vehicles

MDPI

Article

Short-Term Metro OD Passenger Flow Prediction Based on Origin–Destination Bidirectional Graph Convolution and Dynamic Gated Fusion

Chao Chen [1,2,*], Jiaming Tang [1], Hao Li [1], Zhengxing He [1] and Xuan Chen [1]

[1] School of Computer Science and Engineering, Sichuan University of Science and Engineering, Yibin 644000, China; 324085406217@stu.suse.edu.cn (J.T.); 323085406210@stu.suse.edu.cn (H.L.); 324085406206@stu.suse.edu.cn (Z.H.); 324085406201@stu.suse.edu.cn (X.C.)
[2] Intelligent Perception and Control Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Yibin 644000, China
* Correspondence: chenchao@suse.edu.cn

Abstract

Accurate short-term origin–destination (OD) passenger flow prediction provides important demand information for metro operation. An OD flow is associated with both origin-side and destination-side spatial dependencies, but their relative predictive value may differ across OD pairs and operating periods. Separately extracting the two representations does not establish whether their fusion should vary with the current state. To investigate this problem, an origin–destination bidirectional graph convolution and dynamic gated fusion model, termed OD-BGCN-GRU, is proposed. Origin-GCN and Destination-GCN propagate information along the two dimensions of the OD matrix, respectively. A state-dependent gate adaptively combines the two spatial representations, while a gated recurrent unit captures temporal evolution. Experiments on Hangzhou Metro automatic fare collection data show that OD-BGCN-GRU achieves an MAE of 0.704 ± 0.007 and an RMSE of 2.253 ± 0.021 on active OD pairs, reducing MAE by 3.03% compared with ODMixer. Moreover, dynamic fusion further reduces MAE by 2.36% relative to OD-specific static fusion, supporting the predictive benefit of state-dependent coordination between origin-side and destination-side spatial information. These findings provide a useful modeling basis for fine-grained short-term metro OD demand forecasting and subsequent demand-oriented operational analysis.

Keywords: metro; origin–destination passenger flow; short-term prediction; graph convolutional network; dynamic gated fusion; gated recurrent unit; vehicles

Check for updates

Received: 12 August 2026
Revised: 14 September 2026
Accepted: 29 September 2026
Published: 1 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

With the continuous expansion of urban rail transit networks, metro passenger flows have become increasingly complex in both time and space. Compared with station entry and exit flows, origin–destination (OD) passenger flow provides a more direct description of travel demand between different origin and destination stations. Short-term OD prediction uses historical automatic fare collection (AFC) records to estimate the distribution of passenger flows among stations in future intervals and can provide demand information for capacity allocation and passenger organization under networked metro operation conditions. These applications depend on how forecasts are subsequently incorporated into operational decisions; prediction accuracy alone does not establish improvements in vehicle deployment or passenger control.

Travel demand also varies with the hour of the day, the day of the week, and the distinction between working days and holidays. In road transport, Macioszek and Kurek [1] documented differences between public-holiday and workday traffic distributions. This evidence motivates attention to calendar-dependent mobility patterns, while the magnitude and timing of those patterns must be examined separately for metro travel. Road traffic conditions may also affect the relative attractiveness of travel modes, but their relationship with metro OD demand is not estimated in the present study. Within a metro network, concentrated departures and arrivals can create distinct spatial relationships at the two ends of an OD trip.

Early passenger flow forecasting studies mainly relied on statistical and machine learning methods to describe historical trends and periodic changes. High-dimensional OD matrices present an additional difficulty because temporal variation must be learned together with relationships among many origin–destination combinations. Recurrent neural networks (RNNs), including long short-term memory (LSTM) networks [2] and gated recurrent units (GRUs) [3], provide trainable temporal representations. Convolutional LSTM (ConvLSTM) combines recurrent learning with convolutional operations [4]. For travel demand and OD prediction, Chu et al. [5] developed a deep multi-scale convolutional LSTM network to jointly learn demand variation at different spatial scales. These methods established the value of combining spatial feature extraction with temporal sequence learning, although a regular convolution depends on the arrangement of the OD matrix and does not directly encode metro topology.

Graph neural networks provide a representation of non-Euclidean spatial relationships. Graph convolutional networks (GCNs) aggregate information through graph connectivity [6], whereas graph attention networks learn neighborhood aggregation weights from node features [7]. In traffic forecasting, the spatio-temporal graph convolutional network (STGCN) [8] combines graph and temporal convolutions, and the diffusion convolutional recurrent neural network (DCRNN) [9] integrates graph diffusion with recurrent temporal modeling. Graph WaveNet [10] introduces adaptive spatial dependencies and dilated temporal convolutions. Attention-based spatial–temporal GCN (ASTGCN) [11] and spatial–temporal synchronous GCN (STSGCN) [12] further investigate selective and joint spatio-temporal aggregation. Adaptive graph convolutional recurrent networks (AGCRN) [13] and multivariate time-series graph neural networks (MTGNN) [14] also show that the choice of graph structure and parameter sharing affects forecasting performance.

Attention-based sequence modeling provides another approach to spatial and temporal interactions [15]. The graph multi-attention network (GMAN) [16] applies spatial and temporal attention to traffic prediction, while the spatio-temporal adaptive embedding Transformer (STAEformer) [17] emphasizes the role of suitable representations in Transformer-based forecasting. The spatial–temporal identity (STID) baseline [18] demonstrates the importance of comparing complex architectures with carefully designed simple models. These studies motivate a broader comparison than recurrent baselines alone. However, station-level or sensor-level architectures require an explicit adaptation before they can predict an entire OD matrix; their original performance cannot be transferred directly to the metro OD task.

OD-specific studies have addressed relationships among origin–destination demands, incomplete observations, and demand heterogeneity. Ke et al. [19] proposed a spatio-temporal encoder–decoder residual multi-graph convolutional network for ride-sourcing OD demand, combining multiple spatial relationships with temporal learning. For urban rail transit, Zhang et al. [20] developed a channel-wise attentive split-convolutional neural network (CAS-CNN). Liu et al. [21] addressed online metro OD prediction through heterogeneous information aggregation, considering the limited availability of recent complete

OD matrices. Ye et al. [22] combined OD completion and representation augmentation to forecast metro OD matrices under limited observable data. These contributions show that the information available at the prediction time is an essential part of the task definition, rather than simply a preprocessing detail.

Recent work has further examined targeted demand, multiple graph relationships, and two-dimensional OD representations. Lv et al. [23] introduced passenger source-based attention for demand associated with oversaturated sections. Xing et al. [24] combined multiple relationship graphs with spatiotemporal fusion for urban rail OD prediction. Physics-guided OD modeling has also considered sparsity, high dimensionality, and disrupted demand [25]. ODformer [26] models spatial dependencies associated with common origins and destinations within a Transformer-based OD forecasting framework. ODMixer [27] uses a multi-view mixer to capture origin-side and destination-side relations together with short-term and longer-term temporal information. Therefore, separately representing the two OD dimensions is an established direction and is not, by itself, the novelty claimed in this study.

Nevertheless, obtaining origin-side and destination-side spatial representations does not directly determine how their relative contributions should be coordinated for a specific OD pair and time interval. When the destination is fixed, flows from adjacent origins may vary jointly, whereas when the origin is fixed, flows toward adjacent destinations may exhibit coordinated changes. The predictive value of these two spatial relationships may therefore differ across OD pairs and vary over time.

To address this issue, this study develops OD-BGCN-GRU, which integrates origin–destination bidirectional graph convolution, dynamic gated fusion, and GRU-based temporal modeling. Origin-GCN and Destination-GCN propagate spatial information along the two dimensions of the OD matrix using the metro topology. A shared gating function then generates channel-wise fusion coefficients from the current spatial representations, and the fused features are passed to a GRU for temporal modeling. The main contribution of this study is the explicit modeling and controlled evaluation of state-dependent coordination between the two spatial views. Its effectiveness is further examined through comparisons with recent OD forecasting methods and appropriately controlled fusion alternatives.

This paper is structured as follows: Section 2 presents the prediction problem and the proposed model; Section 3 describes the study area, data preparation, and experimental design; Section 4 presents the evaluation framework and results; and Section 5 discusses the main findings, summarizes the conclusions, and identifies the limitations and future research directions.

2. Methodology

2.1. Problem Definition

Let the metro network be represented as

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A}), \qquad N = |\mathcal{V}|, \tag{1}$$

where $\mathcal{V}$ denotes the station set, $\mathcal{E}$ denotes the physical connections between stations, and $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix. An element $A_{ij} = 1$ if stations i and j are directly connected and $A_{ij} = 0$ otherwise.

At time interval t, the OD passenger flow is represented as

$$\mathbf{X}_t = [x_{ij}^t]_{N \times N}, \qquad x_{ij}^t \geq 0, \tag{2}$$

where x_{ij}^t denotes the passenger flow from origin station i to destination station j during interval t. Row i describes the destination distribution associated with origin i, whereas

column j describes the origin distribution associated with destination j. Same-station trips are excluded from the prediction target, and therefore

$$x_{ii}^{t} = 0. \tag{3}$$

Given P consecutive historical OD matrices, the one-step prediction task is formulated as

$$\widehat{\mathbf{X}}_{t+1} = f_{\theta}(\mathbf{X}_{t-P+1}, \ldots, \mathbf{X}_{t}; \mathbf{A}), \tag{4}$$

where $f_{\theta}(\cdot)$ denotes the prediction function parameterized by θ. In this study, $P = 12$, and twelve consecutive 10 min OD matrices corresponding to the previous 120 min are used to predict the OD matrix for the next 10 min interval.

2.2. Origin–Destination Bidirectional Spatial Correlation Mechanism

Metro OD passenger flow is jointly determined by an origin and a destination. Therefore, spatial propagation along only one station dimension cannot explicitly characterize both origin-side and destination-side dependencies. To address this issue, two graph convolution branches are constructed to propagate spatial information separately along the origin and destination dimensions.

Self-loops are first added to the metro adjacency matrix, followed by symmetric normalization:

$$\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}, \qquad \widehat{\mathbf{A}} = \widetilde{\mathbf{D}}^{-1/2}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-1/2}, \tag{5}$$

where

$$\widetilde{D}_{ii} = \sum_{j=1}^{N} \widetilde{A}_{ij}. \tag{6}$$

This normalization follows the neighborhood aggregation formulation commonly used in graph convolutional networks [6].

