MULTI-KNOWLEDGE FUSION NETWORK FOR TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Forecasting the behavior of complex dynamical systems such as interconnected sensor networks characterized by high-dimensional multivariate time series(MTS) is of paramount importance for making informed decisions and planning for the future in a broad spectrum of applications. Graph forecasting networks(GFNs) are well-suited for forecasting MTS data that exhibit spatio-temporal dependencies. However, most prior works of GFN-based methods on MTS forecasting rely on domain expertise to model the nonlinear dynamics of the system but neglect the potential to leverage the inherent relational-structural dependencies among time series variables underlying MTS data. On the other hand, contemporary works attempt to infer the relational structure of the complex dependencies between the variables and simultaneously learn the nonlinear dynamics of the interconnected system but neglect the possibility of incorporating domain-specific prior knowledge to improve forecast accuracy. To this end, we propose a novel hybrid architecture that combines explicit prior knowledge with implicit knowledge of the relational structure within the MTS data. It jointly learns intra-series temporal dependencies and inter-series spatial dependencies by encoding time-conditioned structural spatio-temporal inductive biases to provide more accurate and reliable forecasts. It also models the time-varying uncertainty of the multi-horizon forecasts to support decision-making by providing estimates of prediction uncertainty. In doing so, the proposed scalable framework and utilizing a cost-effective GPU hardware to improve the training efficiency have shown promising results on multiple large-scale spatio-temporal graphs datasets. It outperforms state-of-the-art forecasting methods by a significant margin. We report and discuss the ablation studies to validate our forecasting architecture.

1 INTRODUCTION

Accurate multivariate time series forecasting(MTSF) is critical for a broad spectrum of domains that have significant financial or operational impacts, including retail and finance, intelligent transportation systems, logistics and supply chain management, and many others. However, MTSF can be challenging due to the complexity of the relationships between time series variables and the unique characteristics of the MTS data, such as non-linearity, heterogeneity, sparsity, and non-stationarity. In this context, Spatial-temporal graph neural networks(STGNNs) have been widely studied for modeling the long-range intra-temporal dependencies and complex inter-dependencies among the variables in the MTS data for improved multi-horizon forecast accuracy. The explicit relationships among variables are based on prior knowledge provided by human experts in the form of a predefined or explicit graph, while implicit relationships among variables within the MTS data are obtained through neural relational inference methods(Deng & Hooi (2021); Kipf et al. (2018)). The implicit relationships are highly-complex and non-linear, can change over time, and uncover hidden relationships unknown to human experts which are not obvious. The existing "human-in-the-loop" STGNNs(Yu et al. (2017), Li et al. (2017), Guo et al. (2020)) incorporate domain-specific knowledge of the relational-structural dependencies among the interdependent variables while simultaneously learning the dynamics from the MTS data. However, arguably, the explicit graph structures in most real-world scenarios are either unknown, inaccurate, or partially available, thus resulting in suboptimal forecasting. Even if available, the explicit graph structure represents a simplified view of dependencies and often fails to capture the non-static spatial-temporal dependencies within the MTS data. Precisely it falls short of accurately inferring the latent time-conditioned underlying relations that drive the co-movements among variables in the substantial MTS data. On the contrary, a recent class of STGNNs(Shang et al. (2021); Deng & Hooi (2021); Wu et al. (2020);

