Geographical Cellular Traffic Prediction with Multivariate Spatio-Temporal Modeling

ChungYi Lin[1,2], Shen-Lung Tung[2] and Winston H. Hsu[1]

[1] *National Taiwan University,* [2] *Chunghwa Telecom laboratories*

Abstract

This paper presents a novel approach for evaluating road traffic usage using multi-type Geographical Cellular Traffic (GCT). Working with a major telecom company, we propose a new prediction task for transportation traffic using GCT data. To accurately tackle this task, we propose a model that effectively integrates multivariate relation exploration and spatio-temporal modeling across multiple regions. Furthermore, we develop a new core as the foundation of each modeling component, efficiently improving the incorporation of attention mechanisms in the CNN-based architecture. Extensive experiments demonstrate the superior performance of our model in successfully handling the prediction task and reveal the influence of various GCT combinations. It is worth noting that our proposed data and model can pave a new path for intelligent transportation systems and urban planning.

Keywords

Geographical Cellular Traffic, Multivariate Spatio-Temporal Modeling, Graph Neural Network

1. Introduction

Recently, traffic prediction has become increasingly important for intelligent transportation systems [1, 2]. Accurate traffic prediction can help alleviate traffic congestion [3] and optimize traffic signal control [4]. However, traditional traffic prediction approaches rely on dedicated sensors, which require costly maintenance and development, have limited deployment coverage, and are susceptible to insufficient usable information.

To tackle the limitations of traditional traffic prediction, we leverage large-scale and widely-distributed mobile user data integrated with road network information to analyze traffic usage conditions. Collaborating with Taiwan's major telecom provider, Chunghwa Telecom, we utilize geolocated cellular traffic, named *Geographical Cellular Traffic (GCT)*, which is further classified into vehicle, pedestrian, and stationary types. Accumulating GCT at fixed intervals offers insights into human activity patterns and road network usage, defined as GCT flow. Consequently, we propose a new task of forecasting specific vehicular GCT (V-GCT) flow in various regions, which is highly related to road traffic conditions and differs from predicting cellular traffic usage for mobile networks in previous studies [5, 6, 7, 8]. Hence, our proposed task and dataset using mobile user data offer new insights into road network usage and traffic conditions.

To address this new task, we propose a model with Multivariate, Temporal, and Spatial View Modelings that integrates multi-type GCT and utilizes spatio-temporal correlations to predict V-GCT flow accurately. Additionally, we present a novel core for each View Modeling, designed to enhance the efficiency of the attention mechanism when processing convolution-encoded representations. Our experiments demonstrate the superior performance of our model compared to representative and state-of-the-art baselines, underscoring the importance of incorporating multi-type GCT data. Overall, this work makes the following key contributions:

- **Novel data:** We collected over 30 million GCT records from diverse road segments, analyzing spatial correlations, relationships among GCT types, and their evolution over time.
- **Prediction task and model:** Our novel task predicts V-GCT, which provides valuable transportation insights for city authorities and is being employed in a proof-of-concept area. Meanwhile, our multivariate spatio-temporal model effectively captures dependencies and relationships between GCT types for accurate predictions.
- **Experiments and analysis:** Extensive evaluations demonstrate our model's superior performance against baselines for diverse prediction intervals. Ablation and sensitivity analyses highlight the importance of model components and GCT flow combinations for adaptability and real-world potential.

2. Data Processing

This section describes the definitions of geographical cellular traffic (GCT), data preprocessing, analysis, and potential applications.

STRL'23: Second International Workshop on Spatio-Temporal Reasoning and Learning, 21 August 2023, Macao, S.A.R

Table 1
Example of GCT with essential data fields.

IMEI*	Latitude	Longitude	Time	Type
gn...mE	24.78585	120.98825	09/25 17:59:58	vehicle
gn...GI	24.78601	120.98838	09/25 18:00:00	pedestrian
gn...mU	24.78608	120.98829	09/25 18:00:05	stationary
		...		

*IMEI numbers were hashed to protect privacy before processing.

2.1. Definitions

Multi-Type Geographical Cellular Traffic (GCT). GCT is cellular traffic with estimated GPS coordinates obtained from triangulation, indicating where the traffic was generated. Each GCT is classified[1] into three categories: vehicle, pedestrian, and stationary.

GCT Flow. We define GCT flow as the total quantity of GCT within a fixed interval (e.g., 5 min) as in previous vehicular flow studies [9]. With multi-type GCT, there are various GCT flows, including vehicle (V-GCT), pedestrian (P-GCT), and stationary (S-GCT) flows.

Road Segments. Road segments are defined as 20m x 20m areas, based on the average road size in our proof-of-concept (POC) area in Hsinchu, Taiwan. These segments geographically interconnect, forming a road network.

