting passenger flow exchanges between metro and bus systems. Integrating these external variables into predictive models can improve both accuracy and generalization.

## 2.3. Multimodal passenger flow prediction

In urban transportation systems, passenger behavior often leads to interconnected traffic modes rather than isolated ones. Sun et al. (2017) identified a complementary relationship between metro and bus systems, demonstrating that active metro stations positively influence bus passengers. This finding underscores the necessity of considering the interactions between different traffic modes when predicting passenger flow in metro and bus systems. Zhong et al. (2017) developed a citywide passenger flow prediction model based on multi-task learning, which incorporates various types of heterogeneous data to enhance prediction accuracy. Peng et al. (2020) employed a spatiotemporal incident dynamic graph structure to model the relationships between stations, introducing a dynamic recursive GCN to capture spatiotemporal features. Although these studies address the interactions among metro, bus, and taxi systems, there remains a lack of exploration into an intrinsic relationship, which leads to the oversimplified assumption that passenger flows in each mode are identical.

To bridge this gap, some researchers have utilized static correlation matrices to describe the relationships among various traffic modes. Toman et al. (2020) applied time series methods to investigate the dynamic demand relationships between Uber/Lyft, taxis, shared bicycles, and the metro, revealing the mobility patterns and correlations among different traffic modes. Ke et al. (2021) designed a multiple GCN framework for ride-hailing systems, distinguishing between single-person and shared ride-hailing services. However, as multimodal transportation systems become increasingly complex, research on multimodal passenger flow prediction is progressively shifting towards dynamic spatiotemporal correlations. Several studies have employed multi-task learning and knowledge-sharing techniques to capture the interaction mechanisms between traffic modes by integrating fixed or dynamic correlation matrices (Li et al., 2021; Yang et al., 2023). Zhang et al. (2024) proposed a bidirectional temporal convolutional network with a graph sparse attention mechanism for joint prediction. This approach enhances temporal feature correlation by multiplying multimodal graphs with self-attention weights to capture spatial local features.

In contrast, global spatial features are obtained through a Top-U sparse attention mechanism. Yang et al. (2024) introduced a multi-task learning model that integrates adaptive multiple GCNs and an attention mechanism. This encoder-decoder architecture aims to predict short-term inflow across multiple traffic modes. The encoder learns complex dynamic spatiotemporal correlations between modes, while the decoder focuses on feature extraction and knowledge sharing among modes.

Despite these advancements, several significant challenges remain in multimodal passenger flow prediction. First, more research is needed to establish more accurate dynamic spatiotemporal relationships within multimodal transportation systems that effectively reflect the complex interactions between traffic modes. Second, while joint predictions across traffic modes have been attempted, limitations persist in fully leveraging spatial correlations for interactive modeling. Lastly, researchers should consider incorporating additional external factors in dynamic interactions to enhance the adaptability and accuracy of prediction models, thus improving the prediction of actual traffic demand. Fang et al. (2024) highlighted that many existing passenger flow prediction methods rely on distance and adjacency to construct empirical graphs, which may not fully capture the correlations among metro stations. Therefore, integrating adjacency, travel distance, and passenger flow transfer into a unified framework could uncover hidden spatial dependencies within the traffic network. In this regard, this study establishes the interaction relationships between metro and bus systems through hypergraphs and employs the HGK-Means algorithm to construct an adjacency matrix, facilitating a more

comprehensive understanding of the high-order relationships among different traffic modes.

#### 2.4. Hypergraph embedding

Hypergraphs represent a significant extension of graph theory, effectively capturing complex, high-order relationships among objects. Unlike traditional graphs, which only facilitate pairwise relationships, hypergraphs connect multiple vertices through hyperedges, thus enabling the representation of intricate interactions and dependencies (Arya and Worring, 2018). The concept of the hypergraph network was first introduced by Feng et al. (2019), followed by the development of the HGCN by Yadati et al. (2019). This approach naturally extends the spectral methods of GCN to hypergraphs, allowing features to be aggregated by each node from its connected hyperedges.

The propagation efficiency of multimodal data is enhanced by HGCNs, which effectively capture complex high-order relationships among various nodes. As a result, increasing attention has been garnered in passenger flow prediction. Yi and Park (2020) integrated HGCN with RNN to learn high-order spatiotemporal relationships within traffic networks. Wang et al. (2021) constructed a primary hypergraph based on the topology of the metro system, revealing and expanding advanced hyperedges derived from travel modes over multiple temporal spans. However, the construction of these hypergraphs often relies on prior knowledge or static traffic data, which limits their adaptability to dynamically changing traffic environments.

Subsequently, research efforts have been made to investigate the construction and application of dynamic hypergraphs, enabling a better capture of spatiotemporal features. Zhao et al. (2023) employed an adaptive K-Means clustering algorithm to identify multiple related nodes, which were then used to construct hyperedges that evolved into hypergraphs. This method emphasizes adaptability and dynamism, allowing real-time updates to hypergraph structures based on fluctuations in traffic data. Wang et al. (2022) introduced a method for primary hypergraph construction that leverages metro topology, using clustering algorithms to identify stations with similar traffic modes and creating hidden hypergraphs. Sun et al. (2022) adaptively generated hypergraph structures through dual transformations between graphs and hypergraphs, using HGCN to learn high-order spatiotemporal relationships in traffic networks. While these approaches primarily focus on the formation of hyperedges by linking nodes through edges in the graph, they may overlook the temporal dependencies inherent in time-varying data. To address this limitation, Wu et al. (2024) proposed a hypergraph attention recurrent network that captures dynamic pairwise and higher-order relationships through adaptive hypergraph construction methods, incorporating the temporal Hawkes attention mechanism to manage long-term temporal dependencies. Shang et al. (2024) integrated the hypergraph convolution and hyperedge attention mechanism into the Transformer architecture to enable predictions across multi-scale time ranges.

However, directly extending existing hypergraph research to multimodal passenger flow prediction may treat different traffic modes as a single category. This approach would involve constructing hyperedges without distinguishing between the various modes, thereby failing to capture the inter-modal differences. Even with the application of the hypergraph attention mechanism, such a model would only capture the spatiotemporal dependencies within individual modes and would still be unable to observe the disparities between different modes. This study uses the hypergraph attention mechanism to model intra-modal higher-order relationships, and the HGK-Means algorithm is employed to capture the inter-modal dynamic interactions. As a result, the STHCN framework is proposed, which is specifically designed to address the spatiotemporal dependencies inherent in joint passenger flow prediction.

### 3. Preliminaries

Some relevant concepts are provided in this section. Then, a dynamic hypergraph construction method tailored for multimodal transportation systems is proposed. This method comprises two primary components: the static hypergraph representing single-modal transportation systems and the dynamic interactions that characterize multimodal transportation systems.

#### 3.1. Related concepts

**Definition 1.** (*Multimodal transportation systems*). *The multimodal transportation systems refer to utilizing two or more traffic modes to facilitate the movement of passengers or goods within a system. This study focuses primarily on two traffic modes: The metro system, represented as  $V_s = \{v_1, v_2, \dots, v_n\}$ , and the bus system, denoted as  $V_b = \{v_1, v_2, \dots, v_{n^*}\}$ . The metro system comprises  $|V_s| = n$  stations. The entering metro passengers  $v_i \in V_s$  ( $i = 1, 2, \dots, n$ ) at time t is represented by a vector  $x_i^t \in \mathbb{R}$ . The total entering metro passengers  $V_s$  at time t is given by  $X_s^t = \{x_1^t, x_2^t, \dots, x_n^t, \dots, x_{n^*}^t\}$ . The bus system comprises  $|V_b| = n^*$  stations. The boarding bus passengers  $v_{i^*} \in V_b$  ( $i^* = 1, 2, \dots, n^*$ ) at time t is represented by a vector  $x_{i^*}^t \in \mathbb{R}$ . The total boarding bus passengers  $V_b$  at time t is expressed as  $X_b^t = \{x_1^t, x_2^t, \dots, x_{i^*}^t, \dots, x_{n^*}^t\}$ .*

**Definition 2.** (*Intra-modal pairwise relationships*). *The intra-modal pairwise relationships are introduced to describe the correlation between bus or metro systems, denoted as  $G_{\text{intra}} = \{G_s, G_b\}$ . Specifically, the metro correlation is represented as  $G_s = (V_s; A_s)$ , where  $A_s \in \mathbb{R}^{n \times n}$  indicates the adjacency matrix for the relationships among metro stations. Similarly, the bus correlation is represented as  $G_b = (V_b; A_b)$ , where  $A_b \in \mathbb{R}^{n^* \times n^*}$  denotes the adjacency matrix for the relationships among bus stations.*

**Definition 3.** (*Inter-modal pairwise relationships*). *The inter-modal pairwise relationships describe the correlation between metro and bus systems, represented as  $G_{\text{inter}} = (V_s, V_b; A_{s,b})$ . Where  $A_{s,b} \in \mathbb{R}^{n \times n^*}$  denotes the adjacency matrix that captures the cross mode relationships. Due to the heterogeneity of multimodal transportation systems,  $A_{s,b}$  is generally not a square matrix.*

**Definition 4.** (Joint passenger flow): The joint passenger flow  $X^{t \rightarrow t+H}$  is introduced to characterize the evolving relationships between metro and bus passenger flows over future H time steps. While inter-modal information sharing and feature fusion mechanisms are incorporated, independent prediction outcomes for each traffic mode are maintained. The joint passenger flow is formally expressed as  $X^{t \rightarrow t+H} = \{X_n^{t \rightarrow t+H}, X_{n^*}^{t \rightarrow t+H}\}$ , where  $X_n^{t \rightarrow t+H}$  and  $X_{n^*}^{t \rightarrow t+H}$  indicate the predicted passenger flow for metro and bus systems, respectively.

**Problem** (*Joint passenger flow prediction*). The joint passenger flow prediction for multimodal transportation systems aims to jointly predict metro and bus passenger flows for a future time interval, utilizing historical passenger flow data. Given the metro and bus passenger flows for the historical  $T$  time steps, and considering the intra-modal pairwise relationships  $G_{\text{intra}}$  and inter-modal pairwise relationships  $G_{\text{inter}}$ , the joint passenger flow  $X^{t \rightarrow t+H}$  can be expressed as a mapping function  $f(\bullet)$ . Specifically, these relationships are formulated as follows:

$$X^{t \rightarrow t+H} = \{X_n^{t \rightarrow t+H}, X_{n^*}^{t \rightarrow t+H}\} = \int (X_n^{t-T \rightarrow t}, X_{n^*}^{t-T \rightarrow t}, G_{\text{intra}}, G_{\text{inter}}) \quad (1)$$

where  $X_n^{t-T \rightarrow t}$  and  $X_n^{t \rightarrow t}$  represent the historical passenger flow for metro and bus systems, respectively.

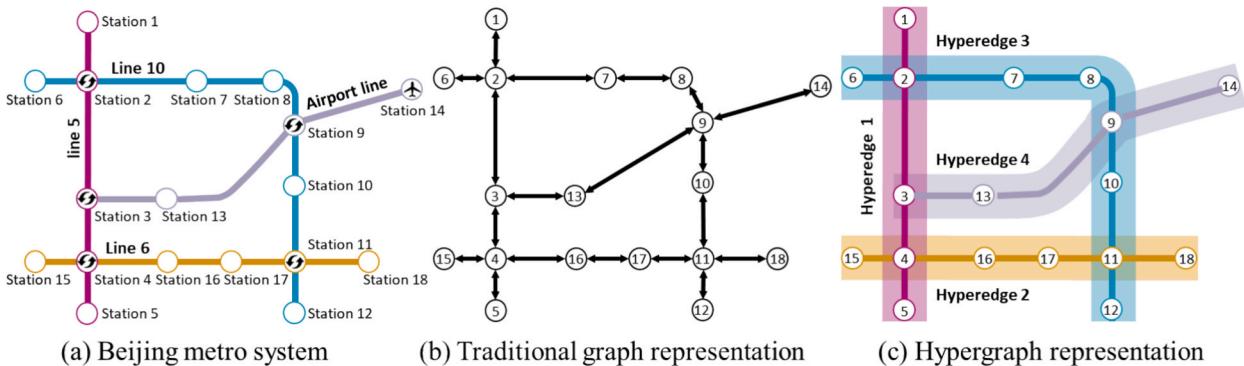
### 3.2. High-order relationship

A pairwise relationship within the same traffic mode is defined to connect adjacent stations with similar functions and flow characteristics. This representation employs an adjacency matrix to map the correlation between two sites as edges, where each edge signifies a pairwise relationship. However, the relationships between stations extend beyond simple pairwise connections, encompassing more complex ternary or quaternary higher-order relationships (Wu et al., 2024). For instance, in a metro system, each station is part of different lines, which exhibit significant variations in departure intervals, operating hours, and the frequency of departures. Consequently, the station connection cannot be regarded as merely a bilateral relationship.

**Fig. 1(a)** illustrates a metro system consisting of Beijing Metro Lines 5, 6, 10, and the Airport Line. Different line attributes can dynamically and asymmetrically influence a series of stations along each line. For instance, while the Airport Line operates with a departure interval of 20 to 30 min, Lines 5 and 10 operate at a much higher frequency of 3 to 5 min during peak hours. These disparities in operational schedules lead to congestion at transfer stations, such as Stations 3 and 9, which further cascade and impact passenger flow at other stations on Lines 5 and 10. Moreover, modeling based solely on pairwise relationships fails to capture the inherent commonalities when multiple stations share similar attributes. For example, Station 4, located in a commercial area, experiences simultaneous increases in passenger flow at neighboring stations on Lines 5 and 6 during shopping hours. Such scenarios highlight the limitations of traditional graph approaches in representing the complex interactions within the metro system, as shown in **Fig. 1(b)**. To address these limitations, static hypergraphs are introduced for the simultaneous connection of multiple stations. This approach transcends the constraints of pairwise relationships and enables a more nuanced representation of the high-order relationships inherent in passenger flow dynamics, as depicted in **Fig. 1(c)**.

Static hypergraphs are based on the fixed topology structure of metro and bus system lines. Each line is regarded as a hyperedge, and its operational attributes, such as departure intervals, running times, and frequencies, directly impact passenger flow. These attributes exhibit high stability over specific periods, enabling static hypergraphs to describe the topology of transportation networks accurately.

Static hypergraphs are particularly suitable for metro and bus systems, where each line represents a coherent and continuous spatial unit serving a group of stations in a synchronous operation mode. The spatial adjacency relationships and operational dependencies between stations are integrated into a unified framework by representing lines as hyperedges. This modeling approach represents the relationships between directly connected adjacent stations and captures complex interactions beyond simple paired connections. These interactions include broader networks formed by shared operational features, such as unified schedules, fare systems, and coordinated transfers.



**Fig. 1.** Graph and hypergraph representation of a metro system.

Although static hypergraphs are employed to model higher-order relationships between stations within the same line, stations across different lines can also indirectly form higher-order relationships through shared transfer stations or dependencies between lines. For example, a single hyperedge may not directly connect two stations on other lines, such as Metro Stations 2 and 13 in Fig. 1(a). However, they can be indirectly linked through transfer stations such as Metro Stations 3 and 9, which mediate between the two hyperedges. Transfer stations are allowed to be part of multiple hyperedges in static hypergraphs, enabling the hypergraph to effectively capture complex multi-path interactions between stations, even without direct line connections between these stations.

### 3.3. Hypergraph construction

This study uses the same static hypergraphs to establish high-order relationships within metro and bus stations. For metro stations, the pairwise relationships are defined as  $G_s = (V_s; A_s)$ , where  $V_s$  represents the set of metro stations and  $A_s$  denotes the adjacency matrix, based on a traditional graph structure. The adjacency matrix  $A_s$  is derived from the set of edges  $E_s = \{e_1, e_2, \dots, e_s\}$ , allowing us to reformulate the pairwise relationships as  $G_s = (V_s, E_s; A_s)$ . In this context,  $E_s$  signifies the set of edges, and  $s$  indicates the total number of edges. Let  $g_{i,j}$  denote the connection relationships represented by the edge  $e_{i,j}$  formed by the metro stations  $v_i \in V_s$  and  $v_j \in V_s$ . Consequently, the adjacency matrix can be defined as follows:

$$g_{i,j} = \begin{cases} 1 & e_{i,j} \in E_s \\ 0 & \text{otherwise} \end{cases}, i, j = 1, 2, \dots, n \quad (2)$$

In contrast to pairwise relationships, constructing higher-order relationships using static hypergraph structures involves the adjacency matrix  $A_s^{(h)}$ , which is associated with the hyperedge set  $E_s^{(h)} = \{r_1, r_2, \dots, r_m\}$ . In this framework, hyperedges are defined as metro lines, and the set of hyperedges represents the collection of all metro lines. Therefore, the higher-order relationships among metro stations can be expressed as  $G_s^{(h)} = (V_s, E_s^{(h)}; A_s^{(h)})$ . It is essential to derive the adjacency matrix  $A_s^{(h)} \in R^{n \times n}$  for this representation.

To begin, the incidence matrix  $I_s^{(h)} \in R^{n \times m}$  is defined to capture the connectivity between metro stations and lines. Specifically, let  $h_{i,k}$  denote the connection relationships between a metro station  $v_i \in V_s$  and a metro line  $r_k \in E_s^{(h)}$ . The value of  $h_{i,k}$  is determined as follows:

$$h_{i,k} = \begin{cases} 1 & v_i \in r_k \\ 0 & \text{otherwise} \end{cases}, i = 1, 2, \dots, n \text{ and } k = 1, 2, \dots, m \quad (3)$$

Based on this definition, the incidence matrix  $I_s^{(h)}$  is subsequently constructed to represent this relationships as follows:

$$I_s^{(h)} = \begin{array}{ccccccccc} r_1 & r_2 & \cdots & r_k & \cdots & r_m \\ \vdots & \vdots & & \vdots & & \vdots \\ v_1 & & & h_{1,1} & h_{1,2} & \cdots & h_{1,k} & \cdots & h_{1,m} \\ v_2 & & & h_{2,1} & h_{2,2} & h_{2,3} & \ddots & \cdots & \vdots \\ \vdots & & & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ v_i & & & h_{i,1} & h_{i,2} & \cdots & h_{i,k} & \ddots & h_{i,m} \\ \vdots & & & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \\ v_n & & & h_{n,1} & h_{n,2} & \cdots & h_{n,k} & \cdots & h_{n,m} \end{array} \quad (4)$$

Next, the station hyperdegree within the metro system is defined. In traditional graph structures, the station degree  $d_s(v_i)$  indicates the number of metro stations directly adjacent to station  $v_i$ . The station degree distribution  $P_s(d_s = u)$  represents the probability of randomly selecting a metro station with a degree  $d_s = u$ . In the static hypergraphs, the station hyperdegree  $d_s^{(h)}(v_i)$  can be defined as the sum of the weights of the metro lines directly adjacent to station  $v_i$ . The station hyperdegree distribution  $P_s^{(h)}(d_s^{(h)} = u)$  indicates the probability that a randomly selected metro station has a hyperdegree  $d_s^{(h)} = u$ , with expressions as follows:

$$d_s^{(h)}(v_i) = \sum_{r_k \in E_s^{(h)}} \omega_s(r_k) h_{i,k} \quad (5)$$

$$P_s^{(h)}(d_s^{(h)} = u) = \frac{1}{n} \sum_{v_i \in V_s} p_i \quad (6)$$

$$p_i = \begin{cases} 1 & d_s^{(h)}(v_i) = u \\ 0 & \text{otherwise} \end{cases}, u = 1, 2, \dots \quad (7)$$

where  $\omega_s(r_k)$  denotes the influence weight of the metro line  $r_k$  on station  $v_i$ .