Kipf et al. (2018)) jointly infer the discrete dependency graph structure describing the implicit relations between variables while simultaneously learning the dynamics in MTS data. Despite the success, these approaches neglect to exploit the predefined graph of the inter-relationships among variables obtained from the domain-expertise knowledge resulting in suboptimal performance on the graph time-series forecasting. In addition, implicit graph structure learning from MTS data suffers from inherent limitations of pairwise associations. While in contrast, the relations within the complex dynamical systems of interconnected networks could go beyond pairwise connections. Hypergraph, a generalization of a graph, offers a natural fit for modeling the higher-order structural relations underlying the interconnected networks in complex high-dimensional data. Moreover, the standard STGNNs focus on learning pointwise forecasts but do not provide uncertainty estimates of forecasts. To overcome the challenges, we propose an explicit-implicit knowledge fusion neural network(EIKF-Net) framework with a joint learning paradigm on the explicit-implicit interaction structure for a thorough understanding of the underlying dependencies between time series variables, while simultaneously learning the complex dynamics of the MTS data for better forecast accuracy and to provide reliable uncertainty estimates of forecasts. The proposed framework consists of two main components: spatial and temporal learning components. We adopt a space-then-time(STT, Gao & Ribeiro (2022)) approach, where spatial message-passing schemes are performed prior to the temporal-encoding step. The spatial learning component is further composed of an implicit hypergraph and explicit graph learning modules. The former infers the implicit hypergraph structure, which captures the hierarchical interdependencies among variables in MTS data. Simultaneously it performs hypergraph representation learning schemes to encode the spatio-temporal dynamics underlying the hypergraph-structured MTS data into the latent hypernode-level representations. The latter performs the graph representation learning schemes to encode the pair-wise spatial relations between the multiple co-evolving variables to capture the spatio-temporal dynamics within the graph-structured MTS data into the latent node-level representations. We perform convex combination(i.e., "mix up") of the latent graph and hypergraph representations through a gating mechanism. It leads to more accurate latent representations of the complex non-linear dynamics of the MTS data. The mixup representations allow the framework to capture different types of dependencies that exist at different observation scales(i.e., correlations among variables could potentially differ in the short and long-term views in the MTS data). The temporal learning component focusses on learning the time-evolving dynamics of interdependencies among the variables present in the MTS data to provide accurate multi-horizon forecasts with predictive uncertainty estimates. To summarize, our work presents an end-to-end methodological framework to infer the implicit interaction structure from MTS data. It simultaneously learns the spatio-temporal dynamics within the explicit graph and implicit hypergraph structured MTS data using graph and hypergraph neural networks, respectively, to capture the evolutionary and multi-scale interactions among the variables in the latent representations. It performs inference over these latent representations for downstream MTSF task and models the time-varying uncertainty of the forecasts in order to provide more accurate risk assessment and better decision making by estimating predictive uncertainty. The framework is designed to offer better generalization and scalability with reduced computational requirements for large-scale spatio-temporal MTS data-based forecasting tasks as those found in real-world applications.

2 Problem Definition

Lets us assume a historical time series data, with n-correlated variables, observed over T training steps is represented by $\mathbf{X}=(\mathbf{x}_1,\ldots,\mathbf{x}_{\mathrm{T}})$, where the subscript refers to time step. The observations of the n-variables at time point t are denoted by $\mathbf{x}_t=(\mathbf{x}_t^{(1)},\mathbf{x}_t^{(2)},\ldots,\mathbf{x}_t^{(n)})\in\mathbb{R}^{(n)}$, where the superscript refers to variables. Under the rolling-window method for multi-step forecasting, where at the current time step t, we predefine a fixed-length look-back window to include the prior τ-steps of historical MTS data to predict for the next υ-steps. In the context of MTSF, the learning problem can be formalized using the rolling window method. The goal is to use a historical window of n-correlated variables, represented by the $\mathbf{X}_{(t-\tau:\,t-1)}\in\mathbb{R}^{n\times\tau}$, which have been observed over previous τ-steps prior to current time step t, to predict about the future values of n variables for the next υ-steps denoted as $\mathbf{X}_{(t:t+\upsilon-1)}\in\mathbb{R}^{n\times\upsilon}$. The MTSF problem is further formulated on the graph and hypergraph structure to capture the spatial-temporal correlations among multitudinous correlated time series variables. We represent the historical inputs as continuous-time spatial-temporal graphs, denoted as $\mathcal{G}_t=(\mathcal{V},\mathcal{E},\mathbf{X}_{(t-\tau:\,t-1)},\mathrm{A}^{(0)})$. $\mathcal{G}_t$ is composed of a set of nodes($\mathcal{V}$), edges($\mathcal{E}$) that describe the connections among the variables and node feature matrix $\mathbf{X}_{(t-\tau:\,t-1)}$ that changes over time, where t is the current time step. The adjacency matrix, $\mathrm{A}^{(0)}\in\{0,1\}^{|\mathcal{V}|\times|\mathcal{V}|}$, describes the explicit fixed-graph structure based on prior knowledge of time-series relationships. In addition, we treat historical MTS

