ST-BayesianNet: Spatiotemporal Bayesian Convolution Neural Networks for Multivariate Time Series Forecasting

Lei Wang, *Student Member, IEEE*, Huaming Wu, *Senior Member, IEEE*, Keqiu Li, *Fellow, IEEE*, and Wei Yu

***Abstract*—Multivariate time series forecasting has extensive applications across various domains, including economics, finance, bioinformatics, and intelligent transportation. The inherent spatiotemporal data is characterized by pronounced nonlinearity and stochastic uncertainty. However, current deep learning-based methods all employ deterministic parameters to characterize data features. This approach fails to effectively capture the temporal and spatial uncertainty inherent in data, resulting in limited model capability to extract data features and reduced analytical prediction accuracy. To solve this problem, this paper proposes Spatiotemporal Bayesian Convolution Neural Networks, referred to as ST-BayesianNet, for enhancing multivariate time series forecasting. Specifically, we decompose the uncertainty of spatiotemporal data into space-time dimensions, thus facilitating the prediction of multivariate spatiotemporal sequences. First, we leverage a self-adaptive uncertainty adjacency matrix to model intricate uncertain spatial relationships, while the acquisition of knowledge for this uncertain matrix hinges upon judicious a priori assumptions. Then, a non-deterministic Temporal Bayesian Convolutional Neural Network (TBCN) is constructed to adeptly capture temporal uncertainty. The optimization of model parameters, comprising both deterministic and probabilistic aspects, is achieved through variational inference. Finally, the experimental results obtained from seven real-world datasets confirm that ST-BayesianNet is more accurate than baseline methods at making predictions.**

***Index Terms*—Multivariate time series forecasting, Bayesian neural network, variational inference, uncertainty modeling.**

I. Introduction

FORECASTING multivariate spatiotemporal data is a critical task for learning systems operating in dynamic environments, as noted in [1], attracting significant attention from the deep learning community. Deep learning has become popular for extracting spatiotemporal features hidden in the data due to its robust fitting capabilities, as it can capture the overall trends across a set of dynamically changing variables. This task is vital across various domains, including autonomous vehicle operations [2], energy and smart grid optimization [3], supply chain management [4] and industrial processes [5], [6], fueling extensive research interest.

Spatiotemporal data inherently combines spatial and temporal dimensions. The spatial dimension is typically represented on graph domains, shaped by the complex topological structures of spatial networks, such as road layouts influencing traffic flow data [7]. The temporal dimension captures how data evolves over time, reflecting dynamic patterns and trends. Most multivariate spatiotemporal forecasting methods model interdependencies among variables, where each variable depends on its own historical values and those of others [8]. Effectively capturing both spatial and temporal dependencies simultaneously is a critical focus in this field [6].

Recent advancements in spatiotemporal graph modeling have generally followed two main approaches: modeling the temporal dimension and modeling the spatial dimension. For the temporal aspect, techniques aim to extract dynamic information embedded in the time dimension. For instance, convLSTM [9] enhances the traditional fully connected Long Short-Term Memory (LSTM) architecture by integrating convolutional operations, proving effective for feature extraction from spatiotemporal data. ASTGCN [10] employs pure convolutional layers to derive temporal features, while leveraging Graph Convolutional Networks (GCN) for spatial information. AST-MAGCN [11] combines Generative Adversarial Networks (GAN) with GCN, enabling real-time extraction of spatiotemporal states, with forecasting outputs refined by the GAN framework.

Conversely, the spatial dimension is addressed by synchronously integrating spatial information [12] for multivariate spatiotemporal prediction. Models like TC-GCN [13], a GCN-based approach, utilize the spatial relationship graph inherent in the data. HistGNN [14] captures multi-scale spatiotemporal dynamics, improving accuracy in complex weather forecasting across regions and timescales. LSTTN [15] introduces a transformer-based neural network for traffic flow prediction, while Beyond Spatial [16] proposes a graph-based model using multivariate transfer entropy to enhance interpretability beyond spatial neighbors. However, static spatial graph structures

Received 28 November 2024; revised 21 June 2025 and 4 October 2025; accepted 16 November 2025. This work was supported in part by the Emerging Frontiers Cultivation Program of Tianjin University Interdisciplinary Center, the Tianjin Intelligent Manufacturing Special Fund Project under Grant 20211093, and in part by Tianjin Science and Technology Planning Project under Grant 22ZYQYSN00110 and Grant 22ZYYYJC00020. *(Corresponding author: Huaming Wu.)*

Lei Wang and Keqiu Li are with the College of Intelligence and Computing, Tianjin University, Tianjin 300350, China (e-mail: wanglei2019@tju.edu.cn; keqiu@tju.edu.cn).

Huaming Wu is with the Center for Applied Mathematics, Tianjin University, Tianjin 300072, China (e-mail: whming@tju.edu.cn).

Wei Yu is with the School of International Business, Zhejiang Yuexiu University, Shaoxing 312069, China (e-mail: weiyu@zyufl.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIV.2025.3634557.

Digital Object Identifier 10.1109/TIV.2025.3634557

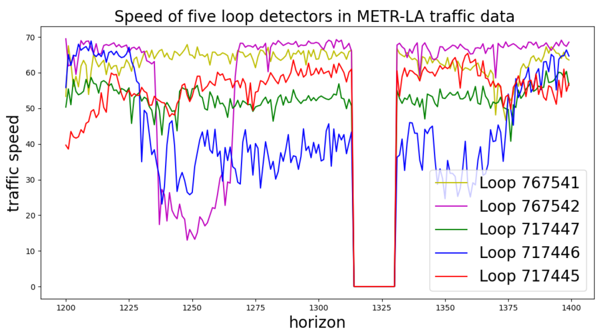


Fig. 1. The uncertainty in spatiotemporal data using speed rates from five loop detectors in the METR-LA traffic dataset. Around horizon 1325, random factors caused inaccurate measurements, with speed rates for all five detectors dropping to zero, highlighting the aleatory uncertainty inherent in real-world loop detector data.

may not always reflect true dependencies, prompting a shift toward adaptive graph modeling as a research focus. Graph-WaveNet [7] derives spatial relationships through node embeddings, while Bai et al. introduce two adaptive GCN modules: a Node Adaptive Parameter Learning (NAPL) module [17] to capture node-specific patterns, and a data-adaptive graph generation module to infer dependencies across diverse traffic series. AutoSTG [18] uses meta-learning to generate adjacency matrices for both spatial GCN and temporal CNN, modeling relationships between network parameters and meta-knowledge within the attribute graph.

Modeling and predicting multivariate spatiotemporal data through spatiotemporal models requires precision that depends on the intricate relationships and evolving characteristics of such data. The aforementioned deep learning models employ deterministic parameter operations when modeling spatiotemporal data, meaning parameters are fixed at specific values after model training. However, the complex relationships within spatiotemporal data exhibit randomness and uncertainty [19], stemming from measurement accuracy issues and the challenge of precisely fitting data features. Aleatoric uncertainty of data is usually caused by imprecise sensing instruments or data logging. As depicted in Fig. 1, all loop points recorded zero velocity around horizon 1325, which is a phenomenon commonly observed in real spatiotemporal data.

Concurrently, deterministic models such as convLSTM [9], Graph-WaveNet [7], and AutoSTG [18] often produce smooth predictions. However, real-world data frequently exhibits discontinuous, non-smooth states, indicating that deterministic models possess representational uncertainty when fitting data characteristics. The contingent uncertainty of spatio-temporal data and the representational uncertainty of model fitting underscore the importance of uncertainty modeling for multivariate spatio-temporal sequence data. Such uncertainty spatiotemporal data models must not only capture complex spatiotemporal relationships but also effectively represent the underlying uncertainty within the data, ultimately providing predictive confidence measures.

To tackle the challenges outlined earlier, this study introduces a spatiotemporal graph learning framework based on Bayesian probability, named the Spatiotemporal Bayesian Inference Network or ST-BayesianNet. This framework uses a deep neural network to model spatiotemporal interdependencies, capturing both deterministic and uncertain elements. To address latent uncertainties in the data, we separate uncertainty into temporal and spatial dimensions, applying variational inference to each. For spatial uncertainty, we develop a Bayesian Graph Convolution Network (BGCN) that enables end-to-end supervised training to learn a self-adaptive uncertainty adjacency matrix directly from the data. For temporal uncertainty, we design a Temporal Bayesian Convolutional Neural Network (BTCN) to capture temporal uncertainty while also regularizing the parameters of the entire network.

ST-BayesianNet effectively captures aleatoric uncertainties within complex spatial relationships, which is optimized globally using variational inference, delivering accurate time series predictions while simultaneously quantifying prediction uncertainty. The main contributions of this work are as follows:

- ST-BayesianNet introduces a novel deep learning framework to capture both deterministic and uncertain components of the spatiotemporal dependencies. We decompose it into temporal and spatial dimensions of uncertainty and employ variational inference methods to approximate the optimal solution for training parameters for characterizing the inherent uncertainty in spatiotemporal data. This dual-dimensional uncertainty modeling module is then integrated with deterministic spatiotemporal modeling modules to construct a globally optimized framework that simultaneously achieves uncertainty representation and enhances prediction accuracy.
- We propose a Bayesian Graph Convolutional Network (BGCN) that automatically models spatial uncertainty. This module employs a self-adaptive uncertainty adjacency matrix learned directly from the data through end-to-end supervised training. Additionally, we have designed a nondeterministic Bayesian Temporal Convolutional Network (BTCN) that captures uncertainty in the temporal dimension and regularizes the parameters of the entire network. Integrating these components enables ST-BayesianNet to effectively model uncertainty in complex spatiotemporal relationships.
- We comprehensively evaluated ST-BayesianNet on seven real-world spatiotemporal datasets. The results show a prediction error reduction of 1.2% to 4% compared to benchmark models. Additionally, visualizations of the model's output demonstrate that ST-BayesianNet generates more plausible distribution predictions, a capability not achievable by prior deterministic models.