2.2. Data Collection and Preprocessing

Data sourcing. We extracted essential data fields from the telecom company's Geographical Cellular Traffic Storage Database to reduce storage and computational requirements, as shown in Table 1. Each row represents one GCT, comprising five data fields: International Mobile Station Equipment Identity (IMEI, a unique mobile phone identifier), latitude and longitude coordinates, recording time, and GCT type.

Road Segment Selection for GCT Collection. We selected geographically connected road segments as the scope for GCT data collection to capture the mobility of mobile users across different areas, as shown in Figure 1. We collaborated with transportation authorities to identify 21 road segments for analysis, each with unique functional locations nearby (e.g., universities, science parks, and shopping areas). To ensure relevance to road usage conditions, we extracted GCTs located within these road segments and included them in Table 1.

Preprocessing for GCT Flow. We preprocessed the data to ensure accuracy by eliminating inconsistencies, outliers, and missing values. Next, we removed duplicate GCT record with identical IMEI and timestamps to prevent distortion and ensure reliability. Finally, we calculated the total GCTs for every 5-minute interval

[1]The algorithm is the telecom company's confidential trade secret.

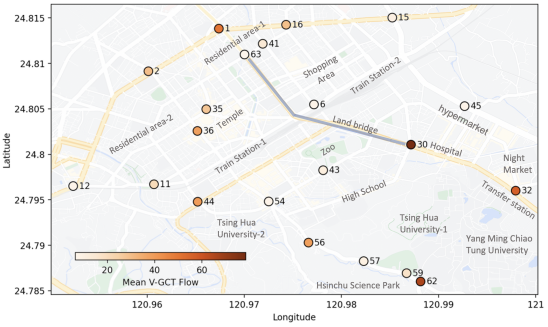


Figure 1: The distribution of road segments in the POC area, with their average V-GCT flows influenced by the functionality of nearby regions.

to maintain consistency, and assigned unique V-GCT, P-GCT, and S-GCT flows to each road segment.

Data Privacy Protection. In compliance with strict personal data protection laws, we established a collaboration agreement with the company to outline data sharing terms and ensure adherence to privacy regulations. We processed GCT data in a secure intranet environment and hashed IMEI numbers to protect the privacy users.

2.3. Data Analysis

2.3.1. Time Evolving Spatial Correlations

Spatial Correlation. Building on the approach used in [10, 11], we utilized the Pearson correlation coefficient to assess the spatial correlation between road segments in the POC area. Specifically, when using the historical one-hour V-GCT as the series variable for each segment in Figure 2, a Pearson correlation coefficient is assigned to each road segment pair, ranging from -1 to 1. Road segments in closer proximity tend to exhibit similar V-GCT flow patterns, leading to higher Pearson correlation coefficients. This highlights the spatial correlation between road segments based on V-GCT.

Time Evolving Correlation. We use the Pearson correlation coefficient to explore spatial correlations over time, as shown in Figure 2. Notable observations include:

- At 18:00, road segments near the Hsinchu Science Park (ID: 57, 59, 62) exhibit high Pearson coefficients, indicating similar movement patterns as commuting users leave the workspace.

- At 19:00, highly correlated regions emerge along the route (ID: 44, 35, 6, 30, 43, 45) from the work area to residential areas.

- By 20:00, users gradually return home or dine out, resulting in high Pearson coefficients for road segments near residential-commercial mixed areas (ID: 1, 16, 41, 43) reflecting similar V-GCT patterns. Road segments near the Hsinchu Science Park (ID: 54, 56, 57, 59, 62)

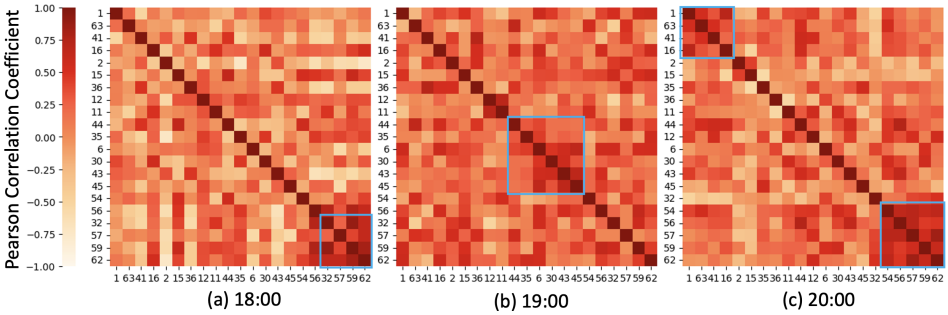


Figure 2: Pearson Correlation Coefficients are calculated between pairs of road segments based on their V-GCT flows from the previous hour. As traffic conditions evolve over time, areas with high Pearson coefficients (highlighted with a blue frame) shift among different segments, reflecting the movement of mobile users with vehicle drivers' properties.

also exhibit high Pearson coefficients again, as users who finish work later leave the area.