data as a sequence of dynamic hypergraphs, denoted as $\mathcal{HG}_t=(\mathcal{HV}, \mathcal{HE}, \mathbf{X}_{(t-\tau:\, t-1)}, \mathrm{I})$. The hypergraph is represented by a fixed set of hypernodes($\mathcal{HV}$) and hyperedges($\mathcal{HE}$), where time series variables denote the hypernodes and hyperedges capture the latent higher-order relationships between hypernodes. The time-varying hypernode feature matrix is given by $\mathbf{X}_{(t-\tau:\, t-1)}$. The implicit hypergraph structure is learned through an embedding-based similarity metric learning approach. The incidence matrix, $\mathbf{I}\in\mathbb{R}^{n\times m}$, describes the hypergraph structure, where $\mathbf{I}_{p,\,q}=1$ if the hyperedge q incident with hypernode p and otherwise 0. The number of hyperedges(m) in a hypergraph determines the sparsity of the hypergraph. Given a $\mathcal{G}_t$ and $\mathcal{HG}_t$, the novel framework is designed to learn a function $F(\theta)$ that maps historical MTS data, $\mathbf{X}_{(t-\tau:\, t-1)}$, to their respective future values, $\mathbf{X}_{(t:t+\upsilon-1)}$ values defined as follows,

$$\left[\mathbf{X}_{(t-\tau)}, \cdots, \mathbf{X}_{(t-1)}; \mathcal{G}_t, \mathcal{HG}_t\right] \xrightarrow{F(\theta)} \left[\mathbf{X}_{(t+1)}, \cdots, \mathbf{X}_{(t+\upsilon-1)}\right] \tag{1}$$

Simply, the MTSF task formulated on the explicit graph($\mathcal{G}_t$) and implicit hypergraph($\mathcal{HG}_t$) is described as follows:

$$\min_{\theta} \mathcal{L}\left(\mathbf{X}_{(t:t+\upsilon-1)}, \hat{\mathbf{X}}_{(t:t+\upsilon-1)}; \mathbf{X}_{(t-\tau:\, t-1)}, \mathcal{G}_t, \mathcal{HG}_t\right) \tag{2}$$

where θ denotes all the learnable parameters for trainable function $F(\theta)$. $\hat{\mathbf{X}}_{(t:t+\upsilon-1)}$ denotes the model predictions, $\mathcal{L}$ is the loss function. The loss function to train our learning algorithm is mean absolute error(MAE) loss.

$$\mathcal{L}_{\text{MAE}}(\theta)=\frac{1}{\upsilon}\left|\mathbf{X}_{(t:t+\upsilon-1)} - \hat{\mathbf{X}}_{(t:t+\upsilon-1)}\right| \tag{3}$$