Finally, the line hyperdegree of the metro system is defined as follows. In the traditional graph structure, the edge degree is described as the sum of the nodes connected by an edge  $e_k$ , represented as  $\delta(e_k) = 2$ . Similarly, in the hypergraph structure, the line hyperdegree is introduced, defined as the total number of metro stations served by a line  $r_k$ , denoted as  $\delta_s^{(h)}(r_k)$ . The formal expression is given as:

$$\delta_s^{(h)}(r_k) = \sum_{v_i \in r_k} h_{i,k} \quad (8)$$

The adjacency matrix  $A_s^{(h)}$ , which represents the higher-order relationships among metro stations, can be derived, with its calculation expressed as follows:

$$A_s^{(h)} = I_s^{(h)} W_s (I_s^{(h)})^T - D_s^V \quad (9)$$

where  $W_s \in R^{m \times m}$  represents a diagonal matrix composed of line weights  $\omega_s(r_k)$ ;  $D_s^V \in R^{n \times n}$  indicates a diagonal matrix consisting of the hyperdegrees  $d_s^{(h)}(v_i)$ .

For bus stations, the same method is used to construct high-order relationships between bus stations by associating the adjacency matrix  $A_b^{(h)}$  with the hyperedge set  $E_b^{(h)} = \{r_1, r_2, \dots, r_m\}$ . The higher-order relationships among bus stations can be represented as  $G_b^{(h)} = (V_b, E_b^{(h)}; A_b^{(h)})$ .

The adjacency matrix  $A_b^{(h)}$ , which represents the higher-order relationships among bus stations, can be derived, with its calculation expressed as follows:

$$A_b^{(h)} = I_b^{(h)} W_b (I_b^{(h)})^T - D_b^V \quad (10)$$

where  $W_b \in R^{m^* \times m^*}$  represents a weight matrix;  $D_b^V \in R^{n^* \times n^*}$  indicates a hyperdegree matrix.

#### 4. Methodology

To enhance joint passenger flow prediction, an STHCN framework is developed, as shown in Fig. 2. The framework is composed of three primary components:

**Input:** The input consists of historical passenger flow  $X^{t-T \rightarrow t}$  of multimodal transportation systems, intra-modal high-order relationships  $G_{\text{intra}}^{(h)} = \{G_s^{(h)}, G_b^{(h)}\}$  constructed from static hypergraphs, and inter-modal high-order relationships  $G_{\text{inter}}^{(h)}$  derived from dynamic hypergraphs using the HGK-Means algorithm.

**Spatiotemporal Aggregation:** A four-layer spatiotemporal convolution module is designed to extract features of stations and lines, organized in a “sandwich” structure. The first and fourth layers implement a temporal gated convolutional network to capture temporal dependencies. The second and third layers employ a spatial hypergraph convolutional network to model spatial relationships effectively.

**Output:** The output comprises the future passenger flow  $X^{t \rightarrow t+H}$  of multimodal transportation systems.

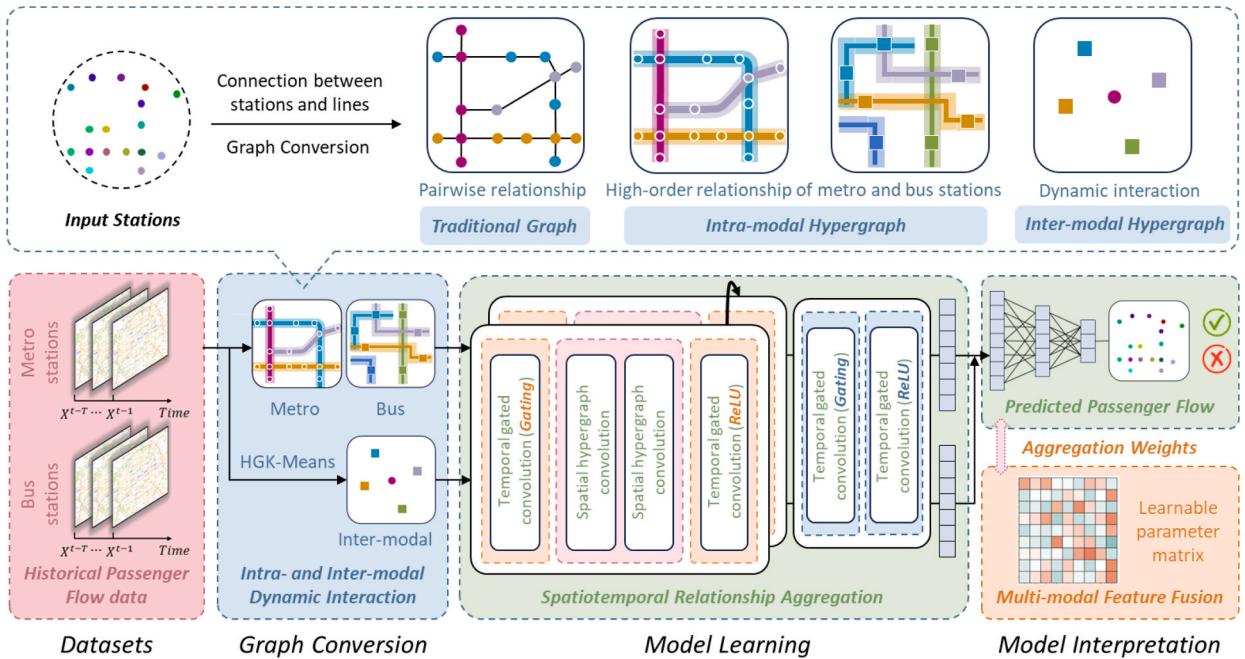


Fig. 2. Joint passenger flow prediction model framework.

#### 4.1. Dynamic interaction

Definition 3 provides inter-modal pairwise relationships, represented as  $G_{\text{inter}} = (V_s, V_b; A_{s,b})$ . The adjacency matrix  $A_{s,b} \in R^{n \times n^*}$  indicates the connectivity between metro station  $v_i$  and bus station  $v_{i^*}$ . When using dynamic hypergraphs to construct high-order relationships between multimodal transportation systems, the adjacency matrix  $A_{s,b}^{(h)}$  is associated with the dynamic hyperedge set  $E_{s,b}^{(h)} = \{w_1, w_2, \dots, w_n\}$ . Consequently, the high-order relationships can be defined as  $G_{\text{inter}}^{(h)} = (V_s, V_b, E_{s,b}^{(h)}; A_{s,b}^{(h)})$ .

The dynamic hypergraph construction of multimodal transportation systems is illustrated in Fig. 3. The number of metro stations is significantly lower than that of bus stations, so the bus stations are clustered around metro station  $v_i$ , with metro station  $v_i$  serving as the cluster center. This forms an inter-modal dynamic hypergraph  $G_i^{(h)} \in G_{\text{inter}}^{(h)}$ , where the hyperedge is denoted as  $w_i \in E_{s,b}^{(h)}$ . The hypergraph  $G_i^{(h)}$  can be explained as follows:

(1) The hypergraph  $G_i^{(h)}$  represents the transfer reachable area of the metro station  $v_i$ . Each hypergraph  $G_i^{(h)}$  comprises a metro station  $v_i$  and multiple bus stations  $B_i = \{v_1, v_2, \dots, v_{i^*}\}$ .

(2) It is assumed that passengers transfer along the same bus line and select only one bus station for their transfer. Consequently, the bus stations  $B_i = \{v_1, v_2, \dots, v_{i^*}\}$  included in the hypergraph  $G_i^{(h)}$  are sourced from different bus lines.

(3) To maintain the high-order relationships among stations within the same metro line  $r_k$ , the clustering of bus stations is categorized into  $|E_s^{(h)}| = m$  orders, where each order corresponds to a metro line  $r_k$ .

To enhance the attraction of bus stations to specific metro lines, the HGK-Means algorithm is introduced, which is based on hypergraph game theory (Rezaee et al., 2021). In this framework, metro stations are conceptualized as individual participants, with the number of participants in the  $k$ -th order corresponding to the line hyperdegree  $\delta_s^{(h)}(r_k)$ . According to Nash bargaining game theory, for any metro line  $r_k$ , the two dynamic hypergraphs  $G_i^{(h)}$  and  $G_j^{(h)}$ , formed by the clustering centers of metro stations  $v_i \in r_k$  and  $v_j \in r_k$ , can be viewed as two participants. The objective function is defined as follows:

$$\max(I - u_i)(J - u_j) \quad (11)$$

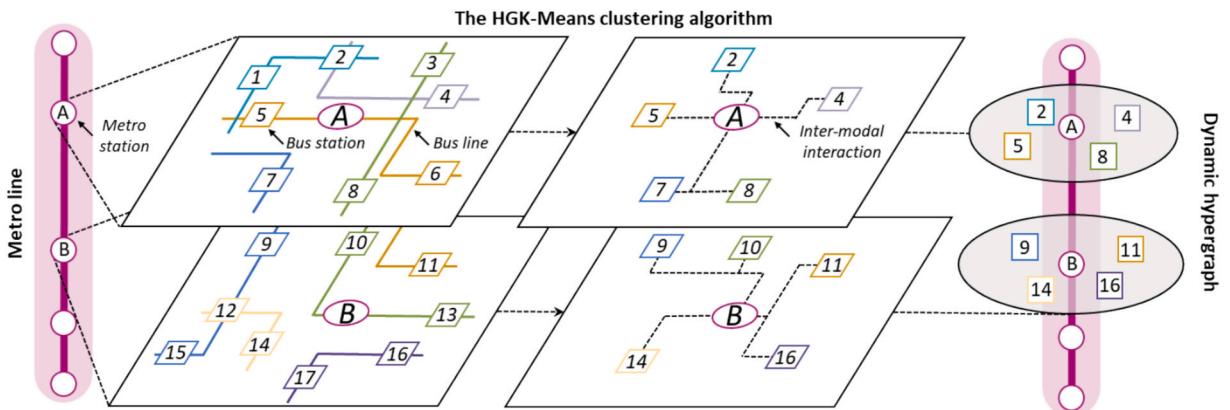
where  $I$  and  $J$  denote the utility functions of  $G_i^{(h)}$  and  $G_j^{(h)}$ ;  $u_i$  and  $u_j$  represent the utility collapse points, defined as follows:

$$u_i = \frac{1}{2}\sigma_i \left( \frac{1}{d_i} + \frac{1}{d_j} \right) \quad (12)$$

$$u_j = \frac{1}{2}\sigma_j \left( \frac{1}{d_i} + \frac{1}{d_j} \right) \quad (13)$$

where  $\sigma_i$  and  $\sigma_j$  indicate the coverage proportion of bus stations in  $G_i^{(h)}$  and  $G_j^{(h)}$ ;  $d_i$  and  $d_j$  denote the minimum trade-off function of transfer distances in  $G_i^{(h)}$  and  $G_j^{(h)}$ , defined as follows:

$$d_i = \min \left[ \gamma \max(D_i) + (1 - \gamma) \frac{1}{|B_i|} \sum_{k=1}^{|B_i|} D_i(k) \right] \quad (14)$$



**Fig. 3.** Dynamic hypergraph of multimodal transportation systems.

$$d_j = \min \left[ \gamma \max(D_j) + (1 - \gamma) \frac{1}{|B_j|} \sum_{k=1}^{|B_j|} D_j(k) \right] \quad (15)$$

where  $\gamma$  represents weight coefficient;  $B_i$  and  $B_j$  indicate the sets of bus stations in  $G_i^{(h)}$  and  $G_j^{(h)}$ ;  $D_i$  and  $D_j$  denote the sets of transfer distances in  $G_i^{(h)}$  and  $G_j^{(h)}$ .

The trade-off function  $\tilde{D}_i$  and  $\tilde{D}_j$  representing the weighted evaluation quantities for transfer distance, can be obtained as follows:

$$\tilde{D}_i = \gamma \max(D_i) + (1 - \gamma) \frac{1}{|B_i|} \sum_{k=1}^{|B_i|} D_i(k) \quad (16)$$

$$\tilde{D}_j = \gamma \max(D_j) + (1 - \gamma) \frac{1}{|B_j|} \sum_{k=1}^{|B_j|} D_j(k) \quad (17)$$

Let  $I = 1/\tilde{D}_i$  and  $J = 1/\tilde{D}_j$ , the objective function of the HGK-Means algorithm can be expressed as follows:

$$\max \left( \frac{1}{\tilde{D}_i} - u_i \right) \left( \frac{1}{\tilde{D}_j} - u_j \right) \quad (18)$$

Further extend Eq. (18) to  $n$  stations of  $m$  orders, with the objective function being:

$$\max \sum_{k=1}^m \prod_{i=1}^{\delta_s^{(h)}(r_k)} \left( \frac{1}{\tilde{D}_i} - u_i \right) \quad (19)$$

$$u_i = \sigma_i \frac{1}{\delta_s^{(h)}(r_k)} \sum_{l=1}^{\delta_s^{(h)}(r_k)} \frac{1}{d_l} \quad (20)$$

Upon deriving the dynamic hypergraph  $G_i^{(h)}$  through the HGK-Means algorithm, the adjacency matrix  $A_{s,b}^{(h)}$  is obtained. Within the hypergraph  $G_i^{(h)}$  is a single metro station  $v_i$  with a hyperdegree of 1. The relationships between the bus station  $v_{i^*} \in V_b$  and the hyperedge  $w_i \in E_{s,b}^{(h)}$  is denoted as  $h_{i^*,i}$ . Consequently, bus station hyperdegree can be articulated as follows:

$$d_{s,b}^{(h)}(v_i) = \sum_{w_i \in E_{s,b}^{(h)}} \omega_{s,b}(w_i) h_{i^*,i} \quad (21)$$

$$h_{i^*,i} = \begin{cases} 1 & v_{i^*} \in w_i, i^* = 1, 2, \dots, n^* \text{ and } i = 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

where  $\omega_{s,b}(w_i)$  represents the influence weight of the transfer distance, diminishing with increasing distance, let  $\omega_{s,b} \in [0, |B_j|]$ .

The adjacency matrix  $A_{s,b}^{(h)}$ , which represents the higher-order relationships among metro and bus stations, can be derived, with its calculation is expressed as follows:

$$A_{s,b}^{(h)} = \left[ A_1^{(h)} \mid A_2^{(h)} \mid \dots \mid A_i^{(h)} \mid \dots \mid A_n^{(h)} \right] \quad (23)$$

$$A_i^{(h)} = I_i^{(h)} W_i \left( I_i^{(h)} \right)^T - D_i^V \quad (24)$$

where  $I_i^{(h)} \in R^{1 \times n^*}$  indicates the incidence matrix between bus stations and metro station  $v_i$ ;  $W_i \in R^{n^* \times n^*}$  denotes a diagonal matrix composed of transfer distance weights  $\omega_{s,b}(w_i)$ ;  $D_i^V \in R^{1 \times n^*}$  represents a matrix consisting of the station hyperdegrees  $d_{s,b}^{(h)}(v_i)$ .

#### 4.2. Temporal gated convolution

RNNs have demonstrated considerable success in time series analysis (He et al., 2022; Huang et al., 2024). However, their performance is often constrained by limited memory capacity when processing long sequences, which can result in gradient vanishing or exploding problems. Additionally, the longer training times associated with RNNs render them less suitable for applications involving passenger flow data. Conversely, CNNs offer several notable advantages, including faster training speeds, simpler architectures, and the ability to leverage parallel processing. To overcome the limitations of RNNs, this study builds upon the temporal gated convolutional network introduced by Yu et al. (2018). By adjusting the width of the convolutional kernel, a one-dimensional temporal convolutional network is implemented using two-dimensional causal convolution, as depicted in Fig. 4.

The passenger flow  $X^{t-T \rightarrow t}$  serves as the input to the temporal gated convolutional network. This network utilizes gating mecha-

nisms to effectively differentiate the importance of various information features and applies causal convolution operations for efficient integration of sequential information. After aggregating the information, the data undergoes residual calculation and Sigmoid activation, producing output matrices  $P \in R^{(T-K_t+1) \times C_0}$  and  $Q \in R^{(T-K_t+1) \times C_0}$ , both with the same number of channels  $C_0$ . The final output, representing a passenger flow time series of length  $T - K_t + 1$ , is calculated as follows:

$$\alpha^* \tau X^{t-T-t} = P \odot \text{Sigmoid}(Q) \quad (25)$$

where  $\alpha$  indicates the temporal gated convolution kernel,  $\alpha = \{\alpha_1, \alpha_2\}$ ,  $\alpha_1$  denotes the first layer convolution kernel and  $\alpha_2$  represents the fourth layer convolution kernel;  ${}^*\tau$  indicates the temporal gated convolution operator;  $K_t$  denotes the width of the convolution kernel corresponding to the time step  $t$ ;  $C_0$  represents the number of output channels;  $\odot$  indicates the Hadamard product;  $\text{Sigmoid}(\bullet)$  denotes the Sigmoid activation function.

After the data is processed by the temporal gated convolution and spatial convolution layers, subsequent temporal convolutions no longer require the gating mechanism. Instead, the ReLU activation function can be employed as the activation function for the temporal convolution network. This adjustment enhances the model's convergence speed while significantly reducing computational complexity, ensuring a more efficient prediction process.

#### 4.3. Spatial hypergraph convolution

The metro system offers substantial advantages in large capacity and high efficiency for passenger flow transfer. However, its accessibility remains relatively limited. To address this limitation, the bus system is vital in enhancing accessibility and extending the overall travel chain. A spatial hypergraph convolutional network is developed to explore the complex dynamic interactions between traffic modes resulting from passenger flow transfers, featuring a two-layer architecture. This network integrates upper and lower time-gated convolutional networks to enable efficient feature transfer and allocation across stations and hyperedges. The spatial hypergraph convolutional network operates through a two-stage feature aggregation and allocation process. This design allows for the simultaneous capture of local features specific to individual stations and global features associated with transportation lines, while accounting for the spatial relationships across different traffic modes.

