

Irregular Traffic Time Series Forecasting Based on Asynchronous Spatio-Temporal Graph Convolutional Networks

Weijia Zhang
HKUST(GZ)
Guangzhou, China
vegazhang3@gmail.com

Le Zhang
Baidu Research
Beijing, China
zhangle0202@gmail.com

Jindong Han
HKUST
Hong Kong, China
jhanao@connect.ust.hk

Hao Liu*
HKUST(GZ) & HKUST
Guangzhou, China
liuh@ust.hk

Yanjie Fu
Arizona State University
Phoenix, United States
yanjie.fu@asu.edu

Jingbo Zhou
Baidu Research
Beijing, China
zhoujingbo@baidu.com

Yu Mei
Baidu Inc.
Beijing, China
whqyqy@hotmail.com

Hui Xiong*
HKUST(GZ) & HKUST
Guangzhou, China
xionghui@ust.hk

ABSTRACT

Accurate traffic forecasting is crucial for the development of Intelligent Transportation Systems (ITS), playing a pivotal role in modern urban traffic management. Traditional forecasting methods, however, struggle with the irregular traffic time series resulting from adaptive traffic signal controls, presenting challenges in asynchronous spatial dependency, irregular temporal dependency, and predicting variable-length sequences. To this end, we propose an Asynchronous Spatio-tEmporal graph convolutional nEtwoRk (ASEER) tailored for irregular traffic time series forecasting. Specifically, we first propose an Asynchronous Graph Diffusion Network to capture the spatial dependency between asynchronously measured traffic states regulated by adaptive traffic signals. After that, to capture the temporal dependency within irregular traffic state sequences, a personalized time encoding is devised to embed the continuous time signals. Then, we propose a Transformable Time-aware Convolution Network, which adapts meta-filters for time-aware convolution on the sequences with inconsistent temporal flow. Additionally, a Semi-Autoregressive Prediction Network, comprising a state evolution unit and a semi-autoregressive predictor, is designed to predict variable-length traffic sequences effectively and efficiently. Extensive experiments on a newly established benchmark demonstrate the superiority of ASEER compared with twelve competitive baselines across six metrics.¹

*Corresponding author.

¹This project is available at <https://github.com/usail-hkust/ASEer>.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '24, August 25–29, 2024, Barcelona, Spain.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0490-1/24/08

<https://doi.org/10.1145/3637528.3671665>

CCS CONCEPTS

• Information systems → Spatial-temporal systems.

KEYWORDS

Traffic forecasting; irregular time series analysis; convolutional networks; spatio-temporal modeling

ACM Reference Format:

Weijia Zhang, Le Zhang, Jindong Han, Hao Liu, Yanjie Fu, Jingbo Zhou, Yu Mei, and Hui Xiong. 2024. Irregular Traffic Time Series Forecasting Based on Asynchronous Spatio-Temporal Graph Convolutional Networks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), August 25–29, 2024, Barcelona, Spain*. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3637528.3671665>

1 INTRODUCTION

Recent years have witnessed significant advancements in traffic forecasting, which plays a pivotal role in underpinning Intelligent Transportation Systems (ITS) [20], facilitating emergency response and management [29], and is integral to the development of autonomous driving [18]. In particular, timely and accurate traffic forecasting is of great importance to help the Intelligent Traffic Signal Control Systems (ITSCS) to anticipate future traffic state variations, thereby providing crucial insights to support the systematic analysis, informed decisions, and optimal control optimization of ITSCS to enhance the overall transportation system efficiency [40].

In practice, the traffic dynamics of the road network is jointly decided by the vehicles on the road and the intervention of traffic signals, e.g., intersection traffic lights, ramp metering lights, and lane allocation signals [5, 36, 39, 41]. On the one hand, the traffic signal adaptively adjusts its control cycles in response to real-time traffic flow variations [41]. On the other hand, traffic flows are dynamically regulated by these adaptive signal control strategies with varying cycle lengths. As a result, the urban traffic states, entangling both length-varying traffic signal cycles and the corresponding traffic flows, exhibit significant irregularity and render more complex traffic dynamics, as depicted in Figure 1.

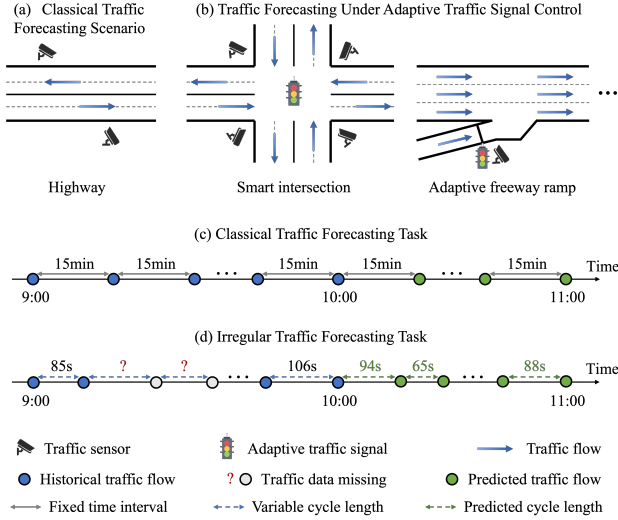


Figure 1: The distinction between classical traffic forecasting on the highway and irregular traffic forecasting under adaptive traffic signal control. The forecasting task aims to predict future traffic variations (10:00-11:00) based on past observations (9:00-10:00).

However, existing studies [16, 22, 44, 47] on urban traffic forecasting primarily focus on capturing spatiotemporal dependencies among geo-distributed time series with evenly spaced temporal variables, largely overlooking the irregularity of traffic time series induced by the interplay between traffic dynamics and adaptive traffic control policies. These approaches model spatiotemporal dependencies and forecast traffic variations within fixed time intervals, which are misaligned with the traffic control cycles, e.g., a fixed 5-minute window may always fail to cover multiple complete control cycles, leading to substantial fluctuations and inaccuracies in understanding traffic dynamics and forecasting outcomes. In this study, we investigate the irregular traffic time series forecasting (*a.k.a.* irregular traffic forecasting) task, aiming to predict a variable-length sequence of traffic states, encompassing traffic signal cycle lengths and the corresponding traffic flows, in the future time window based on incomplete historical traffic state sequences characterized by variable time intervals.

It is a non-trivial task due to the following three major challenges: (1) *Asynchrony in spatial dependency modeling.* Traffic time series has obvious spatial dependency due to the traffic state’s diffusion nature on road network [22]. However, due to the time-misaligned traffic signal cycles (diverged cycle beginning time and length) between sensors, their traffic state measurements under the adaptive control policy would be observed asynchronously. Such asynchrony hinders correlating and integrating these sensors’ traffic states synchronously favored by classical traffic forecasting methods [22, 47], presenting a substantial challenge to model their spatial dependency. (2) *Irregularity in temporal dependency modeling.* The future traffic states are correlated with their historical values. Unlike previous traffic forecasting research [22, 47] that deals with regular traffic time series, we need to handle irregular traffic state sequences characterized by variable time intervals between successive measurements. These irregularities stem from fluctuating cycle lengths

and data missing in sensor measurements, leading to inconsistent temporal flow within the time series, challenging classical traffic forecasting methods to capture the underlying temporal dynamics and dependencies precisely. (3) *Variable-length sequence to be predicted.* Our goal is to predict the complete traffic state sequences in a future time window (e.g., the next hour). However, due to signal cycle lengths varying across different sensors and times, the lengths of sequences to be predicted also vary. While an autoregressive prediction model may seem plausible for predicting sequences of variable lengths, significant error accumulation and poor prediction efficiency issues are pronounced with longer sequences [26, 27]. This presents a significant challenge in effectively and efficiently forecasting variable-length traffic state sequences.

To tackle the above challenges, we present an **Asynchronous Spatio-tEmporal graph convolutional nEtwoRk (ASEER)** for irregular traffic forecasting. Specifically, by linking traffic sensors via a traffic diffusion graph, we propose an **Asynchronous Graph Diffusion Network** to model the spatial dependencies among nodes with time-misaligned traffic state measurements. It allows each node to asynchronously diffuse its traffic measurements to neighbors and store received traffic information with a message buffer. The stored messages are then integrated through an asynchronous graph convolution for the spatial node representation. To capture temporal dependencies in irregular sequences, a learnable personalized time encoding is first devised to embed the continuous time of traffic measurements. Then, we propose a **Transformable Time-aware Convolution Network** that learns meta-filters to derive time-aware convolution filters with transformable filter sizes, which are applied for efficient temporal convolution on sequences with inconsistent temporal flow. Lastly, we design a **Semi-Autoregressive Prediction Network** to iteratively predict variable-length traffic state sequences effectively and efficiently. It incorporates a state evolution unit to evolve traffic hidden state with elapsed time and a semi-autoregressive predictor to predict a sequence of consecutive traffic states at each prediction step.

Our major contributions can be summarized as follows: (1) We investigate a novel irregular traffic forecasting problem, which imposes three critical new challenges for traffic forecasting from spatial, temporal, and predicted sequence length perspectives. (2) We propose an **Asynchronous Graph Diffusion Network** to model spatial dependency among asynchronous time series data. (3) We propose a **Transformable Time-aware Convolution Network** with personalized time encoding to efficiently capture temporal dependency within irregular time series. (4) We design a **Semi-Autoregressive Prediction Network** to empower effective and efficient prediction for variable-length time series. (5) We meticulously collect and develop two novel real-world datasets of irregular traffic time series from two leading pilot cities for ITSCS in China, and establish a systematic evaluation scheme comprising six metrics, setting a new benchmark in the field and potentially fostering advancements in related areas. Extensive experiments demonstrate the superiority of ASEER compared with twelve competitive baseline approaches.