The structure of this paper is organized as follows. In Section II, we provide an overview of related works concerning approaches to traffic prediction. Section III delves into the details of ST-BayesianNet. The performance evaluation of ST-BayesianNet is presented in Section IV, encompassing prediction results and an analysis of its resilience to perturbations. Finally, Section V concludes the paper, summarizing the findings and contributions.

II. Related Work

As a key area of multivariate time series analysis, research on spatio-temporal forecasting models has received significant attention due to its ability to handle complex nonlinear data patterns. This paper focuses on studies of deterministic and uncertain spatio-temporal forecasting models. This section highlights breakthrough achievements in the relevant research progress of deterministic spatio-temporal neural network models and uncertain Bayesian neural network models, which form the basis of this study.

A. Spatiotemporal Neural Network Models

Spatio-temporal neural network models hold significant theoretical importance as they aim to capture the intrinsic relationships between future data points and historical observations within spatio-temporal datasets, thereby enabling high-precision spatio-temporal forecasting. Such models typically integrate spatio-temporal information through joint modeling and represent widely recognized and extensively studied deterministic spatio-temporal forecasting frameworks.

Existing deep learning models employ various architectures tailored to different attributes of the spatiotemporal dimension to extract latent feature information from the data and facilitate accurate prediction. For instance, CNN [20] or RNN [21], [22] based methods have been widely utilized to capture temporal patterns. More recently, GCNs have gained popularity in modeling spatial relationships, where the adjacency matrix, often based on distance information, delineates the spatial connections between monitoring points [23], [24].

Graph neural networks have a wide range of applications in the field of spatio-temporal data forecasting. For instance, Wu et al. [7] proposed a novel Graph Convolutional Neural Network architecture, termed Graph-WaveNet, designed specifically for spatiotemporal graph modeling. The methodology integrates adaptive dependency matrices derived from node embeddings, thereby enhancing the model's ability to discern and leverage the intrinsic spatial dependencies embedded within the input data. DCRNN [25] is a model that represents traffic flow as a diffusion process across a directed graph, and introduced a convolutional recursive neural network architecture that is based on diffusion principles. This deep learning framework is established as a robust approach for traffic forecasting, adeptly capturing and intertwining the spatial and temporal inter-dependencies that are characteristic of traffic flow patterns. The Spatiotemporal Graph Convolutional Network(ST-GCN) [26], integrates graph convolutional modules to model spatial dependencies and temporal dynamics for accurate traffic prediction.

Additionally, GMAN [27] employed a multi-graph attention mechanism within a deep network architecture, which executes attentional computations across spatial as well as temporal domains, thereby enabling a comprehensive analysis of multi-dimensional data. Nevertheless, these methods suffer from limitations in accuracy and applicability due to their neglect of modeling the uncertainty and data drift characteristics inherent in spatiotemporal data.

B. Bayesian Neural Network Models

The inherent spatio-temporal relationships within spatiotemporal data often exhibit high complexity and uncertainty. Consequently, some researchers have turned to Bayesian neural network models. These approaches address the challenge of quantifying uncertainty in spatiotemporal data by incorporating neural networks with probabilistic model parameters. For instance, Gal et al. [28] proposed Bayesian Convolutional Neural Networks (CNNs), which is the first Bayesian approach used to CNNs that leverages Bernoulli variational inference to combat over-fitting in small datasets, providing a robust framework for uncertainty estimation and improved classification accuracy. Chandra et al. [29] introduced Bayesian graph CNNs that leverage tempered Markov chain Monte Carlo (MCMC) sampling via parallel computation, employing Langevin gradient proposal distributions to address the quantification of uncertainty in the analyzed sample data. This innovative approach extends the traditional application of graph CNNs by integrating Bayesian inference to model the inherent uncertainty in spatial data more effectively.

DeepAR [30] employed an RNN architecture for probabilistic forecasting, utilizing simplified temporal convolutional layers to reduce parameter count and LSTM-units to capture temporal dynamics. The method employs an auto-regressive approach that incorporates Gaussian-distributed stochastic error terms to reduce prediction errors. DeepAR, an efficient forecasting technique, leverages these random features to capture and analyze the inherent uncertainty in temporal data.

In summary, the intricate interplay between time and space in spatiotemporal data is characterized by nonlinear and uncertain relationships. However, the previously discussed Spatiotemporal methods are inherently deterministic, meaning that they yield a fixed output for a given input once the model parameters have been determined. Furthermore, existing Bayesian neural network methods can only handle uncertainty analysis for small-scale data or focus on uncertainty in either the temporal or spatial dimension during data analysis and prediction tasks. Consequently, mainstream spatio-temporal and uncertainty methods have failed to fully capture the uncertainty arising from the joint temporal and spatial dimensions, leaving room for improvement in the data analysis and prediction performance of these models.

III. Methods

A. Problem Definition

Definition 1: **Spatial Network** G. A weighted undirected graph $G = (V, E, A)$ is used to describe the spatial topological structure or semantic relationship in the spatiotemporal data, where $V = \{v_0, v_1, \ldots, v_N\}$ is treated as N monitoring vertices, and E is expressed as a set of edges. We use $A \in R^{N\times N}$ to represent the adjacency matrix of G, which is the weight matrix in this paper. In some cases, there is more than one spatial network, i.e., we will have multiple adjacency matrices $\{A_1, A_2, \ldots, A_k\}$.

Definition 2: **Feature Matrix** X. The information on the spatial network G is regarded as the node attribute features V,

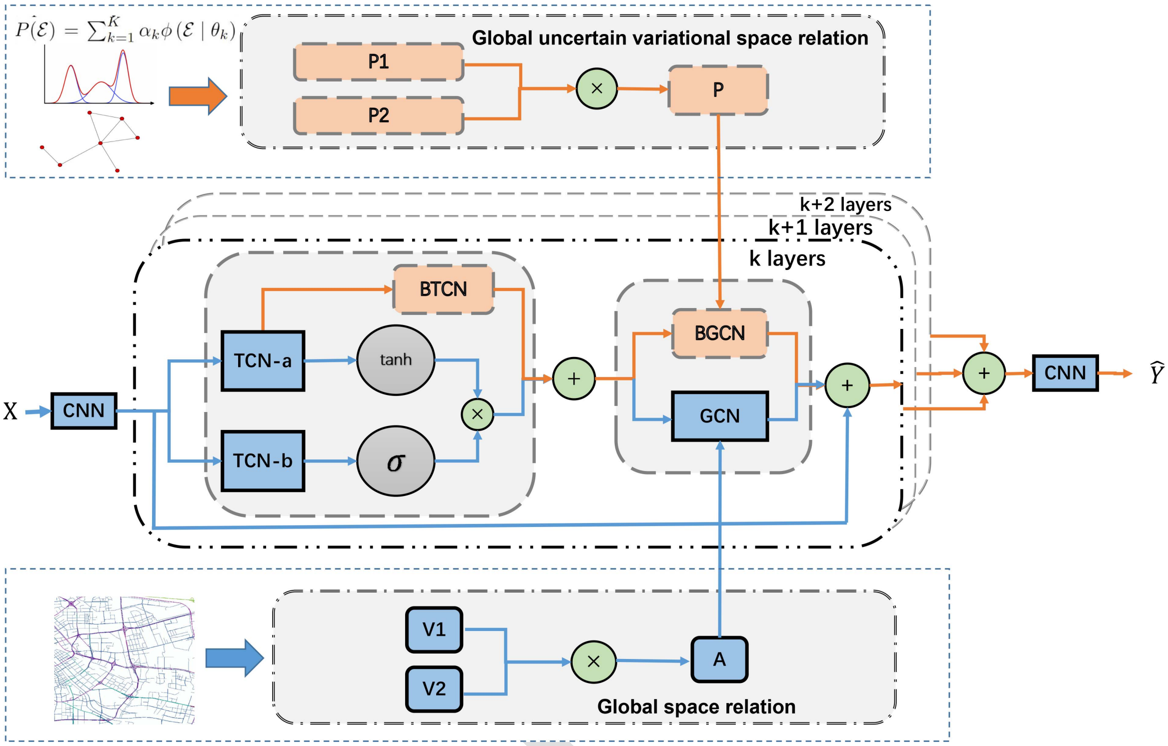


Fig. 2. The overall architecture of ST-BayesianNet. It comprises two main branches: the deterministic computing branch (blue), which captures deterministic patterns in the data, and the non-deterministic computing branch (orange), which models uncertainty. Each branch is composed of multiple spatiotemporal dependency blocks, designed to independently extract deterministic and stochastic information from the spatiotemporal data.

which is indicated by $X \in R^{T\times N\times C}$, where C, T and $N = |V|$ are the number of node attribute features such as traffic speed, number of traffic vehicles, the length of the historical times series and the number of spatial sensor nodes, respectively.