The operational workflow of the spatial hypergraph convolutional network is illustrated in Fig. 5 and is summarized as follows:

**Data Mapping:** By leveraging the high-order relationships  $G_{\text{intra}}^{(h)}$  and  $G_{\text{inter}}^{(h)}$ , the adjacency matrices  $A_s^{(h)}$ ,  $A_b^{(h)}$  and  $A_{s,b}^{(h)}$  are constructed. Subsequently, the passenger flow  $X^{t-T-t}$  aggregated at various temporal granularities, are mapped to their corresponding stations.

**Feature Extraction:** Spatial characteristics are leveraged through the application of spatial hypergraph convolution, enabling the extraction of hyperedge features from the high-order relationships  $G_{\text{intra}}^{(h)}$  and  $G_{\text{inter}}^{(h)}$ .

**Feature Allocation:** After hyperedge features are aggregated, passenger flow is redistributed to metro and bus systems based on specified temporal granularities. This process facilitates dynamic updates of passenger flow at each station, ensuring accurate and timely representation of flow variations.

Assuming the passenger flow transfer roles of metro and bus lines are equally significant, let  $\omega_s(r_k) = \omega_b(r_k) = 1$ . Following the methodology established by Feng et al. (2019), the station features for the high-order relationships can be articulated as follows:

$$S^{t-T-t} = \beta^* \vartheta X^{t-T-t} \approx (D^v)^{-1/2} (A^{(h)} + D^v) \beta (D^e)^{-1} (D^v)^{-1/2} X^{t-T-t} \Theta \quad (26)$$

where  $\beta$  represents the spatial hypergraph convolution kernel;  ${}^*\vartheta$  indicates the spatial hypergraph convolution operator;  $D^v$  denotes a matrix composed of station hyperdegrees,  $D^v = \{D_s^v, D_b^v, D_{s,b}^v\}$ ;  $D^e$  represents a matrix composed of line hyperdegrees,  $D^e = \{D_s^e, D_b^e, D_{s,b}^e\}$ ;  $A^{(h)}$  indicates the adjacency matrix,  $A^{(h)} = \{A_s^{(h)}, A_b^{(h)}, A_{s,b}^{(h)}\}$ ;  $\Theta$  denotes the filter parameters associated with each station.

To generate the hyperedge features of the line where a station is located, the extracted station features are multiplied by the

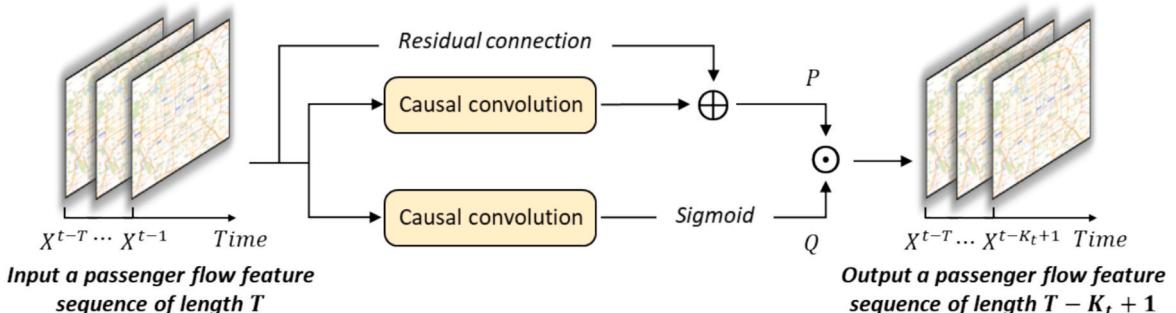


Fig. 4. Architecture of temporal gated convolutional network.

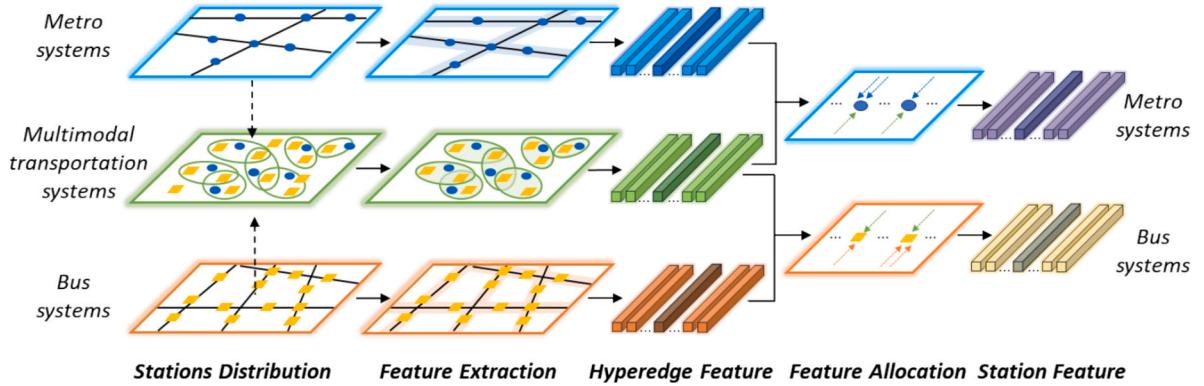


Fig. 5. Architecture of spatial hypergraph convolutional network.

transpose of the adjacency matrix, as follows:

$$E^{t-T \rightarrow t} = \text{ReLU}\left[\left(A^{(h)}\right)^T S^{t-T \rightarrow t} \Theta\right] \quad (27)$$

While static hypergraphs can connect stations with similar functions or traffic characteristics along the same operating line through hyperedges, effectively modeling high-order relationships within traffic modes, the high-order relationships between stations are dynamically influenced by changes in passenger flow. The adjacency matrix derived from static hypergraphs represents fixed hyperedges and fails to capture the dynamic correlations inherent in passenger flow data. To address this limitation, a hypergraph self-attention mechanism (Bai et al., 2021; Wu et al., 2024) is introduced to transform static hypergraphs into dynamic hypergraphs. This mechanism dynamically adjusts the weight matrix between stations using self-attention, based on real-time traffic flow data and passenger demand. This approach more effectively captures the high-order relationships between metro and bus systems across varying periods and traffic states.

Taking the metro system as an example, for each station  $v_i \in V_s$  and line  $r_k \in E_s^{(h)}$ , the weights of the adjacency matrix are updated using station and hyperedge features, as expressed by:

$$\tilde{A}^{(h)} = \frac{\exp\left(\text{ReLU}\left(\vec{a}\left[\left(S^{t-T \rightarrow t} W_s\right)_{v_i \in V_s} \parallel \left(E^{t-T \rightarrow t} W_s\right)_{r_k \in E_s^{(h)}}\right]\right)\right)}{\sum_{v_j \in r_k} \exp\left(\text{ReLU}\left(\vec{a}\left[\left(S^{t-T \rightarrow t} W_s\right)_{v_j \in V_s} \parallel \left(E^{t-T \rightarrow t} W_s\right)_{r_k \in E_s^{(h)}}\right]\right)\right)} \quad (28)$$

where  $\vec{a}$  represents the weight vector used to calculate the similarity of output vectors,  $\parallel$  indicates the concatenation operation.

#### 4.4. Multimodal feature fusion

In the STHCN, the input historical passenger flow  $X^{t-T \rightarrow t}$  comprises three components: metro passenger flow  $X_n^{t-T}$ , bus passenger flow  $X_{n^*}^{t-T}$ , and joint passenger flow  $X_{n,n^*}^{t-T} = X_n^{t-T} + X_{n^*}^{t-T}$ . After applying temporal and spatial convolution operations to aggregate the spatiotemporal relationships, the predicted passenger flow output is formulated as follows:

$$X^{t \rightarrow t+H} = \alpha_2 * \tau \text{ReLU}\left[\beta * \vartheta\left(\alpha_1 * \tau X^{t-T \rightarrow t}\right)\right] \quad (29)$$

where  $X^{t \rightarrow t+H}$  represents the predicted passenger flow output, which is defined as:

$$X^{t \rightarrow t+H} = \left\{ X_n^{t+H}, X_{n^*}^{t+H}, \tilde{X}_n^{t+H}, \tilde{X}_{n^*}^{t+H} \right\} \quad (30)$$

where  $X_n^{t+H}$  indicates the predicted metro passenger flow;  $X_{n^*}^{t+H}$  denotes the predicted bus passenger flow;  $\tilde{X}_n^{t+H}, \tilde{X}_{n^*}^{t+H}$  represent the jointly predicted passenger flow of multimodal transportation systems.

When employing the HGK-Means algorithm for bus station clustering, certain stations do not contribute to the construction of inter-modal high-order relationships. The spatiotemporal relationships derived from the combined passenger flow  $X_n^{t-T} + X_{n^*}^{t-T}$  are not universally applicable to all stations. This observation underscores that the aggregation of spatiotemporal relationships, both intra-modal and inter-modal, influences each station to varying degrees. Such variations necessitate optimization and adjustment through the training process. To address this, the predicted passenger flow output is updated as follows:

**Table 2**

The training process of the STHCN framework.

**Model:** STHCN

---

**Input:** Historical passenger flow: metro passenger flow  $X_n^{t-T}$ , bus passenger flow  $X_{n,n}^{t-T}$ , and joint passenger flow  $X_{n,n}^{t-T}$ .

High-order relationships: intra-modal relationships  $G_{\text{intra}}^{(h)} = \{G_s^{(h)}, G_b^{(h)}\}$  and inter-modal relationships  $G_{\text{inter}}^{(h)}$ .

**Output:** The STHCN with well-trained parameters  $W$ .

// Construct a set of input-output instances  $D$

Initialize a null set for datasets:  $D \leftarrow \emptyset$

for a time interval  $t$  ( $1 \leq t \leq T$ ) do

Get spatiotemporal features of all passenger flow at historical time step:  $\tilde{X}^{t-T \rightarrow t}$

//  $X^t$  is the prediction target at time  $t$

Put training sample into test datasets:  $D \leftarrow D + (\tilde{X}^{t-T \rightarrow t}, X^t)$

end

Divide  $D$  into training, validation, and test datasets:  $D_{\text{train}}$ ,  $D_{\text{valid}}$ , and  $D_{\text{test}}$ .

// Training HGK-Means algorithm

Determine the clusters: the number of metro stations  $\delta_s^{(h)}(r_k)$  for metro line  $r_k$

Randomly assign bus stations to metro stations, and calculate the initial transfer matrix

for  $k = 1 \rightarrow \text{number of metro lines } m$  do

repeat

Calculate the trade-off function  $\tilde{D}_i$  and utility collapse points  $u_i$  by Eqs. (12) and (19)

Construct the objective function  $G$  (Eq. (20)):  $G \leftarrow \max \prod_{i=1}^{\delta_s^{(h)}(r_k)} \left( \tilde{D}_i^{-1} - u_i \right)$

// Particle swarm optimization algorithm for optimization

if  $G_{\text{current}} < G_{\text{best}}$

Update the velocity and position of particles

Reassign bus stations to metro stations

Update the trade-off function  $\tilde{D}_i$  and utility collapse points  $u_i$

Calculate the new transfer distance matrix.

end

until: maximum iterations or objective function convergence

end

// Training the STHCN framework

Initialize model parameters: hidden state, all weights, biases

Concatenate the dynamic hypergraphs:  $G \leftarrow [G_s^{(h)}, G_b^{(h)}, G_{\text{inter}}^{(h)}]$

Calculate adjacency matrix  $A^{(h)}$  according to Eqs. (9), (10), and (24)

for  $n = 0 \rightarrow \text{number of epochs}$  do

Randomly select a batch of sample  $D_e$  from  $D_{\text{train}}$ , where  $e = 1, 2, \dots$

for each iteration  $i = 1, 2$  do

Obtain the output of the temporal gated convolutional layer by passing the input  $D_e$  (Eq. (25)):  $D_e^{\text{tgtc}-1} \leftarrow \alpha_1 * \tau D_e = P \odot \text{Sigmoid}(Q)$

Obtain the station features of the spatial hypergraph convolutional layer by passing the input  $D_e^{\text{tgtc}}$  (Eq. (26)):  $D_e^{\text{hgc}-s} \leftarrow S^{t-T \rightarrow t} = (D^V)^{-1/2} (A^{(h)} + D^V) \beta (D^e)^{-1} (D^V)^{-1/2} D_e^{\text{tgtc}} \Theta$

Obtain the hyperedge features of the spatial hypergraph convolutional layer by passing the input  $D_e^{\text{hgc}-s}$  (Eq. (27)):  $D_e^{\text{hgc}-h} \leftarrow E^{t-T \rightarrow t} = \text{ReLU}\left[\left(A^{(h)}\right)^T D_e^{\text{hgc}-s} \Theta\right]$

Update the adjacency matrix  $\tilde{A}^{(h)}$  of the spatial hypergraph convolutional layer by passing the input  $D_e^{\text{hgc}-s}$  and  $D_e^{\text{hgc}-h}$  (Eq. (28)):  $\tilde{A}^{(h)} = \frac{\exp(\text{ReLU}(\vec{a}[D_e^{\text{hgc}-s} W_s \| D_e^{\text{hgc}-h} W_s]))}{\sum_{v_j \in r_k} \exp(\text{ReLU}(\vec{a}[D_e^{\text{hgc}-s} W_s \| D_e^{\text{hgc}-h} W_s]))}$

Update the output of the spatial hypergraph convolutional layer by passing the input  $\tilde{A}^{(h)}$  and  $D_e^{\text{tgtc}}$  (Eq. (26)):  $D_e^{\text{hgc}} \leftarrow S^{t-T \rightarrow t} = (D^V)^{-1/2} (\tilde{A}^{(h)} + D^V) \beta (D^e)^{-1} (D^V)^{-1/2} D_e^{\text{tgtc}} \Theta$

Obtain the output of the temporal gated convolutional layer by passing the input  $D_e^{\text{hgc}}$  (Eq. (29)):  $D_e^{\text{tgtc}-2} \leftarrow \alpha_2 * \tau \text{ReLU}[\beta * \vartheta(\alpha_1 * \tau D_e^{\text{hgc}})]$

end

Obtain the output  $D_e^{\text{st}}$  of two temporal gated convolutional layers by passing the input  $D_e^{\text{tgtc}}$  (Eqs. (25) and (29))

Obtain the outputs  $X_n^{t+H}, X_{n'}^{t+H}, \tilde{X}_n^{t+H}, \tilde{X}_{n'}^{t+H}$  after  $D_e^{\text{st}}$  passing through the fully connected layer:  $X_n^{t+H}, X_{n'}^{t+H}, \tilde{X}_n^{t+H}, \tilde{X}_{n'}^{t+H} \leftarrow \sigma[W_1 \text{Flatten}(D_e^{\text{st}}) + b]$

Obtain the final outputs  $\hat{X}_n^{t+H}, \hat{X}_{n'}^{t+H}$  through the multimodal feature fusion (Eq. (31)):  $[\hat{X}_n^{t+H}, \hat{X}_{n'}^{t+H}]^T \leftarrow \Gamma \odot \begin{bmatrix} X_n^{t+H} & X_{n'}^{t+H} \\ \tilde{X}_n^{t+H} & \tilde{X}_{n'}^{t+H} \end{bmatrix}$

Optimize  $W$  by minimizing loss function (Eq. (35)):  $\text{MSE} = [\sum_{i=1}^{n_s} (y_i - y_i^*)^2] / n_s$

end

---

$$\begin{bmatrix} \hat{\mathbf{X}}_n^{t+H}, \hat{\mathbf{X}}_{n^*}^{t+H} \end{bmatrix}^T = \Gamma \odot \begin{bmatrix} \mathbf{X}_n^{t+H} & \mathbf{X}_{n^*}^{t+H} \\ \tilde{\mathbf{X}}_n^{t+H} & \tilde{\mathbf{X}}_{n^*}^{t+H} \end{bmatrix} \quad (31)$$

where  $\Gamma$  denotes a learnable parameter matrix that adjusts the influence of the aggregated spatiotemporal relationships for each station.

#### 4.5. Training strategy

The training process of the STHCN framework is designed to capture the complex spatiotemporal relationships in multimodal passenger flow data. The framework takes historical passenger flow data as input, along with intra- and inter-modal higher-order relationships.

A dataset of input–output instances is constructed by pairing spatiotemporal features at each historical time step with their corresponding prediction targets, which is then divided into training, validation, and test subsets. During training, the dynamic hypergraphs and their associated adjacency matrices are used to model complex interactions. A multi-stage convolutional process is employed: the first temporal gated convolution extracts temporal features, then spatial hypergraph convolution to capture high-order spatial relationships, and a second temporal gated convolution to refine feature representations. The outputs are passed through a fully connected layer to generate preliminary predictions, which are further refined through multimodal feature fusion. The model parameters are optimized by minimizing the Mean Squared Error (MSE) loss function, ensuring alignment between predicted and actual passenger flow. This iterative training process enables the STHCN to effectively learn and model dynamic spatiotemporal dependencies, providing accurate and robust predictions across multimodal transportation systems.

The loss function of MSE is shown as follows:

$$MSE = \frac{1}{n_s} \sum_{i=1}^{n_s} (y_i - y_i^*)^2 \quad (32)$$

where  $n_s$  represents the sample size of passenger flow data;  $y_i$  indicates the ground truth;  $y_i^*$  denotes the predicted data.

The training process of the STHCN framework is summarized in Table 2.

## 5. Experiments

In this section, the passenger SCD dataset of multimodal transportation systems in the Haidian District of Beijing is employed to validate the effectiveness of the STHCN framework. The experiments are implemented using Python 3.7 and PyTorch 1.12.1, executed on a system equipped with an Intel® Core™ i7-12700H CPU @ 2.30 GHz, an NVIDIA GeForce RTX 3060 GPU, and 16 GB of RAM. This computational setup ensures the efficient execution of the experiments and accurate evaluation of the proposed framework's performance.