2 PRELIMINARIES

Consider a set of N traffic sensors, denoted as $\mathbf{V} = \{v^1, v^2, \dots, v^N\}$, positioned on lanes governed by adaptive traffic signals, e.g., lanes

connecting to smart intersections. Each sensor gathers real-time traffic data specific to a lane.

Definition 1: Traffic State Measurement. The n -th chronological traffic state measurement of a sensor v^i is defined as $x_n^i = \langle p_n^i, f_n^i \rangle$, where p_n^i denotes the traffic signal cycle length under the adaptive control, and f_n^i is the traffic flow during this signal cycle. We further define b_n^i and t_n^i as the beginning and end timestamps (in second) of this signal cycle, and we have $t_n^i = b_n^i + p_n^i - 1$.

As traffic signal cycles occur consecutively in real-world scenarios, we have $b_{n+1}^i = t_n^i + 1$ in case of no missing traffic states between x_n^i and x_{n+1}^i . Due to the unpredictable systematic failures of sensors, there could be multiple traffic states missing between two successive observed measurements.

Problem: Irregular Traffic Forecasting. Given a historical time window \mathcal{T} , e.g., one hour, before current timestamp t , and a set of historical traffic state measurements $\mathbf{X}_{[t-\mathcal{T}+1:t]} = \{[x_n^i]_{n=1}^{T^i}\}_{i=1}^N$ of all sensors \mathbf{V} obtained during \mathcal{T} , our problem is to predict the complete traffic states $\mathbf{Y}_{[t+1:t+\tau]} = \{[x_n^i]_{n=T^i+1}^{T^i+L^i}\}_{i=1}^N$ for all sensors in the next τ time window, e.g., the next hour, formalized as:

$$\mathcal{F}(\mathbf{X}_{[t-\mathcal{T}+1:t]}) \longrightarrow \mathbf{Y}_{[t+1:t+\tau]}, \quad (1)$$

where T^i is the number of observed historical measurements of v^i , L^i is the number of ground truth traffic states of v^i during the predicted time window, $\mathcal{F}(\cdot)$ represents the forecasting model we aim to learn. Note that we use the subscript $[a : b]$ to indicate a time window spanning from timestamp a to b .

3 METHODOLOGY

Framework Overview. Figure 2 shows the framework overview of ASEER, which consists of three major components. Specifically, Asynchronous Graph Diffusion Network (AGDN) models asynchronous spatial dependency based on a traffic diffusion graph. When a node (i.e., traffic sensor) has a traffic state measurement, AGDN asynchronously diffuses the node's traffic measurement to its neighbors, which receive and store the diffused traffic state into their message buffers. Next, the node performs an asynchronous graph convolution to obtain spatial representation through attentively integrating the stored traffic messages, and then the buffer will be cleared. After that, a Transformable Time-aware Convolution Network (TTCN) is adopted to model the temporal dependency within irregular traffic state sequences. TTCN learns meta-filters to derive time-aware convolution filters with transformable filter sizes based on spatial representations obtained from AGDN and traffic measurements along with personalized time encoding. Then the derived time-aware convolution filters are applied for efficient temporal convolution on irregular traffic state sequences to acquire the spatiotemporal representation for each node. Finally, a Semi-Autoregressive Prediction Network (SAPN) is devised to iteratively predict variable-length traffic state sequences. In each prediction step, a State Evolution Unit (SEU), whose hidden state is initialized by spatiotemporal representations, is first introduced to evolve each node's future traffic hidden state with the elapsed time, then a Semi-Autoregressive Predictor (SAP) is adopted to predict a sequence of consecutive traffic states based on both evolutionary and initial traffic hidden states, as well as predicted elapsed time.

3.1 Asynchronous Spatial Dependency Modeling

Previous traffic forecasting studies model spatial dependency by introducing graph neural networks to synchronously diffuse and aggregate time-aligned traffic states between different sensor nodes [22, 47]. However, in our problem, the observed traffic state measurements of different sensors cannot be aligned due to the distinct timestamps of their traffic signal cycles and the data missing issue, which causes severe asynchrony in spatial dependency modeling.

To this end, by linking sensors via a traffic diffusion graph, we propose an Asynchronous Graph Diffusion Network (AGDN), as illustrated in Figure 2, to model asynchronous spatial dependency between time-misaligned traffic measurements. We detail it below. **Diffusion Graph Construction.** To model spatial dependency between traffic sensors, we construct a traffic diffusion graph $\mathcal{G} = (\mathcal{V}, \mathcal{X}_{\mathcal{V}}, \mathcal{E}, \mathcal{X}_{\mathcal{E}})$, where the graph nodes $\mathcal{V} = \mathbf{V}$ represents a set of sensors, $\mathcal{X}_{\mathcal{V}} = \mathbf{X}_{[t-\mathcal{T}+1:t]}$ denotes features of nodes \mathcal{V} , \mathcal{E} are a set of edges indicating proximity between nodes, and $\mathcal{X}_{\mathcal{E}}$ are features in edges \mathcal{E} . Specifically, we define proximity $e^{ij} \in \mathcal{E}$ as: $e^{ij} = 1$ if the geographical distance between node v^i and v^j is smaller than a threshold ϵ , otherwise $e^{ij} = 0$, and there is no self-loop for each node. We also define some edge features $x_e^{ij} \in \mathcal{X}_{\mathcal{E}}$ between nodes v^i and v^j , including geographical distance and the direct reachability in the lane-level road network. Note that it is not limited to geographical proximity and reachability, other graph construction approaches can also be embraced.

Asynchronous Diffusion and Storage. Assume a traffic state measurement x_{n-}^j of node v^j is observed at timestamp t_{n-}^j , then v^j will diffuse x_{n-}^j as a traffic message to its adjacent nodes $v^i \in \mathcal{N}_j$ in terms of edges \mathcal{E} , which can be more formally denoted as:

$$\text{AsynDiff}(v^j \xrightarrow{x_{n-}^j} \{v^i : \forall v^i \in \mathcal{N}_j\}). \quad (2)$$

For each node $v^i \in \mathcal{N}_j$, it receives the traffic message x_{n-}^j and then stores it into its message buffer \mathcal{B}^i for later use:

$$\text{Store}(x_{n-}^j \longrightarrow \{\mathcal{B}^i : \forall v^i \in \mathcal{N}_j\}). \quad (3)$$

Since the timestamps of traffic state measurements are misaligned for different nodes, the traffic messages' diffusion and storage processes perform in an asynchronous way.

Asynchronous Graph Convolution. An immediate problem is how to exploit traffic messages stored in the message buffer to enhance each node's spatial perception. We achieve this by enforcing each node v^i asynchronously integrates the traffic messages in its message buffer \mathcal{B}^i via an asynchronous graph convolution operation, which is triggered when a measurement x_n^i is observed.

Specifically, we first employ x_n^i to query the message buffer \mathcal{B}^i for the proximity weights computation with each traffic message $x_{n-}^j \in \mathcal{B}^i$ via the following attention operation:

$$\alpha_{nn-} = \frac{\exp(\beta_{nn-})}{\sum_{x_{n-}^j \in \mathcal{B}^i} \exp(\beta_{nn-})}, \quad (4)$$

$$\beta_{nn-} = \mathbf{v}^\top \tanh \left(\mathbf{W}_a [x_n^i \oplus x_{n-}^j \oplus \phi^i(t_n^i - t_{n-}^j) \oplus x_e^{ij}] \right),$$

where \oplus indicates concatenation operation, \mathbf{v} and \mathbf{W}_a are learnable parameters, and $\phi^i(\cdot)$ is a learnable time encoding function to

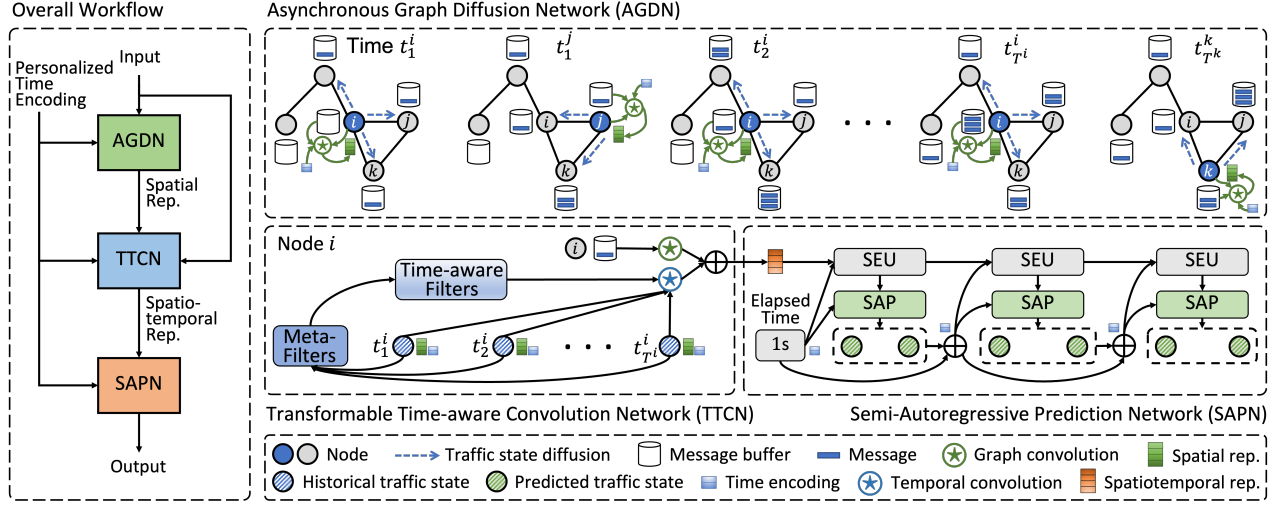


Figure 2: The framework overview of ASEER, which consists of three major components: Asynchronous Graph Diffusion Network (AGDN), Transformable Time-aware Convolution Network (TTCN), and Semi-Autoregressive Prediction Network (SAPN). The traffic states are first inputted to AGDN to obtain spatial representations, which are incorporated by TTCN to acquire the spatiotemporal representations. After that, SAPN predicts the variable-length traffic state sequence based on the spatiotemporal representations. Throughout the entire process, personalized time encoding is used to embed continuous time.

embed cycle-related patterns for each node that will be detailed in the next section.