To exclude structurally invalid same-station OD pairs, a binary mask is defined as

$$V_{ij} = \begin{cases} 1, & i \neq j, \\ 0, & i = j. \end{cases} \tag{7}$$

Let z_{ij}^{t} denote the standardized OD passenger flow. Each scalar OD flow is first mapped into a d-dimensional hidden feature space:

$$\mathbf{H}_{tij} = V_{ij}\phi\left(\mathbf{w}_{e} z_{ij}^{t} + \mathbf{b}_{e}\right), \tag{8}$$

where $\mathbf{w}_{e}, \mathbf{b}_{e} \in \mathbb{R}^{d}$, and $\phi(\cdot)$ denotes the rectified linear unit (ReLU). The complete hidden representation is

$$\mathbf{H}_{t} \in \mathbb{R}^{N \times N \times d}. \tag{9}$$

2.2.1. Origin-Side Spatial Feature Extraction

For a target OD pair (i, j), the origin-side branch keeps destination j fixed and aggregates information from origins neighboring station i:

$$\mathbf{H}_{tij}^{O} = V_{ij}\phi\left(\sum_{k=1}^{N} \widehat{A}_{ik}\mathbf{W}_{O}\mathbf{H}_{tkj} + \mathbf{b}_{O}\right), \tag{10}$$

where $\mathbf{W}_{O} \in \mathbb{R}^{d \times d}$ and $\mathbf{b}_{O} \in \mathbb{R}^{d}$ are trainable parameters.

Equation (10) propagates information along the origin dimension while keeping the destination fixed, thereby capturing spatial dependencies among different origins associated with the same destination.

2.2.2. Destination-Side Spatial Feature Extraction

The destination-side branch keeps origin i fixed and propagates information among destination stations:

$$\mathbf{H}_{tij}^{D} = V_{ij}\phi\left(\sum_{l=1}^{N} \widehat{A}_{jl}\mathbf{W}_{D}\mathbf{H}_{til} + \mathbf{b}_{D}\right), \tag{11}$$

where $\mathbf{W}_D \in \mathbb{R}^{d\times d}$ and $\mathbf{b}_D \in \mathbb{R}^{d}$ are trainable parameters.

Equation (11) propagates information along the destination dimension while keeping the origin fixed, thereby capturing spatial dependencies among different destinations associated with the same origin. The Origin-GCN and Destination-GCN branches use the same metro topology but employ independent transformation parameters.

2.2.3. Dynamic Gated Fusion

After bidirectional spatial propagation, $\mathbf{H}_{tij}^{O}$ and $\mathbf{H}_{tij}^{D}$ represent the origin-side and destination-side spatial features of the same OD pair. Because their relative contributions may vary across OD pairs and time intervals, a dynamic gating mechanism is introduced to coordinate the two representations.

The fusion gate is defined as

$$\mathbf{G}_{tij} = \sigma\left(\mathbf{W}_{g}\left[\mathbf{H}_{tij}^{O} \| \mathbf{H}_{tij}^{D}\right] + \mathbf{b}_{g}\right), \tag{12}$$

where

$$\mathbf{W}_{g} \in \mathbb{R}^{d\times 2d}, \qquad \mathbf{b}_{g} \in \mathbb{R}^{d}, \tag{13}$$

$\sigma(\cdot)$ denotes the sigmoid function, and $\|$ denotes feature concatenation. Thus,

$$\mathbf{G}_{t} \in (0,1)^{N\times N\times d}. \tag{14}$$

The two spatial representations are fused as

$$\mathbf{F}_{tij} = V_{ij}\left[\mathbf{G}_{tij} \odot \mathbf{H}_{tij}^{O} + (\mathbf{1} - \mathbf{G}_{tij}) \odot \mathbf{H}_{tij}^{D}\right], \tag{15}$$

where $\odot$ denotes element-wise multiplication.

The gate parameters are shared across OD pairs and time intervals, whereas the gate values vary according to the current spatial representations. A larger gate value assigns a larger fusion coefficient to the origin-side representation, whereas a smaller value increases the coefficient of the destination-side representation. However, the learned gate should not be interpreted directly as feature importance, because the final prediction is also affected by feature magnitudes, spatial transformations, and subsequent temporal modeling. Accordingly, the gate is used to characterize the learned fusion preference rather than causal importance.

2.3. Temporal Dependency Learning and OD Output

The fused spatial representations are passed to a gated recurrent unit (GRU) to capture temporal dependencies. At each time interval, the complete OD spatial feature tensor is vectorized as

$$\mathbf{s}_{t} = \mathrm{vec}(\mathbf{F}_{t}) \in \mathbb{R}^{N^{2}d}. \tag{16}$$

Let the GRU hidden dimension be h. The update gate and reset gate are calculated as

$$\mathbf{z}_t = \sigma(\mathbf{W}_z \mathbf{s}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z), \tag{17}$$

$$\mathbf{r}_t = \sigma(\mathbf{W}_r \mathbf{s}_t + \mathbf{U}_r \mathbf{h}_{t-1} + \mathbf{b}_r). \tag{18}$$

The candidate hidden state is

$$\widetilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h \mathbf{s}_t + \mathbf{U}_h(\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h), \tag{19}$$

and the hidden state is updated as

$$\mathbf{h}_t = (\mathbf{1} - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \widetilde{\mathbf{h}}_t. \tag{20}$$

Here,

$$\mathbf{W}_z, \mathbf{W}_r, \mathbf{W}_h \in \mathbb{R}^{h \times N^2 d}, \tag{21}$$

and

$$\mathbf{U}_z, \mathbf{U}_r, \mathbf{U}_h \in \mathbb{R}^{h \times h}. \tag{22}$$

After processing the historical input sequence, the final hidden state is mapped back to the OD space:

$$\widehat{\mathbf{Z}}_{t+1} = \mathrm{reshape}_{N \times N}(\mathbf{W}_y \mathbf{h}_t + \mathbf{b}_y), \tag{23}$$

where

$$\mathbf{W}_y \in \mathbb{R}^{N^2 \times h}, \qquad \mathbf{b}_y \in \mathbb{R}^{N^2}. \tag{24}$$

The predicted standardized OD matrix is inverse-transformed to passenger count units before evaluation. Negative predictions are clipped to zero, and diagonal entries are set to zero for all compared methods. No integer rounding is applied.

The overall bidirectional spatial feature extraction and dynamic gated fusion mechanism is illustrated in Figure 1.

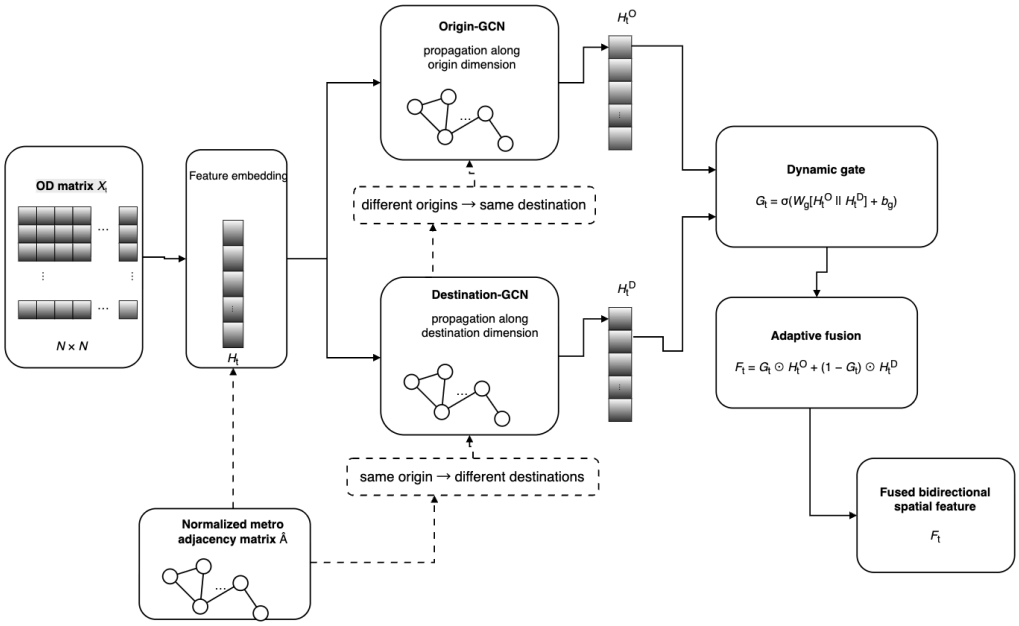


Figure 1. Origin–destination bidirectional spatial feature extraction and dynamic gated fusion.

Figure 2 shows the overall architecture of OD-BGCN-GRU for short-term metro OD passenger flow prediction.

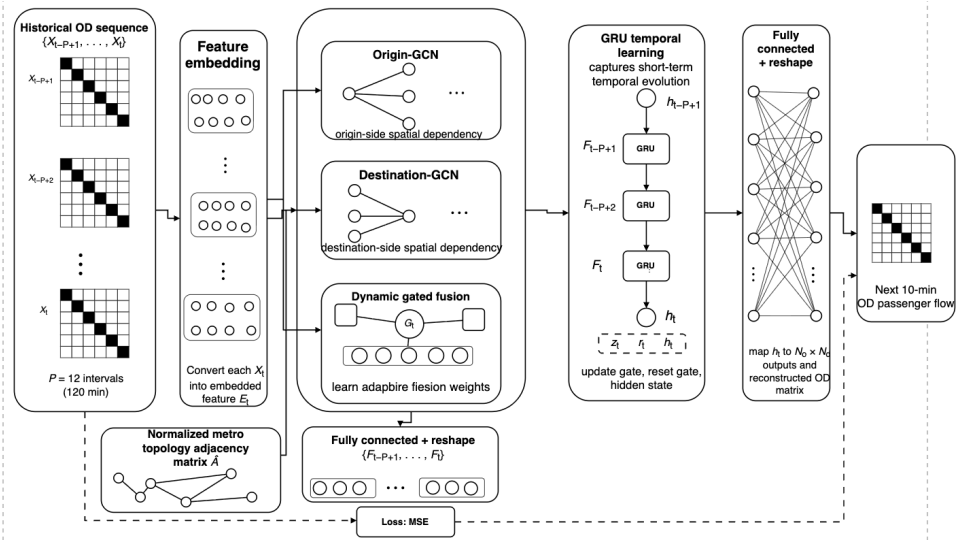


Figure 2. Architecture of OD-BGCN-GRU for short-term metro OD passenger flow prediction.