The problem of multivariate time series forecasting in this paper is considered as predicting future features $\hat{Y} = (\hat{x}_{T+1}, \ldots, \hat{x}_{T+Q})$ from current data $X = (x_1, \ldots, x_T)$. Through the above definition, The mapping function f_θ from X to $\hat{Y}$ should be learned, that is:

$$\hat{Y} = f_\theta(G, X), \tag{1}$$

where θ is the parameter of the model. The real future data is $Y = (x_{T+1}, \ldots, x_{T+Q})$ and the training process is to make the distance between predicted output $\hat{Y}$ and Y smaller and smaller.

B. Overview of ST-BayesianNet

Fig. 2 illustrates the overall architecture of the ST-BayesianNet model, which comprises two primary branches: a deterministic computing branch and a non-deterministic computing branch. The framework is structured into multiple blocks. Vertically, it is segmented into a time-dependent module (BTCN) and a space-dependent module (BGCN), aligning with the structure commonly found in existing deep spatiotemporal prediction frameworks.

In the deterministic computing branch, deterministic spatiotemporal features are extracted. In contrast, the non-deterministic branch captures stochastic patterns in the spatiotemporal data using the BTCN and BGCN modules. Finally, the model output integrates both the deterministic and uncertain components, providing a comprehensive representation of the spatiotemporal dynamics.

In the remainder of this section, we present a comprehensive overview of the core components that make up ST-BayesianNet. The framework is organized into multiple blocks, each containing deterministic temporal and spatial models. To illustrate the design and functionality, we examine a representative block in detail, noting that the structure and operation of the remaining blocks follow a similar repetitive pattern.

C. Deterministic Spatiotemporal Modeling

Spatiotemporal data inherently comprises deterministic components in both its temporal and spatial dimensions. For instance, the power consumption within distinct regions exhibits discernible periodicity, influenced by seasonal fluctuations [31]. Similarly, in the domain of traffic forecasting, the intricate spatial interconnections of road networks play a pivotal role in governing traffic flow dynamics [7]. Therefore, formulating effective methodologies to appropriately capture and model these deterministic spatiotemporal constituents stands as a pivotal determinant of the predictive accuracy and quality in the realm of multivariate time series forecasting.

The ST-BayesianNet framework enables multidimensional uncertainty modeling of spatiotemporal data by integrating BGCN for spatial randomness and BTCN for temporal randomness. This combined approach allows for the generation of predictions comprising deterministic and uncertain elements.

The dual-branch architecture ensures that the model can accurately predict multivariate time series while also assessing prediction uncertainty in complex spatiotemporal tasks. Through variational inference, the training process determines the optimal solution for the infinite approximation parameter equation, enhancing the model's robustness and practicality for real-world applications. We provide a detailed introduction to the model's components next.

In addressing the temporal dimension, ST-BayesianNet employs Gated TCN [7] due to their lean parameter count and straightforward architectural design. Let $X \in R^{T\times N\times C}$ represent the historical input spatiotemporal data. To ensure consistent feature extraction across the entire model, ST-BayesianNet initiates a process of dimension normalization using a CNN with a 1×1 kernel size. This normalization step is outlined as follows:

$$X_s = CNN(X) \in R^{T\times N\times D}, \tag{2}$$

where CNN refers to a two-dimensional convolution operation, and D signifies the standardized feature dimension. Notably, owing to the utilization of the 1×1 convolution operation, the described process does not compromise the inherent spatiotemporal information within the original data.

To extract meaningful temporal information, we incorporate the Gated TCN into our approach. Gated TCN is chosen for its ability to effectively manage the flow of information across layers within temporal convolutional networks. For an input tensor $X_s \in \mathbb{R}^{T\times N\times D}$, the output of the Gated TCN, using time-dilated causal convolution, is computed as follows:

$$X_{dT} = g(X_s \star K_a + a) \odot \sigma(X_s \star K_b + b) \in R^{T_{dT}\times N\times D}, \tag{3}$$

where $\star$ represents a one-dimensional convolution along the time direction with parameters K_a, K_b, a and b. $\odot$ is the element-wise product. $\sigma(.)$ is the sigmoid function and $g(.)$ is the activation function, where $\tanh$ serves as the specific activation function in this paper. In order to increase the perception field of the temporal dimension and enhance the computational efficiency of the model, we employ dilated causal convolution. Consequently, the output dimensionality of the temporal aspect becomes $T_{dT} < T$.

In comparison to univariate prediction, the integration of spatial network information into prediction models is pivotal for capturing the intricacies of spatiotemporal data. Nonetheless, in real-world scenarios, many spatiotemporal datasets struggle to effectively capture spatial topological relationships. Hence, ST-BayesianNet employs the adaptive adjacency matrix method [7], wherein the following spatial relations are defined:

$$A = SoftMax(ReLu(E_1 E_2^T)), \tag{4}$$

where E_1 and E_2 are of dimensions $N \times d$ and represent the embeddings of source nodes in the spatial graph, and d signifies the embedding dimension. Subsequently, this allows for the capturing of hidden spatial dependencies, which can be expressed

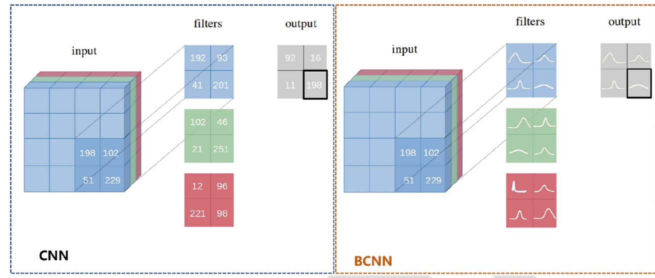


Fig. 3. The difference between BCNN and CNN. The convolution kernel of CNN (left) is a definite value, while the convolution kernel of BCNN (right) is the distribution of data. value [32].

as follows:

$$X_{dS} = \sum_{k=0}^{K} A^k X_{dT} W_k, \tag{5}$$

where X_{dT} is the input of GCN, which is also the output of Gate TCN in (3), K is the GCN order, and W_k is the parameter of each GCN layer.

D. Uncertainty Modeling

The real world is inherently imbued with uncertainty, and this characteristic extends to the realm of spatiotemporal data. Prevailing deep learning models, including LSTM, CNN, and transformers, are deterministic in nature and susceptible to overfitting. Thus, addressing the intrinsic uncertainty within spatiotemporal data becomes imperative. The framework of ST-BayesianNet is tailored to this challenge, wherein we undertake the modeling of temporal and spatial uncertainties distinctively.

1) Temporal Uncertainty Modeling: In order to model time uncertainty, the ST-BayesianNet employs the Bayesian Convolution Network (BCNN) [32], which is the core part of BTCN, underpinned by variational inference principles. A noteworthy distinction between BCNN and CNN lies in the convolution parameters, wherein BCNN introduces an element of randomization, as visually illustrated in Fig. 3.

Let X_s denote the input to the BCNN. Consequently, the output of BCNN takes the form of a random variable.

$$\mathcal{X}_{uT} = X_s \star \mathcal{K} + \mathcal{B} = BCNN_{\mathcal{W}}(X_s), \tag{6}$$

where $\mathcal{X}$ and $\mathcal{W} = \{\mathcal{K}, \mathcal{B}\}$ serve as random variables. $\mathcal{W}$ represents the parameter of the BCNN, adhering to a Gaussian prior distribution, i.e., $p(\mathcal{W}) = N(\mu, \sigma)$. As the direct derivation of the posterior distribution $p(\mathcal{W} \mid D)$ is impractical, we pivot towards optimizing the variational posterior distribution $q(\mathcal{W}|\theta_t)$ to minimize the divergence $KL(p||q)$ until it approaches zero, where $D = \{(x_i, y_i)\}_{i=0}^{n}$ stands for the training data, $KL(.||.)$ represents the Kullback-Leibler divergence between two distributions, and θ_t corresponds to the controlling parameter of $q(\mathcal{W}|\theta_t)$. According to the variational inference principle [33], the optimization objective is formulated as follows:

$$\mathcal{L}(\theta_t) = -\mathbb{E}_{q(\mathcal{W}|\theta_t)}\left[\log\left[\frac{q(\mathcal{W} \mid \theta_t)}{P(D \mid \mathcal{W})P(\mathcal{W})}\right]\right]$$

$$= KL(q(\mathcal{W} \mid \theta_t)||P(\mathcal{W})) + \mathbb{E}_{q(\mathcal{W}|\theta_t)}(P(D \mid \mathcal{W})). \tag{7}$$

Since determining the uncertainty of the temporal dimension is challenging, the posterior distribution $q(\mathcal{W} \mid \theta_t)$ still obeys a Gaussian distribution, similar to the approach in [33].

2) Deterministic and Uncertainty Temporal Dependence (TBN): The temporal dimension is determined by both the provided information and non-deterministic information. Thus, we can combine these two sources to generate the time dimension information feature output. This involves fusing information for (3) and (6), resulting in the following expression:

$$\mathcal{X}_T = X_{dT} + \alpha \mathcal{X}_{uT}, \tag{8}$$

where α is the fusion parameter and can be learned in training.

3) Spatial Uncertainty Modeling: Due to the inherent complexity of spatial relationships in spatiotemporal data, existing models often exhibit model uncertainty (i.e., epistemic uncertainty) [34] when capturing spatial dependencies. Traditional GCNs, such as those defined in (5), are particularly susceptible to overfitting in such settings. To address these challenges, we propose an uncertainty-aware GCN model named BGCN, within the ST-BayesianNet framework. BGCN explicitly models spatial uncertainty, improving robustness and generalization in spatiotemporal learning tasks.