Once the proximity weights are obtained, we asynchronously integrate node’s stored traffic messages received from neighbors via an attentive graph convolution to obtain the spatial representation:

$$\tilde{h}_n^i = \text{MLP} \left(\sum_{x_{n^-}^j \in \mathcal{B}^i} \alpha_{nn^-} \cdot \left[x_{n^-}^j \oplus \phi^i(t_n^i - t_{n^-}^j) \oplus x_e^{ij} \right] \right), \quad (5)$$

where MLP represents a multi-layer perceptron. It is noteworthy that after each asynchronous graph convolution operation on \mathcal{B}^i , all the traffic messages in it will be cleared. It indicates that each node only integrates adjacent traffic messages from its last traffic measurement’s timestamp to the current measurement’s timestamp t_n^i , which guarantees each message is utilized exactly once to avoid redundant information and computation.

There could be some messages received and stored in the message buffer \mathcal{B}^i after timestamp $t_{T_i}^i$ of the last observed traffic state measurement of node v^i during historical time window \mathcal{T} . Thus, we perform a similar asynchronous graph convolution operation for these remaining messages by adding a virtual measurement $\bar{x}_{T_i}^i$ at timestamp $t_{T_i}^i$ without traffic state values. The obtained spatial representation is denoted as $\tilde{h}_{T_i}^i$.

3.2 Irregular Temporal Dependency Modeling

Convolutional Neural Network (CNN) [21] is widely applied to classical traffic forecasting tasks for its both efficiency and effectiveness in temporal dependency modeling [10, 19, 44, 47]. However, applying CNN to our task faces two problems. First, CNN fails to directly process irregular traffic sequences with variable sequence lengths. Second, CNN is incompetent to model temporal dependency in the sequence with varying time intervals, as its filter parameters

are fixed and cannot adaptively adjust according to the inconsistent temporal flow of sequence, which leads to distinct patterns of sequence dependencies.

To tackle the above problems, we propose a Transformable Time-aware Convolution Network (TTCN) that enables to model irregular sequences with transformable time-aware convolution filters, and further devise a personalized time encoding function to embed the unique cycle-related patterns for each node. Specifically, given the historical time window \mathcal{T} before t , for each node v^i , we first concatenate the traffic state measurement x_n^i during \mathcal{T} with the corresponding spatial representation \tilde{h}_n^i and the encoding of time intervals $\phi^i(t_{T_i}^i - t_n^i)$ from timestamp t_n^i to timestamp $t_{T_i}^i$ of the last observed traffic measurement:

$$z_n^i = \left[x_n^i \oplus \tilde{h}_n^i \oplus \phi^i(t_{T_i}^i - t_n^i) \right]. \quad (6)$$

3.2.1 Personalized Time Encoding. The desired time encoding should not only indicate the absolute time interval but also imply the unique cycle-related patterns of traffic dynamics in different nodes. For example, a time interval may signify a distinct number of traffic signal cycles for different sensors, which is important for temporal dependency modeling, especially when the time interval spans multiple missing traffic states. Inspired by the positional encoding in Transformer [37], we introduce a personalized time encoding by adopting a learnable trigonometric function to embed the time interval Δt for each node:

$$\phi_p^i(\Delta t)[s] = \begin{cases} \Delta t, & \text{if } s = 0 \\ \sin(\omega_k^i \Delta t), & \text{if } s = 2k + 1 \\ \cos(\omega_k^i \Delta t), & \text{if } s = 2k + 2 \end{cases}, \quad (7)$$

where the above equation denotes s -th element of the time encoding $\phi_p^i(\Delta t) \in \mathbb{R}^{d_\phi+1}$, and ω_k^i are learnable parameters, indicating the cyclical characteristics of this function. Each node has an individual

time encoding function with separate parameters so that can learn its unique cycle-related patterns.

Due to the data missing problem, some nodes may have too sparse measurement data to learn a satisfactory unique time encoding function. Hence, we also jointly learn a generic time encoding $\phi_g(\Delta t)$, which has a similar function expression to Eq. (7) but is shared by all nodes. Then, we introduce a learnable weight λ_i for each node to adaptively integrate the above two time encoding:

$$\phi^i(\Delta t) = (1 - \exp(-\lambda_i^2)) \cdot \phi_p^i(\Delta t) + \exp(-\lambda_i^2) \cdot \phi_g(\Delta t). \quad (8)$$

λ_i is initialized to be close to zero so that the nodes with limited or even no available data can weigh more on generic time encoding.

3.2.2 Transformable Time-aware Convolution Network. This section assumes all the operations performed on node v^i , thus we omit the superscript i to ease the presentation. We first define $\mathbf{z}_{[t-\mathcal{T}+1:t]} = [z_1, \dots, z_T]$ and T as the sequence length. As illustrated in Figure 2, we leverage meta-filters to derive the time-aware convolution filters with dynamic parameters and transformable filter size T based on sequence inputs, formulated as:

$$\begin{aligned} \mathbf{f}_d &= [\text{Norm}(\mathbf{F}_d(z_1)), \dots, \text{Norm}(\mathbf{F}_d(z_T))], \\ \text{Norm}(\mathbf{F}_d(z_n)) &= \frac{\exp(\mathbf{F}_d(z_n))}{\sum_{z_{n'} \in \mathbf{z}_{[t-\mathcal{T}+1:t]}} \exp(\mathbf{F}_d(z_{n'}))}, \end{aligned} \quad (9)$$

where $\mathbf{f}_d \in \mathbb{R}^{T \times D_{in}}$ is the derived filter for d -th feature map, and \mathbf{F}_d denotes a meta-filter that can be instantiated by learnable neural networks. We normalize the derived filter parameters along the temporal dimension to ensure consistent scaling of the convolution results for variable-length sequences.

With D filters derived according to Eq. (9), we obtain the traffic sequence representation $h_T \in \mathbb{R}^D$ via the following temporal convolution operation:

$$\begin{aligned} h_T &= [\mathbf{z}_{[t-\mathcal{T}+1:t]} \star \mathbf{f}_1, \dots, \mathbf{z}_{[t-\mathcal{T}+1:t]} \star \mathbf{f}_D], \\ \mathbf{z}_{[t-\mathcal{T}+1:t]} \star \mathbf{f}_d &= \sum_{n=1}^T \mathbf{f}_d[n]^\top \mathbf{z}_{[t-\mathcal{T}+1:t]}[n], \end{aligned} \quad (10)$$

where \star denotes the convolution operation. Then we attain the overall spatiotemporal representation for each node via the representations integration: $\mathbf{h}_T = h_T + \hat{h}_T$.

Compared to CNN, TTCN has several advantages in adaptively modeling sequences with inconsistent temporal flow. First, the derived filter is transformable according to sequence length, which enables it to adaptively process variable-length sequences. Moreover, it can derive tailored parameterized filters for sequences with changeable temporal flow or other characteristics. Furthermore, it is worth noting that as the learnable parameters of meta-filter are independent of sequence length, TTCN is allowed to directly model the long-term temporal dependency via an arbitrarily large-size convolution filter without increasing any filter parameters.

3.3 Variable-Length Traffic Sequence Prediction

Our goal is to predict the complete traffic state sequences, including a sequence of traffic signal cycle lengths and the corresponding traffic flows, for all nodes in a future time window. However, the sequences to be predicted have variable lengths in terms of the differences in sensors, time windows, or prediction algorithms,

and the sequence lengths cannot be known in advance. While an autoregressive prediction model that iteratively predicts the next step's value based on previously predicted values seems feasible for the variable-length sequence prediction, the prediction for long sequence can lead to severe error accumulation and poor prediction efficiency issues in this approach [26, 27].

To tackle the above problems, as displayed in Figure 2, we design a Semi-Autoregressive Prediction Network (SAPN) to iteratively predict sub-sequences until the complete sequence meets the requirements of the task. It not only enables variable-length sequence prediction in an efficient way but also mitigates the error accumulation issue for long sequence prediction. Since prediction processes are the same for all nodes, we omit the superscript i to ease the presentation as well.