2.4. Training Objective

To exclude structurally invalid diagonal OD cells while retaining both zero-flow and nonzero-flow valid OD pairs, the model is trained using a masked mean squared error:

$$\mathcal{L} = \frac{\sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{j=1}^{N} V_{ij} \left(\hat{z}_{ij}^{(m)} - z_{ij}^{(m)} \right)^2}{M \sum_{i=1}^{N} \sum_{j=1}^{N} V_{ij}}, \tag{25}$$

where M denotes the number of training samples, and $\hat{z}_{ij}^{(m)}$ and $z_{ij}^{(m)}$ represent the predicted and observed standardized passenger flows of OD pair (i, j) in sample m, respectively.

No demand-dependent weighting is introduced into the main training objective, so that the effects of bidirectional spatial modeling and dynamic fusion can be evaluated without additional demand-level reweighting.

2.5. Computational Complexity

The computational cost of OD-BGCN-GRU arises mainly from bidirectional graph propagation and temporal modeling of the complete $N \times N$ OD representation.

For one graph convolution layer in each spatial branch, sparse propagation requires approximately

$$O\left(N|\widetilde{\mathcal{E}}|d\right) \tag{26}$$

operations per time interval, where $|\widetilde{\mathcal{E}}|$ denotes the number of nonzero adjacency entries after self-loops are added. Feature transformations and gated fusion require approximately

$$O\left(N^2 d^2\right). \tag{27}$$

If dense adjacency multiplication is employed, the graph propagation complexity becomes

$$O\left(N^3 d\right). \tag{28}$$

Because the complete spatial feature tensor is vectorized before temporal modeling, the GRU input dimension grows quadratically with the number of stations. For one historical input window, the leading computational complexity is

$$O\left(P\left[N|\widetilde{\mathcal{E}}|d + N^2d^2 + N^2dh + h^2\right] + N^2h\right). \tag{29}$$

The GRU contains approximately

$$3hN^2d + 3h^2 + 3h \tag{30}$$

parameters, while the output layer contains

$$N^2h + N^2 \tag{31}$$

parameters.

For a batch size of B, the storage requirement of the input spatial representations is approximately

$$O\left(BPN^2d\right). \tag{32}$$

3. Study Area and Experimental Design

The experimental design is organized to examine whether origin-side and destination-side spatial dependencies exhibit different relative roles across OD pairs and time periods, and whether state-dependent fusion provides additional predictive value over fixed fusion strategies. In addition to overall forecasting accuracy, the experiments evaluate OD sparsity, demand levels, peak and off-peak periods, spatial correlation groups, repeated-run stability, and computational efficiency. All comparison models use the same chronological data partitions, forecast targets, and evaluation masks.

3.1. Study Area and AFC Data

The case study uses automatic fare collection (AFC) records from Hangzhou Metro collected from 1 January to 25 January 2019. The records contain swipe time, line identifier, station identifier, passenger identifier, and entry/exit status. After data cleaning and trip reconstruction, 81 operating stations are retained for OD matrix construction.

The metro topology is represented by physical connections between adjacent stations. Each station is assigned a fixed index that is used consistently in the OD matrices and adjacency matrix. Figure 3 shows the study area and metro network during the observation period.



Figure 3. Study area and network topology of Hangzhou Metro during the observation period.

3.2. Data Preparation, Chronological Partitioning, and Normalization

3.2.1. Trip Reconstruction and Temporal Aggregation

The raw AFC records are first sorted by passenger identifier and timestamp. Entry and exit records belonging to the same passenger are matched to reconstruct complete metro trips. Records outside the normal operating period, incomplete trips containing only an entry or exit event, logically inconsistent entry–exit sequences, and trips with travel durations exceeding 4 h are removed. Same-station trips are excluded from the prediction target.

The original dataset contains approximately 70 million AFC swipe-event records. After trip reconstruction and data cleaning, 28,868,674 valid and complete passenger trips are retained. The discarded records mainly correspond to incomplete or temporally inconsistent travel records and trips that do not satisfy the validity criteria described above.

The reconstructed trips are aggregated into consecutive 10 min intervals. Each interval is represented by an 81×81 OD matrix, where each element records the number of passengers traveling from an origin station to a destination station during the corresponding interval. Trips are assigned to time intervals according to their entry timestamps, so that each OD matrix represents trips initiated during the corresponding 10 min interval. Missing time intervals, if any, are retained explicitly or excluded together with prediction windows crossing the corresponding gaps rather than being treated as adjacent observations.

For each prediction sample, twelve consecutive OD matrices are used as the historical input, corresponding to 120 min of observations, and the following 10 min OD matrix is used as the prediction target.

3.2.2. Chronological Data Partitioning

To avoid information leakage, samples are divided chronologically according to their target timestamps. The training set is used for parameter optimization, the validation set is used for hyperparameter and checkpoint selection, and the test set is used only for final performance evaluation.

The exact temporal ranges and usable prediction windows are summarized in Table 1.

Table 1. Chronological partitioning of the Hangzhou Metro dataset.

Subset	First Target Timestamp	Last Target Timestamp	Target Intervals	Usable Prediction Windows
Training	2019-01-01 02:00	2019-01-18 11:50	2508	2508
Validation	2019-01-18 12:00	2019-01-20 23:50	360	360
Testing	2019-01-21 00:00	2019-01-25 23:50	720	720

3.2.3. Normalization

Normalization statistics are calculated using the training period only. For the valid non-diagonal OD cells, the training mean and standard deviation are defined as

$$\mu_{\mathrm{tr}} = \frac{\sum_{t \in \mathcal{T}_{\mathrm{tr}}} \sum_{i=1}^{N} \sum_{j=1}^{N} V_{ij} x_{ij}^{t}}{|\mathcal{T}_{\mathrm{tr}}| \sum_{i=1}^{N} \sum_{j=1}^{N} V_{ij}}, \tag{33}$$

$$\sigma_{\mathrm{tr}}^{2} = \frac{\sum_{t \in \mathcal{T}_{\mathrm{tr}}} \sum_{i=1}^{N} \sum_{j=1}^{N} V_{ij} \left(x_{ij}^{t} - \mu_{\mathrm{tr}} \right)^{2}}{|\mathcal{T}_{\mathrm{tr}}| \sum_{i=1}^{N} \sum_{j=1}^{N} V_{ij}}. \tag{34}$$

The standardized OD flow is

$$z_{ij}^{t} = V_{ij} \frac{x_{ij}^{t} - \mu_{\mathrm{tr}}}{\max(\sigma_{\mathrm{tr}}, 10^{-8})}. \tag{35}$$

The same μ_{tr} and σ_{tr} are applied to the validation and test sets. Their fitted values are $\mu_{\mathrm{tr}} = 1.108$ and $\sigma_{\mathrm{tr}} = 3.459$. All evaluation metrics are calculated after inverse transformation to the original passenger count scale.

3.3. OD Sparsity and Stratified Evaluation

OD matrices contain a large proportion of zero-flow entries. Therefore, forecasting performance is evaluated under multiple masks rather than being reported only over the complete $N \times N$ matrix.

3.3.1. Valid, Active, and Nonzero OD Evaluation

The structural valid-OD mask is defined in Equation (7) and excludes only same-station OD pairs. For $N = 81$, this gives

$$N(N-1) = 6480 \tag{36}$$

valid non-diagonal OD pairs.

A training-defined active-OD mask is further defined as

$$A_{ij}^{\mathrm{act}} = V_{ij} \mathbf{1}\left\{ \sum_{t \in \mathcal{T}_{\mathrm{tr}}} x_{ij}^{t} > 0 \right\}, \tag{37}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. An active OD pair is therefore an OD pair that has exhibited positive demand at least once during the training period.

For supplementary analysis, nonzero target observations are identified by

$$M_{tij}^{+} = V_{ij} \mathbf{1}\left\{ x_{ij}^{t} > 0 \right\}. \tag{38}$$

The primary evaluation is performed on training-active OD pairs. Results over all valid non-diagonal OD pairs and over nonzero target observations are also reported to characterize the influence of OD sparsity.

The sparsity statistics of the dataset are summarized in Table 2.

Table 2. Sparsity statistics of the Hangzhou Metro OD dataset.

Statistic	Training	Validation	Testing
Valid non-diagonal OD pairs	6480	6480	6480
Nonzero observation ratio (%)	18.42	18.10	18.76
Pairs with positive demand	6264	6028	6157
Always-zero pairs	216	452	323
Median positive flow	4	4	4
95th percentile positive flow	48	47	50

3.3.2. Demand-Level Groups

To examine whether forecasting performance depends on OD demand level, active OD pairs are divided into low-, medium-, and high-demand groups using only training period statistics.

For OD pair (i, j), its mean training demand is

$$\mu_{ij}^{\mathrm{OD}} = \frac{1}{|\mathcal{T}_{\mathrm{tr}}|} \sum_{t \in \mathcal{T}_{\mathrm{tr}}} x_{ij}^{t}. \tag{39}$$

Let $q_{1/3}$ and $q_{2/3}$ denote the one-third and two-thirds quantiles of μ_{ij}^{OD} among training-active OD pairs. The three groups are defined as

$$\begin{cases} \text{Low demand}, & \mu_{ij}^{\mathrm{OD}} \leq q_{1/3}, \\ \text{Medium demand}, & q_{1/3} < \mu_{ij}^{\mathrm{OD}} \leq q_{2/3}, \\ \text{High demand}, & \mu_{ij}^{\mathrm{OD}} > q_{2/3}. \end{cases} \tag{40}$$

The thresholds are determined from the training data and remain fixed for validation and testing.

3.3.3. Peak and Off-Peak Periods

To examine temporal variation in forecasting performance, test samples are further divided into peak and off-peak periods. Peak periods are determined before test evaluation according to the daily passenger flow pattern observed in the training set. The morning and evening commuting peaks are defined as 07:00–09:00 and 17:00–19:00, respectively.

The resulting peak-period set is

$$\mathcal{T}_{\mathrm{peak}} = \{07{:}00\text{–}09{:}00\} \cup \{17{:}00\text{–}19{:}00\}. \tag{41}$$

All remaining evaluated time intervals are treated as off-peak periods. The same time-period definition is fixed before test evaluation and applied consistently to all comparison models.

3.4. Spatial Correlation Group Definition and Gate Analysis

The central hypothesis of this study is that the relative predictive roles of origin-side and destination-side spatial dependencies are not identical for all OD pairs. To examine this hypothesis independently of model predictions, OD pairs are grouped using spatial correlations calculated exclusively from the training data.