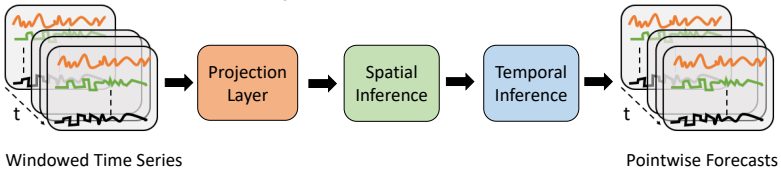

Figure 1: Overview of **EIKF-Net** framework

3 Our Approach

The overall neural forecasting architecture of our framework is illustrated in Figure 1. It consists of three main components: The projection layer, spatial learning, and temporal learning components. The spatial learning component includes two modules: graph and hypergraph learning modules. The hypergraph learning module infers the discrete dependency hypergraph structure to capture the interrelations between time-series variables. It also performs higher-order message-passing schemes to learn the time-conditioned optimal hypernode-level representations by modeling the hypergraph-structured MTS data. The graph learning module utilizes the predefined graph, which represents the relational structure of the variables obtained from domain-expertise knowledge to obtain the graph-structured MTS data. It performs spatial graph-filtering through neighborhood aggregation schemes to compute the optimal node-level representations that better capture the underlying dynamics within the MTS data. The temporal inference component performs a convex combination of the latent explicit-graph and implicit-hypergraph representations and learns the time-evolving interdependencies to provide pointwise forecasts and uncertainty estimations. Overall, joint optimization of different learning components of the proposed framework effectively captures the complex relationships between time-series variables and makes accurate forecasts.

3.1 Projection Layer

The projection layer utilizes a gated linear networks(GLN, Dauphin et al. (2017)) to learn the non-linear representations of the input data, $\mathbf{X}_{(t-\tau:\, t-1)}\in\mathbb{R}^{n\times\tau}$ through a gating mechanism to compute a transformed feature matrix, $\bar{\mathbf{X}}_{(t-\tau:\, t-1)}\in\mathbb{R}^{n\times d}$ as follows,

$$\bar{\mathbf{X}}_{(t-\tau:\, t-1)}=\left(\sigma(\mathrm{W}_0\mathbf{X}_{(t-\tau:\, t-1)}) \otimes \mathrm{W}_1\mathbf{X}_{(t-\tau:\, t-1)}\right)\mathrm{W}_2$$

where $\mathrm{W}_0, \mathrm{W}_1, \mathrm{W}_2\in\mathbb{R}^{\tau\times d}$ are trainable weight matrices, $\otimes$ denotes the element-wise multiplication. σ is the non-linear activation function.

3.2 Spatial-Inference

The spatial inference component of our framework is illustrated in Figure 2. The spatial-learning component encodes non-linear input data, $\bar{\mathbf{x}}_{(t-\tau:\, t-1)}$ to obtain graph and hypergraph representations using two modules: the hypergraph learning module and the graph learning module. The hypergraph learning module performs joint hypergraph inference and representation learning, while the graph learning module performs graph representation learning. The outputs of these two modules are fused using a convex combination approach to regulate the flow of information encoded by each module. The details of each module are discussed in subsequent sections.

3.2.1 Hypergraph Inference and Representation Learning

The hypergraph learning module is composed of two units: the hypergraph inference(HgI) unit and the hypergraph representation learning(HgRL) unit. The HgI unit is a structural modeling approach

that aims to infer the discrete hypergraph topology capturing the hierarchical interdependence relations among time-series variables for a hypergraph-structured representation of the MTS data. We described HgI unit in the technical appendix in great detail. The downstream forecasting task acts as the indirect supervisory information for revealing the high-order structure,i.e., the hypergraph relation structure behind the observed data. The MTS data is represented as hypernode-attributed spatio-temporal hypergraphs. A hypergraph representation learning unit(HgRL) is used to compute optimal hypernode-level representations by capturing the spatio-temporal dynamics within the hypergraph-structured MTS data. These representations are then used to perform inference on the downstream multi-horizon forecasting task. In short, HgRL is a neural network architecture that utilizes both Hypergraph Attention Network(HgAT) and Hypergraph Transformer(HgT) as its backbone. The HgT uses multi-head self-attention mechanisms to learn the latent hypergraph representations, $\mathbf{h}_{\mathbf{i}}^{\prime(t)}$, at time t without leveraging any prior knowledge about the structure of the hypergraph. The HgAT performs higher-order message-passing schemes on the hypergraph to aggregate information and compute the latent hypernode representations, $\mathbf{h}_i^{(t)}$, at time t. The HgAT and HgT form a powerful backbone for HgRL, allowing it to effectively learn hypergraph representations by capturing complex relationships and dependencies within the hypergraph-structured MTS data. We provide the implementation details and a more in-depth explanation in the appendix for further information. We regulate the information flow from $\mathbf{h}_{\mathbf{i}}^{\prime(t)}$ and $\mathbf{h}_i^{(t)}$ by applying a gating mechanism to produce a weighted combination of representations as described below,