Let the attribute values of the spatial detection points $V = \{v_0, \ldots, v_N\}$ be random vectors $\mathcal{E}_1, \mathcal{E}_2 \in R^{N\times d}$, where $d < N$ is the embedded dimension, then the uncertain spatial relationship in spatiotemporal data is defined as follows

$$\mathcal{P} = softMax(Relu(\mathcal{E}_1 \mathcal{E}_2^T)) \in R^{N\times N}. \tag{9}$$

Due to the complexity of spatial relationship, we assume that $\mathcal{E}_1$ and $\mathcal{E}_2$ priori obey mixed Gaussian distribution, that is $\mathcal{E}_1, \mathcal{E}_2 \sim P(\mathcal{E}) = \sum_{k=1}^{K} \alpha_k \phi(\mathcal{E} \mid \theta_k)$, where $\theta_k = (\mu_k, \sigma_k)$ are the parameters of Gaussian distribution $\mathcal{N}(\mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp(-\frac{(x-\mu_k)^2}{2\sigma_k^2})$, and α_k are the mixed parameters. Based on the above concept, ST-BayesianNet can be expressed as

$$\mathcal{X}_{uS} = \sum_{k=0}^{K} \mathcal{P}^k \mathcal{X}_T W_k, \tag{10}$$

where W_k is the parameter of ST-Bayesian, and $\mathcal{X}_T$ is the output of (8).

The inference of the randomization parameter required by ST-BayesianNet is denoted as $\mathcal{E} = (\mathcal{E}_1, \mathcal{E}_2)$. However, expressing the posterior probability $P(\mathcal{E} \mid D)$ directly is challenging. To address this, we utilize a variational distribution $q(\mathcal{E} \mid \theta_s) \sim \mathcal{N}(\mu_w, \sigma_w)$ to approximate the actual posterior distribution $P(\mathcal{E} \mid D)$. Here, θ_s represents the parameters of the variational distribution.

Applying the principles of variational inference, we can derive the following equation:

$$\begin{aligned}
\theta_w^* &= \underset{\theta_s}{\operatorname{argmin}} \mathrm{KL}[q(\mathcal{E} \mid \theta_s) \| P(\mathcal{E} \mid D)] \\
&= \underset{\theta_s}{\operatorname{argmin}} \mathbb{E}_{q(\mathcal{E}|\theta_s)} \left[\log \left[\frac{q(\mathcal{E} \mid \theta_s)}{P(\mathcal{E} \mid D)} \right] \right] \\
&= \underset{\theta_s}{\operatorname{argmin}} \mathbb{E}_{q(\mathcal{E}|\theta_s)} \left[\log \left[\frac{q(\mathcal{E} \mid \theta_s) P(D)}{P(D \mid \mathcal{E}) P(\mathcal{E})} \right] \right] \\
&= \underset{\theta_s}{\operatorname{argmin}} \mathbb{E}_{q(\mathcal{E}|\theta_s)} \left[\log \left[\frac{q(\mathcal{E} \mid \theta_s)}{P(D \mid \mathcal{E}) P(\mathcal{E})} \right] \right] \\
&= \underset{\theta_s}{\operatorname{argmin}} \left[\mathrm{KL}(q(\mathcal{E} \mid \theta_s) || P(\mathcal{E})) - \mathbb{E}_{q(\mathcal{E}|\theta_s)} \left[\log P(D \mid \mathcal{E}) \right] \right].
\end{aligned} \tag{11}$$

Thus, the loss function of ST-BayesianNet is

$$\begin{aligned}
\mathcal{L}(\theta_s) = \underset{\theta_s}{\operatorname{argmin}} & \left[\mathrm{KL}(q(\mathcal{E} \mid \theta_s) || P(\mathcal{E})) \right. \\
& \left. - \mathbb{E}_{q(\mathcal{E}|\theta_a)} \left[\log P(D \mid \theta_s) \right] \right].
\end{aligned} \tag{12}$$

4) Spatial Deterministic and Uncertain Information Fusion: According to (5) and (10), the final spatial feature is the combination of the deterministic and uncertain information from the spatial dimension, which can be written as follows:

$$\mathcal{X} = X_{dS} + \mathcal{X}_{uS}. \tag{13}$$

As indicated by (3), (5), (6), (10), and (13), $\mathcal{X}_{gcn}$ combines the spatiotemporal deterministic and uncertain information present in the spatiotemporal data. To facilitate the training of the model, a shortcut has been incorporated into ST-BayesianNet, denoted as:

$$\mathcal{X}^{(b)} = X_s + \mathcal{X}_{gcn}, \tag{14}$$

where X_s represents the input of this block (as shown in (2)), and $\mathcal{X}^{(b)}$ denotes the output of the current block b.

E. The Output of ST-BayesianNet

Since the ST-BayesianNet model consists of multiple blocks, the output of the model needs to integrate the outputs of multiple blocks. The output of all blocks is represented by:

$$\mathcal{X}^{(o)} = \sum_{b=0}^{B} \sigma(conv(\mathcal{X}^{(b)})), \tag{15}$$

where σ is the activation function, such as ReLU, and B represents the number of blocks in ST-BayesianNet. Therefore, the random output of ST-BayesianNet is given by:

$$\hat{\mathcal{Y}} = conv(\mathcal{X}^{(o)}). \tag{16}$$

In essence, ST-BayesianNet is a randomized model, so the output of ST-BayesianNet is a random variable $\hat{\mathcal{Y}}$. All random parameters of the ST-BayesianNet model are in (6) and (9), and the randomization parameters in all blocks are recorded as ξ. All random parameters ξ are sampled from their corresponding posterior distribution $q(\xi)$ and sent to the final expectation in the network as the deterministic output of the most general model, i.e.

$$\hat{Y} = \mathbb{E}_{\xi \sim q(\xi)} \hat{\mathcal{Y}}(\xi) \approx \sum_{s=0}^{S} \hat{\mathcal{Y}}(\xi_s), \tag{17}$$

where S is the number of samples.

F. The Whole Loss

According to (7) and (12), we can obtain the loss function of the whole random parameter ξ

$$\mathcal{L}(\theta, w) = \arg\min_{\theta, w} \left[\mathrm{KL}(q(\xi \mid \theta) || P(\xi)) - \mathbb{E}_{q(\xi|\theta)} \left[\log P(D \mid \theta)\right]\right], \quad (18)$$

where $\theta = (\theta_t, \theta_s)$ represents the control parameters of the posterior probability distribution, and w denotes the network parameters of ST-BayesianNet. The loss function comprises two components. The first term quantifies the proximity between the prior distribution and the variational prior distribution, while the second term accounts for the data likelihood. For the sake of training convenience, we express the complete loss function in the following manner:

$$\mathcal{L}(\theta, w) = \arg\min_{\theta, w} \left[\mathrm{KL}(q(\xi \mid \theta) || P(\xi)) + \beta * L_{Huber}(Y, \hat{Y})\right], \quad (19)$$

where β is the hyperparameter that balances the influence of likelihood and KL divergence and

$$L_{Huber}(\hat{X}, Y) = \begin{cases} \frac{1}{2}(Y - \hat{X})^2 & |Y - \hat{X}| \leq \delta, \\ \delta|Y - \hat{X}| - \frac{1}{2}\delta^2 & otherwise. \end{cases} \quad (20)$$

IV. Performance Evaluation

A. Experimental Setup

1) Datasets: The predictive performance of ST-BayesianNet is evaluated using two publicly available traffic spatiotemporal datasets, i.e., METR-LA and PEMS-BAY. These datasets are accessible through the open-source code provided in the literature.

- **METR-LA [7]**: It comprises real-time traffic speed data collected from loop detectors installed on highways in Los Angeles County. The dataset covers the time period from March 1 to March 7, 2012, and consists of data from 207 sensors. The traffic speed readings are taken at 5-minute intervals, and the adjacency matrix is constructed based on the spatial distances between the sensors in the traffic network.
- **PEMS-BAY [7]**: This data was meticulously collected through the California Department of Transportation's performance measurement system, covering the period from January 1 to May 31, 2017. It incorporates data from a total of 325 sensors, with each data point sampled at precise 5-minute intervals. The 325×325 adjacency matrix is constructed based on the spatial relationships among roads in the network.
- **solar[1]**: It captures solar power output and environmental conditions from two solar power plants over a 34-day period, specifically from May 15 to June 17, 2020. Collected on an hourly basis, the data includes variables such as temperature, humidity, solar irradiance, and power output.
- **traffic[2]**: It provides hourly traffic flow data collected from automated sensors at 4 key junctions, capturing variations in vehicle counts over an unspecified period.
- **PSM04&PSM08[3]**: A comprehensive collection of real-time traffic data gathered from loop detectors on California's State Route 4 (PSM04)/ State Route 8 (PSM08)highway, offering a robust resource for analyzing and forecasting traffic patterns.

2) Metrics: To evaluate the performance of ST-BayesianNet, we employ two established metrics, Mean Squared Error (MSE) and Mean Absolute Error (MAE), to quantify relative prediction error. Smaller values of MSE and MAE correspond to improved prediction accuracy.