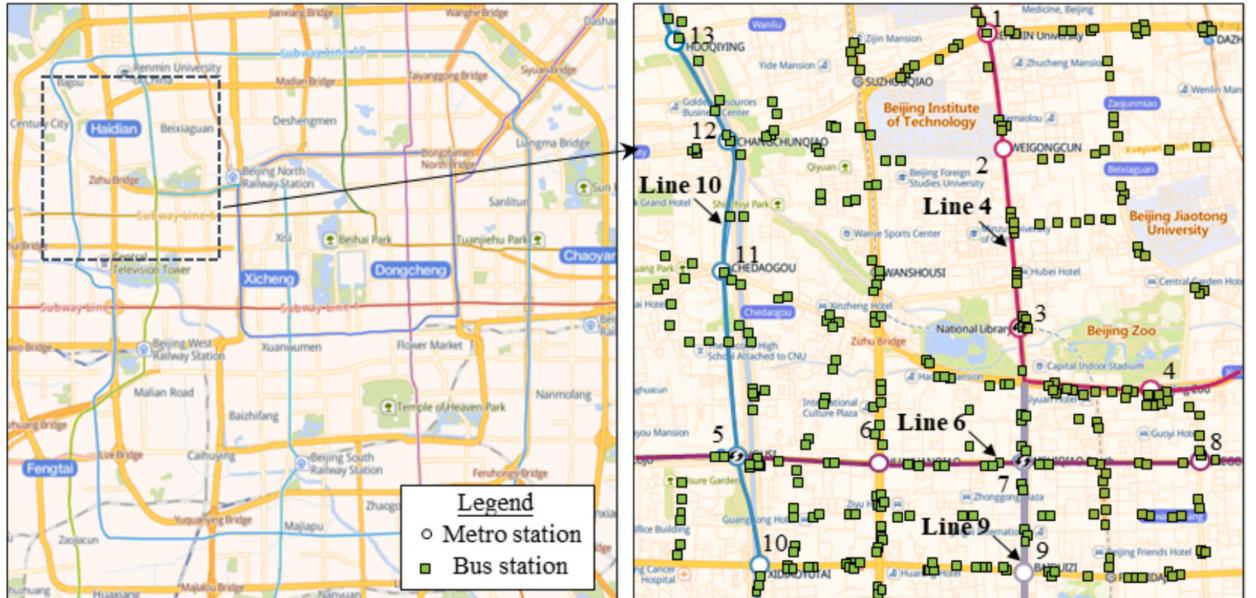


Fig. 6. The collection scope of the passenger SCD dataset.

### 5.1. Data description

The passenger SCD dataset includes 45 bus lines with 128 bus stations and four metro lines with 13 metro stations, as depicted in Fig. 6. Covering a densely populated urban area in the Haidian District of Beijing, this dataset provides an extensive spatial scope, capturing the intricate interactions between metro and bus systems.

The dataset comprises passenger smart card records collected from March 7 to March 13, 2016, specifically for multimodal transportation systems. This study is the first to collect synchronized passenger flow data from metro and bus systems within the same region and time frame, creating a pioneering multimodal dataset. Compared to datasets focused on a single traffic mode, this dataset stands out for its broad regional coverage and inclusion of a substantial number of stations, significantly enhancing its value for multimodal transportation research, prediction tasks, and real-world applications.

The critical fields of the dataset, including entry and exit times, station names, and station IDs, are detailed in Table 3.

Considering the operational hours of the public transportation systems, the analysis and predictions are concentrated on passenger flow between 5:00 AM and 12:00 PM at each station. Passenger counts are recorded at 15-minute intervals, yielding 76 time slots per day and 532 data entries for the analysis. This granularity allows for a detailed examination of spatiotemporal patterns in passenger flow, providing robust data for prediction and modeling.

### 5.2. Feature analysis

Fig. 7 illustrates the statistical distribution and temporal variation of passenger flow across selected bus and metro stations in multimodal transportation systems, while also highlighting significant differences in passenger flow patterns across regions and traffic modes. The left panel presents box plots of the daily passenger volumes for four bus stations (Stations 32, 41, 101, and 128) and four metro stations (Stations 1, 2, 8, and 10). These boxplots prominently display the median, interquartile range, and outliers of passenger flow at each station, revealing the overall scale and variability of passengers. The analysis shows that metro passenger flow predominates, whereas bus passenger flow is comparatively lower. Furthermore, substantial differences in total passenger flows across metro stations reflect the varying demand levels at different locations..

The right panel depicts time series plots of passenger flow recorded at 15-minute intervals throughout the day (5:00 AM to 12:00 PM). These plots reveal the temporal distribution characteristics of bus and metro systems, which typically exhibit a bimodal pattern, although their peak times differ. Metro passenger flow tends to peak within a narrower time window with less fluctuation, whereas bus passenger flow demonstrates greater variability and more dispersed peak periods. Additionally, the shaded regions in the time series plots reflect variations in passenger flow across different days, further emphasizing the dynamic nature of demand.

Overall, the passenger flow dynamics in multimodal transportation systems exhibit significant spatiotemporal heterogeneity, which is influenced by various factors and displays distinct trends. The metro system primarily handles large-scale, centralized passenger flows with high efficiency, while the bus system provides flexible services that extend coverage to more areas. Integrating statistical distributions and temporal variations offers a comprehensive perspective for understanding passenger flow dynamics, providing critical insights for optimizing transportation planning and predicting future demand.

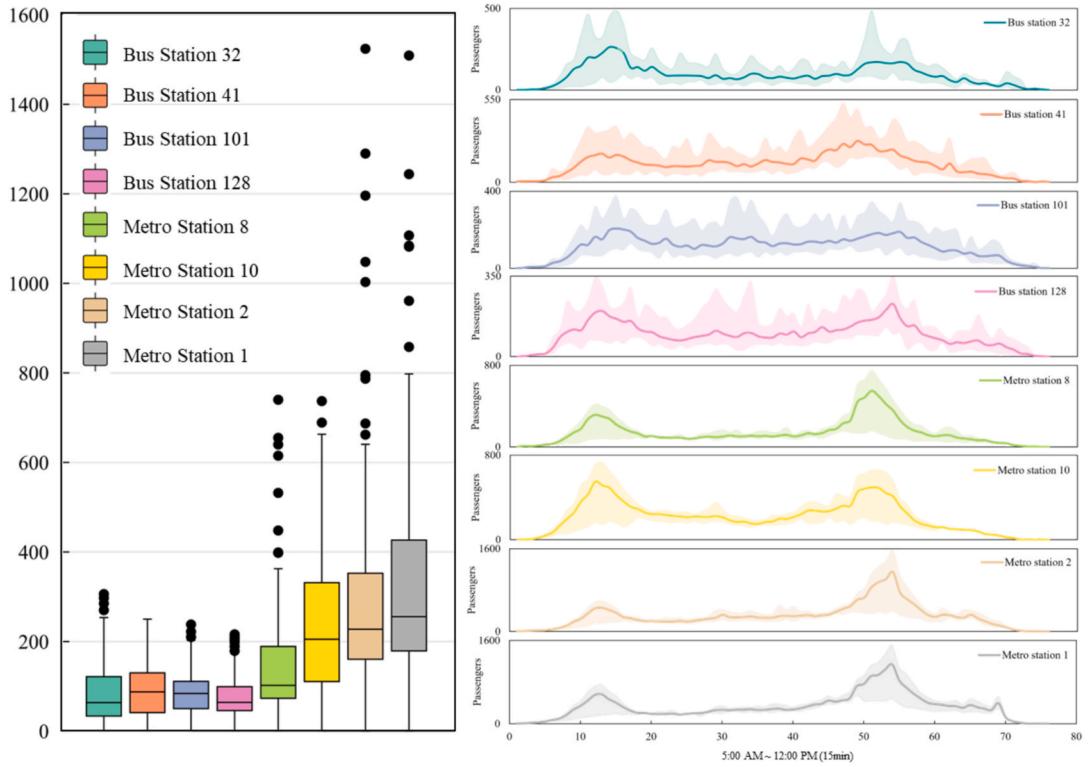
Fig. 8 and 9 Depict the temporal passenger flow distribution across selected metro and bus stations, respectively, highlighting variations in travel patterns over the course of a day. notably, the selected metro stations are all part of the same Metro Line 10, and the selected bus stations are all part of the same Bus Line 61.

In Fig. 8, the passenger flow patterns at Metro Stations 10, 5, 11, 12, and 13 demonstrate a clear bimodal distribution, with peaks corresponding to morning and evening rush hours. These peaks are more pronounced in the morning (around 8:00 AM) and evening (around 6:00 PM), reflecting the typical commuting behavior of passengers using the metro system. Station 5 shows the highest peak in passenger flow, indicating its significance as a major hub or transfer station. In contrast, Station 13 exhibits a unique peak during the morning hours, suggesting that it primarily serves passengers commuting to work or school. The overall temporal patterns of metro

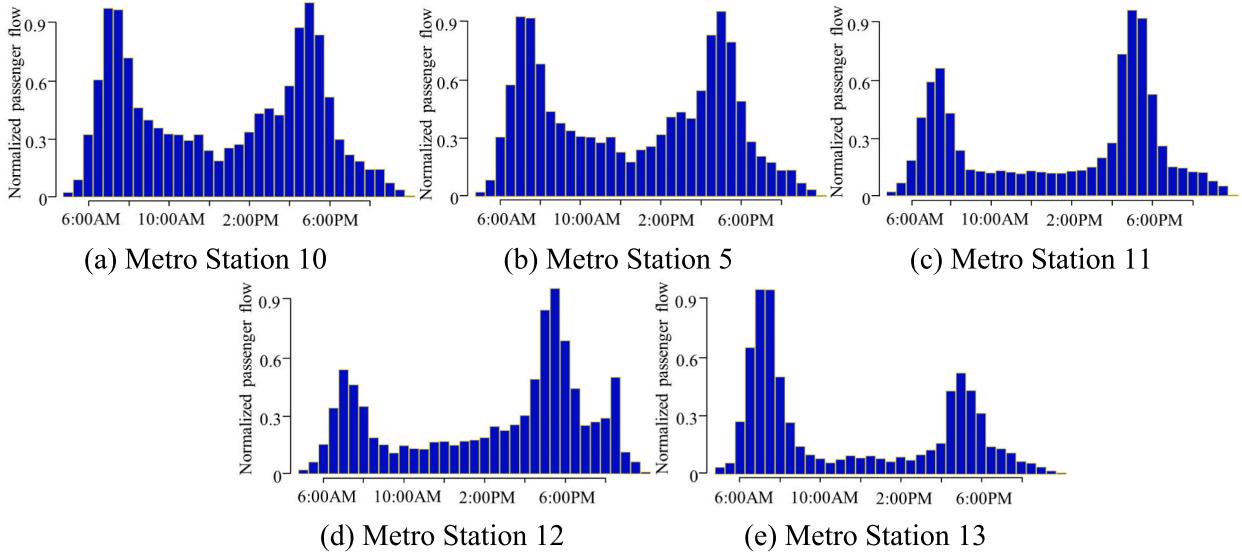
**Table 3**

Passenger SCD dataset of multimodal transportation systems.

Type	Variable	Description	Value
Metrosystem	GRANT_ID	Grant card ID	10000 ~ 99999999
	ENTRY_TIME	Entry station time	2016/03/07 00:00:00 ~ 2016/03/13 23:59:59
	DEAL_TIME	Deal time	2016/03/07 00:00:00 ~ 2016/03/13 23:59:59
	LINE_CODE	Line code	4, 6, 9, 10
	ENTRY_NUM	Entry station number	1 ~ 13
	EXIT_NUM	Exit station number	1 ~ 13
	ENTRY_STATION	Entry station name	—
Bus system	EXIT_STATION	Exit station name	—
	BUSDATA_ID	Bus card ID	18... (11 digits total)
	DEAL_TIME	Deal time	2016/03/07 00:00:00 ~ 2016/03/13 23:59:59
	UP_TIME	Update time	2016/03/07 00:00:00 ~ 2016/03/13 23:59:59
	LINE_CODE	Line code	114, 121, ... (45 lines total)
	VEHICLE_CODE	Vehicle code	—
	ON_STATION	Getting on station number	1 ~ 128
	OFF_STATION	Getting off station number	1 ~ 128



**Fig. 7.** Passenger flow distribution and temporal variation across selected bus and metro stations.

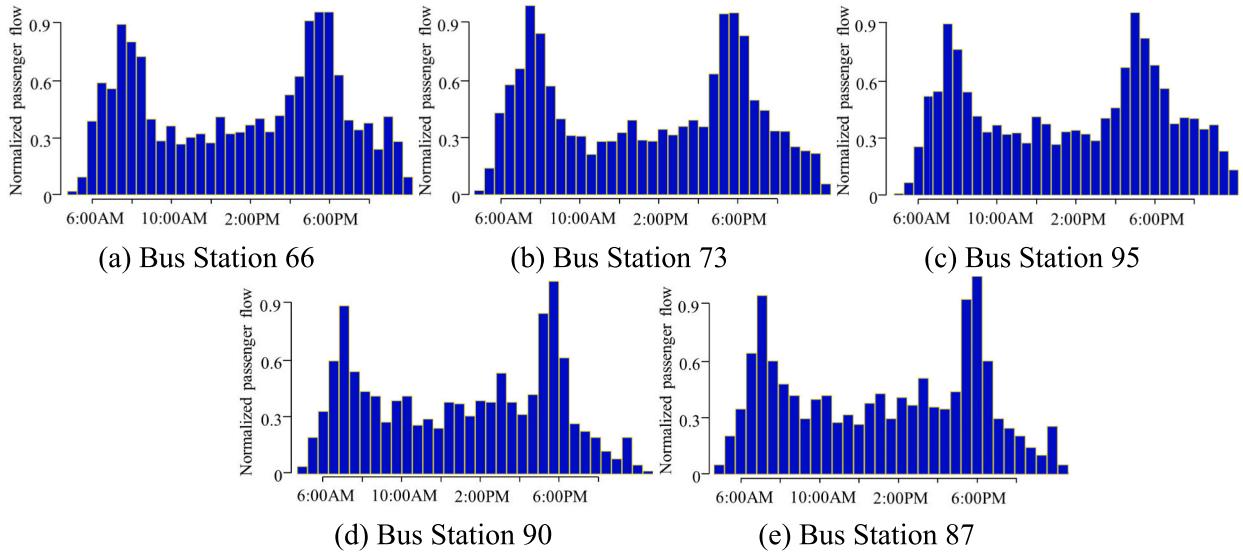


**Fig. 8.** Temporal passenger flow distribution across selected metro stations.

stations exhibit lower variability during non-peak hours, reflecting the metro system's stable passenger flow throughout the day.

In Fig. 9, the passenger flow distributions at Bus Stations 66, 73, 95, 90, and 87 display a similar bimodal pattern, though the peaks are less distinct compared to metro stations. These stations exhibit greater variability in passenger flow, likely due to the more localized and flexible nature of bus services. For example, Station 66 shows a more evenly distributed flow throughout the day, indicating consistent demand, while Station 87 demonstrates higher evening peaks, suggesting its importance for returning commuters.

The comparison between Fig. 8 and Fig. 9 underscores the differences in passenger flow dynamics between metro and bus systems.

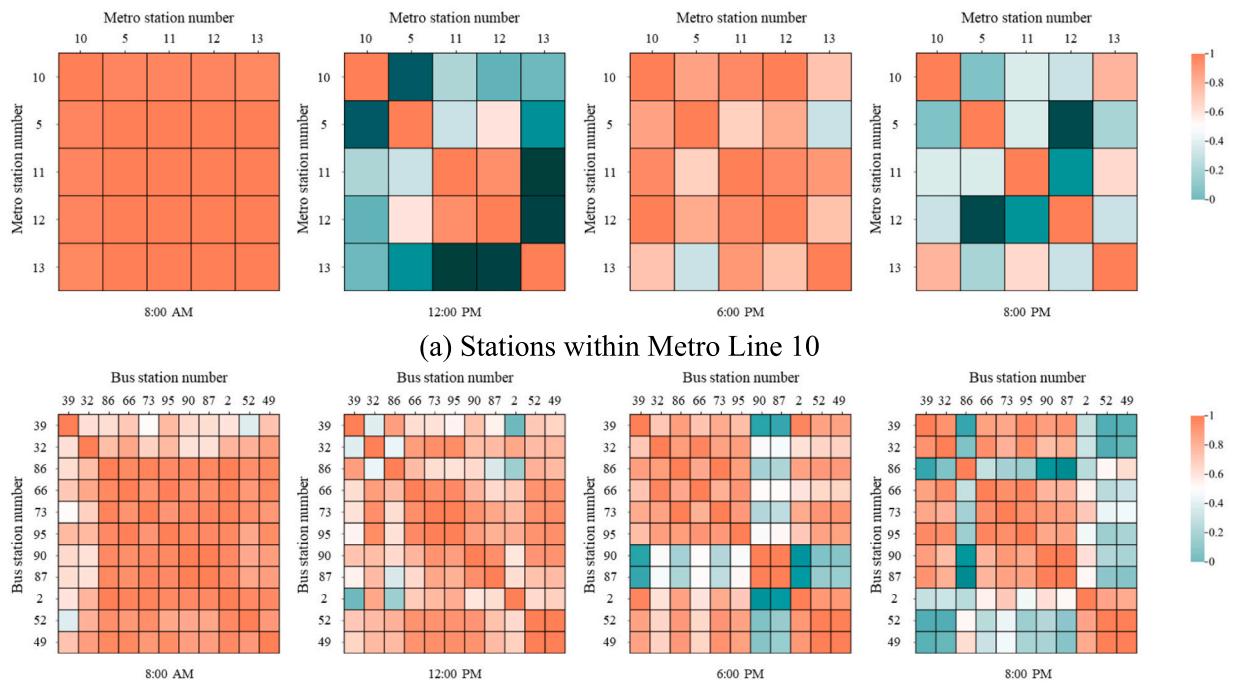


**Fig. 9.** Temporal passenger flow distribution across selected bus stations.

Metro stations exhibit more concentrated and consistent peak times, while bus stations show greater temporal variability and more dispersed peaks, reflecting their complementary roles in multimodal transportation systems. This analysis highlights the necessity of integrating these modes to meet diverse passenger needs efficiently.

Fig. 10 illustrates the spatial passenger flow distribution across selected metro and bus lines, focusing on analyzing the correlation between stations along the same line. Specifically, the horizontal and vertical axes represent the station numbers along each line, with the colors in the heatmaps indicating the degree of spatial correlation between the stations at different times.

In Fig. 10(a), the stations of Metro Line 10 exhibit a strong correlation during peak hours, but show a degree of independence during off-peak periods. Specifically, during peak hours in the morning and evening (around 8:00 AM and 6:00 PM), the metro system



**Fig. 10.** Spatial passenger flow distribution across selected metro and bus lines.

primarily serves commuting traffic to and from work and school. The changes in passenger flow at Metro Stations 5, 11, 12, and 13 are highly correlated with Metro Station 10 during these times. However, during off-peak hours (around 12:00 PM and 8:00 PM), the focus of metro service shifts to cater to other travel needs, such as leisure and shopping trips, which reduces the correlation between Metro Stations 5, 11, and 12 and Metro Station 10.

In Fig. 10(b), the stations of Bus Line 61 show strong correlation throughout the day, reflecting the short-distance travel demand typical of bus lines. However, there are subtle variations in station interdependence at different times of day. For example, at 8:00 AM and 12:00 PM, strong dependencies are observed between Bus Station 66 and Bus Station 49, whereas at 6:00 PM and 8:00 PM, the dependency shifts toward Bus Station 37 and Bus Station 89. This indicates that the spatial correlation is influenced by the fluctuations in passenger flow at different times.