$$\mathbf{h}_{\mathbf{i}}^{\prime\prime(t)}=\sigma\big(g^{\prime}(\mathbf{h}_{\mathbf{i}}^{\prime(t)})\big)+(1-g^{\prime})(\mathbf{h}_i^{(t)})\big);\, g^{\prime}=\sigma\big(f_s^{\prime}(\mathbf{h}_{\mathbf{i}}^{\prime(t)})+f_g^{\prime}(\mathbf{h}_i^{(t)})\big) \tag{4}$$

where $f_s^{\prime}$ and $f_g^{\prime}$ are linear projections. Fusing representations can be beneficial for modeling the multi-scale interactions underlying the spatio-temporal hypergraph data and also help to mitigate overfitting. The spatio-temporal data often contains correlations between variables that change over time or at different observation scales. By fusing representations, the proposed framework incorporates the most relevant information to capture the time-evolving underlying patterns in the MTS data, which leads to more accurate and robust forecasts. In brief, the hypergraph learning module optimizes the discrete hypergraph structure through the similarity metric learning technique. It formulates the posterior forecasting task as message-passing schemes with hypergraph neural networks to learn the optimal hypergraph representations, which leads to more accurate and expressive representations of the MTS data for better forecast accuracy.

3.2.2 Graph Representation Learning(GRL)

We represent the MTS data as continuous-time spatio-temporal graphs based on predefined graphs obtained from domain-specific knowledge. We utilize the Temporal Graph Convolution Neural Network(T-GCN, Zhao et al. (2019)) to compute optimal node-level representations by modeling the graph topology dependencies and feature attributes within the graph-structured MTS data. These graph representations are further processed by the downstream temporal inference component for learning the non-linear temporal dynamics of inter-series correlations among the variables. In short, the T-GCN performs neighborhood aggregation schemes on predefined graph topology to compute the optimal node-level representations, $\mathbf{h}_{\mathbf{i}}^{\prime\prime\prime(t)}$, at a specific time t. It effectively captures fine-grained, data-source specific patterns accurately. We discuss in the appendix a more detailed description of the technique.

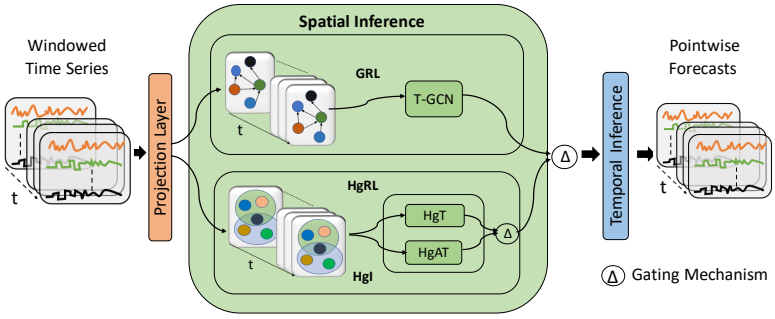

Figure 2: Overview of spatial inference component.