- MSE (Mean Squared Error): It emphasizes larger errors due to its squaring nature and can be computed as:

$$MSE = \frac{1}{QN} \sum_{t=1}^{Q} ||Y_t - \hat{Y}_t||_2^2. \quad (21)$$

- MAE (Mean Absolute Error): It gives an average magnitude of errors without squaring, and can be computed as:

$$MAE = \frac{1}{QN} \sum_{t=1}^{Q} |Y_t - \hat{Y}_t|. \quad (22)$$

3) Baselines: In order to verify the performance of ST-BayesianNet, we introduce several baselines in different approaches:

- **FNN [35]**: A feed-forward neural network designed for time series prediction, capable of capturing complex patterns and dependencies in sequential data.
- **GRU [36]**: A variant of RNNs that mitigates the vanishing gradient problem through gating mechanisms, enhancing its ability to learn long-term dependencies.
- **AGCRN [17]**: A model that dynamically adapts to traffic patterns by learning node-specific parameters and generating data-driven graphs for improved traffic forecasting.
- **GCN [37]**: A neural network architecture designed to process graph-structured data, where nodes and edges are associated with features.
- **STGCN [38]**: A Spatio-Temporal Graph Convolutional Network that integrates graph convolutional modules to model spatial dependencies and temporal dynamics for accurate traffic prediction.
- **TCGCN [39]**: An advanced model that combines community-enhanced graph convolutional networks with attention mechanisms to capture complex spatiotemporal patterns in traffic data.
- **STHSL [40]**: A Spatial-Temporal Self-Supervised Hypergraph Learning framework that addresses label scarcity in crime prediction by capturing cross-region dependencies and temporal patterns.

[1] [Online]. Available: https://www.kaggle.com/datasets/anikannal/solar-power-generation-data

[2] [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/traffic-prediction-dataset

[3] [Online]. Available: https://gitcode.com/open-source-toolkit/06a2f/overview

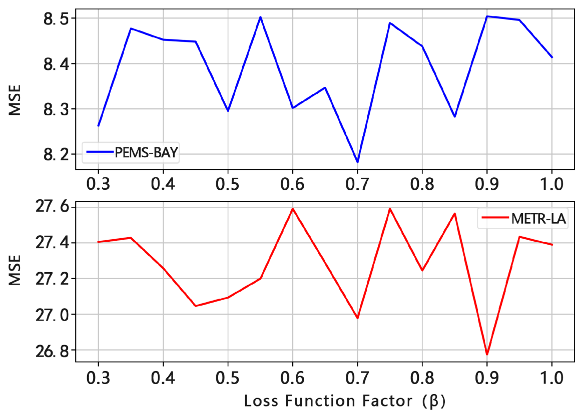


Fig. 4. The impact on MSE of changes in hyperparameter $beta$ within datasets PEMS-BAY(top) and METR-LA(bottom). This indicates that $beta$ is sensitive to these two datasets, and adjusting it can help optimize model performance.

- **Crossformer [41]**: A Transformer-based model that leverages cross-dimensional dependencies for multivariate time series forecasting.
- **Graph-WaveNet [7]**: A GNN-based spatiotemporal model that uses a learnable dependency matrix to capture both spatial and temporal correlations, enabling the modeling of long-range dependencies.

The selection of baseline methods in comparative trials highlights the unique characteristics and application domains of each approach. Feedforward Neural Networks (FNN) and Gated Recurrent Units (GRU) demonstrate proficiency in handling sequential data by capturing intricate patterns and long-term dependencies, making them suitable for general time series analysis and speech processing tasks. Graph-based models such as Graph Convolutional Networks (GCN), Spatial-Temporal Graph Convolutional Networks (STGCN), Temporal Convolutional Graph Convolutional Networks (TCGCN), and Graph WaveNet integrate spatial-temporal dynamics and attention mechanisms, making them particularly effective for traffic forecasting and complex spatiotemporal modeling. The Adaptive Graph Convolutional Recurrent Network (AGCRN) further enhances performance by dynamically adapting to evolving traffic conditions. Additionally, models like Spatial-Temporal Hierarchical Self-Supervised Learning (STHSL) and Crossformer are designed to address challenges such as label scarcity and multivariate time series forecasting, supporting a wide range of regional prediction tasks. All baseline models are implemented using the optimal hyperparameter settings as specified in their original publications.

4) The Model Parameters: The convolution layers utilize a kernel size of 3 with 1 padding to maintain the output shape. ST-BayesianNet consists of 3 blocks, and the weight β in the loss function of (19) is set to 0.7 for the METR-LA dataset and 0.5 for the PEMS-BAY and other datasets, as determined through experimentation.

Fig. 4 presents the parameter selection experiment for β across the two databases. It's evident that β significantly impacts the test MSE of the ST-BayesianNet model. The optimal β varies for different databases, as the model necessitates adjusting parameters based on the database to strike a balance between prior fitting and data likelihood.

5) Other Settings: To ensure a fair comparison, the loss functions employed for baselines are all Huber loss (as specified in (20)), chosen for their demonstrated effectiveness. Additionally, a batch size of 64 is consistently used across all models. During the experimental setup, the dataset is partitioned into training, validation, and test sets at a ratio of $8 : 1 : 1$. Following 50 epochs of training, the model with the best performance on the validation set is selected for testing on the test set.

B. Comparison of Prediction Accuracy

Table I presents a comprehensive comparison of various methods for predicting three, six and twelve time steps (e.g., 15, 30 and 60 minutes) on the specified datasets. The horizontal axis of the table shows the comparison methods and testing metrics, and the vertical axis shows the datasets and time steps. To address concerns about the depth, refinement and applicability of the evaluation to large-scale scenarios, extensive experiments were conducted across six diverse spatiotemporal datasets. These datasets encompass real-world traffic, energy and power system data, with scale varying from hundreds to potentially thousands of nodes when considering interconnected systems. The evaluation metrics include mean squared error (MSE) and mean absolute error (MAE), with additional analyses on fault tolerance (robustness to noise), computational complexity (inference time and space), ablation studies and uncertainty visualisation providing a more refined and in-depth assessment. Upon inspection, the following conclusions can be drawn:

ST-BayesianNet has been evaluated on five spatiotemporal traffic datasets, including METR-LA, PEMS-BAY, and several others. Across these benchmarks, it consistently outperforms most competing methods. Notably, ST-BayesianNet demonstrates a significant performance advantage over widely adopted temporal models such as STGCN, TCGCN, Crossformer, and AGCRN, achieving substantial improvements on the majority of datasets. It is worth emphasizing that these baseline models are well-established and commonly used deep learning approaches for multivariate time series analysis, further underscoring the effectiveness of ST-BayesianNet.

The model maintains a high level of performance, particularly in prediction tasks involving large-scale traffic datasets containing nearly a thousand node information points. This demonstrates that ST-BayesianNet also possesses outstanding predictive capabilities for large-scale practical applications. This is thanks to its Bayesian Time-Convolutional Network (BTCN) and Bayesian Graph Convolutional Network (BGCN), which are highly effective at representing and analysing the uncertainty inherent in real-time application data.

Comparative experimental results also indicate that models that integrate spatiotemporal information, such as TCGCN, ST-BayesianNet and Graph-WaveNet, perform better than models that rely solely on temporal or spatial approaches. ST-BayesianNet achieves an MSE reduction of over 4.2% compared to other spatio-temporal fusion baselines across all step-length prediction tests on real-world traffic datasets such as METR-LA

TABLE I

THE PERFORMANCE COMPARISON RESULTS OF THE ST-BAYESIANNET FOR FORECASTING ON METR-LA, PEMS-BAY, SOLAR, TRAFFIC, PSM04 AND PSM08 DATASETS (THE HORIZONS ARE 3 STEPS, 6 STEPS, AND 12 STEPS, RESPECTIVELY). AND **BOLD** INDICATES THE BEST ACCURACY, UNDERLINED INDICATES THE SECOND BEST, THE ITALICIZED NUMBERS SIGNIFY THE FOURTH AND '-' INDICATES THAT THE MODEL DOES NOT CONVERGE