To be specific, we employ the spatiotemporal representation \mathbf{h}_T acquired from AGDN and TTCN as the initial traffic hidden state. In each prediction step, a semi-autoregressive predictor predicts a sequence of consecutive traffic states based on the evolutionary and initial traffic hidden states, as well as the predicted elapsed time encoding, formulated as:

$$[\hat{p}_n, \hat{u}_n]_{n=T+m\xi+1}^{T+(m+1)\xi} = \text{SAP} \left([\hat{\mathbf{h}}_{T+m\xi+1} \oplus \mathbf{h}_T \oplus \phi(\hat{\delta}_{T+m\xi+1}) \right], \quad (11)$$

where ξ is the prediction step size, $m \geq 0$ denotes m -th prediction step, $[\hat{p}_n, \hat{u}_n]_{n=T+m\xi+1}^{T+(m+1)\xi}$ respectively represent a sequence of consecutive cycle lengths and unit time (per second) traffic flows, and $\hat{\delta}_{T+m\xi+1}$ indicates the elapsed time to the timestamp t_T of node's last observed measurement. $\hat{\delta}_{T+m\xi+1}$ is initialized to 1 when $m = 0$, and iteratively updates based on the accumulation of predicted cycle lengths:

$$\hat{\delta}_{T+(m+1)\xi+1} = \hat{\delta}_{T+m\xi+1} + \sum_{k=1}^{\xi} \hat{p}_{T+m\xi+k}. \quad (12)$$

Since the underlying traffic state is actually dynamically evolving with passage of time, we introduce a state evolution unit that learns to evolve each node's traffic hidden state with the elapsed time:

$$\hat{\mathbf{h}}_{T+m\xi+1} = \text{SEU} \left(\hat{\mathbf{h}}_{T+(m-1)\xi+1}, \phi(\hat{\sigma}_{T+m\xi+1}) \right), \quad (13)$$

where $\hat{\mathbf{h}}_{T+(m-1)\xi+1} = \mathbf{h}_T$ and $\hat{\sigma}_{T+m\xi+1} = 1$ if $m = 0$, otherwise $\hat{\sigma}_{T+m\xi+1} = \sum_{k=1}^{\xi} \hat{p}_{T+(m-1)\xi+k}$, representing the elapsed time to last update of traffic hidden state. Next, we can obtain the corresponding traffic flows of predicted traffic signal cycles by multiplying the predicted unit time traffic flows with cycle lengths:

$$\left[\hat{\mathbf{f}}_n \right]_{n=T+m\xi+1}^{T+(m+1)\xi} = [\hat{u}_n]_{n=T+m\xi+1}^{T+(m+1)\xi} \odot [\hat{p}_n]_{n=T+m\xi+1}^{T+(m+1)\xi}, \quad (14)$$

where \odot denotes Hadamard product. By iteratively performing the above prediction step until the predicted sequence covers the required time window, we can derive the variable-length traffic state sequence we expect.

Compared to autoregressive models, our SAPN predicts a variable-length sequence with fewer prediction steps, which improves prediction efficiency and may reduce the risks of causing prediction error accumulation. It is evident that both the autoregressive and non-autoregressive prediction models can be regarded as a special case of semi-autoregressive model when the prediction step size is set to one or the length of sequence. Thus, our SAPN can also be considered as incorporating both strengths of autoregressive and

non-autoregressive models to predict variable-length sequences. In the implementation, we instantiate SAP via MLP and SEU via Gated Recurrent Unit [7] as its mechanism aligns well with the recurrent update for traffic hidden state.

3.4 Model Training

Due to the data missing problem, we design three masked losses to train our model. The first loss is introduced to optimize the traffic signal cycle length forecasting via the masked Mean Absolute Error (MAE):

$$\mathcal{L}_p = \frac{1}{\sum_{i=1}^N L_{\perp}^i} \sum_{i=1}^N \sum_{l=1}^{L^i} \left| \hat{p}_{T^{i+l}}^i - p_{T^{i+l}}^i \right| \times \zeta_{T^{i+l}}^i, \quad (15)$$

$\zeta_{T^{i+l}}^i$ is a mask term, which equals zero if the ground truth value $p_{T^{i+l}}^i$ is missing, otherwise it equals one, and L_{\perp}^i denotes the number of nonzero mask items for each node.

To further mitigate error accumulation in cycle length prediction, we additionally introduce a timing loss to improve the accuracy of predicted elapsed time accumulated by cycle lengths:

$$\mathcal{L}_{\delta} = \frac{1}{\sum_{i=1}^N L_{\perp}^i} \sum_{i=1}^N \sum_{l=1}^{L^i} \left| \hat{\delta}_{T^{i+l}}^i - \delta_{T^{i+l}}^i \right| \times \zeta_{T^{i+l}}^i. \quad (16)$$

Similarly, we introduce a masked MAE loss to optimize the corresponding traffic flow prediction for each traffic signal cycle:

$$\mathcal{L}_f = \frac{1}{\sum_{i=1}^N L_{\perp}^i} \sum_{i=1}^N \sum_{l=1}^{L^i} \left| \hat{u}_{T^{i+l}}^i \times p_{T^{i+l}}^i - f_{T^{i+l}}^i \right| \times \zeta_{T^{i+l}}^i. \quad (17)$$

Since traffic flow prediction is also based on cycle lengths, to avoid disturbance from the error of predicted cycle lengths, we use the ground truth cycle lengths to calculate the corresponding traffic flows in the training phase.

Consequently, ASEER aims to jointly minimize an overall objective that combines the above three masked losses:

$$\mathcal{L} = \mathcal{L}_p + \mathcal{L}_{\delta} + \mathcal{L}_f. \quad (18)$$

4 EXPERIMENTS

4.1 Experimental Setup

Datasets. We conduct experiments on two real-world datasets, ZHUZHOU and BAODING, which represent two major pilot cities for ITSCS and autonomous driving in China. Both datasets consist of a set of entrance lanes connecting to smart intersections and traffic state measurements of lanes collected by the installed camera sensors. The statistics of the datasets are summarized in Table 1. We take the data from the first 60% of the entire time range as the training set, the following 20% for validation, and the remaining 20% as the test set. For both datasets, we set both historical and predicted time window lengths \mathcal{T} and τ to one hour. Please refer to Section A.1 for more description and analysis of the datasets.

Implementation Details. All experiments are performed on a Linux server with 20-core Intel(R) Xeon(R) Gold 6148 CPU @ 2.40GHz and NVIDIA Tesla V100 GPU. We calculate spherical distance as the geographical distance and choose distance threshold $\epsilon = 1\text{km}$ and prediction step size $\xi = 12$. We adopt three layers MLPs for asynchronous graph convolution, semi-autoregressive predictor, and meta-filters. The dimension for time encoding is set to $d_{\phi} = 16$, and dimensions for convolution filters, state evolution

Table 1: Statistics of datasets.

Description	ZHUZHOU	BAODING
# of measurements	19,824,504	13,093,975
# of sensors	620	264
Time range	2022/07/20-2022/10/02	2021/12/01-2022/02/25
Missing period ratio	44.2%	27.2%
Average / maximal ground truth sequence length to be predicted	57 / 213	64 / 155

unit, and hidden layers of the above MLPs are all set to 64. To reduce parameter magnitude, in the implementation, we individualize the last layer's parameters but share the other parameters of MLP for meta-filters. We employ Adam optimizer to train our model, set learning rate to 0.001. ASEER and all learnable baselines are trained with an early stop criterion if the loss doesn't decrease lower on the validation set over 10 epochs.

Evaluation Metrics. We define six metrics to comprehensively evaluate the forecasting performance, including C-MAE, C-RMSE, and C-MAPE to evaluate the accuracy of predicted traffic signal cycle lengths, and F-MAE, F-RMSE, and F-AAE for the traffic flow prediction evaluation. Lower is better for all these metrics.

For the evaluation of signal cycle length prediction, we quantify the predicted errors of both the beginning timestamps and cycle lengths via masked Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE):

$$\begin{aligned} \text{C-MAE} &= \frac{1}{2 \sum_{i=1}^N K_{\perp}^i} \sum_{i=1}^N \sum_{k=1}^{K^i} \left(|\hat{b}_k^i - b_k^i| + |\hat{p}_k^i - p_k^i| \right) \times \zeta_k^i, \\ \text{C-RMSE} &= \sqrt{\frac{1}{2 \sum_{i=1}^N K_{\perp}^i} \sum_{i=1}^N \sum_{k=1}^{K^i} \left((\hat{b}_k^i - b_k^i)^2 + (\hat{p}_k^i - p_k^i)^2 \right) \times \zeta_k^i}, \\ \text{C-MAPE} &= \frac{100\%}{2 \sum_{i=1}^N K_{\perp}^i} \sum_{i=1}^N \sum_{k=1}^{K^i} \left(\frac{|\hat{b}_k^i - b_k^i|}{\delta_k^i} + \frac{|\hat{p}_k^i - p_k^i|}{p_k^i} \right) \times \zeta_k^i. \end{aligned} \quad (19)$$

where K^i denotes the number of ground truth traffic states of each sensor for evaluation, K_{\perp}^i is the number of observed measurements.

Since traffic flows depend on the corresponding traffic signal cycles, it is incomparable between the predicted and ground truth traffic flows if they are misaligned in signal cycles. Thus, we introduce two types of metrics to evaluate the prediction accuracy of traffic flows from multiple aspects. First, we assume all the traffic signal cycle lengths can be accurately predicted and use the ground truth cycle lengths for calculation, then we can directly compare the predicted and ground truth traffic flows via the following masked MAE and RMSE metrics:

$$\begin{aligned} \text{F-MAE} &= \frac{1}{\sum_{i=1}^N K_{\perp}^i} \sum_{i=1}^N \sum_{k=1}^{K^i} \left| \hat{f}_k^i - f_k^i \right| \times \zeta_k^i, \\ \text{F-RMSE} &= \sqrt{\frac{1}{\sum_{i=1}^N K_{\perp}^i} \sum_{i=1}^N \sum_{k=1}^{K^i} \left(\hat{f}_k^i - f_k^i \right)^2 \times \zeta_k^i}. \end{aligned} \quad (20)$$

Second, without the above assumption for cycle lengths, by computing traffic flow density at any timestamp, we calculate the masked Accumulative Absolute Error (AAE) between predicted and ground

Table 2: Overall performance evaluated by C-MAE, C-RMSE, C-MAPE, F-MAE, F-RMSE, and F-AAE on ZHUZHOU and BAODING. The best-performing results are highlighted in bold, and the second-best results are highlighted by underline.