By comparing Fig. 10(a) and Fig. 10(b), it is evident that the spatial correlation between metro and bus systems differs significantly. The passenger flow pattern of metro lines demonstrates a strong station-to-station correlation during peak hours, with a noticeable weakening of this dependence during off-peak periods due to varying travel demands and the broader coverage of metro services. In contrast, stations along bus lines show persistent interdependence across all times of the day, reflecting the more localized and flexible nature of bus services. Overall, the spatial correlation between stations on the same line extends beyond simple pairwise relationships between adjacent stations, revealing more significant higher-order relationships.

Fig. 11 presents the temporal alignment of passenger flow across selected metro and bus stations, derived using the HGK-Means algorithm to construct hypergraphs that represent inter-modal interactions. Fig. 11(a) shows Metro Station 2 and its associated Bus Stations 101, 102, 122, and 123, while Fig. 11(b) depicts Metro Station 8 and its associated Bus Stations 9, 15, 47, and 63.

In Fig. 11(a), the passenger flow distribution at Metro Station 2 closely aligns with the combined flow trends of the bus stations in the hypergraph, particularly during morning and evening peak hours. This correspondence highlights significant inter-modal interactions, indicating effective passenger flow transfer between metro and bus systems at this location. The evening peak at Metro Station 2 is more prominent, suggesting its pivotal role in handling returning commuters during rush hours. Fig. 11(b) demonstrates that while the passenger flow at Metro Station 8 is relatively lower in magnitude, its temporal distribution remains consistent with the trends observed at the associated bus stations. This alignment underscores the interconnected nature of passenger flow across traffic modes, even at stations with lower individual demand. The peaks during the morning and evening indicate the shared commuting patterns within the multimodal transportation systems.

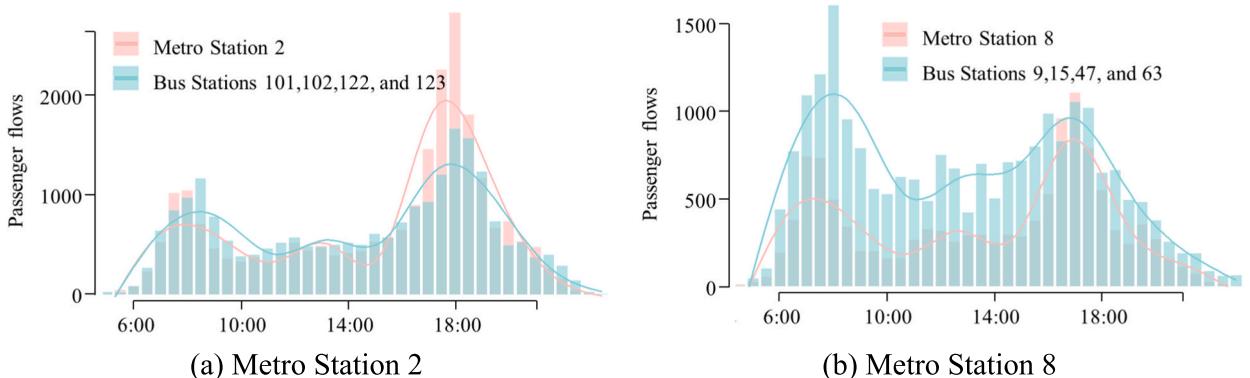
These findings emphasize the importance of dynamic hypergraph-based analysis in capturing inter-modal interactions. The results reveal that metro and bus systems complement each other in terms of spatial coverage and exhibit synchronized temporal patterns, driven by direct passenger flow transfers. This strong inter-modal connectivity reinforces the necessity of incorporating dynamic inter-modal interactions when analyzing and optimizing multimodal transportation systems, particularly for enhancing operational efficiency and passenger experience.

### 5.3. Evaluation metrics

To effectively evaluate the performance of the STHCN framework in joint passenger flow prediction, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson Correlation Coefficient (PCC) are selected as the evaluation metrics, defined as follows:

$$MAE = \frac{1}{n_s} \sum_{i=1}^{n_s} |y_i - y_i^*| \quad (33)$$

$$RMSE = \sqrt{\frac{1}{n_s} \sum_{i=1}^{n_s} (y_i - y_i^*)^2} \quad (34)$$



**Fig. 11.** Temporal alignment of passenger flow across selected metro and bus stations in multimodal transportation systems.

$$PCC = \frac{\sum_{i=1}^{n_s} [(y_i - \bar{y}_i)(y_i^* - \bar{y}_i^*)]}{\sqrt{\sum_{i=1}^{n_s} (y_i - \bar{y}_i)^2} \sqrt{\sum_{i=1}^{n_s} (y_i^* - \bar{y}_i^*)^2}} \quad (35)$$

where  $n_s$  represents the total number of data;  $\bar{y}_i$  indicates the average of the ground truth;  $\bar{y}_i^*$  denotes the average of the predicted data.

The combination of MAE, RMSE, and PCC provides a comprehensive evaluation of the STHCN framework from multiple perspectives. While MAE and RMSE assess the absolute prediction errors, PCC specifically evaluates the consistency of trends in replicating passenger flow patterns (Ye et al., 2019; Zhang et al., 2024). This multiple faceted approach ensures a thorough analysis of model performance, capturing both the accuracy of predictions and the alignment with observed trends.

#### 5.4. Baseline models

Several baseline models are selected for comparison with the STHCN framework, including traditional, single, and hybrid models based on RNNs, CNNs, and GCNs architectures. The parameters of the baseline models are determined using the control variable method to ensure consistency and fairness in evaluation. The details of these baseline models are outlined as follows:

**HA**: This method predicts future passenger flow by calculating the average passenger count during corresponding periods in historical data based on observed seasonal patterns.

**ARIMA** (Zhang et al., 2011): A widely adopted time series method that identifies trends and seasonal fluctuations in historical data to predict future passenger flow at specific time intervals.

**RF** (Filipovska and Mahmassani, 2020): A machine learning approach for regression and classification tasks. It generates predictions by constructing an ensemble of decision trees, where the number of trees is set to 10, and the maximum depth of each tree is determined automatically based on the purity of leaf nodes or sample size constraints.

**MLP** (Bratsas et al., 2019): This model predicts future passenger flow based on historical data from multiple stations. The network comprises two fully connected layers with 256 and  $2 \times 4 \times$  stations neurons, respectively.

**LSTM** (Zhao et al., 2017): A variant of RNN, LSTM effectively models temporal dependencies in traffic data. This model utilizes a two-layer LSTM to capture passenger flow variations across traffic modes.

**GRNN** (Tan et al., 2021): This model enhances clustering performance by refining the K-Means clustering metric with a between-within proportion similarity. It integrates GRNN with parameter optimization through a genetic algorithm to predict passenger flow.

**T-GCN** (Zhao et al., 2020b): This model captures spatiotemporal dependencies by combining GCN and GRU. TGCN uses 5000 training epochs, a  $3 \times 3$  convolution kernel in the GCN layer, and a single GRU layer with 4 hidden units.

**STGCN** (Yu et al., 2018): This model captures both spatial and temporal dependencies by employing a spatial GCN layer and a temporal gated convolutional layer, making it suitable for passenger flow prediction.

**DCRNN** (Li et al., 2018): This model combines diffusion-based GCN with GRU to model spatiotemporal dependencies. The diffusion order is set to 2, and other hyperparameters are aligned with those in STHCN.

**Conv-GCN** (Zhang et al., 2020): This model first processes passenger flow using multiple GCN, followed by a 3D CNN to integrate these flows. The final predictions are made through a fully connected layer.

**Multi-STGCnet** (Ye et al., 2020): This model incorporates temporal and spatial components. The temporal component utilizes three LSTM layers to capture temporal dependencies, while the spatial component integrates three GCN layers with LSTM to model spatial correlations.

**Ada-MSHyper** (Shang et al., 2024): This model introduces an adaptive hypergraph learning module and a multi-scale interaction module, utilizing node and hyperedge constraint mechanisms to enable the Transformer to model grouped pattern interactions at different scales.

All baseline models are based on a single-mode, independent prediction approach. In metro prediction, only metro passenger flows and their adjacency relationships are used as inputs, while bus prediction tasks rely exclusively on bus data. This configuration is

**Table 4**  
Model performance comparison of multimodal transportation systems.

Models	Metro system			Bus system		
	MAE	RMSE	PCC	MAE	RMSE	PCC
HA	50.22	82.06	0.8392	33.07	40.65	0.7898
ARIMA	47.76	79.53	0.8059	29.18	36.35	0.7349
RF	49.61	83.97	0.8039	27.98	38.58	0.6636
LSTM	44.42	78.21	0.9281	33.50	40.44	0.8932
GRNN	43.52	78.59	0.8805	28.83	39.80	0.7842
T-GCN	46.15	67.74	0.8744	26.47	36.49	0.7007
STGCN	38.92	71.07	0.9204	21.26	33.73	0.7468
DCRNN	36.82	65.18	0.9303	19.89	31.44	0.8717
Conv-GCN	37.71	64.90	0.9395	20.68	33.15	0.9322
Multi-STGCnet	36.67	63.94	0.9603	17.42	27.01	0.9383
Ada-MSHyper	34.16	59.66	0.9681	15.97	24.78	0.9429
<b>STHCN</b>	<b>31.84</b>	<b>56.08</b>	<b>0.9727</b>	<b>15.43</b>	<b>23.93</b>	<b>0.9490</b>

designed to evaluate the improvement in prediction performance resulting from the multimodal dynamic interaction mechanism, rather than attributing the performance enhancement solely to an increase in data scale.

## 6. Results and discussions

This study utilizes historical passenger flow from metro and bus systems over 3 h as input to predict short-term (15 min) and medium- to long-term (30 min, 60 min and 120 min) passenger flow. The experimental model is configured with the following hyperparameters: both the temporal convolution and spatial hypergraph convolution kernels are set to a size of 3, with the channel configurations of [1, 32, 32, 64] and [64, 32, 32, 128], respectively. During training, the number of epochs is set to 20, the batch size is set to 32, and the initial learning rate is set to 0.001. The learning rate decays by a factor of 0.7 after every 5 epochs. In the passenger SCD datasets, 71 % is used for training and validation, and 29 % is used for testing. The datasets are available at <https://github.com/luodongyu-bjtu/Joint-passenger-flow-data>.

### 6.1. Performance comparison of different models

**Table 4** presents the short-term passenger flow prediction performance of the baseline models and the STHCN on the selected dataset. The evaluation metrics represent the average results obtained from 10 independent runs for each model, with a prediction interval of 15 min. Overall, the STHCN demonstrates superior prediction accuracy in multimodal transportation systems, outperforming all baseline models and achieving the best results across all evaluation metrics.

Metro passenger flow in the metro system is generally more predictable and consistent than in the bus system, resulting in better evaluation metric performance for metro predictions than bus predictions. Across the board, the STHCN achieves significantly higher prediction accuracy in multimodal systems. Compared to Ada-MSHyper, the best-performing baseline model among all the 11 evaluated methods, the STHCN delivers substantial improvements: reductions in MAE of at least 6.79 % and 3.38 %, reductions in RMSE of at least 6.00 % and 3.43 %, and increases in PCC of at least 3.55 % and 0.64 % for metro and bus systems, respectively. These results underscore the model's effectiveness and accuracy in joint passenger flow prediction tasks, particularly in capturing the complexity of multimodal transportation systems.

Traditional time series models, such as HA and ARIMA, face notable limitations in handling nonlinear and complex transportation data. Passenger flow is influenced by numerous intricate and nonlinear factors that these models cannot adequately capture. While flexible, neural network-based approaches, such as MLP and LSTM, often involve many parameters and are prone to overfitting, especially when learning noise or local features, resulting in reduced generalization ability on test data. Among the machine learning models considered, RF performs relatively well, with some metrics comparable to those achieved by spatiotemporal graph convolution-based methods. This performance can be attributed to RF's capability to enhance accuracy by increasing the number of trees. However, its inability to account for spatial interactions in passenger flow limits its predictive accuracy to individual stations.

Graph-based models demonstrate strong performance in predicting passenger flow across both traffic modes by integrating temporal and spatial features. Despite their strengths, the STHCN surpasses these models because it can effectively capture and model complex spatiotemporal interactions across multiple traffic modes. While existing models often focus exclusively on internal features within individual modes, the STHCN uniquely incorporates dynamic inter-modal interactions. The model leverages multiple graph convolution layers and has also achieved significant results, similar to [Zhang et al. \(2020\)](#). Conventional graph structures remain constrained in their ability to model intra-modal passenger flow characteristics and inter-modal dynamic interactions simultaneously. The fusion of hypergraph and Transformer architectures exhibits remarkable potential in enabling multi-scale interactions, similar to [Shang et al. \(2024\)](#). However, this integration primarily conceptualizes the high-level relationships between intra- and inter-modal interactions as a single category without a detailed distinction between the various traffic modes.

Additionally, a comparison with models adopting a sandwich structure, similar to [Ye et al., \(2020\)](#), highlights the STHCN's advantages. While the sandwich structure excels in capturing spatiotemporal features through its aggregation module, the STHCN further enhances prediction accuracy by incorporating dynamic hypergraph-based structures. This approach captures the characteristics of individual traffic modes and accounts for inter-modal dynamic interactions, making it highly effective for multimodal transportation systems.

To assess the impact of changes in passenger flow on prediction outcomes, **Table 5** divides the multimodal transportation systems into distinct regions. Specifically, Regions 1 to 4 represent the passenger flow transfer for four metro lines (Metro Lines 4, 6, 9, and 10).

**Table 5**  
Regional division of multimodal transportation systems.

Regions	Description	Metro passenger flow	
		Standard deviation	Mean
Region 1	Metro Stations 1, 2, 3, and 4 on Metro Line 4 and the bus stations that interact with them	(300.30, 307.83, 238.86, 172.47)	254.95
Region 2	Metro Stations 5, 6, 7, and 8 on Metro Line 6 and the bus stations that interact with them	(136.11, 220.94, 269.96, 163.98)	202.74
Region 3	Metro Stations 3, 7, and 9 on Metro Line 9 and the bus stations that interact with them	(238.86, 269.96, 158.23)	222.35
Region 4	Metro Stations 10, 5, 11, 12, and 13 on Metro Line 10 and the bus stations that interact with them	(179.53, 136.11, 225.66, 214.49, 84.76)	168.11
Region 5	Other bus stations that do not participate in the inter-modal dynamic interactions	—	—

Each region encompasses the metro stations along the respective lines and the bus stations that interact dynamically with them. Notably, there are variations in the scale of passenger flow between Regions 1 and 4, with the mean and standard deviation of passenger flow used to quantify these differences. Furthermore, some bus stations do not engage in the dynamic interactions between the metro and bus systems and are therefore categorized separately in Region 5.

**Tables 6 and 7** present a performance comparison of metro and bus systems in different regions, and **Fig. 12** shows the average and interval variation values of the experimental results. Since PCC, as an overall evaluation metric for multimodal transportation systems, is not well-suited for regional analysis, MAE and RMSE are adopted as the evaluation metrics for a more granular comparative assessment. The results indicate that STHCN demonstrates outstanding predictive performance across all regions of the multimodal transportation system, with average values outperforming all baseline models on every evaluation metric, particularly excelling in managing passenger flow fluctuations and capturing dynamic interactions.

The results indicate that the STHCN exhibits superior prediction performance in all regions of the multimodal transportation system, outperforming all baseline models across all evaluation metrics, particularly in handling passenger flow fluctuations and capturing dynamic interactions.

In **Table 6**, the metro predictive performance is closely tied to regional passenger flow volatility. The greater the fluctuation in passenger flow, the more pronounced the prediction errors. For instance, the mean passenger flow in Region 1 is 86.84 higher than in Region 4, leading to increases of 34.09 % in MAE and 24.51 % in RMSE for the STHCN in Region 1. This result underscores that regions with significant fluctuations in passenger flow are more challenging to predict compared to regions with stable flow patterns. These fluctuations are primarily driven by substantial differences between peak and off-peak periods. Unlike the bus system that primarily cater to short-distance travel, the metro system experiences a stark contrast in travel demand between peak and off-peak hours due to the distribution of long-distance stations, which causes significant variation in passenger flow. Despite these challenges, the STHCN consistently demonstrates strong predictive ability across all regions. Compared to the best-performing baseline models, the STHCN reduced MAE by at least 4.61 %, 7.33 %, 6.86 %, and 11.19 % in Regions 1 to 4, and RMSE by at least 6.99 %, 3.44 %, 4.71 %, and 8.75 %, respectively. This highlights STHCN's ability to mitigate the impact of passenger flow fluctuations on prediction accuracy through dynamic interaction learning.

In **Table 7**, the bus results exhibit relatively consistent prediction performance across regions 1 to 4, with minor prediction differences for the STHCN. However, in Region 5, the MAE and RMSE for the STHCN increase by at least 6.18 % and 7.64 %, respectively, compared to other regions, emphasizing the importance of modeling dynamic interactions between different traffic modes. The baseline models, which do not account for the dynamic interactions between the metro and bus systems, show no significant variation in prediction performance across Regions 1 to 5. In contrast, the STHCN shows limited improvement in Region 5, with MAE reduced by only 2.23 % and RMSE by 1.65 %. However, in Regions 1 to 4, the STHCN demonstrates substantial improvements in prediction accuracy, with MAE reduced by at least 9.34 %, 8.99 %, 9.13 %, and 9.87 %, respectively, and RMSE reduced by at least 8.77 %, 9.52 %, 7.86 %, and 5.11 %, respectively. These results illustrate that the STHCN effectively captures the travel and fluctuation patterns of bus passenger flow, significantly improving prediction accuracy in regions with high passenger flow variability. This further confirms the efficacy and robustness of inter-modal dynamic interactions learning in enhancing the prediction performance.

In **Fig. 12(a)**, the overall error range remained stable, despite a surge in traffic within the Region 1 of the metro system, which led to a significant increase in MAE and RMSE indicators. This demonstrates that the STHCN exhibits strong adaptability to high traffic fluctuations, and its predictive stability has been thoroughly validated. In **Fig. 12(b)**, the STHCN shows even greater robustness within the bus system, maintaining high prediction accuracy across different regions. Despite the inherent instability of the bus system, the model's predictive performance remained largely unaffected. This highlights that the STHCN framework possesses both strong adaptability and stability in the comprehensive prediction of multimodal transportation systems.