Methods		ST-bayesianet		FNN		GRU		AGCRN		GCN		STGCN		TCGCN		STHSL		Crossformer		Graph-Wavenet	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
METR-LA	15m.	**20.032**	2.448	28.425	3.051	32.238	2.952	24.046	2.587	56.336	4.756	95.019	3.849	20.538	**2.446**	32.120	2.974	46.352	2.577	20.490	2.504
	30m.	**27.254**	**2.706**	36.120	3.307	43.302	3.290	30.795	2.822	66.272	5.071	103.960	5.211	27.383	2.716	43.686	3.281	72.592	3.926	29.055	2.774
	1h.	**36.995**	3.044	46.612	3.680	59.338	3.860	40.884	3.154	83.939	5.613	119.961	5.359	37.112	**3.033**	64.772	4.251	96.196	4.437	37.977	3.117
PEMS-BAY	15m.	5.016	1.099	11.647	1.833	6.741	1.226	5.522	1.143	44.936	2.933	131.055	5.344	**4.909**	**1.097**	6.337	1.213	8.443	1.422	5.153	1.121
	30m.	8.359	1.327	15.354	2.019	12.536	1.536	8.730	1.346	46.573	3.056	115.063	5.036	**8.281**	**1.309**	11.972	1.513	16.656	1.811	9.689	1.393
	1h.	**13.102**	**1.576**	19.639	2.202	19.294	1.892	13.170	1.588	49.810	3.250	162.193	5.776	13.154	1.606	20.581	1.989	-	-	13.922	1.637
solar	30m.	**4.132**	**1.011**	12.642	2.268	11.741	2.312	5.262	1.224	58.600	4.982	166.111	10.051	5.013	1.145	14.736	2.577	16.405	2.500	10.136	1.942
	1h.	**6.472**	**1.360**	12.884	2.310	17.766	2.855	8.637	1.702	103.157	6.822	371.089	13.841	8.787	1.636	14.878	2.523	38.394	3.923	19.868	2.889
	2h.	**11.540**	**1.952**	16.765	2.644	30.924	3.700	15.267	2.373	98.217	6.730	211.283	10.065	15.665	2.337	129.028	8.194	44.714	4.035	36.795	4.233
traffic	3h.	0.001	**0.015**	0.001	0.015	0.002	0.019	0.002	0.027	0.002	0.029	0.006	0.048	**0.001**	0.015	0.002	0.026	-	-	0.001	0.015
	6h.	0.002	0.018	**0.001**	**0.017**	0.002	0.024	0.003	0.034	0.003	0.031	0.005	0.042	0.002	0.018	0.003	0.027	-	-	0.002	0.019
	12h.	0.002	0.021	**0.002**	**0.019**	0.002	0.021	0.002	0.022	0.003	0.030	0.004	0.041	0.002	0.021	0.003	0.027	-	-	0.002	0.022
PSM04	15m.	**823.702**	**18.002**	1.5K	25.167	1.1K	20.870	940.633	19.575	4.5K	40.942	7.7K	72.171	999.370	19.877	951.134	19.894	1.2K	22.615	983.639	20.261
	30m.	**850.817**	**18.376**	1.7K	27.176	1.2K	22.538	1.0K	20.645	4.7K	42.580	7.2K	70.217	1.0K	20.554	947.983	19.721	2.0K	30.386	1.1K	21.907
	1h.	**920.785**	**18.909**	1.7K	27.550	1.5K	24.818	1.0K	21.002	5.0K	44.788	7.9K	72.972	973.879	19.915	1.0K	20.938	2.4K	33.705	1.1K	21.293
PSM08	15m.	**497.753**	**14.326**	1.1K	23.073	631.344	16.374	542.285	15.212	3.3K	37.382	5.7K	65.827	598.585	15.884	582.550	16.016	774.455	19.768	572.157	15.728
	30m.	**514.480**	**14.450**	1.3K	24.639	780.775	17.956	631.951	16.358	3.4K	38.847	8.7K	81.033	572.612	15.341	604.886	16.024	1.8K	28.380	678.878	17.260
	1h.	**551.402**	**14.734**	1.5K	25.887	932.173	19.422	686.404	17.068	3.6K	40.522	10.2K	83.297	608.375	15.791	644.150	16.489	2.0K	30.260	620.514	16.048

and Traffic. This demonstrates its robustness in large-scale applications where data uncertainty arises from sensor failures or environmental factors.

Our approach not only matches Graph-WaveNet's performance consistently, but also surpasses it, with equivalent or better MSE/MAE across all horizons and datasets. Unlike deterministic models such as TCGCN and Graph-WaveNet, ST-BayesianNet generates predictive distributions to capture inherent uncertainties, providing probabilistic outputs that are absent from the baselines. It learns spatial topologies adaptively and models uncertainty via variational inference, yielding superior expectation-based predictions. Furthermore, ST-BayesianNet is scalable, efficiently handling large-scale applications with extensive spatial and temporal datasets while maintaining robust performance in diverse, high-volume real-world scenarios.

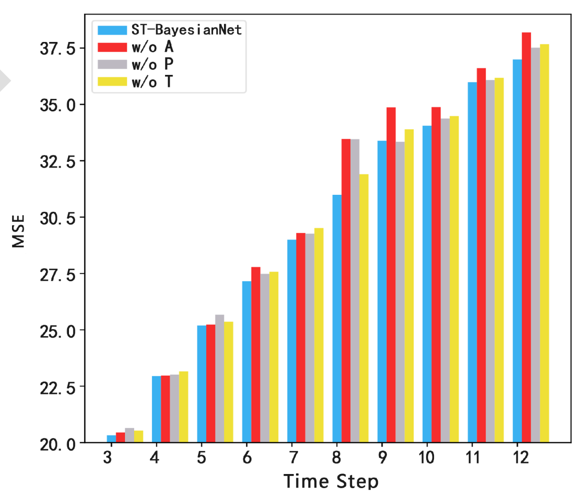


Fig. 5. The MSE result of ST-BayesianNet and its variants w/o A,w/o T and w/o P under different time steps. It can be seen that each module plays an important role in the entire ST Bayesian model.

C. Ablation Study

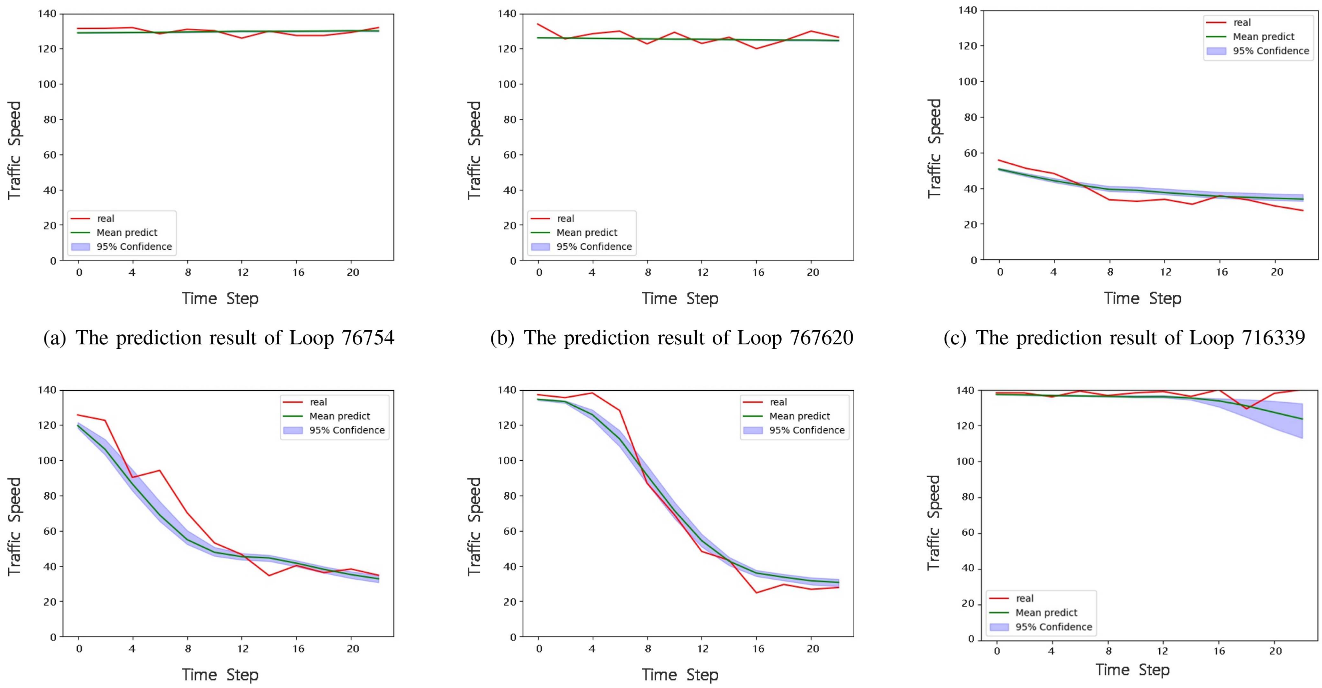
To further explore the impact of key components in our ST-BayesianNet model, we conducted an ablation study using the METR-LA dataset. We have labeled the various model variants as follows:

- **w/o** A: In this configuration, we exclude the dynamic graph convolution from ST-BayesianNet as described in (5).
- **w/o** T: In this configuration, we exclude the BTCN component, as defined in (6) and (8), from every layer of ST-BayesianNet without a non-deterministic time model.
- **w/o** P: In this configuration, we exclude the BGCN component, as defined in (6) and (8), from every layer of ST-BayesianNet without a non-deterministic spatial model.

Fig. 5 presents the MSE for each prediction horizon of ST-BayesianNet along with the other variants on the METR-LA dataset. It demonstrates that ST-BayesianNet generally outperforms the variants **w/o** A, **w/o** T, and **w/o** $\mathcal{P}$, particularly when dealing with longer sequence predictions. This suggests the effectiveness of dynamic graph convolution, non-deterministic time model, and non-deterministic spatial model in improving the predictive performance of ST-BayesianNet.

D. Randomized Predictive Output

Unlike existing spatiotemporal prediction methods such as Graph-WaveNet, DCRNN, and TCGCN, which exclusively



(a) The prediction result of Loop 76754 (b) The prediction result of Loop 767620 (c) The prediction result of Loop 716339

(d) The prediction result of Loop 767542 (e) The prediction result of Loop 767573 (f) The prediction result of Loop 767454

Fig. 6. Prediction results of ST-BayesianNet on the METR-LA dataset. Since the output of ST-BayesianNet is a probability distribution, the prediction results include a 95% confidence interval, enhancing the model's reliability and interpretability. In (a) and (b), the predictions are highly accurate, resulting in very narrow confidence intervals that are nearly invisible in the plots.

yield deterministic outcomes, ST-BayesianNet has the capacity to provide both temporal and spatial uncertainties. Consequently, its output is in the form of a probability distribution. The advantage of this form of output lies in its ability to furnish not only a plausible prediction but also a range of prediction probabilities. The visualization outcomes of several prediction outputs from the ST-BayesianNet model are depicted in Fig. 6. In this experiment, predictions for 12 steps are conducted using a 12-horizon input on the METR-LA dataset, employing 100 samples. The figure illustrates the true value (in red), the mean value (in green), and the 95% confidence region (in blue) derived from the 100 predictions. It can be seen that the prediction results of ST-BayesianNet are relatively accurate, which reflects that our proposed model architecture is relatively reasonable and can fully obtain spatiotemporal information.