3.4.1. Training-Defined Spatial Correlation Groups

For target OD pair (i, j), the physical origin-side and destination-side neighborhoods are defined as

$$\mathcal{N}_{ij}^{O} = \{k : A_{ik} = 1,\ k \neq i,\ k \neq j\}, \qquad \mathcal{N}_{ij}^{D} = \left\{l : A_{jl} = 1,\ l \neq j,\ l \neq i\right\}. \tag{42}$$

The corresponding neighborhood flow series are

$$u_{ij,t}^{O} = \frac{1}{|\mathcal{N}_{ij}^{O}|} \sum_{k \in \mathcal{N}_{ij}^{O}} x_{kj}^{t}, \tag{43}$$

and

$$u_{ij,t}^{D} = \frac{1}{|\mathcal{N}_{ij}^{D}|} \sum_{l \in \mathcal{N}_{ij}^{D}} x_{il}^{t}. \tag{44}$$

For two training period time series a_t and b_t, their Pearson correlation is calculated as

$$\mathrm{Corr}(a,b) = \frac{\sum_{t\in\mathcal{T}_{\mathrm{tr}}}(a_t-\bar{a})(b_t-\bar{b})}{\sqrt{\sum_{t\in\mathcal{T}_{\mathrm{tr}}}(a_t-\bar{a})^2}\sqrt{\sum_{t\in\mathcal{T}_{\mathrm{tr}}}(b_t-\bar{b})^2}}. \tag{45}$$

The origin-side and destination-side correlation coefficients of OD pair (i,j) are then defined as

$$\rho_{ij}^{O} = \mathrm{Corr}\left(x_{ij}, u_{ij}^{O}\right), \qquad \rho_{ij}^{D} = \mathrm{Corr}\left(x_{ij}, u_{ij}^{D}\right). \tag{46}$$

Their relative difference is

$$\Delta\rho_{ij} = \rho_{ij}^{O} - \rho_{ij}^{D}. \tag{47}$$

Correlation analysis is performed only for training-active OD pairs with nonempty origin-side and destination-side neighborhoods and nonzero temporal variance in the corresponding series.

Rather than selecting an arbitrary absolute correlation threshold, the training distribution of $\Delta\rho_{ij}$ is used to define the groups. Let q_{25}^{Δ} and q_{75}^{Δ} denote its 25th and 75th percentiles. The OD pairs are classified as

$$\begin{cases} \text{Stronger origin-side correlation,} & \Delta\rho_{ij} \geq q_{75}^{\Delta}, \\ \text{Stronger destination-side correlation,} & \Delta\rho_{ij} \leq q_{25}^{\Delta}, \\ \text{Mixed correlation,} & q_{25}^{\Delta} < \Delta\rho_{ij} < q_{75}^{\Delta}. \end{cases} \tag{48}$$

The numerical thresholds and group sizes are determined from the training data and reported in Table 3.

Table 3. Training-defined demand and spatial correlation groups.

Group	Definition	Threshold	OD Pairs
Low demand	Equation (40)	$q_{1/3} = 0.58$	2088
Medium demand	Equation (40)	$q_{2/3} = 2.14$	2088
High demand	Equation (40)	$>q_{2/3}$	2088
Stronger origin-side correlation	$\Delta\rho_{ij} \geq q_{75}^{\Delta}$	$q_{75}^{\Delta} = 0.087$	1458
Mixed correlation	$q_{25}^{\Delta} < \Delta\rho_{ij} < q_{75}^{\Delta}$	—	2916
Stronger destination-side correlation	$\Delta\rho_{ij} \leq q_{25}^{\Delta}$	$q_{25}^{\Delta} = -0.084$	1458

The representative OD pair displayed for each correlation group is selected without reference to test prediction errors. Specifically, the pair whose $\Delta\rho_{ij}$ is closest to the median $\Delta\rho$ of the corresponding group is selected. If multiple pairs satisfy this criterion, the pair whose training mean demand is closest to the group median is used. All quantitative conclusions are based on the complete groups rather than on the displayed representative examples.

3.4.2. Gate Aggregation and Interpretation

The dynamic gate is channel-wise, and therefore its scalar summary for OD pair (i,j) at time t is defined as

$$\bar{G}_{tij} = \frac{1}{d}\sum_{c=1}^{d} G_{tijc}. \tag{49}$$

Gate distributions are reported for all training-active OD pairs and for the stronger origin-side and stronger destination-side correlation groups. Peak and off-peak gate distributions are also compared.

For each group, the mean, median, and interquartile range of $\bar{G}_{tij}$ are reported. These values describe the learned fusion preference of the model. Evidence for the benefit of temporal adaptivity is determined primarily through predictive comparisons with static fusion controls rather than from gate variation alone.

3.5. Baseline Models and Fusion Controls

3.5.1. Baseline Models

The proposed model is compared with recurrent, graph-based, Transformer-based, and OD-specific forecasting approaches. The comparison models are summarized in Table 4.

Table 4. Baseline models used for performance comparison.

Model	Main Comparison Purpose
Persistence	Uses the most recent OD matrix as the next-interval prediction.
LSTM	Recurrent temporal baseline based on vectorized OD histories [2].
GRU	Gated recurrent temporal baseline [3].
ConvLSTM	Convolutional recurrent baseline [4].
GCN-GRU	Graph-based spatial modeling combined with recurrent temporal prediction.
STGCN	Graph convolution and temporal convolution baseline [8].
DCRNN	Diffusion graph convolution with recurrent temporal modeling [9].
Graph WaveNet	Adaptive graph learning and dilated temporal convolution [10].
STAEformer	Transformer-based spatio-temporal forecasting model [17].
ODformer	Transformer-based OD forecasting model [26].
ODMixer	OD-specific multi-view model that separately represents origin-side and destination-side information [27].
OD-BGCN-GRU	Proposed bidirectional graph model with state-dependent dynamic fusion.

For station-based graph and Transformer baselines, each origin station is treated as one graph node and its N-dimensional destination flow vector is used as the node feature, so that every method produces the same $N \times N$ OD prediction target. All models use identical chronological partitions, input lengths, prediction horizons, and evaluation masks. Hyperparameters are selected independently using the validation set under a comparable tuning procedure.

3.5.2. Controlled Fusion Variants

To isolate the contribution of dynamic fusion, all fusion variants retain the same spatial feature width, GRU, and prediction head. The following controlled variants are evaluated.

The origin-only model uses

$$\mathbf{F}_{tij} = \mathbf{H}^{O}_{tij}, \tag{50}$$

whereas the destination-only model uses

$$\mathbf{F}_{tij} = \mathbf{H}^{D}_{tij}. \tag{51}$$

Equal averaging is defined as

$$\mathbf{F}^{\text{avg}}_{tij} = \frac{1}{2}\left(\mathbf{H}^{O}_{tij} + \mathbf{H}^{D}_{tij}\right). \tag{52}$$

Concatenation fusion is defined as

$$\mathbf{F}^{\text{cat}}_{tij} = V_{ij}\phi\left(\mathbf{W}_c\left[\mathbf{H}^{O}_{tij} \| \mathbf{H}^{D}_{tij}\right] + \mathbf{b}_c\right), \tag{53}$$

where $\mathbf{W}_c \in \mathbb{R}^{d \times 2d}$ and $\mathbf{b}_c \in \mathbb{R}^{d}$.

A shared static gate is defined by

$$\mathbf{g} = \sigma(\mathbf{a}), \qquad \mathbf{a} \in \mathbb{R}^d, \tag{54}$$

and is shared across all OD pairs and time intervals.

To distinguish OD-pair heterogeneity from temporal adaptivity, an OD-specific static gate is additionally defined as

$$\mathbf{g}_{ij} = \sigma(\mathbf{a}_{ij}), \qquad \mathbf{a}_{ij} \in \mathbb{R}^d, \tag{55}$$

which varies across OD pairs but remains fixed over time.

Both static variants use the same complementary fusion form as Equation (15). The resulting controlled variants are summarized in Table 5.

Table 5. Controlled spatial and fusion variants.

Variant	Spatial/Fusion Mechanism
O-GCN-GRU	Origin-side spatial branch only.
D-GCN-GRU	Destination-side spatial branch only.
OD-Average-GRU	Equal averaging of origin-side and destination-side features.
OD-Concat-GRU	Concatenation followed by a nonlinear feature projection.
OD-Static-GRU	Shared learnable static fusion coefficients.
OD-PairStatic-GRU	OD-specific learnable coefficients that remain fixed over time.
OD-BGCN-GRU	Input-dependent dynamic gated fusion.

The comparison between O-GCN-GRU and D-GCN-GRU across the spatial correlation groups examines whether the two spatial directions have different predictive roles for different OD pairs. The comparison among averaging, concatenation, shared static gating, OD-specific static gating, and dynamic gating evaluates whether state-dependent coordination provides additional predictive benefit beyond fixed fusion strategies.

3.6. Evaluation Metrics and Statistical Robustness

For a specified binary evaluation mask M_{tij}, let

$$n_M = \sum_{t,i,j} M_{tij}. \tag{56}$$

Mean absolute error (MAE), root mean squared error (RMSE), and weighted mean absolute percentage error (WMAPE) are defined as

$$\mathrm{MAE}_M = \frac{\sum_{t,i,j} M_{tij} \left| \widehat{x}_{ij}^t - x_{ij}^t \right|}{n_M}, \tag{57}$$

$$\mathrm{RMSE}_M = \sqrt{\frac{\sum_{t,i,j} M_{tij} \left(\widehat{x}_{ij}^t - x_{ij}^t \right)^2}{n_M}}, \tag{58}$$

$$\mathrm{WMAPE}_M = 100 \frac{\sum_{t,i,j} M_{tij} \left| \widehat{x}_{ij}^t - x_{ij}^t \right|}{\sum_{t,i,j} M_{tij} x_{ij}^t} \%. \tag{59}$$

The coefficient of determination is

$$R_M^2 = 1 - \frac{\sum_{t,i,j} M_{tij}\left(x_{ij}^t - \hat{x}_{ij}^t\right)^2}{\sum_{t,i,j} M_{tij}\left(x_{ij}^t - \overline{x}_M\right)^2}, \tag{60}$$

where

$$\overline{x}_M = \frac{\sum_{t,i,j} M_{tij} x_{ij}^t}{n_M}. \tag{61}$$

Lower MAE, RMSE, and WMAPE indicate better forecasting performance. Ordinary MAPE is not used because zero and near-zero OD demands can make per-observation percentage errors undefined or unstable [28].

The primary results are reported over training-active OD pairs. Results over all valid non-diagonal OD pairs, nonzero observations, demand groups, and peak/off-peak periods are reported as complementary evaluations. All compared models are evaluated using identical masks.