**Fig. 13** provides an intuitive demonstration of the STHCN's performance in predicting passenger flow, presenting the actual and predicted passenger flow values for three metro stations (Stations 3, 5, and 10) and bus stations (Stations 20, 66, and 89), as well as the error between the actual and predicted values. For metro passenger flow, the patterns are relatively stable and predictable, characterized by distinct morning and evening peaks, which correspond to typical commuting behaviors. The STHCN excels in capturing

**Table 6**  
Model performance comparison of metro systems in different regions.

Models	Region 1		Region 2		Region 3		Region 4	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
HA	58.35	92.48	48.98	78.92	49.02	85.86	45.21	73.22
ARIMA	53.37	87.68	48.46	81.25	47.51	77.52	42.01	71.14
RF	56.28	94.69	48.65	81.82	49.16	85.82	44.49	74.60
LSTM	48.44	86.19	41.81	74.73	46.94	84.08	40.02	67.74
GRNN	49.99	88.22	41.93	77.22	43.39	80.45	39.73	70.36
T-GCN	50.93	79.02	46.52	69.18	47.06	69.01	40.80	58.23
STGCN	46.62	76.36	37.90	70.90	37.88	73.04	37.01	64.91
DCRNN	44.22	72.91	36.21	66.13	35.81	64.24	32.62	58.43
Conv-GCN	45.99	72.67	35.92	65.40	35.94	66.62	35.05	57.17
Multi-STGCnet	42.79	71.89	35.96	61.13	35.46	66.81	32.43	57.35
Ada-MSHyper	41.85	65.98	32.47	59.38	33.26	62.21	29.66	50.77
<b>STHCN</b>	<b>39.92</b>	<b>61.37</b>	<b>30.09</b>	<b>57.34</b>	<b>30.98</b>	<b>59.28</b>	<b>26.34</b>	<b>46.33</b>

**Table 7**

Model performance comparison of bus systems in different regions.

Models	Region 1		Region 2		Region 3		Region 4		Region 5	
	MAE	RMSE								
HA	33.45	39.98	32.80	40.28	33.88	41.10	32.54	39.62	33.95	42.41
ARIMA	28.94	35.10	29.20	35.99	29.70	36.9±	27.51	35.13	30.64	38.36
RF	28.05	38.42	27.85	38.08	28.42	39.15	26.59	36.78	29.01	40.85
LSTM	33.73	40.69	33.64	40.01	32.71	39.39	32.57	39.41	34.69	42.42
GRNN	28.90	39.29	29.01	40.19	28.60	39.11	27.90	38.65	29.75	41.47
T-GCN	26.79	36.87	26.33	36.16	26.81	36.96	25.15	35.04	27.61	37.30
STGCN	21.23	33.97	21.47	34.11	21.81	33.62	20.75	32.62	21.43	34.56
DCRNN	19.90	30.99	19.58	31.12	19.63	31.89	19.45	30.49	20.84	32.94
Conv-GCN	20.90	33.49	20.80	32.87	20.79	33.34	20.16	32.13	21.11	34.23
Multi-STGCnet	17.91	26.48	17.70	26.82	17.48	26.91	16.50	26.64	17.81	28.33
Ada-MSHyper	17.24	26.01	17.01	26.06	17.08	26.34	15.19	23.09	17.04	26.72
STHCN	<b>15.63</b>	<b>23.73</b>	<b>15.48</b>	<b>23.58</b>	<b>15.52</b>	<b>24.27</b>	<b>13.69</b>	<b>21.91</b>	<b>16.66</b>	<b>26.28</b>

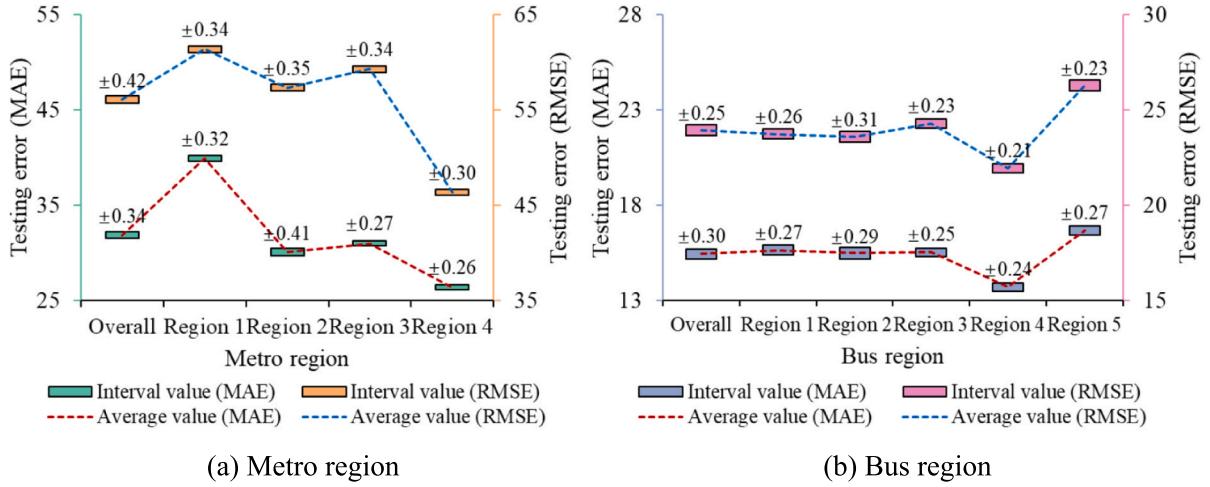


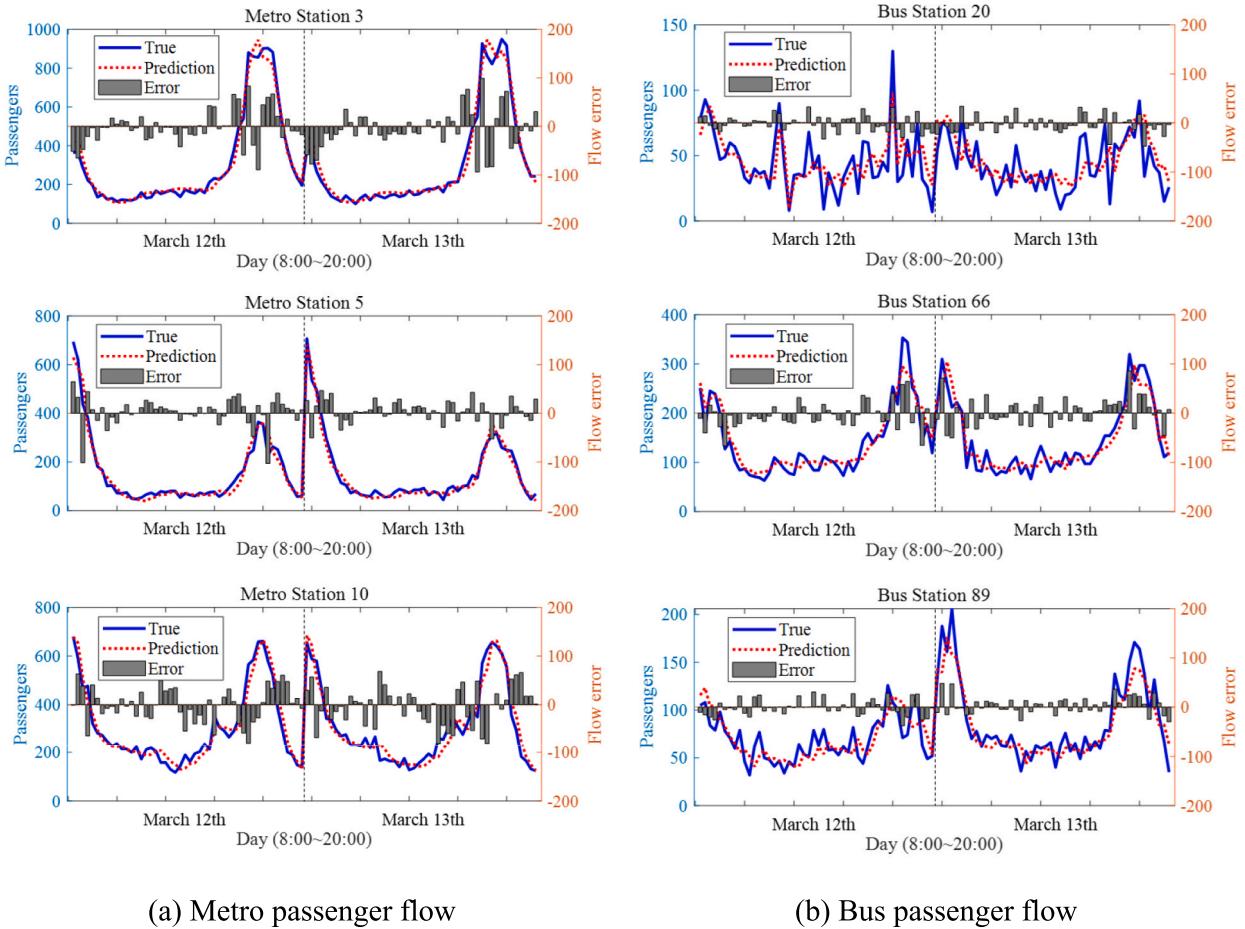
Fig. 12. Error comparison in different regions.

these trends, as seen in the minimal deviations between actual and predicted values in Fig. 13(a). For instance, Metro Station 10 demonstrates consistent prediction accuracy, with errors largely confined to minor deviations during peak hours. While Metro Station 3 shows slightly larger errors during the evening peak, these discrepancies are reflected in the flow error (black bars) and may be attributed to sudden passenger surges or localized events not captured by the model.

In contrast, bus passenger flow is more dynamic and irregular, particularly during off-peak hours, as shown in Fig. 13(b). This variability can be attributed to the diverse travel purposes of passengers during non-commuting times, resulting in less predictable flow patterns. For example, Bus Station 66 exhibits strong prediction accuracy during peak periods but shows larger deviations during off-peak hours, particularly on March 13. Similarly, Bus Station 20 displays significant fluctuations and larger errors during midday hours. Nevertheless, the STHCN effectively captures the general flow trends and accurately predicts regular patterns during high-traffic periods, and the main trends during off-peak times. Bus Station 89, while exhibiting consistent performance, shows slightly higher errors during the evening peak. Overall, the results emphasize the robustness of the STHCN in handling multimodal passenger flow predictions. The model performs exceptionally well in the metro system, where flow is more stable and predictable, while successfully adapting to the higher variability of the bus system. Despite the inherent challenges of irregular passenger flow during off-peak periods, the STHCN effectively captures both regular and fluctuating patterns, demonstrating its suitability for multimodal transportation systems.

To further evaluate the predictive performance of the models, both the baseline models and the STHCN are applied to medium- and long-term passenger flow prediction tasks within the multimodal transportation system, as shown in Fig. 14. The experimental results reveal that models such as Conv-GCN and Multi-STGCnet, which leverage multiple GCNs, outperform other baseline models in terms of predictive accuracy. However, the STHCN achieves superior performance by integrating a novel dynamic hypergraph structure, with particularly significant improvements observed in predicting bus passenger flow.

In the 120-minute metro passenger flow prediction, a decline in evaluation metrics is observed. This decrease can be attributed to the dynamic hypergraph constructed using the HGK-Means algorithm, which effectively captures critical inter-modal dynamic interaction information. For metro passenger flow, the high volatility and complexity of interactions, especially over longer prediction



**Fig. 13.** Passenger flow prediction results across selected metro and bus stations in multimodal transportation systems.

horizons, increase variability in the dynamic information. This added complexity may negatively affect the model's stability, potentially leading to counterproductive effects. Despite these challenges, the STHCN demonstrates its robustness by delivering more accurate and comprehensive predictions across both traffic modes, further underscoring its effectiveness in capturing spatiotemporal and inter-modal dynamics.

## 6.2. Ablation experiment

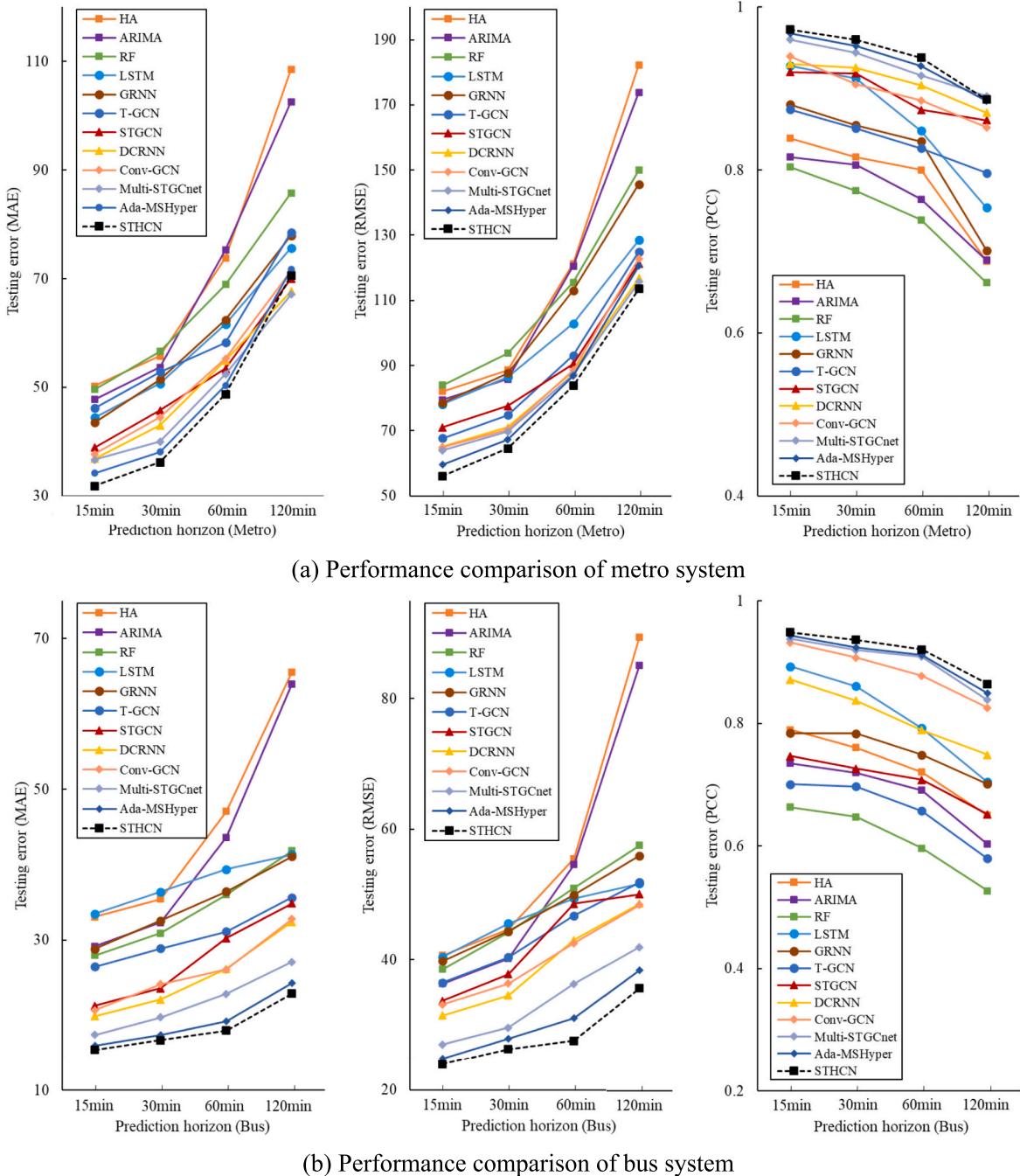
To comprehensively evaluate the contribution of each module within the STHCN framework, a series of variant models are constructed and their performance is validated through ablation experiments. Ablation experiments systematically assess the impact of specific components or features on overall model performance by progressively removing or modifying them. The specific experimental settings are as follows:

**Ablation 1** (*– Hypergraph self-attention mechanism*): In this variant, the hypergraph self-attention mechanism is removed, inter-modal high-order relationships are constructed using only the static hypergraph, and inter-modal dynamic interactions are preserved. This experiment explores the advantages of dynamic hypergraphs, formed by adding the attention mechanism, in capturing high-order relationships within the modality compared to static hypergraphs.

**Ablation 2** (*– Intra-modal high-order relationship*): This variant replaces the intra-modal high-order relationships with a traditional graph-based approach, removes the hypergraph attention mechanism, while retaining the inter-modal dynamic interactions constructed using dynamic hypergraphs. This experiment investigates the advantages of high-order relationships in capturing intra-modal passenger flow patterns compared to conventional pairwise relationships.

**Ablation 3** (*– Inter-modal dynamic interaction*): In this variant, the inter-modal dynamic interactions are removed, and only the spatiotemporal convolution results from a single-mode are used for prediction. This setup evaluates the influence of direct passenger flow exchanges between traffic modes on the overall predictive performance.

**Ablation 4** (*– Both*): Serving as the baseline for the ablation study, this variant eliminates both the hypergraph self-attention mechanism, intra-modal high-order relationships, and inter-modal dynamic interactions. It relies solely on a traditional graph-



**Fig. 14.** The medium- and long-term passenger flow prediction of multimodal transportation systems.

based approach to construct the multimodal transportation systems, providing a reference point to assess the contribution of the removed components.

These ablation experiments systematically evaluate each module's role in enhancing the STHCN framework's predictive accuracy. The results, summarized in Table 8, provide critical insights into the importance of incorporating intra-modal high-order relationships and inter-modal dynamic interactions, highlighting their essential contributions to the overall performance of the framework.

Table 8 demonstrates that all modules proposed in this study significantly enhance the accuracy of passenger flow prediction. Given the complexity of multimodal transportation systems, hypergraphs are more effective than traditional graphs in capturing local and global network characteristics. By modeling the transformation relationships between nodes and hyperedges, hypergraphs accurately reflect passenger flow dynamics at individual stations and across the entire route, offering superior predictive capabilities. Ablation 1

**Table 8**

Ablation analysis of STHCN components in multimodal transportation systems.

Experiments	Metro system			Bus system		
	MAE	RMSE	PCC	MAE	RMSE	PCC
Ablation 1	33.88	58.19	0.9708	15.73	24.40	0.9461
Ablation 2	37.38	63.61	0.9451	16.17	24.83	0.9448
Ablation 3	36.24	62.46	0.9546	16.53	26.01	0.9364
Ablation 4	38.55	70.54	0.9277	17.71	29.43	0.9305
STHCN	<b>31.84</b>	<b>56.08</b>	<b>0.9727</b>	<b>15.43</b>	<b>23.93</b>	<b>0.9490</b>

indicates that removing the hypergraph self-attention mechanism led to a 6.02 % and 1.91 % increase in MAE, a 3.62 % and 1.93 % increase in RMSE, and a 1.95 % and 0.31 % decrease in PCC for the metro and bus systems, respectively. Despite these changes, Ablation 1 outperformed the best results of the baseline model shown in Table 4, demonstrating the significant benefits of learning dynamic interactions. Furthermore, replacing hypergraphs with traditional graphs in Ablation 2 increased MAE and RMSE, while PCC experienced a significant decrease. A similar trend is observed in Ablation 3, which removed dynamic interactions. These findings underscore the indispensability of inter-modal interactions and the self-attention mechanism in modeling multimodal transportation systems.

When all modules are removed, as in Ablation 4, the prediction error increases significantly, leading to the lowest PCC values and the highest MAE and RMSE across all experiments. This result underscores the complementary nature of the two modules in improving the STHCN framework's performance. The dynamic hypergraph captures high-order relationships within traffic modes and effectively integrates inter-modal interactions, resulting in a more comprehensive representation of multimodal transportation systems. These insights reinforce the importance of each module in enhancing the overall accuracy and robustness of the STHCN, particularly in handling the complexity and variability of multimodal passenger flow.