Algorithm	ZHUZHOU					BAODING						
	C-MAE	C-RMSE	C-MAPE	F-MAE	F-AAE	C-MAE	C-RMSE	C-MAPE	F-RMSE	F-AAE		
LAST	50.5386	135.2616	5.54%	1.6669	3.0995	0.9192	42.8037	106.6547	4.79%	1.7521	2.8031	0.9557
HA	52.1532	135.3569	5.76%	1.4502	2.6567	0.7998	49.7496	114.8265	5.53%	1.5449	2.4594	0.8427
TCN	43.7838	110.1670	5.01%	1.3950	2.5824	0.7818	35.8318	95.7333	4.18%	1.3815	2.2060	0.7635
GRU	40.6209	99.8693	4.82%	1.3623	2.5553	0.7524	30.4621	83.4349	3.82%	1.3576	2.1655	0.7423
T-LSTM	39.1882	87.3458	5.38%	1.3641	2.5494	0.7539	29.0845	82.5219	3.76%	1.3673	2.1887	0.7475
GRU-D	37.8531	84.6255	5.23%	1.3486	2.5333	0.7449	28.9117	82.5226	3.67%	1.3611	2.1735	0.7456
mTAND	37.5762	86.3045	3.93%	1.3563	2.5282	0.7498	27.2703	78.1066	<u>2.86%</u>	1.3575	2.1641	0.7487
Warpformer	<u>35.7369</u>	85.3125	5.02%	1.3399	2.5662	0.7405	27.8527	78.7168	3.89%	1.3554	2.1720	0.7427
DCRNN	38.5976	90.3190	4.36%	1.3318	<u>2.4438</u>	0.7348	31.0564	76.3693	3.86%	1.3681	2.1601	0.7467
GWNet	38.9913	106.6415	4.52%	1.3834	2.7915	0.7618	<u>26.4988</u>	84.3211	3.05%	1.3925	2.2482	0.7903
STAEformer	40.4448	<u>78.2176</u>	4.84%	1.3503	2.4501	0.7468	28.1453	73.5397	3.74%	1.3801	2.1518	0.7585
PDFormer	36.4779	86.1173	4.66%	<u>1.3170</u>	2.4575	<u>0.7269</u>	27.1969	<u>68.5613</u>	3.82%	<u>1.3540</u>	<u>2.1496</u>	<u>0.7413</u>
ASEER	32.5803	72.1835	<u>4.10%</u>	1.2913	2.3864	0.7151	19.1188	54.4919	2.80%	1.3062	2.0827	0.7219

truth traffic flow density at identical timestamps:

$$F\text{-AAE} = \frac{1}{Z} \sum_{i=1}^N \sum_t |\hat{\rho}_t^i - \rho_t^i| \times \eta_t^i, \quad (21)$$

where $\hat{\rho}_t^i = \hat{f}_k^i / \hat{p}_k^i$, $t \in [\hat{b}_k^i, \hat{t}_k^i]$ and $\rho_t^i = f_k^i / p_k^i$, $t \in [b_k^i, t_k^i]$ are the predicted and ground truth traffic flow densities of sensor v^i at timestamp t , respectively. η_t^i is the mask term at t , which equals one if ρ_t^i can be obtained from observed measurement, and zero otherwise. In our experiments, the timestamp is in seconds, and we use a normalization term Z to obtain the average result in minutes. **Baselines.** We compare our approach with the following twelve baselines, including two heuristic approaches (LAST, HA), two classical sequence modeling approaches (TCN [2], GRU [7]), four competitive irregular time series modeling approaches (T-LSTM [3], GRU-D [4], mTAND [34], Warpformer [49]), and four competitive classical traffic forecasting approaches (DCRNN [22], GWNet [44], STAEformer [25], PDFormer [15]). For fair comparison, all learnable baseline models are set to predict the cycle lengths and unit time traffic flows by optimizing the hybrid loss function in Eq. (18) like ASEER. In addition, except for autoregressive models (GRU, T-LSTM, GRU-D, DCRNN), other baselines predict in a semi-autoregressive manner with the same prediction step size as ASEER. We carefully tune major hyper-parameters of each baseline based on their recommended settings for better performance on our datasets. Please refer to Section A.2 for more details of baselines.

4.2 Overall Performance

Table 2 reports the overall performance of ASEER and all compared baselines on two datasets *w.r.t.* six metrics. As can be seen, ASEER achieves the best overall performance among all the compared approaches on two datasets, which demonstrates our model’s superiority in irregular traffic forecasting task. Besides, we have several observations. Firstly, all learnable approaches outperform the statistical approaches (*i.e.*, LAST, HA), which validates that the data-driven approaches to learn complex non-linear interactions within traffic data is helpful for this task. Secondly, we find CNN-based baselines TCN and GWNet do not achieve a desired performance for the reason that classical CNN with the fixed parameterized filters is incompetent to model the temporal dependency

in irregular sequences. Thirdly, we observe ASEER obtains a superior overall performance than approaches (*i.e.*, GRU-D, T-LSTM, mTAND, and Warpformer) for irregular time series, as these approaches fail to model the complex spatial dependencies between large-scale sensors. From these approaches, we notice mTAND has a slight advantage in C-MAPE than ASEER on ZHUZHOU. This is probably because mTAND as a powerful approach for interpolation task performs well in the short-term future cycles’ beginning times prediction. However, ASEER significantly outperforms mTAND in the other metrics. Lastly, we observe a notable performance improvement by comparing ASEER with the state-of-the-art approaches (*i.e.*, DCRNN, GWNet, STAEformer, and PDFormer) for classical traffic forecasting. The improvement can primarily be attributed to the capability of ASEER to effectively model asynchronous spatial dependency and irregular temporal dependency in the irregular traffic forecasting problem.

4.3 Ablation Study

We evaluate the performance of ASEER and its four variants on both ZHUZHOU and BAODING across six metrics. (1) **w/o AGDN** removes the AGDN module; (2) **w/o TTCN** replaces TTCN with a 1D CNN, whose filter size is set to the maximal sequence length in the dataset; (3) **w/o PTE** removes personalized time encoding; (4) **w/o SAPN** replaces SAPN with an autoregressive MLP predictor. The results of ablation study are shown in Figure 3, As can be seen, removing any component causes notable overall performance degradation compared to ASEER, which demonstrates the effectiveness of each component. From these results, we observe **w/o TTCN** almost results in significant performance descent for all metrics on both datasets, which verifies the effectiveness of TTCN to improve classical CNN to model the temporal dependency within irregular traffic sequences. In addition, **w/o AGDN** causes a remarkable accuracy decline for all the metrics *w.r.t.* traffic flow, which validates the effect of AGDN on modeling asynchronous spatial dependency of traffic dynamics. We also observe **w/o AGDN** causes a more obvious accuracy decline on BAODING than ZHUZHOU for three metrics *w.r.t.* cycle lengths. This is probably because the distribution of cycle lengths in BAODING is denser, AGDN’s smoothness induces a more precise prediction. Moreover, we notice that **w/o PTE** leads to a consistent performance reduction for all metrics on

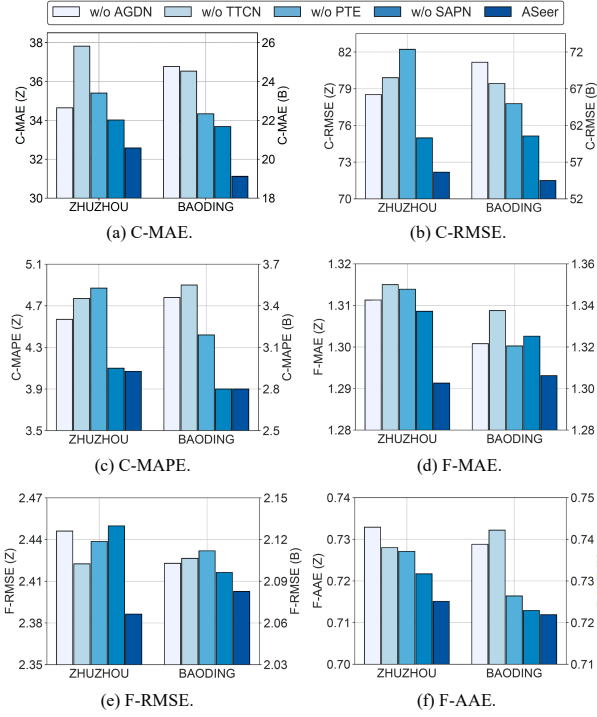


Figure 3: Results of ablation study. "Z" and "B" denote ZHUZHOU and BAODING, respectively.

both datasets, which demonstrates that a well-learned personalized time encoding function to embed continuous time for each sensor can facilitate the prediction of both cycle lengths and traffic flows. Finally, by comparing ASeer with *w/o SAPN*, we observe a more obvious performance degradation on BAODING for metrics *w.r.t.* cycle lengths, which is probably because the sequence is longer on BAODING, an autoregressive model causes a severe error accumulation problem on cycle length prediction. *w/o SAPN* also shows a consistent performance descent for three metrics *w.r.t.* traffic flow, which confirms that SAPN improves the long traffic sequence prediction performance.

4.4 Parameter Sensitivity

We conduct experiments for two important hyper-parameters, *i.e.*, the prediction step size ξ and dimension of all hidden layers, on both ZHUZHOU and BAODING to study the sensitivity of these hyper-parameters. We report experimental results on metrics C-MAE, F-MAE, and F-AAE to evaluate the model's prediction performance on both cycle lengths and traffic flows.

Figure 4 shows the results of varying the prediction step size ξ from 1 to 48. As can be seen, there is a notable overall prediction performance improvement by increasing ξ from 1 (autoregressive model) to 12 (semi-autoregressive model), which demonstrates the effectiveness of SAPN to mitigate error accumulation problem in the autoregressive prediction model. However, we also observe a performance degradation when the prediction step size is too large. This is probably because a too-large prediction step size may result in under-training for SAPN to make predictions based on different elapsed times.