Overall, the time-evolving Pearson correlation analysis not only reveals spatial correlations between road segments by V-GCT flow but also indicates changes in population activity patterns during different periods. Concentrated regions with high Pearson coefficients may shift over time, providing a new insight for understanding user flow and identifying congestion points in traffic management over time.

2.3.2. Understanding Regional Functionality through Multi-Type GCT Flows

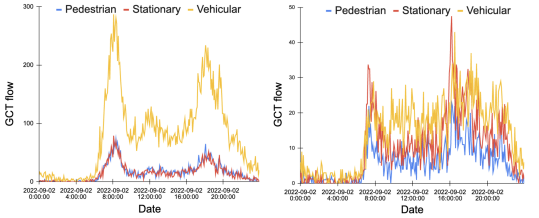
Recognizing regional functionality can aid in prediction tasks [12], but traditionally involves time-consuming manual labeling. By analyzing variations between three types of GCT flows, we can uncover hidden interactions of user groups in different urban areas.

Implicit Interactions among Multi-Type GCT Flows. Figure 3(a) and 3(b) reveal distinct patterns in different urban areas, with commuting areas displaying dominant V-GCT flow and residential-commercial areas showing significant P-GCT and S-GCT flows. Inspired by [13], we subtracted V-GCT from P-GCT and S-GCT to obtain (V-P)-GCT and (V-S)-GCT, respectively, using one day of data, and calculated the Pearson correlation coefficient between V-GCT and these flows. The subtraction types, capturing relative differences between the GCT flows, exhibit higher correlations with V-GCT, highlighting distinct patterns with unique characteristics in different areas. For instance, a high (V-P)-GCT value may indicate a region with more vehicular traffic. These relative differences better capture GCT flow interactions.

Deriving Insights into Model Design. The higher correlations for subtraction types in multi-type GCT flows provide valuable insights for our model. Incorporating these relative differences improves capturing interactions, and distinguishing area functionality.

ID:62	P	S	V-P	V-S
V	0.78	0.75	0.98	0.99

ID:41	P	S	V-P	V-S
V	0.56	0.61	0.91	0.92



(a) Near commuting route (ID:62). (b) Near residential-commercial area (ID:41).

Figure 3: Functional areas exhibit varying multi-type GCT flow patterns, with V-GCT flow prominent in the commuting road (ID: 62), while P-GCT and S-GCT flows are relatively higher in residential-commercial areas (ID:41). To explore the latent relationships among multi-type GCT flows, we subtracted V-GCT from P-GCT (V-P) and S-GCT (V-S) and calculated the Pearson correlation between V-GCT and these GCT types over one day, as shown in the table above. The subtracted GCT types (V-P and V-S) flows, which exhibit higher Pearson correlation coefficients with V-GCT, show more robust correlations than P-GCT and S-GCT, revealing implicit correlations.

2.4. Potential Applications

Multi-type GCT flows provide new insights for urban planning, with possible future applications including:

Transportation Management. GCT flow assists authorities in developing effective traffic strategies, improving flow and reducing travel times.

Public Safety. Real-time monitoring systems using GCT flow aid in understanding crowd density and mobility, ensuring safety during critical incidents.

Urban Planning. Analyzing GCT flow helps planners identify infrastructure needs and optimize city layouts to accommodate growing populations.

3. Multi-View Modeling for V-GCT Prediction

3.1. Definition of Prediction Task

Given N road segments, we collect multi-type GCT flows at fixed intervals. Each historical GCT flow is represented by $X_f = \{X_f^1, X_f^1, ..., X_f^{T_{in}}\}$, where f corresponds to V-GCT, S-GCT, or P-GCT, and $X_f^t \in \mathbb{R}^N$ denotes the values at time step t. Our objective is to predict V-GCT for all segments over the next T_{out} steps using one or multiple GCT flows X_f from the past T_{in} steps. We denote the predicted values as $\{Y^{T_{in}+1}, Y^{T_{in}+2}, \ldots, Y^{T_{in}+T_{out}}\}$, where $Y^{T_{in}+i} \in \mathbb{R}^N$.

3.2. Overview of the Proposed Model

As shown in Figure 4, our model is designed to explore the implicit relationships within complex multivariate spatio-temporal inputs from various view modeling perspectives, with each module presented as follows:

Multivariate View Modeling. This component explores hidden relationships among multivariate features. With a verified robust correlation between V-GCT and the relative difference of other GCT types, this module delves into their relationship, incorporating implicit road usage patterns for enhancing prediction capabilities.

Temporal and Spatial View Modelings. These components explore the relationships among time steps and road segments, forming an integrated TS-Module. Stacked TS-Modules are executed sequentially to capture dependencies across different time scales.

Skip connection and Output module. The skip connection connects the output of temporal view modeling, linking them to the output module. The output module is composed of multiple layers of MLP [14], transforming the connected outputs into multi-step predictions.