3.3 Temporal-Inference

The mixture-of-experts(MOE) mechanism combines the predictions of multiple subnetworks (experts) to produce a final prediction. In this specific framework, the experts are graph and hypergraph learning modules. The expert predictions are combined through a gating mechanism in an input-dependent manner by calculating a weighted sum of their predictions. The goal of training in this

framework is to achieve two objectives: 1) Identifying the optimal distribution of weights for the gating function that precisely captures the underlying distribution within the MTS data, and 2) Training the experts using the distribution weights specified by the gating function. The fused representations are obtained by combining the predictions of the experts using the calculated weights as follows,

$$\mathbf{h}_{\mathbf{i}}^{\prime\prime\prime\prime(t)}=\sigma\big(g^{\prime\prime}(\mathbf{h}_{\mathbf{i}}^{\prime\prime(t)})+(1-g^{\prime\prime})(\mathbf{h}_{\mathbf{i}}^{\prime\prime\prime(t)})\big); g^{\prime\prime}=\sigma\big(f_s^{\prime\prime}(\mathbf{h}_{\mathbf{i}}^{\prime\prime(t)})+f_g^{\prime\prime}(\mathbf{h}_{\mathbf{i}}^{\prime\prime\prime(t)})\big) \tag{5}$$

where $\mathbf{h}_{\mathbf{i}}^{\prime\prime(t)}$, $\mathbf{h}_{\mathbf{i}}^{\prime\prime\prime(t)}$ are computed by the graph and hypergraph learning modules, respectively. $f_s^{\prime\prime}$ and $f_g^{\prime\prime}$ are linear projections. The temporal inference component consists of a stack of 1×1 convolutions. The fused representations are fed as input to the temporal inference component. This component aims to model the non-linear temporal dynamics of inter-series dependencies among variables underlying the spatial-temporal MTS data and predicts the pointwise forecasts, $\hat{\mathbf{X}}_{(t:t+\upsilon-1)}$. Our proposed framework uses the spatial-then-time modeling approach to learn the higher-order structure representation and dynamics in MTS data. This approach first encodes the spatial information of the relational structure, including both explicit graph and implicit hypergraph, which captures the complex spatial dependencies. By incorporating the temporal learning component, the framework then analyzes the evolution of these dependencies over time, which helps to improve the interpretability and generalization of the framework. This approach is beneficial for dealing with real-world applications that involve entangled complex spatial-temporal dependencies within the MTS data, which can be challenging to model using traditional methods. Additionally, by minimizing the negative Gaussian log likelihood, the uncertainty estimates of forecasts can be provided through the **w/Unc-EIKF-Net** framework(i.e., **EIKF-Net** with Local Uncertainty Estimation). Minimizing the negative Gaussian log likelihood is equivalent to maximizing the likelihood of the model's predictions given the true values. This allows the framework to provide more accurate and reliable uncertainty estimates of forecasts. In summary, our proposed methods(**EIKF-Net**, **w/Unc- EIKF-Net**) allow for simultaneously modeling the latent interdependencies and then analyze the evolution of these dependencies over time in the sensor network based dynamical systems in an end-to-end manner.