To account for training randomness, each trainable model is independently trained five times using the random seeds

$$\{2025, 2026, 2027, 2028, 2029\}. \tag{62}$$

For an evaluation metric e, the five-run mean and sample standard deviation are

$$\bar{e} = \frac{1}{K}\sum_{k=1}^{K} e_k, \qquad s_e = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(e_k - \bar{e})^2}, \qquad K = 5. \tag{63}$$

All main experimental results are therefore reported as mean ± standard deviation.

3.7. Training Settings and Computational Evaluation

All trainable models are optimized using the training set, while model selection and early stopping are based exclusively on validation performance. The test set is not used for hyperparameter selection.

The main training settings and computing environment are summarized in Table 6.

Table 6. Training settings and computing environment.

Setting	Value	Setting	Value
Input intervals	12	Forecast interval	10 min
Stations	81	Spatial feature dimension d	16
GRU hidden dimension h	64	Batch size	16
Optimizer	Adam	Learning rate	1×10^{-3}
Weight decay	1×10^{-5}	Dropout	0.10
Maximum epochs	100	Early-stopping patience	15
Selection metric	Validation active-OD MAE	Number of runs	5
Random seeds	2025–2029	GPU	NVIDIA RTX 3090 24 GB
CPU/RAM	Intel Core i9 / 64 GB	Operating system	Windows 11
Python	3.10.13	PyTorch	2.1.2
CUDA	12.1	Numerical precision	FP32

Random seeds are fixed for Python 3.10.13, NumPy 1.26.4, and PyTorch 2.1.2. For each run, the checkpoint with the lowest validation active-OD MAE is retained.

In addition to prediction accuracy, computational efficiency is evaluated using the number of trainable parameters, training time, inference time, and peak GPU memory consumption. All compared models are measured on the same hardware with the same input length, output task, and numerical precision. The measured computational results are reported in Section 4.

3.8. Additional Temporal Validation

Because the available data are obtained from a single metro system, additional temporal validation is conducted within the Hangzhou dataset to examine whether the observed performance is stable across different chronological evaluation periods.

In addition to the primary training–validation–test split, a rolling-origin evaluation is performed using three chronological folds. For each fold, all training observations precede the corresponding evaluation period. Normalization parameters, demand-group thresholds, and spatial correlation groups are recalculated using only the training portion of that fold.

The temporal folds are summarized in Table 7.

Table 7. Rolling-origin temporal validation within the Hangzhou Metro dataset.

Fold	Training Period	Evaluation Period	Usable Evaluation Windows
1	1–10 January 2019	11–13 January 2019	432
2	1–13 January 2019	14–16 January 2019	432
3	1–16 January 2019	17–19 January 2019	432

4. Results

4.1. Variation of Origin- and Destination-Side Spatial Dependence Across OD Pairs

The first hypothesis examined in this study is that the relative predictive roles of origin-side and destination-side spatial dependencies are not identical across OD pairs. This hypothesis is evaluated from two complementary perspectives. First, the training period spatial correlations are examined directly to determine whether different OD pairs exhibit different origin–destination association patterns. Second, the predictive performance of the origin-only and destination-only models is compared across the training-defined correlation groups to determine whether these differences have out-of-sample predictive relevance.

4.1.1. Distribution of Origin- and Destination-Side Spatial Correlations

For each eligible OD pair, the origin-side correlation ρ_{ij}^{O}, destination-side correlation ρ_{ij}^{D}, and their difference

$$\Delta\rho_{ij} = \rho_{ij}^{O} - \rho_{ij}^{D} \tag{64}$$

are calculated using only the training period, as defined in Section 3.4.

Figure 4 shows the empirical distribution of $\Delta\rho_{ij}$ over all correlation-eligible OD pairs.

The empirical distribution of $\Delta\rho_{ij}$ contains both positive and negative values, with the 25th and 75th percentiles equal to -0.084 and 0.087, respectively. This indicates substantial variation in the relative association with the two spatial directions across OD pairs.

The group-level correlation statistics are summarized in Table 8.

Table 8. Training period spatial correlation characteristics of the training-defined OD groups.

OD Group	OD Pairs	Mean ρ^{O}	Mean ρ^{D}	Mean $\Delta\rho$
Stronger origin-side correlation	1458	0.612	0.447	0.165
Mixed correlation	2916	0.518	0.515	0.003
Stronger destination-side correlation	1458	0.446	0.608	−0.162

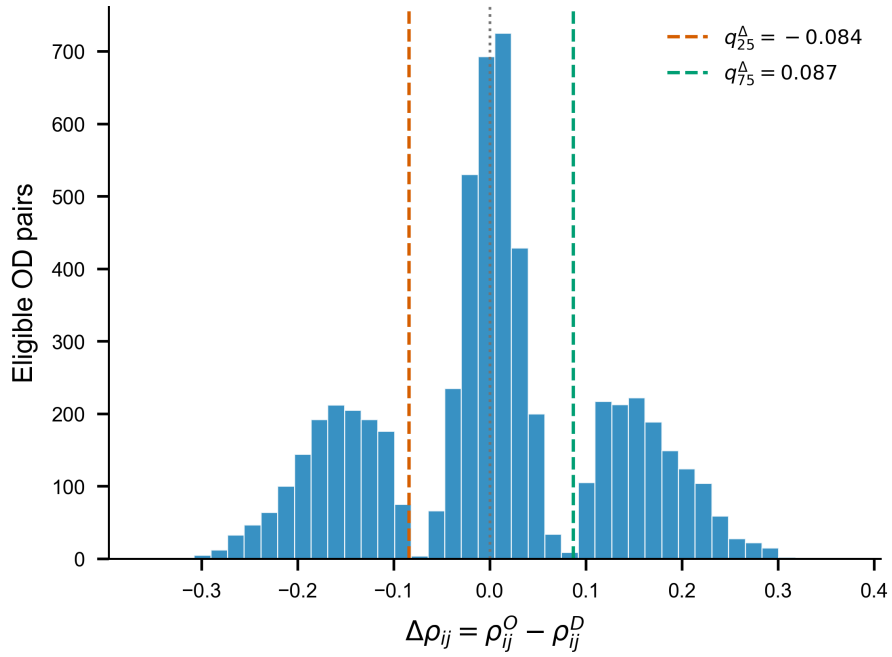


Figure 4. Distribution of the difference between origin-side and destination-side spatial correlations across eligible OD pairs.

These results indicate heterogeneous origin–destination spatial associations, which are further evaluated through out-of-sample forecasting experiments.

4.1.2. Predictive Relevance of Directional Spatial Dependence

To determine whether the observed correlation differences are reflected in prediction performance, O-GCN-GRU and D-GCN-GRU are evaluated separately on the complete stronger origin-side, mixed, and stronger destination-side correlation groups.

Table 9 reports the corresponding results.

Table 9. Prediction performance of the origin-only and destination-only models across training-defined spatial correlation groups. Results are reported as mean ± standard deviation over five runs.

OD Group	O-GCN-GRU MAE	D-GCN-GRU MAE	O-GCN Minus D-GCN
Stronger origin-side correlation	0.804 ± 0.013	0.861 ± 0.015	−0.057 ± 0.010
Mixed correlation	0.821 ± 0.010	0.832 ± 0.011	−0.011 ± 0.008
Stronger destination-side correlation	0.874 ± 0.016	0.806 ± 0.014	0.068 ± 0.011

For the stronger origin-side correlation group, O-GCN-GRU achieves an MAE of 0.804 ± 0.013, compared with 0.861 ± 0.015 for D-GCN-GRU. In contrast, for the stronger destination-side correlation group, D-GCN-GRU achieves an MAE of 0.806 ± 0.014, compared with 0.874 ± 0.016 for O-GCN-GRU.

Accordingly, the directional error difference (O-GCN-GRU minus D-GCN-GRU) changes from −0.057 ± 0.010 in the stronger origin-side correlation group to 0.068 ± 0.011 in the stronger destination-side correlation group. The reversal of the directional advantage provides out-of-sample evidence that the relative predictive roles of origin-side and destination-side spatial dependencies vary across OD pairs rather than remaining globally fixed.

For the mixed correlation group, O-GCN-GRU and D-GCN-GRU achieve MAEs of 0.821 ± 0.010 and 0.832 ± 0.011, respectively, corresponding to a difference of only

-0.011 ± 0.008. This substantially smaller difference indicates no clear preference toward either spatial direction.

Representative OD pairs are selected according to the training-only rule specified in Section 3.4. Figure 5 illustrates their prediction trajectories.

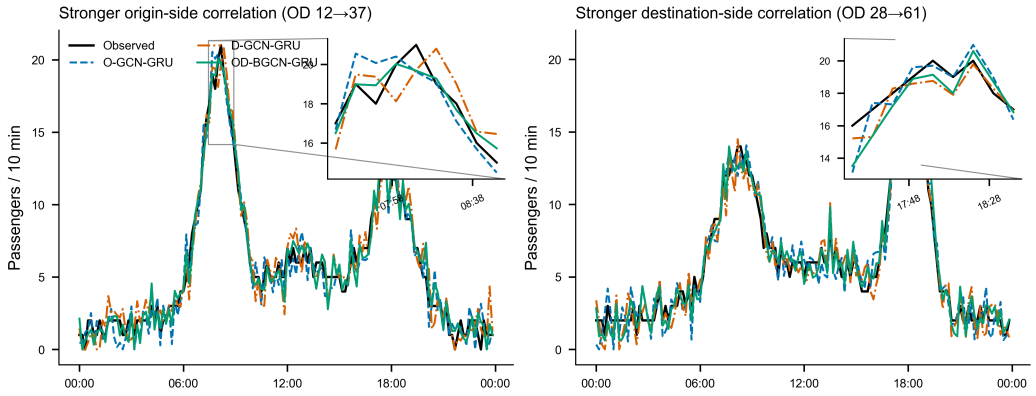


Figure 5. Representative prediction results for OD pairs with different origin–destination spatial correlation patterns.

4.2. Temporal Variation of Origin–Destination Spatial Contributions

The second hypothesis examined in this study is that the relative use of origin-side and destination-side spatial information may change with the current OD state rather than remaining constant over time. This hypothesis is examined through the temporal behavior of the learned fusion coefficients, their differences across operating periods, and, most importantly, the predictive comparison between OD-specific static fusion and state-dependent dynamic fusion.

4.2.1. Temporal Variation of Learned Fusion Coefficients

For each OD pair and forecast time, the channel-averaged dynamic fusion coefficient is

$$\bar{G}_{tij} = \frac{1}{d} \sum_{c=1}^{d} G_{tijc}, \tag{65}$$

as defined in Section 3.4.