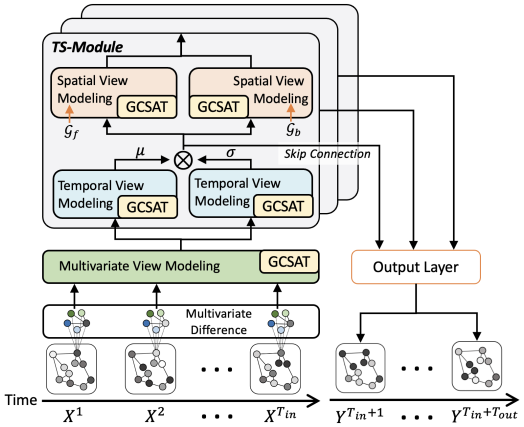


Figure 4: Proposed Multi-View Modeling Architecture. Multivariate View Modeling uncovers hidden functionality by examining interconnections among V-GCT and variations between other GCT flows. A TS-Module comprises of Temporal and Spatial View Modelings, operating on representations for spatio-temporal modeling. Based on GCSAT, the three components are enhanced in terms of attention mechanism efficiency for handling CNN-encoded representations. Each temporal view model links a skip connection to the output layer for multi-step prediction.

3.3. Graph Channel-Specific Attention (GCSAT)

We present a novel core, GCSAT, as the fundamental basis for the multivariate, spatial, and temporal modeling components. This core aims to improve the efficiency of attention mechanisms in handling the multi-channel representation after CNN encoding.

Preliminary. Graph Convolutional Network(GCN)-based models [15, 14, 16, 17] use a 1D CNN to encode input into latent representations in the form of $[C, N, T]$, where C denotes channels, N represents spatial nodes, and T signifies historical observation time steps. While GCN-based methods yield promising spatio-temporal prediction results, they assign equal weights to neighboring nodes, causing suboptimal performance [18]. Integrating Graph Attention Network (GAT) [19] into GCN-based methods enables dynamic weight learning between nodes. However, GAT overlooks channel-specific information, crucial for channels with varying importance or contributions to prediction tasks.

Channel-Specific Attention. To account for channel-specific weights, we use C independent GATs to examine correlations between each channel's attribute nodes. For the c-th channel representation, $H^c = \{h_1^c, h_2^c, \ldots, h_Z^c\} \in \mathbb{R}^{Z \times D}$, where Z is the number of attribute node (e.g., road segments or time steps) and D is node's dimension. We use the *attention coefficient* in GAT [19], $e(h_i^c, h_j^c)$, to determine node j's importance to node i.

We normalize these coefficients across all neighbors of node i based on $\mathcal{G}$ using the *attention function*: $\alpha_{ij}^c = softmax(e(h_i^c, h_j^c))$. We then compute a weighted sum of node i's features and its neighbors and concatenate the results of C independent attention mechanisms:

$$\hat{H}_i = \|_{c=1}^{C}(\sigma(\sum_{j \in N_i} \alpha_{ij}^c h_j^c)), \quad (1)$$

where $i \in \{1, 2, \ldots, Z\}$, and $\sigma(\cdot)$, and $\sigma(\cdot)$ is a nonlinear function. The final output is the concatenation of the aggregated representations $\hat{H}_i$ with the addition of a residual connection. We denote the function of GCSAT:

$$GCSAT(H, \mathcal{G}) \coloneqq \{\hat{H}_1, \hat{H}_2, ..., \hat{H}_Z\} + H, \quad (2)$$

where adding the multichannel representation H is considered a residual connection.

3.4. Multivariate View Modeling

The goal of Multivariate View Modeling is to investigate the relationships among multi-type GCT flows before the TS-Module, and extract implicit relations to enhance V-GCT prediction. Although existing GAT-based models have shown improved task accuracy by exploring feature relationships [20, 21], attention mechanisms may not fully utilize the potential differences within various features during a limited number of epochs [13].

Deriving Insights from Multi-type GCT Flows. We have verified that there is a more robust correlation between V-GCT and the relative differences of P-GCT and S-GCT, as shown by the Pearson coefficients displayed in Figure 3. Inspired by anomaly detection [20, 13, 22], we propose utilizing the magnitude differences between V-GCT flow and P-GCT and S-GCT flows to gain deeper insights into inherent regional functionality and mobility user activity patterns. By learning from these differences, we can effectively extract hidden relationships within multi-type GCT and address the inefficiency of directly modeling features.

Difference Representation. All multi-type GCT features (V-GCT, P-GCT, and S-GCT) are encoded by a 1D CNN into shape $[C, N, T]$. To prevent aggregation of unrelated information that might impact later training results, we process the features at each time step individually. At time step t, the difference between V-GCT and different GCT flows is calculated as: $\Delta_t^f = H_t^{V-GCT} - H_t^f$, where f represents P-GCT or S-GCT. Thus, the difference representation at time t is constituted as follows: $H_t^{\Delta} = \{H_t^{V-GCT}, \Delta_t^{P-GCT}, \Delta_t^{S-GCT}\} \in R^{C\times 3\times N}$, where H_t^{V-GCT} can be represented as: $H_t^{\Delta}[:, 0, :]$.