4 Experimental Results

We conduct experiments to verify the performance of proposed models(**EIKF-Net**, **w/Unc- EIKF-Net**) on large-scale spatial-temporal datasets. Additional details about the benchmark datasets are described in the appendix. Table 1 presents a comparison of the forecast errors of proposed models(**EIKF-Net** and **w/Unc- EIKF-Net**) with those of several baseline models on five different datasets(PeMSD3, PeMSD4, PeMSD7, PeMSD7M, and PeMSD8). The forecast errors are evaluated for a 12(τ)-step-prior to 12(υ)-step-ahead forecasting task which is a popular and well-established benchmark in the MTSF task. The performance of the proposed models was evaluated using multiple metrics such as mean absolute error(MAE), root mean squared error(RMSE), and mean absolute percentage error(MAPE). Using multiple evaluation metrics in multi-horizon prediction tasks provides a comprehensive evaluation of the proposed models performance with the baselines. The results for the baseline models were reported in a previous study by Choi et al. (2022). The proposed models(**EIKF-Net**, **w/Unc- EIKF-Net**) consistently demonstrate state-of-the-art performance compared to baseline models on various benchmark datasets based on multiple evaluation metrics. The results show that the proposed models demonstrate the best performance with lower forecast error on benchmark datasets. Specifically, they report a 12.2%, 14.8%, 8.8%, 10.6% and 8.9% significant drop in the RMSE metric compared to the next-best baseline models on PeMSD3, PeMSD4, PeMSD7, PeMSD8, and PeMSD7(M) datasets, respectively. In addition to the pointwise forecasts, the **w/Unc- EIKF-Net** model(i.e., **EIKF-Net** with local uncertainty estimation) provides time-varying uncertainty estimates. Its performance is slightly worse than the **EIKF-Net** model but still outperforms several strong baselines in the literature, as reflected in the lower prediction error. In brief, the empirical results show the efficacy of the proposed neural forecasting architecture in modeling the complex and nonlinear spatio-temporal dynamics underlying the MTS data to provide better forecasts. The appendix provides more detailed information on the experimental setup, ablation studies, and other additional experimental results on multi-horizon forecasting. Moreover, the appendix discusses the experimental results that support the **EIKF-Net** framework's ability to handle missing data and provide more insights into the **w/Unc- EIKF-Net** framework for estimating uncertainty. Furthermore, the appendix also includes time series visualizations of model predictions with the uncertainty estimates compared to the ground truth. Additionally, the appendix provides more information on the existing works and a short description of baseline models.

5 Conclusion

We propose a framework that combines implicit and explicit knowledge for learning the dynamics of MTS data to provide accurate multi-horizon forecasts. The experimental results on real-world

datasets demonstrate the effectiveness of the proposed framework and have shown improved forecast estimates and reliable uncertainty estimations. By reducing the computational requirements, the framework makes it possible to handle larger datasets and improve the forecast accuracy. Utilizing cost-effective GPU hardware helps to speed up the training process and reduce the memory requirements, making the framework more practical to use in real-world scenarios. For future work, we would endeavor to generalize the framework to handle much larger scale graph datasets for forecasting, synthetic-private data generation, missing-data imputation etc.