### 6.3. Dynamic interaction analysis

The inter-modal dynamic interactions derived through the HGK-Means clustering algorithm effectively account for transfer demand to the greatest extent. Unlike traditional clustering approaches, the proposed algorithm allows a single bus station to be included in multiple hypergraphs simultaneously. This flexibility, typically observed in less busy and remote areas where metro stations share the same bus station with low transfer demand, provides a more adaptable and realistic modeling framework.

To assess the impact of inter-modal dynamic interactions, a hypergraph model is constructed using the K-Means algorithm with a

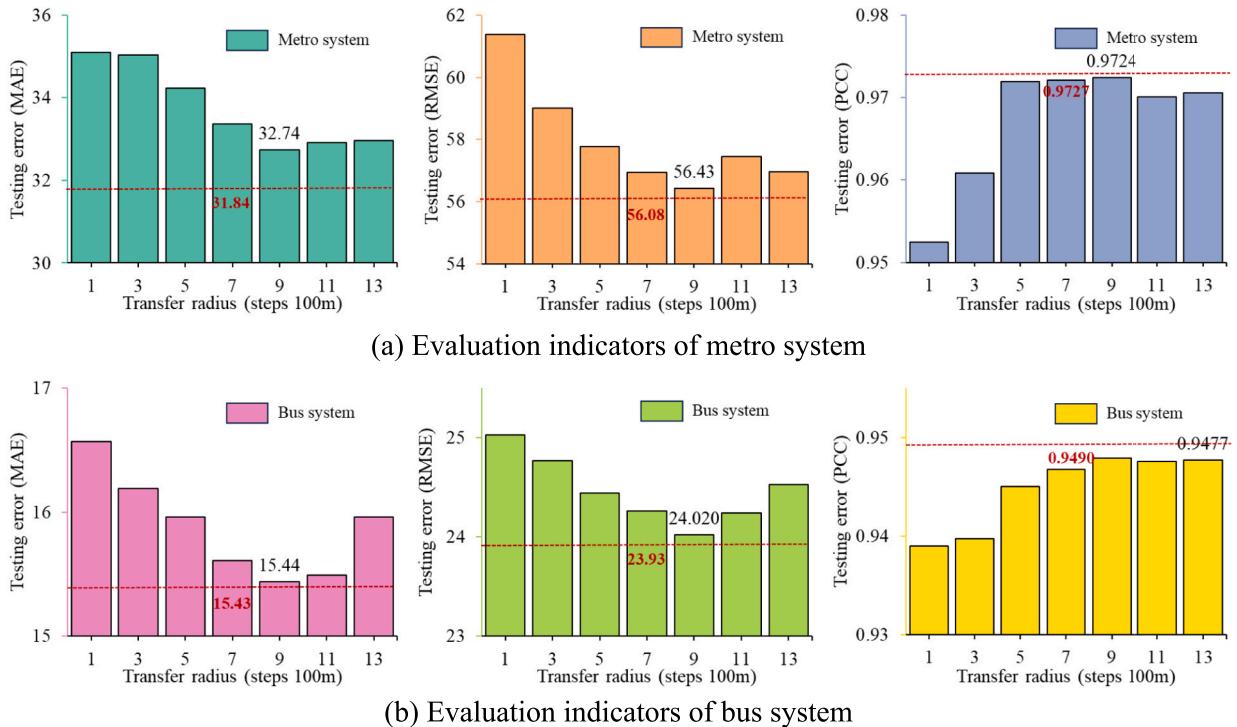


Fig. 15. Dynamic interaction analysis of different transfer radius.

fixed transfer radius, and its application to metro and bus passenger flow prediction is illustrated in Fig. 15. The experimental results show that for the tested dataset, the STHCN achieves optimal performance with a fixed transfer radius of approximately 900 m. However, the proposed method eliminates the need to predefine an optimal transfer radius, enabling more precise and efficient modeling of the dynamic interaction process.

Compared to the STHCN model relying on the optimal transfer radius, the proposed method reduces MAE by 2.75 % and 0.06 % and RMSE by 0.62 % and 0.37 %, while improving PCC by 0.03 % and 0.12 %, respectively. These results underscore the effectiveness of the proposed method in enhancing both the accuracy and efficiency of the model, demonstrating its superiority in capturing complex inter-modal interactions within multimodal transportation systems.

Fig. 16 presents the visual analysis of passenger transfers at metro stations and total passenger transfers at bus stations, as depicted in the inter-modal hypergraph constructed using different algorithms. To enhance the clarity of the analysis, two representative metro stations are selected for the visual analysis: Metro Station 9 (a non-transfer station for Metro Line 9) and Metro Station 3 (a transfer station for Metro Lines 4 and 9). The selection of these stations effectively highlights the dynamic interaction characteristics between different traffic modes.

In Fig. 16(a), the HGK-Means algorithm demonstrates a strong correlation when analyzing passenger transfers between metro and bus systems. Comparing the experimental results with different fixed transfer radius reveals that the HGK-Means algorithm captures the dynamic interactions more effectively. The 100 m, 300 m, and 500 m configurations exhibit poor performance among the various fixed transfer radius settings. In comparison, the 900 m fixed radius method shows the best results, aligning with the evaluation analysis in Fig. 15. In Fig. 16(b), the passenger transfer at the transfer station (Metro Station 3) is influenced by significant fluctuations in passenger flow during the morning and evening peak hours, leading to a slight decline in the accuracy of the constructed inter-modal hypergraph. Despite this, the HGK Means algorithm continues to outperform other traditional methods. This outcome further underscores the superiority of the HGK-Means algorithm in capturing the complex interactions within multimodal transportation systems.

#### 6.4. Hyperparameter analysis

First, discuss the model hyperparameters. In the experiment, the control variable method is employed to optimize the hyperparameters of the STHCN framework. This approach ensures that all other hyperparameter settings remain fixed while optimizing a single hyperparameter. Three hyperparameters require tuning: the convolution kernel width, the spatial hypergraph convolution kernel, and the weight coefficients of the trade-off function, as illustrated in Fig. 17.

In Fig. 17, the convolution kernel width is set to 3. Increasing the kernel width further would lead to the blurring of temporal features, while a smaller kernel width would fail to capture periodic patterns effectively. The spatial hypergraph convolution kernel is empirically validated and set to 3, considering both computational efficiency and model performance. Through ablation studies and grid search, this configuration demonstrated optimal performance in capturing multi-scale spatial dependencies while mitigating the risk of overfitting. The weight coefficient for the trade-off function is set to 0.7. The maximum value in the HGK-Means algorithm has a more significant impact on the results than the average value, suggesting that an increased emphasis on the maximum value is beneficial.

Then, discuss the sandwich structure. The sandwich structure of spatiotemporal aggregation is effective in capturing both temporal and spatial features of passenger demand. By strategically configuring the combination of structural components and the number of

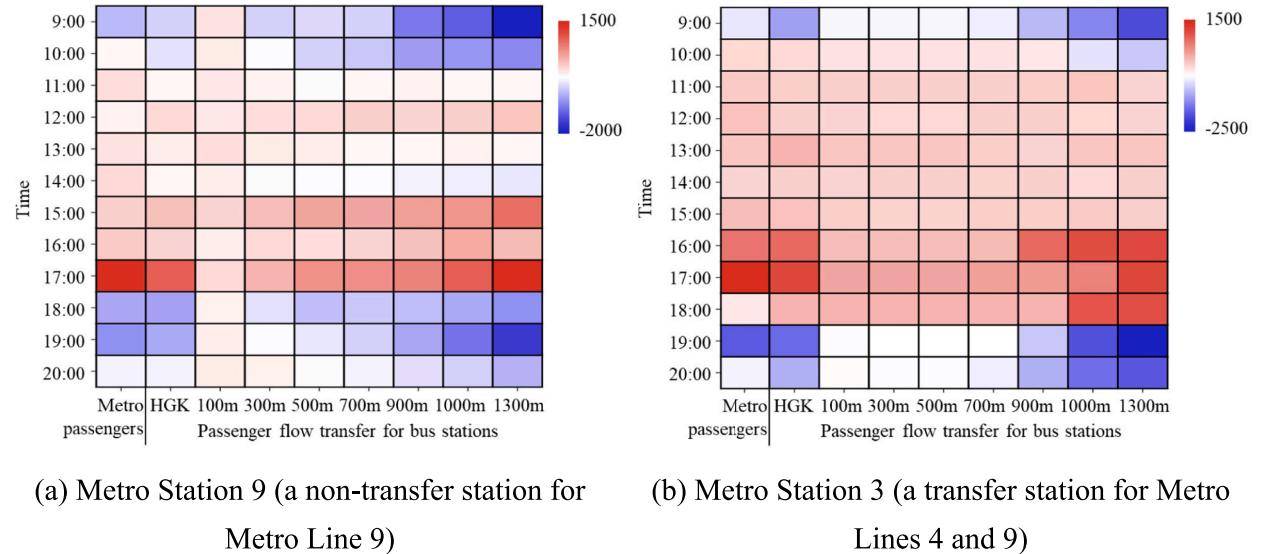


Fig. 16. Visualization of inter-modal dynamic interactions.

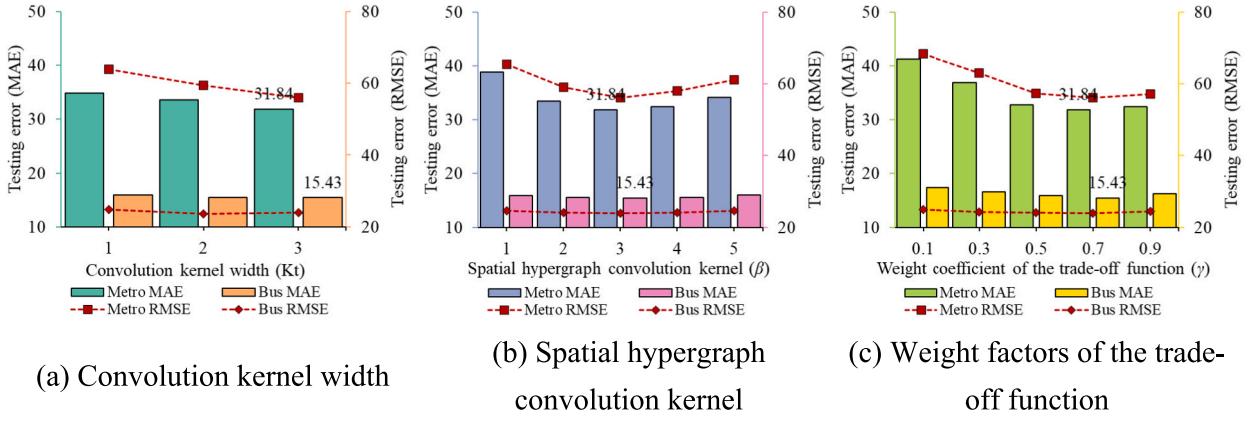


Fig. 17. Hyperparameter tuning of the STHCN framework.

stacking layers, the framework can comprehensively capture the flow of spatiotemporal information. However, the improvement in joint passenger flow prediction is highly dependent on the optimal configuration. A series of comparative experiments highlights that different configurations have varying impacts on the predictive performance, underscoring the importance of carefully choosing the structure, as illustrated in Fig. 18. The horizontal axis represents the number of spatiotemporal aggregation stacks. In contrast, the vertical axis indicates the number of hypergraph convolutional layers.

In Fig. 18, the STHCN achieves its best overall performance when configured with two spatiotemporal aggregation stacks and two hypergraph convolutional layers, denoted as the  $2 \times 2$  configuration. In contrast, configurations using only one spatiotemporal stack or one convolutional layer exhibit insufficient feature representation capacity. For instance, the  $1 \times 1$  configuration in the metro system results in an MAE of 32.7 and RMSE of 63.96, which is 2.62 % and 12.32 % worse than the  $2 \times 2$  configuration, respectively. Moreover, simply increasing the number of layers—either by adding more spatiotemporal aggregation stacks or hypergraph convolutional layers—does not consistently improve performance, and in many cases leads to degradation. For example, in the metro system, the  $3 \times 2$  configuration slightly improves PCC to 0.9730, marginally higher than the optimal setting (PCC = 0.9727). However, this negligible gain comes at the cost of a significantly higher MAE of 33.95 and RMSE of 57.16, indicating possible overfitting or feature over-smoothing, and excessive feature fusion causes loss of informative variance.

Further quantitative comparison underscores the sensitivity differences between the two systems. In the metro system, the performance gap between the best and worst configurations is substantial: MAE rises from 31.84 to 42.03 (+24.24 %), RMSE increases from 56.08 to 67.20 (+16.55 %), and PCC drops from 0.9727 to 0.9571 (-1.60 %). In contrast, the bus system shows lower sensitivity to configuration changes: MAE rises from 15.43 to 19.11 (+19.25 %), RMSE increases from 23.93 to 27.58 (+13.23 %), PCC drops more notably from 0.949 to 0.8365 (-11.85 %). These discrepancies suggest that the metro system exhibits stronger spatiotemporal dependencies, making it more responsive to precise architectural tuning. The bus system, while still benefiting from structured feature extraction, demonstrates relatively weaker spatiotemporal coupling due to higher temporal volatility and more stochastic demand patterns. Over-complicating the model in this setting, particularly by stacking excessive layers, may distort naturally fluctuating passenger flow characteristics rather than enhancing prediction.

The empirical evidence indicates that a moderately complex architecture, specifically a  $2 \times 2$  configuration, is optimal for capturing dynamic interactions in multimodal transportation systems. Both underfitting and overfitting configurations compromise performance, highlighting the necessity of carefully balancing depth and aggregation in spatiotemporal hypergraph models to ensure robustness and generalization in passenger flow prediction.

### 6.5. Efficiency analysis

To ensure consistency in batch size and computing platform, the STHCN is compared with several graph-based baseline models, focusing on analyzing the number of parameters and running times, as presented in Table 9. Specifically, the runtime reported here exclusively accounts for the main body of the model, excluding data processing.

In Table 9, T-GCN and DCRNN leverage GRU to capture time-series dynamics at each time step, which significantly increases the number of model parameters and prolongs the running time. Conv-GCN, which combines GCN and 3D CNN, benefits from good parallel computing capabilities, so it has a relatively smaller parameter count and more efficient running time. Multi-STGCnet, using three LSTMs and three GCNs, significantly increases both model parameters and running time. While the STHCN incorporates the construction of inter-modal dynamic interactions, which increases the number of parameters, the feature extraction module based on hypergraph convolution does not involve complex calculations. As a result, the overall computational efficiency of the STHCN is moderate compared to other models.

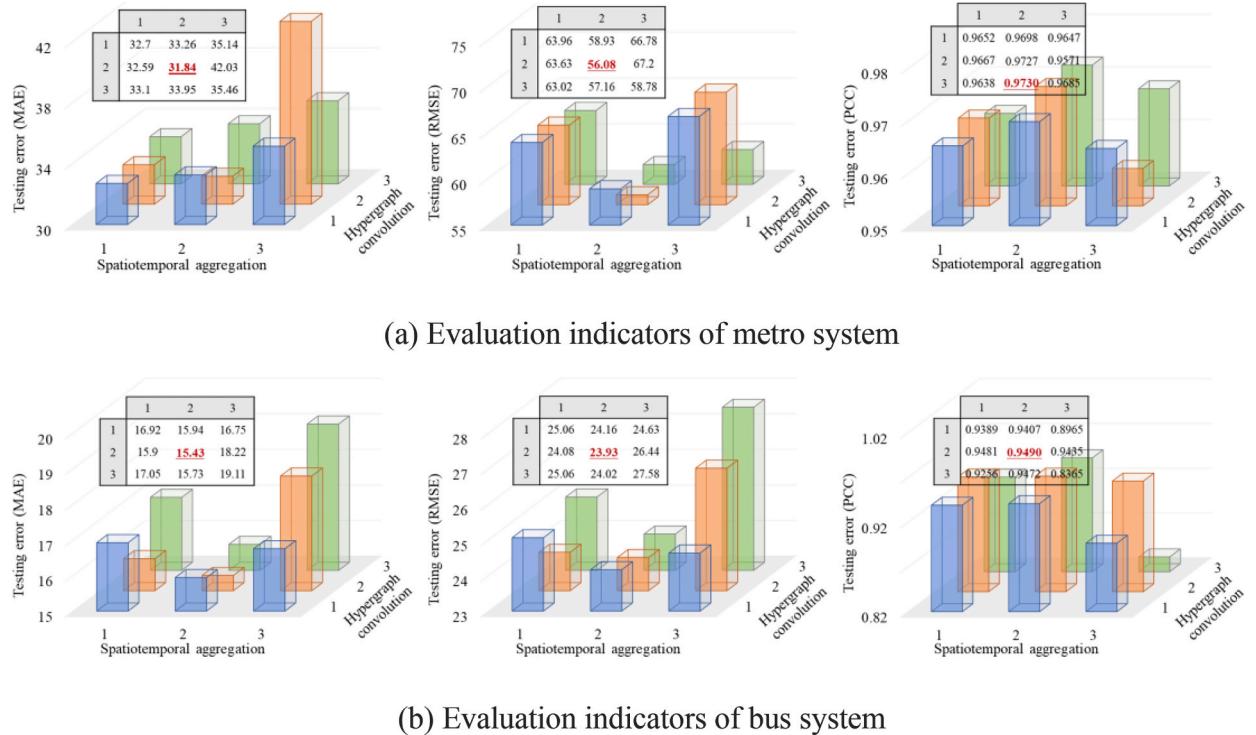


Fig. 18. Spatiotemporal aggregation analysis of different combination methods.