GCSAT for Multivariate Difference Modeling. After establishing the difference representation, H_t^{Δ} is applied to GCSAT in Equation 2, as follows:

$$\hat{H_t^{\Delta}} = GCSAT(H_t^{\Delta}, \mathcal{G}), \tag{3}$$

where each element in H_t^{Δ} is considered as the feature node, $\mathcal{G}$ is a complete graph, and GCSAT is stacked with two layers for more detail extraction.

For the output of GCSAT at each time step, we only extract the first node of $\hat{H_t^{\Delta}}$, $\hat{H_t^{\Delta}}[:, 0, :]$, which is the aggregated V-GCT representation with a shape $[C, 1, N]$. Then, we concatenate the outputs for each time step in Equation 3 along the second dimension, resulting in a shape $[C, T, N]$:

$$H_{mout} = \|_{t=1}^{T}(\hat{H_t^{\Delta}}[:, 0, :]), \tag{4}$$

By capturing complex interactions and relationships between GCT flows, our approach can achieve more accurate V-GCT flow predictions. The output is then forwarded to the next Temporal View Modeling.

3.5. Temporal View Modeling

Temporal View Modeling, as shown in the TS-Module of Figure 4, converts the multi-channel representation into the shape $[C, T, S]$ for GCSAT processing. GCSAT treats input as T nodes with S dimensions, and $\mathcal{G}$ in Equation 2 represents a complete graph.

Time series data in practical scenarios often exhibit both short-term and long-term dependencies. Simultaneously capturing these patterns using attention among time nodes is challenging due to entangled dependencies, making it difficult to identify valuable signals [23, 17].

Extracting Different Time Scales To address the above issue, we adopt two kernels with different sizes, inspired by [14, 17], to extract short-term and long-term temporal patterns. Applying two 2D CNNs with kernel sizes (2,1) and (5,1) to the multi-channel representation H with shape $[C, T, N]$ produces outputs H_2 with shape $[C, (T-1), N]$ and H_5 with shape $[C, (T-4), N]$, respectively. The (2,1) kernel uncovers local temporal relationships, revealing short-term patterns and dependencies crucial for understanding rapid changes. In contrast, the (5,1) kernel captures longer-range temporal relationships, exposing hidden longer-term trends and dependencies by encompassing a broader context of time steps.

GCSAT for Temporal Modeling. After extracting different scale representations H_2 and H_5, we use GCSAT to explore temporal dependencies among them, feeding them into Equation 2 as follows:

$$H_{cat} = GCSAT(H_2, \mathcal{G}) + GCSAT(H_5, \mathcal{G}), \tag{5}$$

where the outputs for different time scale representations in Equation 5 are truncated to match the temporal nodes of the representation with the largest kernel size and then concatenated accordingly.

Gating Mechanism. Leveraging the gating mechanism's benefits in [15, 17], which controls the amount of information passed to the next module, we process the outputs of two Equation 5 separately with distinct activation functions and perform element-wise multiplication:

$$H_{tout} = \sigma(H_{cat}^1) \odot \mu(H_{cat}^2), \tag{6}$$

where H_{cat}^1 and H_{cat}^2 are separately generated in Equation 5, σ denotes the sigmoid function, μ denotes the tangent hyperbolic function, and $\odot$ represents the Hadamard product. H_{out} is the output of the Gating Mechanism and will be fed into the next modeling.

3.6. Spatial View Modeling

As the analysis in Figure 2 demonstrates the existence of spatial correlations between V-GCT flows among road segments. Thus, it is crucial to explore the relationships between road segments. Spatial View Modeling, depicted

in the TS-Module of Figure 4, aims to model these relationships to improve our understanding and prediction of V-GCT flow patterns.

GCSATs for modeling bidirectional flow. Leveraging GCSAT's flexibility, we extract spatial correlations among road segments. First, we transform the representation H from the previous temporal modeling output into a shape of [C,S,T], with S road segments as nodes and T features.

To account for bidirectional V-GCT flow among road segments, we utilize two different GCSATs to explore propagations in both directions. The adjacency matrix A of road segments is constructed using road network distance and a thresholded Gaussian kernel [24]. Following [15], we define the forward transition matrix $\mathcal{G}_f = A/rowsum(A)$ and the backward transition matrix $\mathcal{G}_b = A^T/rowsum(A^T)$. After inputting the two transition matrices into their respective GCSATs, we combine their outputs to obtain the final output H_{sout}, as:

$$H_{sout} = GCSAT(H, \mathcal{G}_f) + GCSAT(H, \mathcal{G}_b) \quad (7)$$

This approach allows us to capture spatial relationships among road segments while considering their bidirectional connections. Incorporating spatial correlation information into our modeling process enables us to better explore dependencies among road segments.