Methods	**PeMSD3**			**PeMSD4**			**PeMSD7**			**PeMSD8**			**PeMSD7(M)**		
	MAE	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**
HA	31.58	52.39	33.78	38.03	59.24	27.88	45.12	65.64	24.51	34.86	59.24	27.88	4.59	8.63	14.35
ARIMA	35.41	47.59	33.78	33.73	48.80	24.18	38.17	59.27	19.46	31.09	44.32	22.73	7.27	13.20	15.38
VAR	23.65	38.26	24.51	24.54	38.61	17.24	50.22	75.63	32.22	19.19	29.81	13.10	4.25	7.61	10.28
FC-LSTM	21.33	35.11	23.33	26.77	40.65	18.23	29.98	45.94	13.20	23.09	35.17	14.99	4.16	7.51	10.10
TCN	19.32	33.55	19.93	23.22	37.26	15.59	32.72	42.23	14.26	22.72	35.79	14.03	4.36	7.20	9.71
TCN(w/o causal)	18.87	32.24	18.63	22.81	36.87	14.31	30.53	41.02	13.88	21.42	34.03	13.09	4.43	7.53	9.44
GRU-ED	19.12	32.85	19.31	23.68	39.27	16.44	27.66	43.49	12.20	22.00	36.22	13.33	4.78	9.05	12.66
DSANet	21.29	34.55	23.21	22.79	35.77	16.03	31.36	49.11	14.43	17.14	26.96	11.32	3.52	6.98	8.78
STGCN	17.55	30.42	17.34	21.16	34.89	13.83	25.33	39.34	11.21	17.50	27.09	11.29	3.86	6.79	10.06
DCRNN	17.99	30.31	18.34	21.22	33.44	14.17	25.22	38.61	11.82	16.82	26.36	10.92	3.83	7.18	9.81
GraphWaveNet	19.12	32.77	18.89	24.89	39.66	17.29	26.39	41.50	11.97	18.28	30.05	12.15	3.19	6.24	8.02
ASTGCN(r)	17.34	29.56	17.21	22.93	35.22	16.56	24.01	37.87	10.73	18.25	28.06	11.64	3.14	6.18	8.12
MSTGCN	19.54	31.93	23.86	23.96	37.21	14.33	29.00	43.73	14.30	19.00	29.15	12.38	3.54	6.14	9.00
STG2Seq	19.03	29.83	21.55	25.20	38.48	18.77	32.77	47.16	20.16	20.17	30.71	17.32	3.48	6.51	8.95
LSGCN	17.94	29.85	16.98	21.53	33.86	13.18	27.31	41.46	11.98	17.73	26.76	11.20	3.05	5.98	7.62
STSGCN	17.48	29.21	16.78	21.19	33.65	13.90	24.26	39.03	10.21	17.13	26.80	10.96	3.01	5.93	7.55
AGCRN	15.98	28.25	15.23	19.83	32.26	12.97	22.37	36.55	9.12	15.95	25.22	10.09	2.79	5.54	7.02
STFGNN	16.77	28.34	16.30	20.48	32.51	16.77	23.46	36.60	9.21	16.94	26.25	10.60	2.90	5.79	7.23
STGODE	16.50	27.84	16.69	20.84	32.82	13.77	22.59	37.54	10.14	16.81	25.97	10.62	2.97	5.66	7.36
Z-GCNETs	16.64	28.15	16.39	19.50	31.61	12.78	21.77	35.17	9.25	15.76	25.11	10.01	2.75	5.62	6.89
STG-NCDE	15.57	27.09	15.06	19.21	31.09	12.76	20.53	33.84	8.80	15.45	24.81	9.92	2.68	5.39	6.76
EIKF-Net	**13.67**	**20.17**	**10.83**	**18.36**	**26.48**	**10.05**	**19.86**	**30.85**	**8.59**	**14.93**	**22.17**	**7.83**	**2.44**	**4.91**	**6.03**
w/Unc- EIKF-Net	13.97	20.59	11.04	18.73	27.18	10.26	-	-	-	15.12	22.84	8.19	2.54	5.09	6.24

Table 1: Pointwise forecast and predictive uncertainty errors on benchmark datasets at horizon@12. "**-**" indicates an OOM(Out Of Memory) error.

REFERENCES

Jeongwhan Choi, Hwangyong Choi, Jeehyun Hwang, and Noseong Park. Graph neural controlled differential equations for traffic forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 6367–6374, 2022.

Yann N Dauphin, Angela Fan, Michael Auli, and David Grangier. Language modeling with gated convolutional networks. In *International conference on machine learning*, pp. 933–941. PMLR, 2017.

Ailin Deng and Bryan Hooi. Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 4027–4035, 2021.

Jianfei Gao and Bruno Ribeiro. On the equivalence between temporal and static equivariant graph representations. In *International Conference on Machine Learning*, pp. 7052–7076. PMLR, 2022.

Kan Guo, Yongli Hu, Zhen Qian, Hao Liu, Ke Zhang, Yanfeng Sun, Junbin Gao, and Baocai Yin. Optimized graph convolution recurrent neural network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 22(2):1138–1149, 2020.

Thomas Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, and Richard Zemel. Neural relational inference for interacting systems. In *International Conference on Machine Learning*, pp. 2688–2697. PMLR, 2018.

Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, 2017.

Chao Shang, Jie Chen, and Jinbo Bi. Discrete graph structure learning for forecasting multiple time series. *arXiv preprint arXiv:2101.06861*, 2021.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 753–763, 2020.

Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*, 2017.

Ling Zhao, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, Tao Lin, Min Deng, and Haifeng Li. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(9):3848–3858, 2019.