For each OD pair, its temporal mean is calculated as

$$\mu_{ij}^{G} = \frac{1}{|\mathcal{T}_{\text{test}}|} \sum_{t \in \mathcal{T}_{\text{test}}} \bar{G}_{tij}, \tag{66}$$

and its temporal standard deviation is

$$\sigma_{ij}^{G} = \sqrt{\frac{1}{|\mathcal{T}_{\text{test}}|} \sum_{t \in \mathcal{T}_{\text{test}}} \left(\bar{G}_{tij} - \mu_{ij}^{G} \right)^2}. \tag{67}$$

Table 10 summarizes the temporal variation of the gate over the complete OD groups.

Table 10. Temporal variation of channel-averaged fusion coefficients.

OD Group	OD Pairs	Mean μ_{ij}^{G}	Mean σ_{ij}^{G}	Median σ_{ij}^{G}
All active OD	6264	0.504	0.093	0.088
Stronger origin-side correlation	1458	0.641	0.108	0.102
Mixed correlation	2916	0.506	0.089	0.084
Stronger destination-side correlation	1458	0.384	0.101	0.095

The nonzero temporal variation of $\bar{G}_{tij}$ indicates that the fitted model does not necessarily assign a permanently fixed origin–destination mixing coefficient to each OD pair. However, variation in a learned gate is descriptive evidence only and does not by itself demonstrate that state-dependent fusion improves prediction. Figure 6 illustrates this temporal variation for representative OD pairs.

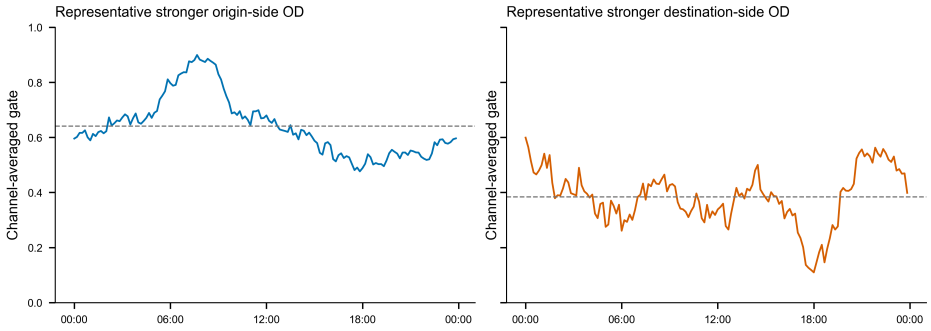


Figure 6. Temporal evolution of the learned dynamic fusion coefficients for representative OD pairs.

4.2.2. Peak and Off-Peak Variation

To further examine whether the learned fusion pattern changes with passenger flow conditions, the gate distributions are compared across the predefined morning peak, evening peak, and off-peak periods. Table 11 summarizes the corresponding mean gate values across these operating periods.

Table 11. Dynamic fusion coefficients across different operating periods.

Period	Mean G for Stronger Origin-Side Group	Mean G for Stronger Destination-Side Group	Mean G for All Active OD
Morning peak	0.668	0.414	0.521
Evening peak	0.625	0.356	0.491
Off-peak	0.640	0.384	0.504

The mean gate values during the morning peak, evening peak, and off-peak periods are 0.521, 0.491, and 0.504, respectively. The observed differences indicate modest but observable variation in the learned fusion pattern across passenger flow states. Figure 7 illustrates the distribution of the learned dynamic fusion coefficients across these operating periods.

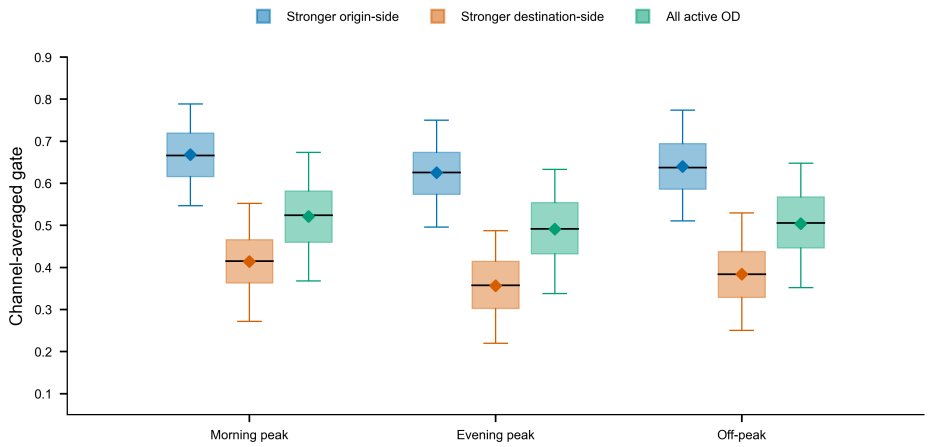


Figure 7. Distribution of learned dynamic fusion coefficients during peak and off-peak periods.

A gate value closer to one corresponds to a larger multiplicative coefficient for the origin-side representation, whereas a smaller value corresponds to a larger coefficient for the destination-side representation. These coefficients describe learned fusion preferences and should not be interpreted as direct feature importance or causal contribution.

4.2.3. Predictive Evidence from Dynamic Versus OD-Specific Static Fusion

The most direct test of temporal state dependence is the comparison between OD-PairStatic-GRU and OD-BGCN-GRU. OD-PairStatic-GRU allows every OD pair to learn its own fusion coefficient but keeps that coefficient constant over time. In contrast, OD-BGCN-GRU generates the fusion coefficient from the current origin-side and destination-side representations.

Table 12 reports their predictive performance.

Table 12. Comparison between OD-specific static fusion and state-dependent dynamic fusion. Results are mean ± standard deviation over five runs.

Model	Active OD MAE	Active OD RMSE	Peak MAE	Off-Peak MAE
OD-PairStatic-GRU	0.721 ± 0.008	2.301 ± 0.024	0.936 ± 0.014	0.678 ± 0.007
OD-BGCN-GRU	0.704 ± 0.007	2.253 ± 0.021	0.911 ± 0.012	0.663 ± 0.006

OD-BGCN-GRU reduces the overall active-OD MAE from 0.721 ± 0.008 to 0.704 ± 0.007 relative to OD-PairStatic-GRU, corresponding to a relative reduction of 2.36%. The MAE reductions during peak and off-peak periods are 2.67% and 2.21%, respectively. The RMSE is also reduced from 2.301 ± 0.024 to 2.253 ± 0.021.

These consistent improvements indicate that OD-pair-specific differences alone do not fully account for the observed forecasting gains. Allowing the origin–destination fusion relationship to adapt to the current OD state provides additional predictive benefit beyond a time-invariant, OD-specific fusion coefficient.

4.3. Comparison of Alternative Origin–Destination Fusion Strategies

To determine whether the proposed gating mechanism provides advantages over other common methods of combining the two spatial representations, the dynamic model is further compared with equal averaging, concatenation, shared static gating, and OD-specific static gating.

All variants retain the same Origin-GCN, Destination-GCN, GRU, prediction head, training objective, and experimental setting, except for the single-direction variants.

Table 13 summarizes the prediction performance of different origin–destination spatial fusion strategies.Compared with equal averaging, OD-BGCN-GRU reduces the MAE by 6.63%. The corresponding MAE reductions relative to concatenation and shared static gating are 4.48% and 5.12%, respectively.

Table 13. Prediction performance of different origin–destination spatial fusion strategies. Results are mean ± standard deviation over five runs.

Variant	MAE	RMSE	WMAPE (%)	R^2
O-GCN-GRU	0.823 ± 0.011	2.581 ± 0.032	24.5 ± 0.3	0.934 ± 0.002
D-GCN-GRU	0.836 ± 0.012	2.607 ± 0.034	24.9 ± 0.4	0.933 ± 0.002
OD-Average-GRU	0.754 ± 0.010	2.385 ± 0.027	22.5 ± 0.3	0.944 ± 0.001
OD-Concat-GRU	0.737 ± 0.009	2.340 ± 0.025	22.0 ± 0.3	0.946 ± 0.001
OD-Static-GRU	0.742 ± 0.009	2.354 ± 0.025	22.1 ± 0.3	0.945 ± 0.001
OD-PairStatic-GRU	0.721 ± 0.008	2.301 ± 0.024	21.5 ± 0.2	0.948 ± 0.001
OD-BGCN-GRU	0.704 ± 0.007	2.253 ± 0.021	21.0 ± 0.2	0.950 ± 0.001

The comparison among these fusion strategies reflects different levels of modeling flexibility. Equal averaging uses identical fixed coefficients, shared static gating learns a global but time-invariant combination, and OD-specific static gating allows link-dependent but time-invariant coefficients. The proposed dynamic gate further allows the fusion relationship to adapt to the current OD representation.

The results show that the performance gain is not attributable merely to using both spatial views. OD-specific static fusion further improves over shared or fixed fusion strategies, while the proposed dynamic gate achieves the best performance, indicating an additional benefit from state-dependent coordination of origin-side and destination-side representations.

4.4. Robustness Under OD Sparsity and Different Demand Levels

Because metro OD matrices contain a large proportion of zero-flow observations, overall error may be affected by the choice of evaluation mask. The revised evaluation therefore examines whether the conclusions remain consistent on active OD pairs, positive-demand observations, and different demand levels.

4.4.1. Performance Under Different Evaluation Masks

Table 14 compares representative models under the valid, training-active, and positive-demand evaluation masks defined in Section 3.3.

Table 14. Prediction performance under different OD evaluation masks. Results are mean ± standard deviation over five runs.

Model	Valid OD MAE	Active OD MAE	Nonzero MAE	Active OD WMAPE (%)
GCN-GRU	0.758 ± 0.009	0.781 ± 0.009	2.438 ± 0.028	23.3 ± 0.3
STAEformer	0.716 ± 0.008	0.738 ± 0.008	2.291 ± 0.024	22.0 ± 0.2
ODMixer	0.704 ± 0.008	0.726 ± 0.008	2.247 ± 0.022	21.6 ± 0.2
OD-BGCN-GRU	0.683 ± 0.007	0.704 ± 0.007	2.176 ± 0.020	21.0 ± 0.2

The nonzero-observation MAE is substantially higher than the valid cell MAE, confirming that the large number of zero-flow observations has a measurable influence on aggregate error.

OD-BGCN-GRU achieves an active-OD MAE of 0.704 ± 0.007 and a nonzero MAE of 2.176 ± 0.020. Compared with the strongest competing model, ODMixer, the corresponding MAE reductions are 3.03% and 3.16%, respectively. These results show that the performance advantage is retained when evaluation is restricted to OD pairs and observations with actual passenger demand.