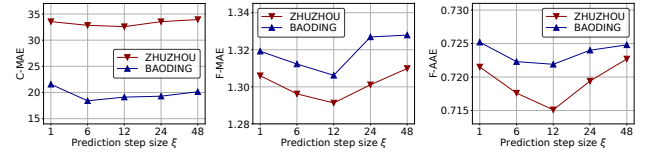


Figure 4: Effect of different prediction step sizes.

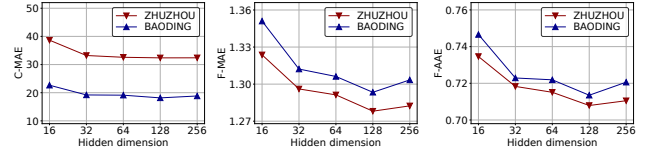


Figure 5: Effect of different hidden dimensions.

We vary the dimension of model's all hidden layers from 16 to 256. The results are shown in Figure 5. We can observe a remarkable prediction performance improvement by increasing the hidden dimension from 16 to 32, and the performance is continuously improving and achieves the best when the dimension is set to 128. However, a larger hidden dimension also takes more expensive computational overhead. Thus, we have to balance the performance and computation cost for the selection of model's hidden dimension.

4.5 Prediction Efficiency Analysis

We conduct experiments to test the prediction efficiency of different models. To ensure a fair comparison, we eliminate the influence of different models on the prediction lengths by standardizing the prediction process. This involves allowing all models to predict the maximum lengths of the corresponding ground truth sequences.

Efficiency of SAPN. To evaluate the effect of SAPN on prediction efficiency, we conduct experiments on both ZHUZHOU and BAODING to specifically test SAPN's average prediction latency based on different prediction step sizes ξ from 1 to 48. We report the respective results of predicting future 1, 4, and 24 hours traffic states in Figure 6. As can be seen, the prediction latency is notably reduced by comparing semi-autoregressive models ($\xi > 1$) with autoregressive model ($\xi = 1$) due to the reduction of total prediction steps. The magnitude of latency reduction even approaches the prediction step size when we predict longer sequences or the step size is not too large, which demonstrates the significant effectiveness of SAPN to improve prediction efficiency. We also observe with the prediction step size increasing, the prediction latency is consistently reduced, and with the predicted hours rising, the model can have a significantly higher prediction efficiency by setting a larger prediction step size. This observation indicates that we can choose a larger prediction step size with the predicted sequence length increasing for higher prediction efficiency.

Efficiency of TTCN. To study TTCN's efficiency, we replace TTCN with several commonly used modules in temporal modeling, *i.e.*, CNN, GRU, and Transformer, and test their running time costs. As illustrated in Figure 7, TTCN achieves more than 40% and 33% faster results than GRU and transformer, respectively, on both datasets. Furthermore, to our surprise, TTCN exhibits even faster than CNN.

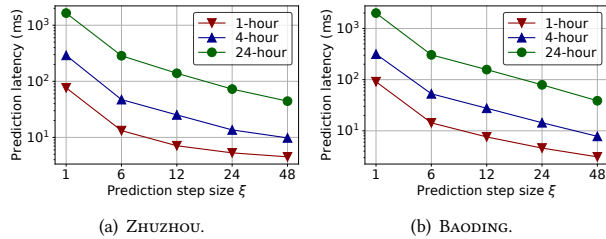


Figure 6: Prediction latency of SAPN with different prediction step sizes.

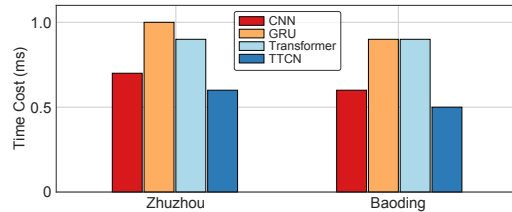


Figure 7: Efficiency of various modules in temporal modeling.

This is probably because TTCN can directly handle variable-length sequences with transformable filter sizes, while CNN is limited to processing fixed-length sequences via padding or clipping, thus it may cost additional time to process longer sequences beyond their original lengths. It demonstrates the efficiency of TTCN.

5 RELATED WORK

Traffic Forecasting. Recently years, deep learning models have dominated the traffic forecasting tasks for their extraordinary capability in modeling the complex spatio-temporal characteristics within traffic data [1, 9, 10, 12, 15, 19, 22, 33, 44, 45, 47, 50, 52, 54]. For spatial modeling, a part of studies [45, 50] first partition a city into a grid-based region map, then utilize Convolutional Neural Networks (CNNs) to capture spatial dependencies between adjacent regions. After that, Graph Neural Networks (GNNs) [8, 17, 38, 43, 48, 51] are widely used to model the non-euclidean spatial dependencies in traffic data [12, 14, 16, 24, 33, 52]. For example, studies [22, 47] employ GNNs to model the traffic flow diffusion process in the road network. Studies [10, 11, 15, 52, 54] incorporate attention mechanism into GNNs to learn the dynamic spatial dependencies between the road network sensors. In addition to the pre-defined relational graph derived from road networks, some works [1, 12, 19, 32, 44, 46] attempt to directly learn the latent graph structure from traffic data. For temporal modeling, CNNs [10, 19, 44, 47, 50] and Recurrent Neural Networks (RNNs) [1, 22, 45] are frequently adopted to capture temporal dependencies within traffic data. Compared to RNNs, CNNs enable parallel computing for all time steps, which exhibits extreme advantages in computational efficiency. However, all the methods of above studies are designed for the time-aligned traffic data with fixed time interval, which fails to handle the challenges of asynchronous spatial dependency and irregular temporal dependency in the irregular traffic forecasting problem.

Irregularly Sampled Time Series. This work is also related to the literature about learning from irregularly sampled time series, which is a kind of time series data characterized by varying time

intervals between temporally adjacent observations [53]. A straightforward approach is to divide the irregularly sampled time series into a regular one with fixed time intervals [23]. However, such a temporal discretization method may lead to information loss and data missing problems [34, 35]. Recent studies tend to directly learn from irregularly sampled time series. Specifically, some studies improve RNNs by using a time gate [28], a time decay term [4], or memory decomposition mechanism [3] to adjust RNNs' memory update for adapting irregular time series. Another line of studies introduces neural Ordinary Differential Equations (ODEs) [6] to model the continuous dynamics in time series, and assume the latent states of time series are continuously evolving through continuous time [30, 31]. Besides, attention mechanism is also applied to model irregularly sampled time series [13, 34, 42, 49, 53]. For example, Shukla and Marlin [34] proposes a multi-time attention network to learn embedding of continuous time. Zhang et al. [49] employs a doubly self-attention to learn representation from the input data unified by a warping module. Zhang et al. [53] introduce GNNs to capture time-varying dependencies between sensors by performing the graph convolution operation at all timestamps when there is an observation at an arbitrary sensor. However, it will be extremely time-consuming once the data is significantly asynchronous across large-scale sensors like us. Furthermore, the above studies primarily focus on solving irregular time series classification instead of forecasting tasks. Finally, to our knowledge, there are no prior studies attempting to modify CNNs to adapt to the irregular time series modeling.

6 CONCLUSION

In this paper, we investigated a new irregular traffic forecasting problem that aims to predict irregular traffic time series resulting from adaptive traffic signal controls, and presented ASEER, an Asynchronous Spatio-Temporal Graph Convolutional Network, to address this problem. Specifically, by representing the traffic sensors as nodes and linking them via a traffic diffusion graph, we first proposed an Asynchronous Graph Diffusion Network to model the spatial dependency between the time-misaligned traffic state measurements of nodes. After that, to capture the temporal dependency within irregular traffic state sequences, we devised a personalized time encoding to embed the continuous time for each node and proposed a Transformable Time-aware Convolution Network to perform efficient temporal convolution on sequences with inconsistent temporal flow. Furthermore, a Semi-Autoregressive Prediction Network was designed to iteratively predict variable-length traffic state sequences effectively and efficiently. Finally, extensive experiments on two newly constructed real-world datasets demonstrated the superiority of ASEER compared with twelve competitive baseline approaches across six metrics.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No.92370204, No.62102110), National Key R&D Program of China (Grant No.2023YFF0725001), Guangdong Basic and Applied Basic Research Foundation (Grant No.2023B1515120057), Guangzhou-HKUST(GZ) Joint Funding Program (Grant No.2023A03J0008), Education Bureau of Guangzhou Municipality, CCF-Baidu Open Fund.