4. Experiments

4.1. Experimental Settings and Baselines

Evaluation Metrics. We use Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Train/Valid/Test data processing. Each type of GCT flow was processed in 5-minute intervals, from August 28, 2022, to September 29, 2022, across 21 road segments. We followed [25], splitting data 70%-20%-10% for training, testing, and validation. Each sequence sample had 24 time steps; the first 12 (T_{in}) as historical input and the remaining 12 (T_{out}) as ground truth.

Baselines. We have selected seven representative traffic prediction baselines for this task. These baselines are categorized as follows, with overviews in Appendix A:

- Temporal Convolution (TCN) [26]
- GCN-based models: Graph Wavenet (GWNet) [15], MTGNN [14], DMGCN [27]
- Attention-based models: Gman [28], MPNet [21]
- State-of-the-art GNN model: ESG [17]

Model Settings. We followed [15] by repeating each model 10 times and reporting the average of the metrics. Our proposed model consists of one Multivariate View Modeling component and three stacked TS-Modules, with each module containing a Temporal View Modeling and a Spatial View Modeling component. We also followed [14] in using skip connection layers and the output module.

4.2. V-GCT Prediction Evaluation

We evaluated our model and various baselines for predicting future V-GCT flows at 15 (3 steps), 30 (6 steps), and 60-minute (12 steps) intervals, and the results are shown in Table 2, including the average MAE, RMSE, and MAPE over 10 repetitions for each method. Our observations are as follows:

Performance comparison across prediction steps. Our model consistently outperformed various baselines, including TCN-based [26], GCN-based [15, 14, 27, 17], and attention-based [28, 21] models, across all prediction steps. This demonstrates our model's superior ability to capture the underlying multivariate relationships and complex spatio-temporal patterns. Although performance decreased as prediction steps increased for all models, our model maintained its superior performance compared to the baselines, even at longer prediction steps.

Performance and impact of multi-type GCT flow. Table 2 shows that attention-based methods (^{2}) outperform both TCN-based (^{0}) and GCN-based (^{1}) methods, highlighting the effectiveness of attention mechanisms in capturing complex relationships between road segments and GCT flows. Moreover, our model further enhances prediction accuracy by effectively capturing human activity patterns and road network usage through the exploration of hidden relationships among multi-type GCT flows. The combination of attention-based mechanisms and multi-type GCT flows improves the model's ability to understand and forecast complex flow patterns.

4.3. Ablation Study of Proposed Model

We conducted an ablation study to assess the impact of our model's components on traffic prediction tasks. Average prediction metrics were calculated for prediction steps 1 (5 min.) through 12 (60 min.). Table 3 compares the full model with three ablated versions: without Spatial View Modeling (w/o S), without Temporal View Modeling (w/o T), and without Multivariate View Modeling (w/o M). Our observations are as follows:

Impact of w/o S. Omitting Spatial View Modeling had the most significant impact on the model's performance, resulting in a decrease in all metrics. This observation emphasizes the high spatial correlations between road segments in V-GCT and suggests that Spatial View Modeling effectively captures these dependencies, leading to improved performance.

Impact of w/o M. The model without Multivariate View Modeling exhibited the second-lowest performance compared to the full model, indicating the importance of

Table 2
Performance comparisons for short-term to long-term V-GCT predictions

Baselines	15 min.			30 min.			60 min.		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
TCN[0]	5.55±0.02	8.82±0.04	34.5%±0.6	5.74±0.02	9.38±0.06	36.9%±0.9	6.58±0.05	11.22±0.13	38.7%±1.4
GWNet[1]	5.46±0.01	8.72±0.04	32.5%±1.3	5.62±0.03	9.08±0.11	32.9%±1.3	6.08±0.06	10.31±0.19	34.6%±1.6
Gman[2]	5.37±0.01	8.61±0.05	32.6%±1.7	5.51±0.04	8.99±0.14	34.1%±1.2	5.77±0.02	9.68±0.11	34.4%±1.0
MTGNN[1]	5.29±0.02	8.52±0.03	32.2%±1.4	5.45±0.01	8.86±0.01	34.2%±1.6	5.74±0.03	9.66±0.13	35.3%±1.9
MPNet[2]	5.30±0.01	8.46±0.04	32.9%±2.0	5.44±0.03	8.84±0.09	34.7%±1.6	5.73±0.03	9.68±0.06	34.8%±2.1
DMGCN[1]	5.28±0.04	8.48±0.11	31.8%±1.7	5.46±0.01	8.81±0.03	33.6%±1.9	5.82±0.02	9.56±0.08	34.6%±1.1
ESG[1]	5.26± 0.03	8.43±0.03	31.1%±1.1	5.40±0.02	8.76±0.06	31.8%±1.3	5.65±0.02	9.46±0.09	32.9%±1.9
Our[2]	**5.23**±0.01	**8.27**±0.06	**29.8%**±0.7	**5.33**±0.02	**8.54**±0.06	**30.5%**±0.8	**5.54**±0.03	**9.25**±0.07	**31.8%**±0.7

[0] denotes the TCN-based methods, [1] denotes the GCN-based methods, and [2] denotes the attention-based methods.