When the predicted true values follow relatively smooth trends, the predicted values from our model exhibit modest variances. However, when the predicted true values display steep variations, our model adeptly identifies and quantifies the inherent uncertainty, leading to predictions accompanied by more substantial variances. In particular, Fig. 6(f) shows the model's capability to deliver comprehensive uncertainty predictions when data carries significant aleatoric uncertainty. This substantiates the ability of ST-BayesianNet to extract and represent uncertainty from spatiotemporal data. This output modality is a distinct advantage over existing spatiotemporal prediction models like Graph-WaveNet, AGCRN, and TCGCN, which generally lack the capacity to offer such probabilistic predictions. Importantly, these probabilistic predictions hold practical value for real-world forecasting scenarios.

E. Fault Tolerance Analysis

In reality, spatiotemporal data often exhibit noise and possess high dimensions. An effective model should exhibit robustness against such noise. In this experiment, we analyze the robustness of ST-BayesianNet in the presence of noise. For the sake of generality, we assume that the training data input is contaminated by Gaussian noise $\mathcal{Z} \in R^{T\times S\times C} \sim \mathcal{N}(0, \sigma^2)$, meaning the input becomes $X + \mathcal{Z}$. The parameter σ signifies the magnitude of the noise level. Here, we employ $\sigma = \{0.0001, 0.001, 0.01, 0.1, 1\}$. We opt to use the METR-LA dataset due to its relative complexity, and for comparison purposes, we select Graph-WaveNet, TCGCN, AGCRN and STGCN as benchmark methods owing to their high prediction accuracy as highlighted in Fig. 7. Specifically, we forecast 12 steps with 12 horizons of input data on the METR-LA dataset as part of this assessment.

Fig. 7 illustrates the performance trends of various methods as noise intensity increases. The MSE curves of ST-BayesianNet, Graph-WaveNet, and AGCRN remain stable despite rising noise levels. This stability can be attributed to specific model characteristics: the Gate-Conv mechanism in Graph-WaveNet acts as a low-pass filter, effectively mitigating noise impact, while the recurrent architecture of AGCRN helps to filter out anomalous data points. ST-BayesianNet, however, demonstrates superior performance in terms of MSE compared to all baseline models,

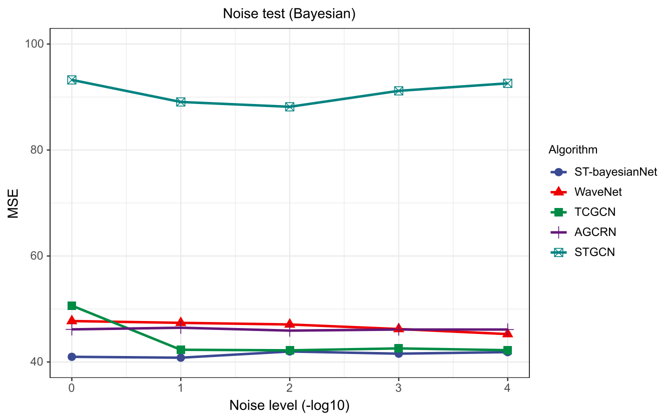


Fig. 7. MSE results under different noise settings on the METR-LA dataset. This comparison evaluates ST-BayesianNet against Graph-WaveNet, TCGCN, STGCN, and AGCRN across varying noise levels. A negative logarithmic scale is applied to the horizontal axis to enhance interpretability. ST-BayesianNet consistently outperforms all baseline models across every noise level and demonstrates superior stability throughout.

TABLE II
COMPARISON OF INFERENCE TIME AND SPACE COMPLEXITY. THE BEST RESULT IS BOLDED, THE SECOND-BEST IS UNDERLINED, THE THIRD-BEST IS MARKED WITH AN ASTERISK, AND THE FOURTH-BEST IS ITALICIZED

Algorithm	Params	MSE	Time	ACC
Graph-Wavenet	271.3K	20.490	*8.946	0.834
AGCRN	751.1K	24.046	**1.462**	*0.839
TCGCN	400.0K	20.538	31.250	0.843
crossformer	2.2M	46.352	2.537	0.830
DGCRN	**199.4K**	21.921	22.298	0.835
DE_NET	409.4K	20.445	58.021	0.834
ST-BayesianNet	*368.0K	**20.032**	*13.885*	**0.846**

highlighting the robustness of its BGCN and BTCN modules in handling uncertainties intrinsic to time series data. This strength is grounded in the overarching Bayesian variational inference framework employed by ST-BayesianNet. Unlike traditional methods that directly solve parameter equations, our model leverages probabilistic trainable parameters within variational inference to approximate optimal solutions during training. This probabilistic approach effectively models uncertainty in spatiotemporal data, thereby enhancing generalization and robustness across diverse application scenarios, especially when dealing with noisy datasets.

F. Complexity Analysis

To assess the practical efficiency and effectiveness of ST-BayesianNet, we further compare its time and space complexity with those of baseline models. All models were trained under the same experimental conditions outlined in Subsection IV-E, and subsequently evaluated in terms of accuracy and inference time. The results are presented in Table II. ST-BayesianNet achieves the best performance in both MSE and accuracy, while maintaining the third-lowest number of parameters and a moderate inference time. Although DGCRN has the smallest parameter count, its inference time ranks third longest and its MSE performance is suboptimal. AGCRN exhibits the shortest inference time but requires roughly twice as many parameters as ST-BayesianNet. Among the baselines, Graph-WaveNet offers the best balance between inference time and parameter count, yet its overall performance remains inferior to that of ST-BayesianNet.

For large-scale applicability, complexity analysis demonstrates that ST-BayesianNet scales efficiently with dataset size, handling expansive systems like urban traffic networks or power grids with 325+ nodes in PEMS-BAY without exponential time growth. Its inference time, with 368.0 K parameters and 73.885 ms on PEMS-BAY, remains competitive, supporting real-time forecasting with uncertainty quantification to aid risk assessment and resource allocation.

The ST-BayesianNet model is characterized by its utilization of spatiotemporal linear models instead of attention mechanisms and large-scale model architectures, rendering it less computationally intensive. Through the application of variational inference techniques, the parameter matrix of the spatial and temporal analysis module is randomized. This approach enables the model to effectively address a wide range of practical applications, outperforming baseline models. Notably, ST-BayesianNet demonstrates superior performance and achieves acceptable inference times when compared to other methods in the analysis.

The efficiency and robustness of ST-BayesianNet are key advantages in practical applications. Its low parameter scale minimizes computational resource demands, facilitating deployment on embedded devices with limited resources. Moreover, its moderate inference time enables efficient real-time prediction, accommodating various real-world scenarios. Furthermore, its superior performance in MSE and Accuracy enhances prediction accuracy, making it suitable for applications like traffic flow forecasting and energy management. These characteristics collectively establish ST-BayesianNet as a functional and practical solution optimized for processing complex spatiotemporal data.

V. CONCLUSION

This paper introduces an innovative multivariate time prediction approach named ST-BayesianNet, designed for spatiotemporal data, and grounded in the principles of variational inference. This method stands out by its capacity to capture uncertainties within both the temporal and spatial dimensions. ST-BayesianNet systematically models the uncertainties inherent in the temporal and spatial facets of the data. To validate its efficacy, comprehensive comparative experiments have been conducted to assess its performance. The experiments, carried out on six publicly available real-world spatiotemporal datasets, including traffic and solar, demonstrate that ST-BayesianNet consistently enhances prediction accuracy and yields predictive confidence estimations. The multivariate time series forecasting based on Bayesian CNNs remains a challenge in the later stages, specifically in real-world applications where data from multiple dimensions needs to be integrated to capture spatiotemporal uncertainties, leading to precise variational inference. Future research will extend the framework's application to larger-scale data scenarios and further explore multi-view learning for spatiotemporal multidimensional joint uncertainty estimation to enhance cross-dimensional uncertainty modeling capabilities.

References

[1] H. Guo, Z. Zhou, D. Zhao, and W. Gaaloul, "EGNN: Energy-efficient anomaly detection for iot multivariate time series data using graph neural network," *Future Gener. Comput. Syst.*, vol. 151, pp. 45–56, 2024, doi: 10.1016/j.future.2023.09.028.

[2] P. Mansourian, N. Zhang, A. Jaekel, and M. Kneppers, "Deep learning-based anomaly detection for connected autonomous vehicles using spatiotemporal information," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, pp. 16006–16017, Dec. 2023, doi: 10.1109/TITS.2023.3286611.

[3] X. Xia et al., "ETD-ConvLSTM: A deep learning approach for electricity theft detection in smart grids," *IEEE Trans. Inf. Forensics Secur.*, vol. 18, pp. 2553–2568, 2023, doi: 10.1109/TIFS.2023.3265884.

[4] K. Sallam, M. Mohamed, and A. W. Mohamed, "Internet of Things (IoT) in supply chain management: Challenges, opportunities, and best practices," *Sustain. Mach. Intell. J.*, vol. 2, pp. 3–1, 2023, doi: 10.61185/SMIJ.2023.22103.