REFERENCES

- [1] Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems* 33 (2020), 17804–17815.
- [2] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* (2018).
- [3] Inci M Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K Jain, and Jiayu Zhou. 2017. Patient subtyping via time-aware LSTM networks. In *Proceedings of the 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 65–74.
- [4] Z Che, S Purushotham, K Cho, D Sontag, and Y Liu. 2018. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Scientific reports* (2018), 6085–6085.
- [5] Jiming Chen, Weixin Lin, Zidong Yang, Jianyuan Li, and Peng Cheng. 2019. Adaptive ramp metering control for urban freeway using large-scale data. *IEEE Transactions on Vehicular Technology* 68, 10 (2019), 9507–9518.
- [6] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. 2018. Neural ordinary differential equations. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 6572–6583.
- [7] Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS Workshop on Deep Learning*.
- [8] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems* (2016), 3844–3852.
- [9] Zheng Fang, Qingqing Long, Guojie Song, and Kunqing Xie. 2021. Spatial-temporal graph ode networks for traffic flow forecasting. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 364–373.
- [10] Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of AAAI conference on artificial intelligence*, 922–929.
- [11] Shengnan Guo, Youfang Lin, Huaiyu Wan, Xiucheng Li, and Gao Cong. 2021. Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting. *IEEE Transactions on Knowledge and Data Engineering* 34, 11 (2021), 5415–5428.
- [12] Jindong Han, Weijia Zhang, Hao Liu, Tao Tao, Naiqiang Tan, and Hui Xiong. 2024. BigST: Linear Complexity Spatio-Temporal Graph Neural Network for Traffic Forecasting on Large-Scale Road Networks. *Proceedings of the VLDB Endowment* (2024), 1081–1090.
- [13] Max Horn, Michael Moor, Christian Bock, Bastian Rieck, and Karsten Borgwardt. 2020. Set functions for time series. In *International Conference on Machine Learning*, 4353–4363.
- [14] Jiahao Ji, Jingyuan Wang, Chao Huang, Junjie Wu, Boren Xu, Zhenhe Wu, Junbo Zhang, and Yu Zheng. 2023. Spatio-temporal self-supervised learning for traffic flow prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37, 4356–4364.
- [15] Jiawei Jiang, Chengkai Han, Wayne Xin Zhao, and Jingyuan Wang. 2023. PDFFormer: Propagation Delay-aware Dynamic Long-range Transformer for Traffic Flow Prediction. In *Proceedings of the AAAI conference on artificial intelligence*, 4365–4373.
- [16] Weiwei Jiang and Jiayun Luo. 2022. Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications* (2022), 117921.
- [17] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations, ICLR*.
- [18] Serjbeet Kohli, Steer Davies Gleave, and Luis Willumsen. 2016. Traffic Forecasting and Autonomous Vehicles. In *2016 European Transport Conference, Barcelona*, 5–7.
- [19] Shiyong Lan, Yitong Ma, Weikang Huang, Wenwu Wang, Hongyu Yang, and Pyang Li. 2022. Dstagnn: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting. In *International Conference on Machine Learning*, 11906–11917.
- [20] Ibai Lana, Javier Del Ser, Manuel Velez, and Eleni I Vlahogianni. 2018. Road traffic forecasting: Recent advances and new challenges. *IEEE Intelligent Transportation Systems Magazine* 10, 2 (2018), 93–109.
- [21] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* (2015), 436–444.
- [22] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *International Conference on Learning Representations*.
- [23] Zachary C Lipton, David Kale, and Randall Wetzel. 2016. Directly modeling missing data in sequences with rnns: Improved classification of clinical time series. In *Machine learning for healthcare conference*, 253–270.
- [24] Fan Liu, Weijia Zhang, and Hao Liu. 2023. Robust Spatiotemporal Traffic Forecasting with Reinforced Dynamic Adversarial Training. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1417–1428.
- [25] Hangchen Liu, Zheng Dong, Renhe Jiang, Jiwen Deng, Jinliang Deng, Quan-jun Chen, and Xuan Song. 2023. Spatio-temporal adaptive embedding makes vanilla transformer sota for traffic forecasting. In *Proceedings of the 32nd ACM international conference on information and knowledge management*, 4125–4129.
- [26] Ivan Marisca, Andrea Cini, and Cesare Alippi. 2022. Learning to Reconstruct Missing Data from Spatiotemporal Graphs with Sparse Observations. In *Advances in Neural Information Processing Systems*.
- [27] Fernando Moreno-Pino, Pablo M Olmos, and Antonio Artés-Rodríguez. 2023. Deep autoregressive models with spectral attention. *Pattern Recognition* (2023), 109014.
- [28] Daniel Neil, Michael Pfeiffer, and Shih-Chii Liu. 2016. Phased LSTM: accelerating recurrent network training for long or event-based sequences. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 3889–3897.
- [29] Rezaur Rahman and Samiul Hasan. 2023. A deep learning approach for network-wide dynamic traffic prediction during hurricane evacuation. *Transportation Research Part C: Emerging Technologies* 152 (2023), 104126.
- [30] Yulia Rubanova, Ricky TQ Chen, and David Duvenaud. 2019. Latent ODEs for irregularly-sampled time series. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 5320–5330.
- [31] Mona Schirmer, Mazin Eltayeb, Stefan Lessmann, and Maja Rudolph. 2022. Modeling irregular time series with continuous recurrent units. In *International Conference on Machine Learning*, 19388–19405.
- [32] Zezhi Shao, Zhao Zhang, Fei Wang, and Yongjun Xu. 2022. Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1567–1577.
- [33] Zezhi Shao, Zhao Zhang, Wei Wei, Fei Wang, Yongjun Xu, Xin Cao, and Christian S Jensen. 2022. Decoupled dynamic spatial-temporal graph neural network for traffic forecasting. *Proceedings of the VLDB Endowment* (2022), 2733–2746.
- [34] Satya Narayan Shukla and Benjamin Marlin. 2021. Multi-Time Attention Networks for Irregularly Sampled Time Series. In *International Conference on Learning Representations*.
- [35] Satya Narayan Shukla and Benjamin M Marlin. 2020. A survey on principles, models and methods for learning from irregularly sampled time series. *arXiv preprint arXiv:2012.00168* (2020).
- [36] Qian Sun, Le Zhang, Huan Yu, Weijia Zhang, Yu Mei, and Hui Xiong. 2023. Hierarchical reinforcement learning for dynamic autonomous vehicle navigation at intelligent intersections. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4852–4861.
- [37] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
- [38] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations*.
- [39] Chen Wang, Bertrand David, René Chalon, and Chuantao Yin. 2016. Dynamic road lane management study: A Smart City application. *Transportation research part E: logistics and transportation review* 89 (2016), 272–287.
- [40] Yizhe Wang, Xiaoguang Yang, Hailun Liang, Yangdong Liu, et al. 2018. A review of the self-adaptive traffic signal control system based on future traffic environment. *Journal of Advanced Transportation* 2018 (2018).
- [41] Hua Wei, Guanjie Zheng, Vikash Gayah, and Zhenhui Li. 2019. A survey on traffic signal control methods. *arXiv preprint arXiv:1904.08117* (2019).
- [42] Yuxi Wei, Juntong Peng, Tong He, Chenxin Xu, Jian Zhang, Shirui Pan, and Siheng Chen. 2023. Compatible transformer for irregularly sampled multivariate time series. In *2023 IEEE International Conference on Data Mining (ICDM)*, IEEE, 1409–1414.
- [43] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. 2020. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems* (2020), 4–24.
- [44] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. 2019. Graph wavenet for deep spatial-temporal graph modeling. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1907–1913.
- [45] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Zhenhui Li, Jieping Ye, and Didi Chuxing. 2018. Deep multi-view spatial-temporal network for taxi demand prediction. In *Proceedings of the AAAI conference on artificial intelligence*, 2588–2595.
- [46] Kun Yi, Qi Zhang, Wei Fan, Hui He, Liang Hu, Pengyang Wang, Ning An, Longbing Cao, and Zhendong Niu. 2023. FourierGNN: Rethinking multivariate time series forecasting from a pure graph perspective. *Advances in Neural Information Processing Systems* 36 (2023).
- [47] Bing Yu, Haoteng Yin, and Zhanxing Zhu. 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 3634–3640.
- [48] Zixuan Yuan, Hao Liu, Yanchi Liu, Denghui Zhang, Fei Yi, Nengjun Zhu, and Hui Xiong. 2020. Spatio-temporal dual graph attention network for query-poi matching. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, 629–638.
- [49] Jiawen Zhang, Shun Zheng, Wei Cao, Jiang Bian, and Jia Li. 2023. Warpformer: A Multi-Scale Modeling Approach for Irregular Clinical Time Series. In *Proceedings*

of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 3273–3285.

- [50] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proceedings of the AAAI conference on artificial intelligence*. 1655–1661.
- [51] Weijia Zhang, Hao Liu, Jindong Han, Yong Ge, and Hui Xiong. 2022. Multi-agent graph convolutional reinforcement learning for dynamic electric vehicle charging pricing. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2471–2481.
- [52] Weijia Zhang, Hao Liu, Yanchi Liu, Jingbo Zhou, and Hui Xiong. 2020. Semi-supervised hierarchical recurrent graph neural network for city-wide parking availability prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 1186–1193.
- [53] Xiang Zhang, Marko Zeman, Theodoros Tsiligkaridis, and Marinka Zitnik. 2022. Graph-Guided Network for Irregularly Sampled Multivariate Time Series. In *International Conference on Learning Representations*.
- [54] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. 2020. Gman: A graph multi-attention network for traffic prediction. In *Proceedings of the AAAI conference on artificial intelligence*. 1234–1241.

A SUPPLEMENTARY EXPERIMENTS

A.1 Data Description and Analysis

A.1.1 Datasets Description. The statistics of the datasets are summarized in Table 1. Specifically, there are total 19,824,504 and 13,093,975 traffic state measurements on ZHUZHOU and BAODING, and the missing period ratios of the two datasets are 44.2% and 27.2%, respectively. Each measurement includes information about the beginning and end timestamps and cycle length of the traffic signal cycle, as well as the lane’s traffic flow, *i.e.*, the number of vehicles passing through the camera of lane, during the signal cycle. Besides, ZHUZHOU has 620 lanes with sensors and ranges from July 20, 2022 to October 2, 2022. BAODING has 264 lanes with sensors and ranges from December 1, 2021 to February 25, 2022. The average, maximal ground truth sequence length to be predicted for the future one hour is 57, 213 on ZHUZHOU, and 64, 155 on BAODING.

A.1.2 Datasets Analysis. The overall distributions of traffic signal cycle lengths on two datasets are depicted in Figure 8, where we can observe the cycle lengths can significantly vary from around 40 to 200 seconds on both datasets, indicating the pronounced irregularity within time series, and BAODING has a denser cycle length distribution than ZHUZHOU.