Table 3
The average predicted metrics for steps 1 through 12

Baselines	MAE	RMSE	MAPE
w/o S	5.47±0.04	8.84±0.09	32.4%±0.4
w/o M	5.45±0.03	8.76±0.02	31.7%±0.5
w/o T	5.41±0.03	8.75±0.09	31.6%±0.4
Full	**5.37**±0.07	**8.66**±0.02	**31.3**±0.5

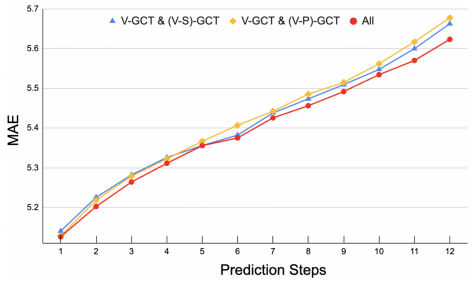


Figure 5: The combinations of V-GCT with all subtraction types consistently yields the lowest prediction error for multi-step predictions.

accounting for the differences in magnitude between pedestrian, vehicular, and stationary GCT flows to make accurate traffic predictions.

Impact of w/o T. The model's performance significantly decreased when Temporal View Modeling was removed, as evidenced by lower metrics across all prediction steps. This result highlights the crucial role of Temporal View Modeling in capturing temporal patterns and dependencies, which contribute to enhanced prediction accuracy.

The complete model consistently outperforms ablated versions, emphasizing the significance of Spatial, Temporal, and Multivariate View Modeling. Each component plays a critical role, and their integration leads to enhanced V-GCT flow predictions.

4.4. Sensitivity Analysis of Multi-Type GCT Feature Combinations

Our model can explore the relationships between multivariate features, namely V-GCT, (V-S)-GCT, and (V-P)-GCT. Thus, we conducted a parameter sensitivity analysis for different combinations of V-GCT and subtraction types to assess their impact on prediction performance. Figure 5 shows the MAE for three feature combination models across 12 prediction steps: V-GCT with (V-S)-GCT, V-GCT with (V-P)-GCT, and a combination of all types including V-GCT, (V-P)-GCT, and (V-S)-GCT. Key observations are:

Impact on Prediction Step: For prediction steps 1 to 4, the model with V-GCT and (V-P)-GCT outperforms the one with V-GCT and (V-S)-GCT. Beyond step 4, the V-GCT and (V-S)-GCT model performs better, indicating that the relative differences between vehicle and pedestrian or stationary GCT flows may vary in importance with the prediction horizon.

Performance of model with V-GCT and all subtraction types. The model incorporating V-GCT and all subtraction types consistently achieves the lowest MAE error over all prediction steps, demonstrating the effectiveness of improving prediction by exploring hidden relative differences among multi-type GCT flows, and our model's capability in modeling complex relations.

5. Conclusion

We proposed and analyzed a multi-type GCT approach that overcomes limitations in traffic prediction. Our predictive model effectively combined multivariate spatio-temporal modeling for V-GCT prediction across multiple road segments, outperforming the baselines. Our experiments highlighted the importance of model components and GCT flow combinations for prediction accuracy. By initially validating the effectiveness of predicting V-GCT, we can further explore other types of GCT predictive results, offering potential applications for improving intelligent transportation.

References

[1] F. Zhu, Y. Lv, Y. Chen, X. Wang, G. Xiong, F.-Y. Wang, Parallel transportation systems: Toward iot-enabled smart urban traffic control and management, IEEE Transactions on Intelligent Transportation Systems 21 (2019).

[2] G. Dong, M. Tang, Z. Wang, J. Gao, S. Guo, L. Cai, R. Gutierrez, B. Campbel, L. E. Barnes, M. Boukhechba, Graph neural networks in iot: A survey, ACM Transactions on Sensor Networks (2023).

[3] Z. Lv, Y. Li, H. Feng, H. Lv, Deep learning for security in digital twins of cooperative intelligent transportation systems, IEEE Transactions on Intelligent Transportation Systems (2021).

[4] P. Xie, T. Li, J. Liu, S. Du, X. Yang, J. Zhang, Urban flow prediction from spatiotemporal data using machine learning: A survey, Information Fusion (2020).

[5] G. Barlacchi, M. De Nadai, R. Larcher, A. Casella, C. Chitic, G. Torrisi, F. Antonelli, A. Vespignani, A. Pentland, B. Lepri, A multi-source dataset of urban life in the city of milan and the province of trentino, Scientific data (2015).