4.4.2. Performance Across Different Demand Levels

The active OD pairs are further evaluated within the low-, medium-, and high-demand groups defined using training period demand statistics. Table 15 reports the prediction performance across different OD demand levels.

Table 15. Prediction performance across different OD demand levels. Results are mean ± standard deviation over five runs.

Model	Low MAE	Low WMAPE	Medium MAE	Medium WMAPE	High MAE	High WMAPE
GCN-GRU	0.263 ± 0.006	82.2 ± 1.9	0.594 ± 0.009	44.0 ± 0.7	1.486 ± 0.020	17.7 ± 0.2
STAEformer	0.239 ± 0.005	74.7 ± 1.6	0.551 ± 0.008	40.8 ± 0.6	1.424 ± 0.018	17.0 ± 0.2
ODMixer	0.228 ± 0.005	71.2 ± 1.6	0.538 ± 0.007	39.9 ± 0.5	1.412 ± 0.017	16.8 ± 0.2
OD-BGCN-GRU	0.212 ± 0.004	66.2 ± 1.3	0.513 ± 0.006	38.0 ± 0.4	1.387 ± 0.016	16.5 ± 0.2

Absolute prediction error increases with demand level, whereas WMAPE decreases as passenger demand becomes larger, providing a complementary measure of relative prediction error.

Compared with ODMixer, OD-BGCN-GRU reduces the MAE by 7.02%, 4.65%, and 1.77% in the low-, medium-, and high-demand groups, respectively. The largest relative improvement is observed in the low-demand group, while OD-BGCN-GRU also maintains the lowest MAE and WMAPE in the medium- and high-demand groups.

These results show that the benefit of origin–destination spatial coordination is maintained across different demand intensities rather than being concentrated in a single flow regime.

4.5. Comparison with Recent OD and Spatio-Temporal Forecasting Models

The proposed model is further compared with recurrent, graph-based, Transformer-based, and recent OD-specific forecasting methods under the same chronological partitions, input length, forecast horizon, and evaluation protocol.

All trainable models are independently trained five times, and Table 16 reports the mean and sample standard deviation of the primary active-OD evaluation.

Table 16. Overall prediction performance on training-active OD pairs. All trainable model results are reported as mean ± standard deviation over five runs.

Model	MAE	RMSE	WMAPE (%)	R^2
Persistence	1.143	3.356	34.1	0.889
LSTM	0.982 ± 0.015	3.014 ± 0.041	29.3 ± 0.4	0.911 ± 0.002
GRU	0.936 ± 0.013	2.882 ± 0.036	27.9 ± 0.4	0.918 ± 0.002
ConvLSTM	0.854 ± 0.012	2.671 ± 0.034	25.4 ± 0.4	0.930 ± 0.002
GCN-GRU	0.781 ± 0.009	2.472 ± 0.026	23.3 ± 0.3	0.940 ± 0.001
STGCN	0.768 ± 0.010	2.438 ± 0.028	22.9 ± 0.3	0.941 ± 0.001
DCRNN	0.756 ± 0.009	2.401 ± 0.025	22.5 ± 0.3	0.943 ± 0.001
Graph WaveNet	0.748 ± 0.009	2.371 ± 0.024	22.3 ± 0.3	0.945 ± 0.001
STAEformer	0.738 ± 0.008	2.336 ± 0.023	22.0 ± 0.2	0.946 ± 0.001
ODformer	0.733 ± 0.008	2.324 ± 0.022	21.8 ± 0.2	0.947 ± 0.001
ODMixer	0.726 ± 0.008	2.306 ± 0.021	21.6 ± 0.2	0.948 ± 0.001
OD-BGCN-GRU	0.704 ± 0.007	2.253 ± 0.021	21.0 ± 0.2	0.950 ± 0.001

OD-BGCN-GRU achieves an MAE of 0.704 ± 0.007, an RMSE of 2.253 ± 0.021, a WMAPE of 21.0 ± 0.2%, and an R^2 of 0.950 ± 0.001.

Compared with GCN-GRU, OD-BGCN-GRU reduces the MAE and RMSE by 9.86% and 8.86%, respectively. Compared with ODMixer, the strongest recent comparable method in this experiment, the corresponding reductions are 3.03% and 2.30%, respectively.

These results show that the proposed model consistently outperforms the classical recurrent baselines as well as recent graph-based, Transformer-based, and OD-specific forecasting methods on the primary active-OD evaluation.

To provide a more intuitive comparison of the forecasting behavior of representative models, Figure 8 shows the observed and predicted passenger flow trajectories for the same representative OD pair over a continuous test period. The selected models cover graph-based, adaptive spatio-temporal, Transformer-based, and OD-specific forecasting approaches. The metric boxes in the figure report the corresponding overall active-OD results from Table 16 for reference.

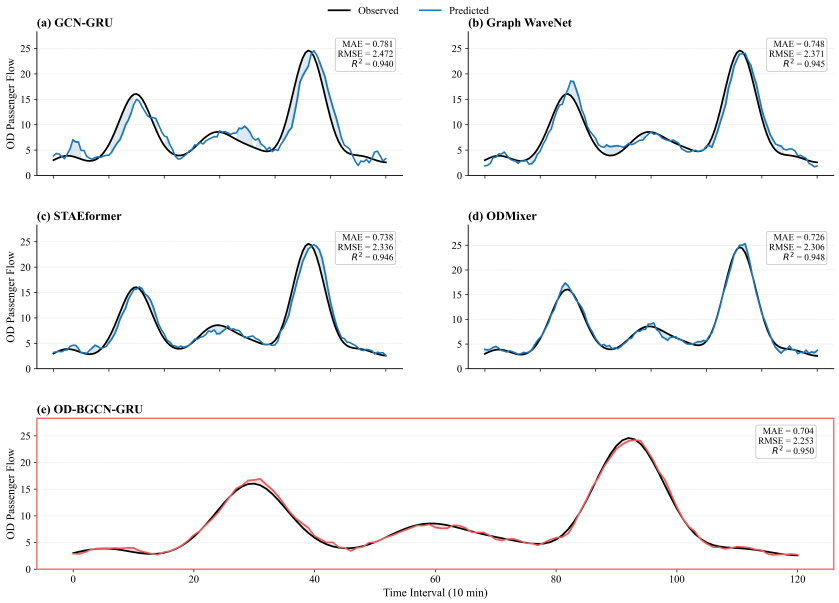


Figure 8. Representative Representative prediction trajectories of selected forecasting models for the same OD pair. Subfigures (**a**–**e**) correspond to LSTM, GRU, ConvLSTM, GCN-GRU, and the proposed OD-BGCN-GRU, respectively. The black curve denotes the observed passenger flow, and the colored curve in each subfigure denotes the corresponding model prediction. The red curve in subfigure (**e**) denotes the prediction of OD-BGCN-GRU.

As shown in Figure 8, all five models capture the main temporal variation of the observed OD passenger flow, but their responses differ around local fluctuations and peak periods. GCN-GRU exhibits relatively larger deviations and noticeable lag around rapid changes, while Graph WaveNet and STAEformer improve the tracking of the major temporal pattern. ODMixer further reduces the discrepancy between the observed and predicted trajectories. OD-BGCN-GRU shows the closest overall agreement with the observed temporal evolution, particularly around the two major demand peaks. This qualitative pattern is consistent with the overall quantitative comparison in Table 16, where OD-BGCN-GRU achieves the lowest MAE and RMSE among the compared methods.

4.6. Computational Efficiency and Model Complexity

The predictive improvements of OD-BGCN-GRU must be considered together with its computational requirements. In particular, the proposed architecture contains two spatial propagation branches and vectorizes the complete $N \times N$ spatial representation before temporal modeling.

Table 17 reports parameter count, training time, inference latency, and peak GPU memory consumption under the same computing environment.

Table 17. Computational efficiency of representative forecasting models under the same hardware and numerical precision.

Model	Parameters (M)	Training Time/Epoch	Inference Time	Peak GPU Memory
GCN-GRU	20.59	13.4 s	7.1 ms	4.9 GiB
Graph WaveNet	2.34	11.8 s	5.9 ms	3.7 GiB
STAEformer	4.86	17.6 s	9.8 ms	6.2 GiB
ODMixer	3.72	14.9 s	7.4 ms	4.6 GiB
OD-BGCN-GRU	20.60	15.8 s	8.2 ms	5.6 GiB

OD-BGCN-GRU contains 20.60 million trainable parameters and requires 15.8 s per training epoch. Its inference latency and peak GPU memory consumption are 8.2 ms and 5.6 GiB, respectively.

Compared with GCN-GRU, OD-BGCN-GRU increases the training time per epoch, inference latency, and peak GPU memory by 17.91%, 15.49%, and 14.29%, respectively, while reducing MAE by 9.86%. The parameter count remains almost unchanged, increasing only from 20.59 M to 20.60 M.

These results indicate that the improvement in forecasting accuracy is accompanied by a moderate increase in computational cost.

4.7. Additional Temporal Validation Within Hangzhou

The primary experiment is based on one chronological partition of a single-city dataset. To examine whether the conclusions depend strongly on that specific split, additional rolling-origin experiments are conducted using different chronological evaluation periods within Hangzhou.

Table 18 reports the rolling-origin temporal validation results within Hangzhou Metro. For each fold, normalization parameters, demand groups, and spatial correlation groups are recalculated using only the corresponding training period.

Table 18. Rolling-origin temporal validation within Hangzhou Metro.

Fold	GCN-GRU MAE	ODMixer MAE	OD-PairStatic-GRU MAE	OD-BGCN-GRU MAE
Fold 1	0.793	0.742	0.736	0.719
Fold 2	0.808	0.753	0.746	0.728
Fold 3	0.784	0.731	0.727	0.708
Mean	0.795	0.742	0.736	0.718

Across the three chronological folds, OD-BGCN-GRU achieves MAEs of 0.719, 0.728, and 0.708, respectively, with an average MAE of 0.718. It consistently outperforms GCN-GRU, ODMixer, and OD-PairStatic-GRU across all three evaluation periods. Based on the average MAE, OD-BGCN-GRU reduces the error by 9.69%, 3.23%, and 2.45% relative to GCN-GRU, ODMixer, and OD-PairStatic-GRU, respectively.

These results indicate that the performance advantage of OD-BGCN-GRU remains stable across different chronological evaluation periods within the Hangzhou dataset.