[5] J. Li, Y. Shi, H. Li, and B. Yang, "TC-GATN: Temporal causal graph attention networks with nonlinear paradigm for multivariate time-series forecasting in industrial processes," *IEEE Trans. Ind. Informat.*, vol. 19, no. 6, pp. 7592–7601, Jun. 2023, doi: 10.1109/TII.2022.3211330.

[6] Y. Yao, M. Yang, J. Wang, and M. Xie, "Multivariate time-series prediction in industrial processes via a deep hybrid network under data uncertainty," *IEEE Trans. Ind. Informat.*, vol. 19, no. 2, pp. 1977–1987, Feb. 2023, doi: 10.1109/TII.2022.3198670.

[7] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph WaveNet for deep spatial-temporal graph modeling," in *Proc. 28th Int. Joint Conf. Artif.* Q2 *Intell.*, 2019, pp. 1907–1913, doi: 10.48550/arXiv.1906.00121.

[8] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang, "Connecting the dots: Multivariate time series forecasting with graph neural networks," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 753–763, doi: 10.1145/3394486.3403118.

[9] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. 29th Int. Conf. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[10] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 922–929, doi: 10.1609/aaai.v33i01.3301922.

[11] J. Wang, W. Wang, X. Liu, W. Yu, X. Li, and P. Sun, "Traffic prediction based on auto spatiotemporal multi-graph adversarial neural network," *Physica A: Stat. Mechan. Appl.*, vol. 590, 2022, Art. no. 126736, doi: 10.1016/j.physa.2021.126736.

[12] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 01, 2020, pp. 914–921, doi: 10.1609/aaai.v34i01.5438.

[13] L. Wang, D. Guo, H. Wu, K. Li, and W. Yu, "TC-GCN: Triple cross-attention and graph convolutional network for traffic forecasting," *Inf. Fusion*, vol. 105, 2024, Art. no. 102229, doi: 10.1016/j.inffus.2024.102229.

[14] M. Ma and Others, "HiSTGNN: Hierarchical spatio-temporal graph neural network for weather forecasting," *Inf. Sci.*, vol. 648, 2023, Art. no. 263, doi: 10.1016/j.ins.2023.119580.

[15] Q. Luo and Others, "LSTTN: A long-short term transformer-based spatiotemporal neural network for traffic flow forecasting," *Knowl.-Based Syst.*, vol. 293, 2024, Art. no. 256, doi: 10.1016/j.knosys.2024.111637.

[16] S. Berkani, A. Bahaj, B. Guermah, and M. Ghogho, "Beyond spatial neighbors: Utilizing multivariate transfer entropy for interpretable graph-based spatio-temporal forecasting," *Eng. Appl. Artif. Intell.*, vol. 146, 2025, Art. no. 110161, doi: 10.1016/j.engappai.2025.110161.

[17] L. Bai, L. Yao, C. Li, X. Wang, and C. Wang, "Adaptive graph convolutional recurrent network for traffic forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 17804–17815, doi: 10.48550/arXiv.2007.02842.

[18] Z. Pan et al., "AutoSTG: Neural architecture search for predictions of spatio-temporal graphs," in *Proc. Web Conf.*, 2021, pp. 1846–1855, doi: 10.1145/3442381.3449816.

[19] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5580–5590, doi: 10.48550/arXiv.1703.04977.

[20] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li, "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, 2019, pp. 5668–5675, doi: 10.48550/arXiv.1803.01254.

[21] C. Zheng, X. Fan, C. Wang, and J. Qi, "GMAN: A graph multi-attention network for traffic prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 01, pp. 1234–1241, doi: 10.48550/arXiv.1911.08415. Q3

[22] L. Zhao et al., "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2019, doi: 10.1109/TITS.2019.2935152.

[23] W. Chen, L. Chen, Y. Xie, W. Cao, Y. Gao, and X. Feng, "Multi-range attentive bicomponent graph convolutional network for traffic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 01, pp. 3529–3536, doi: 10.48550/arXiv.1911.12093.

[24] K. Guo et al., "Optimized graph convolution recurrent neural network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1138–1149, Feb. 2021, doi: 10.1109/TITS.2019.2963722.

[25] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017, *arXiv:1707.01926*. Q4

[26] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," 2017, *arXiv:1709.04875*.

[27] C. Zheng, X. Fan, C. Wang, and J. Qi, "GMAN: A graph multi-attention network for traffic prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 1, pp. 1234–1241, doi: 10.48550/arXiv.1911.08415.

[28] Y. Gal and Z. Ghahramani, "Bayesian convolutional neural networks with Bernoulli approximate variational inference," *Comput. Sci.*, 2015, doi: 10.48550/arXiv.1506.02158. Q5

[29] R. Chandra, A. Bhagat, M. Maharana, and P. N. Krivitsky, "Bayesian graph convolutional neural networks via tempered MCMC," *IEEE Access*, vol. 9, pp. 130353–130365, 2021, doi: 10.1109/ACCESS.2021.3111898.

[30] Y. Liao and C. Liang, "A temperature time series forecasting model based on deepar," in *Proc. 7th Int. Conf. Comput. Commun.*, 2021, pp. 1588–1593, doi: 10.1109/ICCC54389.2021.9674623.

[31] H. Zhou et al., "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 11106–11115, doi: 10.1609/aaai.v35i12.17325.

[32] K. Shridhar, F. Laumann, and M. Liwicki, "Uncertainty estimations by softplus normalization in Bayesian convolutional neural networks with variational inference," 2018, *arXiv:1806.05978*.

[33] K. Shridhar, F. Laumann, and M. Liwicki, "A comprehensive guide to Bayesian convolutional neural network with variational inference," 2019, *arXiv:1901.02731*.

[34] A. D. Kiureghian and O. Ditlevsen, "Aleatory or epistemic? Does it matter?," *Struct. saf.*, vol. 31, no. 2, pp. 105–112, 2009, doi: 10.1016/j.strusafe.2008.06.020.

[35] W. Sun, Y. Lu, X. Zhang, and M. Sun, "DeepGlobal: A framework for global robustness verification of feedforward neural networks," *J. Syst. Archit.*, vol. 128, 2022, Art. no. 102582, doi: 10.1016/j.sysarc.2022.102582.

[36] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," in *Proc. 8th Workshop Syntax, Semantics Struct. Stat. Transl.*, 2014, p. 103, doi: 10.48550/arXiv.1409.1259. Q6

[37] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[38] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," 2017, *arXiv:1709.04875*.

[39] Q. Bi, J. Liu, and J. Tang, "Graph convolutional network representation learning method with community-enhanced temporal features for dynamic graphs," in *Proc. 6th Int. Conf. Artif. Intell. Pattern Recognit.*, 2023, pp. 1336–1341, doi: 10.1145/3641584.3641785.

[40] Z. Li, C. Huang, L. Xia, Y. Xu, and J. Pei, "Spatial-temporal hypergraph self-supervised learning for crime prediction," in *Proc. IEEE 38th Int. Conf. Data Eng.*, 2022, pp. 2984–2996, doi: 10.1109/ICDE53745.2022.00269.

[41] Y. Zhang and J. Yan, "Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting," in *Proc. 11th Int. Conf. Learn. Representations*, 2023, 1–9.

Lei Wang (Student Member, IEEE) received the bachelor's and master's degrees from Nankai University, Tianjin, in 2005 and 2014, respectively. He is currently working toward the Ph.D. degree with the College of Intelligence and Computing, Tianjin University, Tianjin. His research interests mainly include deep learning, traffic data analysis and application, complex spatiotemporal data mining, and multi-source heterogeneous data analysis.

Huaming Wu (Senior Member, IEEE) received the B.E. and M.S. degrees in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 2009 and 2011, respectively, and the Ph.D. degree (with Highest Hons.) in computer science from Freie Universität Berlin, Berlin, Germany in 2015. He is currently a Professor with the Center for Applied Mathematics, Tianjin University, Tianjin, China. His research interests include mobile cloud computing, edge computing, Internet of Things, deep learning, complex networks, and DNA storage.

Keqiu Li (Fellow, IEEE) received the bachelor's and master's degrees from the Department of Applied Mathematics, Dalian University of Technology, Dalian, China, in 1994 and 1997, respectively, and the Ph.D. degree from the Graduate School of Information Science, Japan Advanced Institute of Science and Technology, Nomi, Japan, in 2005. He is currently a Professor with the College of Intelligence and Computing, Tianjin University, Tianjin, China. He has authored or coauthored more than 100 technical papers, such as the IEEE Transactions on Parallel and Distributed Systems, *ACM Transactions on Internet Technology*, INFOCOM, Architectural Support for Programming Languages and Operating Systems, EourSys, and USENIX ATC. His research interests include data center networks, cloud computing, and cybersecurity. He is also an Associate Editor for the IEEE Transactions on Parallel and Distributed Systems and IEEE Transactions on Computers.

Wei Yu received the B.S. degree from Jinggangshan University, Ji'an, China, in 2012, the M.S. degree from Henan Normal University, Xinxiang, China, in 2015, and the Ph.D. degree from the School of College of Intelligence and Computing, Tianjin University, Tianjin, China. He is currently an Associate Professor with Zhejiang Yuexiu University, Shaoxing, China. His research interests include dynamic complex network analysis, large-scale data mining and machine learning.