[6] X. Wang, Z. Zhou, F. Xiao, K. Xing, Z. Yang, Y. Liu, C. Peng, Spatio-temporal analysis and prediction of cellular traffic in metropolis, IEEE Transactions on Mobile Computing (2018).

[7] N. Zhao, A. Wu, Y. Pei, Y.-C. Liang, D. Niyato, Spatial-temporal aggregation graph convolution network for efficient mobile cellular traffic prediction, IEEE Communications Letters (2021).

[8] Y. Yao, B. Gu, Z. Su, M. Guizani, Mvstgn: A multi-view spatial-temporal graph network for cellular traffic prediction, IEEE Transactions on Mobile Computing (2021).

[9] S. Guo, Y. Lin, N. Feng, C. Song, H. Wan, Attention based spatial-temporal graph convolutional networks for traffic flow forecasting, in: Proc. of AAAI, 2019.

[10] C. Zhang, H. Zhang, D. Yuan, M. Zhang, Citywide cellular traffic prediction based on densely connected convolutional neural networks, IEEE Communications Letters (2018).

[11] C. Wang, Y. Zhu, T. Zang, H. Liu, J. Yu, Modeling inter-station relationships with attentive temporal graph convolutional network for air quality prediction, in: Proc. of WSDM, 2021, pp. 616–634.

[12] W. Shao, Z. Jin, S. Wang, Y. Kang, X. Xiao, H. Menouar, Z. Zhang, J. Zhang, F. Salim, Long-term spatio-temporal forecasting via dynamic multiple-graph attention, Proc. of IJCAI, 2022.

[13] A. Deng, B. Hooi, Graph neural network-based anomaly detection in multivariate time series, in: Proc. of AAAI, volume 35, 2021, pp. 4027–4035.

[14] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, C. Zhang, Connecting the dots: Multivariate time series forecasting with graph neural networks, in: Proc. of KDD, 2020.

[15] Z. Wu, S. Pan, G. Long, J. Jiang, C. Zhang, Graph wavenet for deep spatial-temporal graph modeling., in: Proc. of IJCAI, 2019.

[16] C. Tian, W. K. Chan, Spatial-temporal attention wavenet: A deep learning framework for traffic prediction considering spatial-temporal dependencies, IET Intelligent Transport Systems (2021).

[17] J. Ye, Z. Liu, B. Du, L. Sun, W. Li, Y. Fu, H. Xiong, Learning the evolutionary and multi-scale graph structure for multivariate time series forecasting, in: Proc. of KDD, 2022, pp. 2296–2306.

[18] S. Brody, U. Alon, E. Yahav, How attentive are graph attention networks?, in: International Conference on Learning Representations, 2022.

[19] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, in: Proc. of ICLR, 2018.

[20] H. Zhao, Y. Wang, J. Duan, C. Huang, D. Cao, Y. Tong, B. Xu, J. Bai, J. Tong, Q. Zhang, Multivariate time-series anomaly detection via graph attention network, in: Proc. of ICDM, 2020, pp. 841–850.

[21] C.-Y. Lin, H.-T. Su, S.-L. Tung, W. H. Hsu, Multivariate and propagation graph attention network for spatial-temporal prediction with outdoor cellular traffic, in: Proc. of CIKM, 2021.

[22] S. Kim, K. Choi, H.-S. Choi, B. Lee, S. Yoon, Towards a rigorous evaluation of time-series anomaly detection, in: Proc. of AAAI, 2022.

[23] J. Xu, R. Rahmatizadeh, L. Bölöni, D. Turgut, Real-time prediction of taxi demand using recurrent neural networks, IEEE Transactions on Intelligent Transportation Systems 19 (2017) 2572–2581.

[24] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, P. Vandergheynst, The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains, IEEE signal processing magazine (2013).

[25] Y. Li, R. Yu, C. Shahabi, Y. Liu, Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, in: Proc. of ICLR, 2018.

[26] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, arXiv preprint arXiv:1511.07122 (2015).

[27] L. Han, B. Du, L. Sun, Y. Fu, Y. Lv, H. Xiong, Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting, in: Proc. of KDD, 2021.

[28] C. Zheng, X. Fan, C. Wang, J. Qi, Gman: A graph multi-attention network for traffic prediction, in: Proc. of AAAI, 2020.

A. APPENDIX: Overview of baselines

We evaluate our proposed model with the following baselines:

- TCN [26]: A convolutional-based method for time seier modeling.
- GWNet (Graph WaveNet) [15]: A graph-based Wavenet architecture with a spatial diffusion mechanism.
- MTGNN [14]: A graph-based convolutional model with dynamically learned graph structures.
- DMGCN [27]: A model incorporating time-specific spatial dependencies, dynamic graph convolution, and multi-faceted fusion.
- MPNet [21]: A GNN model with propagation attention mechimism.
- Gman [28]: A graph multi-attention model utilizing an encoder-decoder architecture.
- ESG [17]: A model for capturing interactions in time series using evolutionary graph structures.