**Table 9**  
Efficiency comparison of different models.

Models	Metro system					Bus system				
	Parameters ( $\times 10^4$ )	Running times (s)				Parameters ( $\times 10^4$ )	Running times (s)			
		15 min	30 min	60 min	120min		15 min	30 min	60 min	120min
T-GCN	36.71	8.37	7.84	7.03	6.77	50.83	21.89	20.37	19.11	18.61
STGCN	19.58	4.21	3.94	3.73	3.66	26.94	13.09	12.58	12.04	11.12
DCRNN	39.55	8.68	8.2	7.7	7.59	52.46	28.04	26.66	25.53	24.35
Conv-GCN	22.56	5.27	4.86	4.51	4.32	31.85	18.36	16.24	15.03	13.87
Multi-STGCnet	48.35	10.19	9.31	8.54	7.82	64.7	31.68	28.93	27.09	25.88
Ada-MSHyper	56.03	7.86	7.02	6.33	5.41	78.45	23.31	20.36	18.09	15.72
STHCN	<b>28.44</b>	<b>5.21</b>	<b>4.78</b>	<b>4.39</b>	<b>4.21</b>	<b>39.68</b>	<b>14.87</b>	<b>13.94</b>	<b>13.01</b>	<b>12.39</b>

## 7. Conclusions

With the continuous growth in urban transportation demand, accurate passenger flow prediction has become increasingly critical, particularly within multimodal transportation systems. However, achieving joint prediction across multiple traffic modes poses significant challenges due to the inherent differences in passenger flow patterns, the complexity of dynamic inter-modal interactions, and the need to model spatiotemporal dependencies effectively. This study explores the intrinsic relationships between metro and bus passenger flow transfer in multimodal transportation systems for the first time. Firstly, the HGK-Means algorithm is developed to model multimodal transportation systems by encoding intra- and inter-modal high-order relationships, effectively capturing correlations between stations with similar flow patterns. Subsequently, an STHCN model that integrates temporal and spatial networks is built, incorporating hypergraph self-attention mechanism to capture complex spatiotemporal dependencies, enabling accurate joint passenger flow prediction. Finally, a comprehensive dataset, encompassing metro and bus records, is utilized to validate the model's performance. Extensive experiments demonstrate the STHCN's superiority in accurately predicting multimodal passenger flow, significantly outperforming baseline methods.

Some key findings are as follows:

(1) The HGK-Means algorithm effectively models the dynamic interactions between metro and bus systems, capturing passenger flow transfer patterns across modes. It identifies stations with high potential correlation, enabling a deeper understanding of

multimodal transportation dynamics and improving the precision of passenger flow prediction. Compared to the optimal transfer radius method, the proposed approach reduces MAE by 2.75 % and 0.06 %, lowers RMSE by 0.62 % and 0.37 %, and increases PCC by 0.03 % and 0.12 %, respectively.

(2) The STHCN surpasses traditional graph structures by enabling better capture of local and global characteristics within traffic modes while integrating correlations across modes. By transforming inter-modal passenger flow interactions into spatiotemporal dependencies, this framework significantly enhances prediction accuracy for both systems.

(3) Experimental results on the collected multimodal dataset demonstrate the STHCN's superior performance over baseline models. For 15-minute short-term predictions, the STHCN reduces MAE by 6.79 % and 3.38 %, lowers RMSE by 6.00 % and 3.43 %, and increases PCC by 3.55 % and 0.64 %. Overall, the STHCN demonstrates strong competitiveness in multimodal passenger flow prediction.

The findings of this study offer valuable technical insights for the operation and management of urban transportation systems. Future research could integrate additional data sources, such as weather conditions and special events, to enhance the robustness and adaptability of passenger flow prediction models. This study demonstrates the successful application of hypergraph theory to model station interactions within multimodal transportation systems. Moving forward, further exploration of hyperedge interaction behaviors across different traffic modes will be key to advancing prediction performance and system optimization.

#### CRediT authorship contribution statement

**Dongyu Luo:** Writing – original draft, Methodology, Conceptualization. **Jiangfeng Wang:** Methodology. **Wenqi Lu:** Writing – original draft, Formal analysis. **Weidong Ding:** Project administration, Formal analysis. **Xuedong Yan:** Supervision, Funding acquisition, Conceptualization. **Hai Yang:** Supervision, Funding acquisition, Conceptualization.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

This research is supported by the National Key R&D Program of China (2022YFB4300400), and the research grant from the Hong Kong Research Grants Council (Project HKUST16205123).

#### Data availability

Data will be made available on request.

#### References

- Arya, D., Worring, M., 2018. Exploiting relational information in social networks using geometric deep learning on hypergraphs. International Conference on Multimedia Retrieval. 2018, 117–125.
- Baghbani, A., Ranhami, S., Bouguila, N., Patterson, Z., 2025. TMS-GNN: Traffic-aware multistep graph neural network for bus passenger flow prediction. *Transp. Res. Part C Emerging Technol.* 174, 105107.
- Bai, S., Zhang, F., Torr, P., H., S., 2021. Hypergraph convolution and hypergraph attention. *Pattern Recognition.* 110, 107637.
- Bao, J., Kang, J., Yang, Z., Chen, X., 2022. Forecasting network-wide multi-step metro ridership with an attention-weighted multi-view graph to sequence learning approach. *Expert Syst. Appl.* 210, 118475.
- Bratsas, C., Koupidis, K., Salanova, J.M., Giannakopoulos, K., Kaloudis, A., Aifadopoulou, G., 2020. A comparison of machine learning methods for the prediction of traffic speed in urban places. *Sustainability* 12, 142.
- Du, B., Peng, H., Wang, S., Bhuiyan, M.Z.A., Wang, L., Gong, Q., Liu, L., Li, J., 2020. Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction. *IEEE Trans. Intell. Transport. Syst.* 21 (3), 972–985.
- Fang, H., Chen, C., Hwang, F., Chang, C., Chang, C., 2024. Metro station functional clustering and dual-view recurrent graph convolutional network for metro passenger flow prediction. *Expert Syst. Appl.* 247, 122550.
- Feng, Y.F., You, H.X., Zhang, Z.Z., Ji, R.R., Gao, Y., 2019. Hypergraph neural networks. *Proc. Thirty-Third AAAI Conference.* 2019, 3558–3565.
- Filipovska, M., Mahmassani, H.S., 2020. Traffic flow breakdown prediction using machine learning approaches. *Transp. Res. Rec.* 2674 (10), 560–570.
- Hao, S., Lee, D., Zhao, D., 2019. Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system. *Transp. Res. Part C Emerging Technol.* 107, 287–300.
- He, Y., Li, L., Zhu, X., Tsui, K.L., 2022. Multi-graph convolutional-recurrent neural network (MGC-RNN) for short-term forecasting of transit passenger flow. *IEEE Trans. Intell. Transp. Syst.* 23 (10), 18155–18174.
- Hu, S., Chen, J., Zhang, W., Liu, G., Chang, X., 2024. Graph transformer embedded deep learning for short-term passenger flow prediction in urban rail transit systems: a multi-gate mixture-of-experts model. *Inf. Sci.* 679, 121095.
- Huang, H., Mao, J., Kang, L., Lu, W., Zhang, S., Liu, L., 2024. A dynamic graph deep learning model with multivariate empirical mode decomposition for network-wide metro passenger flow prediction. *Comput. Aided Civ. Inf. Eng.* 39 (17), 2596–2618.
- Jia, F., Li, H., Jiang, X., Xu, X., 2019. Deep learning-based hybrid model for short-term subway passenger flow prediction using automatic fare collection data. *IET Intel. Transport Syst.* 13 (11), 1708–1716.
- Ke, J., Feng, S., Zhu, Z., Yang, H., Ye, J., 2021. Joint predictions of multimodal ride-hailing demands: a deep multi-task multi-graph learning-based approach. *Transp. Res. Part C Emerging Technol.* 127, 103063.
- Leng, B., Zeng, J., Xiong, Z., Lv, W., Wan, Y., 2013. Probability tree based passenger flow prediction and its application to the Beijing subway system. *Front. Comp. Sci.* 7, 195–203.

- Li, C., Bai, L., Liu, W., Yao, L., Waller, S.T., 2021. A multi-task memory network with knowledge adaptation for multi-modal demand forecasting. *Transp. Res. Part C Emerging Technol.* 131, 103352.
- Li, P., Wang, S., Zhao, H., Yu, J., Hu, L., Yin, H., Liu, Z., 2023a. IG-Net: an interaction graph network model for metro passenger flow forecasting. *IEEE Trans. Intell. Transportat. Syst.* 24 (4), 4147–4157.
- Li, S., Liang, X., Zheng, M., Chen, J., Chen, T., Guo, X., 2023b. How spatial features affect urban rail transit prediction accuracy: a deep learning based passenger flow prediction method. *J. Intell. Transp. Syst.* 2023, 2279633.
- Li, Y., Yu, R., Shahabi, C., Liu, Y., 2018. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. 6th International Conference on Learning Representations. 2018, 1–16.
- Liang, Y., Huang, G., Zhao, Z., 2022. Joint demand prediction for multimodal systems: a multi-task multi-relational spatiotemporal graph neural network approach. *Transp. Res. Part C Emerging Technol.* 140, 103731.
- Ling, X., Huang, Z., Wang, C., Zhang, F., Wang, P., 2018. Predicting subway passenger flows under different traffic conditions. *PLoS One* 13 (8), e0202707.
- Liu, L., Chen, R., 2017. A novel passenger flow prediction model using deep learning methods. *Transp. Res. Part C Emerging Technol.* 84, 74–91.
- Liu, Y., Lyu, C., Liu, X., Liu, Z., 2021. Automatic feature engineering for bus passenger flow prediction based on modular convolutional neural network. *IEEE Trans. Intell. Transp. Syst.* 22 (4), 2349–2358.
- Lu, Y., Ding, H., Ji, S., Sze, N.N., He, Z., 2021. Dual attentive graph neural network for metro passenger flow prediction. *Neural Comput. Appl.* 33 (20), 13417–13431.
- Luo, D., Zhao, D., Ke, Q., You, X., Liu, L., Ma, H., Zou, X., 2022. Spatiotemporal hashing multigraph convolutional network for service-level passenger flow forecasting in bus transit systems. *IEEE Internet Things J.* 9 (9), 6803–6815.
- Ma, X., Zhang, J., Du, B., Ding, C., Sun, L., 2019. Parallel architecture of convolutional bi-directional LSTM neural networks for network-wide metro ridership prediction. *IEEE Trans. Intell. Transp. Syst.* 20 (6), 2278–2288.
- Mei, Z., Yu, W., Tang, W., Yu, J., Cai, Z., 2023. Attention mechanism-based model for short-term bus traffic passenger volume prediction. *IET Intel. Transport Syst.* 17 (4), 767–779.
- Mulerikkal, J., Thandassery, S., Rejathalal, V., Kunnamkody, D.M.D., 2022. Performance improvement for metro passenger flow forecast using spatio-temporal deep neural network. *Neural Comput. Appl.* 34, 983–994.
- Peng, H., Wang, H., Du, B., Bhuiyan, M.Z.A., Ma, H., Liu, J., Wang, L., Yang, Z., Du, L., Wang, S., Yu, P.S., 2020. Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting. *Inf. Sci.* 521, 277–290.
- Rezaee, M.J., Eshkevari, M., Saberi, M., Hussain, O., 2021. GBK-means clustering algorithm: an improvement to the K-means algorithm based on the bargaining game. *Knowl.-Based Syst.* 213, 106672.
- Sha, S., Li, J., Zhang, K., Yang, Z., Wei, Z., Li, X., Zhu, X., 2020. RNN-based subway passenger flow rolling prediction. *IEEE Access* 8, 15232–15240.
- Shang, Z., Chen, L., Wu, B., Cui, D., 2024. Ada-MSHyper: Adaptive multi-scale hypergraph Transformer for time series forecasting. 38th Conference on Neural Information Processing Systems (NeurIPS 2024), 2024, 2410.23992.
- Sun, S.N., Her, J., Lee, S., Lee, S.L., 2017. Meso-scale urban form elements for bus transit-oriented development: evidence from Seoul, republic of Korea. *Sustainability* 9, 1516.
- Sun, F., Wang, X., Zhang, Y., Liu, W., Zhang, R., 2020. Analysis of bus trip characteristic analysis and demand forecasting based on GA-NARX neural network model. *IEEE Access* 8, 8812–8820.
- Sun, Y., Leng, B., Guan, W., 2015. A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing* 166, 109–121.
- Sun, Y., Jiang, X., Hu, Y., Duan, F., Guo, K., Wang, B., Gao, J., Yin, B., 2022. Dual dynamic spatial-temporal graph convolution network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* 23 (12), 23680–23693.
- Tan, Y., Liu, H., Pu, Y., Wu, X., Jiao, Y., 2021. Passenger flow prediction of integrated passenger terminal based on K-Means-GRNN. *J. Adv. Transp.* 2021, 1055910.
- Toman, P., Zhang, J., Ravishanker, N., Konduri, K.C., 2020. Dynamic predictive models for ride sourcing services in New York city using daily compositional data. *Transp. Res. Part C Emerging Technol.* 121, 102833.
- Wang, J., Zhang, Y., Wei, Y., Piao, X., Yin, B., 2021. Metro passenger flow prediction via dynamic hypergraph convolution networks. *IEEE Intell. Transp. Syst. Mag.* 22 (12), 7891–7963.
- Wang, Z., Pel, A.J., Verma, T., Krishnakumari, P., Brakel, P.V., Oort, N.V., 2022. Effectiveness of trip planner data in predicting short-term bus ridership. *Transp. Res. Part C Emerging Technol.* 142, 103790.
- Wei, Y., Chen, M.C., 2012. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transp. Res. Part C Emerging Technol.* 21, 148–162.
- Wu, J., He, D., Jin, Z., Li, X., Li, Q., Xiang, W., 2024. Learning spatial-temporal pairwise and high-order relationships for short-term passenger flow prediction in urban rail transit. *Expert Syst. Appl.* 245, 123091.
- Xiu, C., Sun, Y., Peng, Q., 2022. Modeling traffic as multi-graph signals: using domain knowledge to enhance the network-level passenger flow prediction in metro systems. *J. Rail Transp. Plann. Manage.* 24, 100342.
- Xu, Z., Lv, Z., Li, J., Sun, H., Sheng, Z., 2023. A novel perspective on travel demand prediction considering natural environmental and socioeconomic factors. *IEEE Intell. Transp. Syst. Mag.* 15 (1), 136–159.
- Yadati, N., Nimishakavi, M., Yadav, P., Nitin, V., Louis, A., Talukdar, P., 2019. HyperGCN: a new method of training graph convolutional networks on hypergraphs. *Adv. Neural Inf. Proces. Syst.* 32, 1509–1520.
- Yang, Y., Zhang, J., Yang, L., Yang, Y., Li, X., 2023. Short-term passenger flow prediction for multi-traffic modes: a Transformer and residual network based multi-task learning method. *Inf. Sci.* 642, 119144.
- Yang, Y., Zhang, J., Yang, L., Gao, Z., 2024. Network-wide short-term inflow prediction of the multi-traffic modes system: an adaptive multi-graph convolution and attention mechanism based multitask-learning model. *Transp. Res. Part C Emerging Technol.* 158, 104428.
- Ye, J., Sun L., Du B., Fu Y., Tong, X., Xiong, H., 2019. Co-prediction of multiple transportation demands based on deep spatio-temporal neural network. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19). 2019, 305–313.
- Ye, J., Zhao, J., Ye, K., Xu, C., 2020. Multi-STGCnet: a graph convolution based spatial-temporal framework for subway passenger flow forecasting. *IEEE Int. J Conf. Neural Network.* 2020, 1–8.
- Ye, J., Park, J., 2020. Hypergraph convolutional recurrent neural network. *Knowledge Discovery and Data Mining.* 2020, 3366–3376.
- Yin, D., Jiang, R., Deng, J., Li, Y., Xie, Y., Wang, Y., Zhou, Y., Song, X., Shang, J.S., 2023. MTMGNN: Multi-time multi-graph neural network for metro passenger flow prediction. *GeoInformatica* 27, 77–105.
- Yu, B., Yin, H., Zhu, Z., 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. *Int. Joint Conf. Artif. Intell.* 2018, 3634–3640.
- Yu, J., Chang, X., Hu, S., Yin, H., Wu, J., 2024. Combining travel behavior in metro passenger flow prediction: a smart explainable stacking-catboost algorithm. *Inf. Process. Manag.* 61, 103733.
- Zeng, J., Tang, J., 2023. Combining knowledge graph into metro passenger flow prediction: a split-attention relational graph convolutional network. *Expert Syst. Appl.* 213, 118790.
- Zhan, S., Cai, Y., Xiu, C., Zou, D., Wang, D., Wong, S.C., 2024. Parallel framework of a multi-graph convolutional network and gated recurrent unit for spatial-temporal metro passenger flow prediction. *Expert Syst. Appl.* 251, 123982.
- Zhang, J., Shen, D., Tu, L., Zhang, F., Xu, C., Wang, Y., Tian, C., Li, X., Huang, B., Li, Z., 2017. A real-time passenger flow estimation and prediction method for urban bus transit systems. *IEEE Trans. Intell. Transport. Syst.* 18 (11), 3168–3178.
- Zhang, D., Yan, J., Polat, K., Alhudhaif, A., Li, J., 2024. Multimodal joint prediction of traffic spatial-temporal data with graph sparse attention mechanism and bidirectional temporal convolutional network. *Adv. Eng. Inf.* 62, 102533.
- Zhang, J., Chen, F., Guo, Y., Li, X., 2020. Multi-graph convolutional network for short-term passenger flow forecasting in urban rail transit. *IET Intel. Transport Syst.* 14 (10), 1210–1217.

- Zhang, J., Chen, F., Shen, Q., 2019. Cluster-based LSTM network for short-term passenger flow forecasting in urban rail transit. *IEEE Access* 7, 147653–147671.
- Zhang, J., Chen, F., Zhu, Y., Guo, Y., 2021. Deep-learning architecture for short-term passenger flow forecasting in urban rail transit. *IEEE Trans. Intell. Transp. Syst.* 22 (11), 7004–7014.
- Zhang, N., Zhang, Y., Lu, H., 2011. Seasonal autoregressive integrated moving average and support vector machine models prediction of short-term traffic flow on freeways. *Transp. Res.* 2215, 85–92.
- Zhang, Z., Han, Y., Peng, T., Li, Z., Chen, G., 2022. A comprehensive spatio-temporal model for subway passenger flow prediction. *ISPRS Int. J. Geo Inf.* 11 (6), 341.
- Zhao, J., Zhang, R., Sun, Q., Shi, J., Zhuo, F., Li, Q., 2023. Adaptive graph convolutional network-based short-term passenger flow prediction for metro. *J. Intell. Transp. Syst.* 2023, 2209913.
- Zhao, Y., Ma, Z., Yang, Y., Jiang, W., 2020a. Short-term passenger flow prediction with decomposition in urban railway systems. *IEEE Access* 34, 4813–4830.
- Zhao, Z., Chen, W., Wu, X., Chen, P.C.Y., Liu, J., 2017. LSTM network: a deep learning approach for short-term traffic forecast. *IET Intel. Transport Syst.* 11 (2), 68–75.
- Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Liu, T., 2020b. T-GCN: a temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* 21 (9), 3848–3858.
- Zhong, C., Batty, M., Manley, E., Wang, J., Wang, Z., Chen, F., Schmitt, G., 2016. Variability in regularity: mining temporal mobility patterns in London, Singapore and Beijing using smart-card data. *PLoS One* 11 (2), e0149222.
- Zhong, R., Lv, W., Du, B., Lei, S., Huang, R., 2017. Spatiotemporal multi-task learning for citywide passenger flow prediction. *International Conference on Ubiquitous Intelligence and Computing*. 2017, 1–8.