Besides, Figure 9 illustrates temporal distributions of traffic signal cycle lengths and traffic flows across different hours on both datasets. We can observe cycle length and traffic flow consistently exhibit higher values during the daytime periods compared to overnight periods. Moreover, they display similar peak patterns

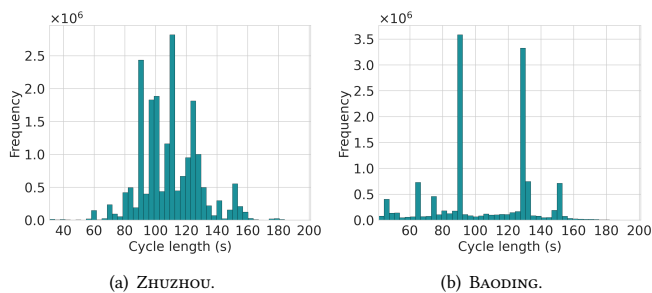


Figure 8: Overall distributions of traffic signal cycle lengths.

during the morning and evening rush hours and tend to vary in a positively correlated manner.

To further investigate the correlations between these two traffic states, we illustrate the variations in traffic flow distributions across different cycle lengths and vice versa in Figure 10. As can be seen in Figure 10(a) and Figure 10(b), traffic flow maintains an upward trend at first along with the increase of cycle length. A similar positive correlation can also be observed in Figure 10(c) and Figure 10(d), which display the variations in cycle length distributions across distinct traffic flows. However, we notice that with a further increase in cycle length, traffic flow tends to decrease. A similar situation is also shown in Figure 10(c). This can be attributed to the fact that although a positive correlation is expected between traffic flow and cycle length for the same lane, the lanes with the longest cycle lengths may not necessarily correspond to the highest traffic flows due to different traffic conditions and signal control strategies among these lanes, and vice versa.

Additionally, Figure 11 displays the spatial distributions of camera sensors and corresponding average traffic flows and cycle lengths on ZHUZHOU as a representative. It can be noticed that both traffic flows and cycle lengths exhibit remarked geographical proximity, indicating that neighboring sensors tend to have similar traffic states. This finding provides partial justification for the effectiveness of the spatial dependency modeling component, AGDN, in the irregular traffic forecasting task.

A.2 Baselines

We compare our approach with the following twelve baselines. These baseline models take the same inputs as ASEER by directly utilizing observed traffic state measurements. All these models aim to predict both traffic flow and cycle lengths by optimizing the hybrid loss function in equation Eq. (18). The autoregressive models, *i.e.*, GRU, T-LSTM, GRU-D, and DCRNN, iteratively predict the next step traffic states based on their previous predictions. Since the other non-autoregressive models require the predicted sequence length to be fixed, to enable variable-length sequence prediction, we allow them to predict in a semi-autoregressive way that they iteratively predict a fixed-length sub-sequence based on observed and previously predicted sequences. The prediction step size is set the same as ours to ensure a fair comparison. We carefully tuned major hyper-parameters of each baseline based on their recommended settings for better performance on our datasets.

LAST predicts future traffic states using the last historical traffic state measurement of each sensor. **HA** predicts future traffic states using the average of each sensor’s historical traffic state measurements. **TCN** [2] is the temporal convolutional network consisting of causal and dilated convolutions. We apply it to our datasets by padding or intercepting all the sequences to a fixed length. We stack 6 temporal convolution layers with filter size of 3. **GRU** [7] is a powerful variant of recurrent neural networks with a gated recurrent unit. **T-LSTM** [3] is a time-aware Long-Short Term Memory (LSTM) model with memory decomposition for irregular time series classification. We modify it to predict traffic states using a LSTM-based decoder. **GRU-D** [4] improves GRU with a time-aware decay mechanism for irregular time series classification. We modify it to predict traffic states using a GRU-based decoder. **mTAND** [34] is a state-of-the-art transformer-based approach for

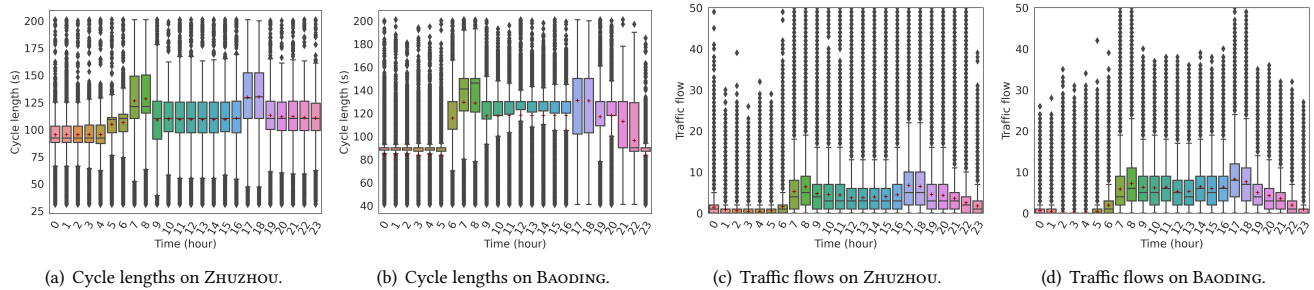


Figure 9: Temporal distributions of traffic signal cycle lengths and traffic flows across time. ‘+’ denotes mean of the box plot.

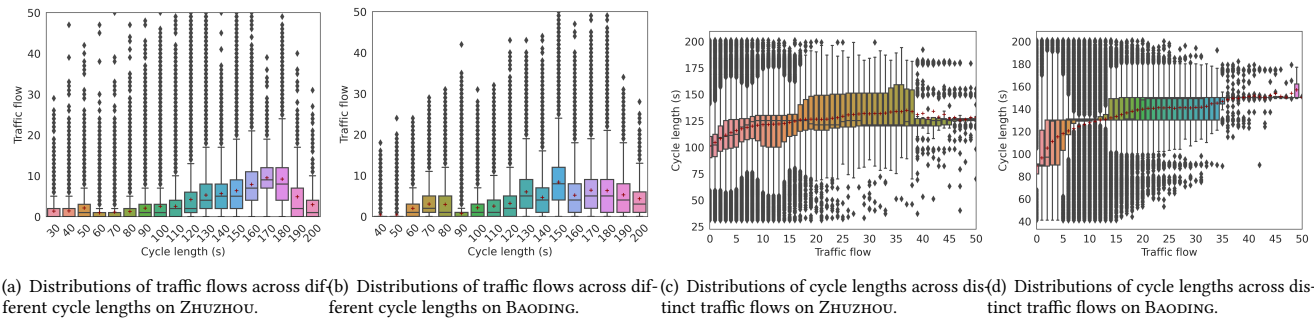


Figure 10: Correlations between traffic flow and traffic signal cycle length. ‘+’ denotes mean of the box plot.

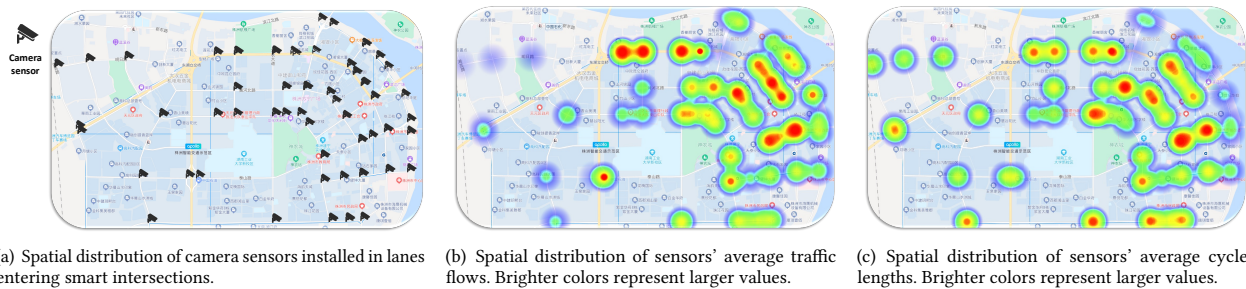


Figure 11: Spatial distributions of camera sensors and corresponding average traffic flows and cycle lengths on ZHUZHOU.

irregularly sampled multivariate time series classification and interpolation tasks. It adopts multi-time attention with time embedding to produce a fixed-length representation of a variable-length time series. The reference point number is set to 64. **Warpformer** [49] is another state-of-the-art transformer-based approach for irregularly sampled multivariate time series classification tasks. It employs a warping module to unify inputs and a doubly self-attention module for representation learning. We set the lengths of warp layers to 0, 24. **DCRNN** [22] is a representative approach based on GNNs and RNNs for classical traffic forecasting tasks, which replaces the matrix multiplications in GRU with a graph convolution operation. The used graph structure is the same as ASEER, and the diffusion step is set to 1. To apply DCRNN to our problem, we pad the input traffic sequences of all nodes to the same length. **GWNet** [44] is a representative approach based on GNNs and CNNs for classical traffic forecasting. It stacks multiple spatial-temporal blocks that are constructed by the graph convolution layer and gated TCN

layer, where the graph convolution is performed on the combination of pre-defined and self-learned adjacency matrix. The pre-defined graph structure is the same as ASEER. We stack 3 blocks with 4 convolution layers and set the convolution filter size to 3. It adopts the same padding strategy as DCRNN. **STAEformer** [25] is a state-of-the-art approach for classical traffic forecasting based on transformer and spatio-temporal adaptive embeddings. We follow the recommended settings the authors give for embedding dimensions and the number of layers and heads and adopts the same padding strategy as DCRNN. **PDFormer** [15] is another state-of-the-art transformer-based approach for classical traffic forecasting. It adopts self-attentions for both spatial and temporal dependencies modeling. A graph-masked self-attention mechanism is employed to capture both geographic and semantic spatial dependencies and a delay-aware feature transformation module is used to model the time delay in spatial information propagation. The depth of encoder layers is set to 2. It adopts the same padding strategy as DCRNN.