5. Discussion and Conclusions

The experiments examine the heterogeneity and dynamic coordination of origin-side and destination-side spatial dependencies in metro OD passenger flow forecasting. The correlation-group results and single-branch comparisons show that the relative predictive roles of the two spatial dependencies differ across OD pairs. Furthermore, the learned fusion coefficients vary across time intervals, while OD-BGCN-GRU achieves better forecasting performance than OD-PairStatic-GRU. These results indicate that the relationship between origin-side and destination-side spatial information is not adequately described by a single fixed fusion pattern. Instead, allowing their relative contributions to vary with the current OD state provides additional predictive benefit beyond OD-specific static fusion.

Compared with recent OD forecasting methods such as ODMixer and ODformer, the main distinction of OD-BGCN-GRU lies in the explicit state-dependent coordination of the two spatial views rather than simply extracting origin-side and destination-side representations separately. The results under active-OD, nonzero-demand, and different demand-level settings further show that the observed performance improvement is not solely attributable to the large proportion of zero entries in the OD matrices. In addition,

the rolling-origin experiments show that the relative performance advantage is maintained across different chronological evaluation periods within the Hangzhou dataset. These results provide complementary evidence for the robustness of the proposed modeling strategy under OD sparsity and temporal variation.

The improvement in forecasting accuracy is accompanied by additional computational cost. The two spatial branches and the recurrent modeling of the complete $N \times N$ OD representation increase training time, inference latency, and memory consumption compared with simpler graph-based baselines. Nevertheless, the additional computational cost remains moderate under the current experimental setting.

Several limitations remain. First, the experiments are conducted on a single metro system over a relatively short observation period, and additional validation on longer periods and other metro networks is required to assess cross-city and long-term generalizability. Second, the current spatial graph is primarily constructed from physical network topology and does not explicitly incorporate functional similarity, passenger transfer relationships, or other data-driven spatial associations. Third, sparse intermittent demand and predictive uncertainty are not explicitly modeled in the current objective. Future work will therefore investigate more diverse datasets, richer spatial relationships, sparse demand-aware learning, probabilistic forecasting, and more scalable representations for large metro OD networks. The forecasting results reported in this study should also be distinguished from direct operational benefits, which would require further evaluation within specific scheduling, capacity-allocation, or passenger flow management applications.

Overall, this study demonstrates that explicitly coordinating origin-side and destination-side spatial information provides an effective modeling strategy for short-term metro OD passenger flow prediction. By combining bidirectional graph convolution with state-dependent dynamic fusion and temporal learning, OD-BGCN-GRU provides a unified framework for capturing both OD-pair heterogeneity and temporal variation in spatial dependence, offering a useful basis for further development of fine-grained metro OD forecasting methods.

Author Contributions: Conceptualization, C.C. and J.T.; methodology, J.T. and C.C.; software, J.T.; validation, J.T., H.L. and Z.H.; formal analysis, J.T. and H.L.; investigation, J.T., H.L. and Z.H.; data curation, J.T. and Z.H.; visualization, J.T.; writing—original draft preparation, J.T.; writing—review and editing, C.C., H.L., Z.H. and X.C.; supervision, C.C.; project administration, C.C. and X.C.; funding acquisition, C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Research and Development Project of Yibin Science and Technology Plan (No. 2025GY013). The authors also acknowledge the computing support provided by the High-Performance Computing Center of the Sichuan University of Science & Engineering.

Institutional Review Board Statement: Not applicable. This study used a previously released metro AFC dataset for secondary data analysis and did not involve direct interaction with human participants or the collection of additional personal information.

Informed Consent Statement: Not applicable. This study was based on previously released AFC records and did not involve direct recruitment or participation of human subjects.

Data Availability Statement: The AFC data used in this study were obtained from the Hangzhou Metro passenger flow dataset released through the Alibaba Tianchi platform. The processed OD data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript, generative AI tools were used solely for language refinement and LaTeX formatting assistance. Specifically, OpenAI ChatGPT (GPT-4o, OpenAI, San Francisco, CA, USA) was used for English language polishing. All scientific content, methodology, data analysis, results, and conclusions were reviewed and verified by the authors, who take full responsibility for the final manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AFC	Automatic Fare Collection
OD	Origin–Destination
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
ConvLSTM	Convolutional Long Short-Term Memory
GCN	Graph Convolutional Network
STGCN	Spatio-Temporal Graph Convolutional Network
DCRNN	Diffusion Convolutional Recurrent Neural Network
ASTGCN	Attention-Based Spatial–Temporal Graph Convolutional Network
STSGCN	Spatial–Temporal Synchronous Graph Convolutional Network
AGCRN	Adaptive Graph Convolutional Recurrent Network
MTGNN	Multivariate Time-Series Graph Neural Network
GMAN	Graph Multi-Attention Network
STAEformer	Spatio-Temporal Adaptive Embedding Transformer
STID	Spatial–Temporal Identity
CAS-CNN	Channel-Wise Attentive Split-Convolutional Neural Network
OD-BGCN-GRU	Origin–Destination Bidirectional GCN with Dynamic Fusion and GRU
ReLU	Rectified Linear Unit
MSE	Mean Squared Error
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
WMAPE	Weighted Mean Absolute Percentage Error
SD	Standard Deviation
IQR	Interquartile Range
CPU	Central Processing Unit
GPU	Graphics Processing Unit
RAM	Random-Access Memory
VRAM	Video Random-Access Memory
CUDA	Compute Unified Device Architecture
cuDNN	CUDA Deep Neural Network Library

References

1. Macioszek, E.; Kurek, A. Road Traffic Distribution on Public Holidays and Workdays on Selected Road Transport Network Elements. *Transp. Probl.* **2021**, *16*, 127–138. [CrossRef]
2. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
3. Cho, K.; van Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [CrossRef]
4. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.k.; Woo, W.c. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; Volume 28.
5. Chu, K.F.; Lam, A.Y.S.; Li, V.O.K. Deep Multi-Scale Convolutional LSTM Network for Travel Demand and Origin-Destination Predictions. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 3219–3232. [CrossRef]
6. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.

7. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
8. Yu, B.; Yin, H.; Zhu, Z. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 3634–3640. [CrossRef] [PubMed]
9. Li, Y.; Yu, R.; Shahabi, C.; Liu, Y. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
10. Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Zhang, C. Graph WaveNet for Deep Spatial-Temporal Graph Modeling. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 1907–1913. [CrossRef] [PubMed]
11. Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. *Aaai Conf. Artif. Intell.* **2019**, *33*, 922–929. [CrossRef]
12. Song, C.; Lin, Y.; Guo, S.; Wan, H. Spatial-Temporal Synchronous Graph Convolutional Networks: A New Framework for Spatial-Temporal Network Data Forecasting. *Aaai Conf. Artif. Intell.* **2020**, *34*, 914–921. [CrossRef]
13. Bai, L.; Yao, L.; Li, C.; Wang, X.; Wang, C. Adaptive Graph Convolutional Recurrent Network for Traffic Forecasting. In Proceedings of the Advances in Neural Information Processing Systems, Online, 6–12 December 2020; Volume 33.
14. Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Chang, X.; Zhang, C. Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, San Diego, CA, USA, 23–27 August 2020; pp. 753–763. [CrossRef]
15. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
16. Zheng, C.; Fan, X.; Wang, C.; Qi, J. GMAN: A Graph Multi-Attention Network for Traffic Prediction. *Aaai Conf. Artif. Intell.* **2020**, *34*, 1234–1241. [CrossRef]
17. Liu, H.; Dong, Z.; Jiang, R.; Deng, J.; Deng, J.; Chen, Q.; Song, X. Spatio-Temporal Adaptive Embedding Makes Vanilla Transformer SOTA for Traffic Forecasting. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, Birmingham, UK, 24–25 October 2023; pp. 4125–4129. [CrossRef]
18. Shao, Z.; Zhang, Z.; Wang, F.; Wei, W.; Xu, Y. Spatial-Temporal Identity: A Simple yet Effective Baseline for Multivariate Time Series Forecasting. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, 17–21 October 2022; pp. 4454–4458. [CrossRef]
19. Ke, J.; Qin, X.; Yang, H.; Zheng, Z.; Zhu, Z.; Ye, J. Predicting Origin-Destination Ride-Sourcing Demand with a Spatio-Temporal Encoder-Decoder Residual Multi-Graph Convolutional Network. *Transp. Res. Part Emerg. Technol.* **2021**, *122*, 102858. [CrossRef]
20. Zhang, J.; Che, H.; Chen, F.; Ma, W.; He, Z. Short-term origin-destination demand prediction in urban rail transit systems: A channel-wise attentive split-convolutional neural network method. *Transp. Res. Part Emerg. Technol.* **2021**, *124*, 102928. [CrossRef]
21. Liu, L.; Zhu, Y.; Li, G.; Wu, Z.; Bai, L.; Lin, L. Online Metro Origin-Destination Prediction via Heterogeneous Information Aggregation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 3574–3589. [CrossRef] [PubMed]
22. Ye, J.; Zhao, J.; Zheng, F.; Xu, C. Completion and Augmentation-Based Spatiotemporal Deep Learning Approach for Short-Term Metro Origin-Destination Matrix Prediction under Limited Observable Data. *Neural Comput. Appl.* **2023**, *35*, 3325–3341. [CrossRef]
23. Lv, S.; Wang, K.; Yang, H.; Wang, P. An origin–destination passenger flow prediction system based on convolutional neural network and passenger source-based attention mechanism. *Expert Syst. Appl.* **2024**, *238*, 121989. [CrossRef]
24. Xing, X.; Wang, B.; Ning, X.; Wang, G.; Tiwari, P. Short-term OD flow prediction for urban rail transit control: A multi-graph spatiotemporal fusion approach. *Inf. Fusion* **2025**, *118*, 102950. [CrossRef]
25. Zhang, S.; Zhang, J.; Yang, L.; Chen, F.; Li, S.; Gao, Z. Physics Guided Deep Learning-Based Model for Short-Term Origin–Destination Demand Prediction in Urban Rail Transit Systems Under Pandemic. *Engineering* **2024**, *41*, 276–296. [CrossRef]
26. Huang, B.; Ruan, K.; Yu, W.; Xiao, J.; Xie, R.; Huang, J. ODformer: Spatial–temporal transformers for long sequence Origin–Destination matrix forecasting against cross application scenario. *Expert Syst. Appl.* **2023**, *222*, 119835. [CrossRef]
27. Liu, Y.; Chen, B.; Zheng, Y.; Cheng, L.; Li, G.; Lin, L. ODMixer: Fine-Grained Spatial-Temporal MLP for Metro Origin-Destination Prediction. *IEEE Trans. Knowl. Data Eng.* **2025**, *37*, 5508–5522. [CrossRef]
28. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.