DarkFarseer: Robust Spatio-temporal Kriging under Graph Sparsity and Noise (Appendix)

A. Analysis and Motivation

In this section, we showcase the sparsity and noise issues in current graph structures. We first investigate two types of graph structures. Subsequently, we analyze the issues of inconsistent graph density and noisy edges present in the two graph construction methods.

A.1 Graphs for ISK

Deep learning-based kriging methods primarily rely on GNNs that necessitate constructing a graph to model the interrelationships of sensors (i.e., nodes in a graph). Heuristic-based Graphs (Jin et al. 2024) based on the distances between nodes is a common and effective approach (Wu et al. 2019; Song et al. 2020; Shang, Chen, and Bi; Liu et al. 2022; Cui et al. 2021; Cini, Mandic, and Alippi; Li et al. 2024b). There are generally two branches to constructing heuristic-based Graphs:

- *Pairwise Connectivity Graphs (PCGs)*: In this approach, the reachability and corresponding distance between two nodes are available according to expert intervention. Thus, the adjacency matrix can be built based on the pairwise distance:

$$\mathbf{A}_{i,j} = \begin{cases} d_{i,j} & \text{if node } i \text{ and } j \text{ are linked,} \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

  where $d_{i,j}$ is the distance between node i and j.

- *Spatial Proximity Graphs (SPGs)*: The reachability between nodes is unknown while the geographical coordinates of each node are available. Thus, it is essential to calculate the distance between any two nodes to construct a graph. The adjacency matrix can be formulated as:

$$\mathbf{A}_{i,j} = \begin{cases} \frac{\exp(-||d_{i,j}||_2)}{\sigma} & \text{if } \frac{\exp(-||d_{i,j}||_2)}{\sigma} < \varepsilon, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

  where $d_{i,j}$ is the node distance, σ is a hyper-parameter to normalize the distance distribution, and ε is the threshold to control the sparsity of the adjacency matrix.

Notably, PCGs require expert intervention for pairwise node distances, while SPGs connect nodes based on geographic distances and a pre-defined threshold ε.

Table 1: Statistical information of PCGs on different datasets.

Scenarios	Traffic Flow		Traffic Speed
Datasets	PEMS03	PEMS04	PEMS-BAY
# of Nodes	358	307	325
Avg Degree	3.05±1.11	2.20±1.01	49.43±27.14

A.2 Limitation Analysis

We conducted comprehensive preliminary experiments multiple times under various mask settings and made observations to confirm our conjectures and motivate the proposal of DARKFARSEER.

A.2.1 Inconsistent Graph Density in PCGs In this section, we analyze the issues caused by inconsistent graph density (sparse or dense) in PCGs. Through preliminary experiments, we conclude that topological prompts can alleviate graph sparsity, and discarding some edges can mitigate graph density.

We investigate three datasets, PEMS03, PEMS04, and PEMS-BAY, across two application scenarios: traffic flow and traffic speed. The statistical information for these datasets is summarized in Table 1. The average degree reflects the average number of connections each node has with others in the graph. It can be observed that the average degree for the graph for traffic flow ranges from 2 to 4. **Conjecture I:** *Fewer neighboring nodes may lead to insufficient interaction between nodes. Enhancing topological information could help alleviate this issue.* In contrast, for traffic speed, the average degree for PEMS-BAY is 49.43. **Conjecture II:** *Not all neighbors of a node are relevant, and an excessive number of neighbors may introduce noise. Reducing the number of edges could mitigate this problem.* To test these conjectures, we conduct two preliminary experiments using the classic graph-based ISK method IGNNK (Wu et al. 2021).

- **Preliminary Experiment I**: To examine **Conjecture I**, we perform edge-adding operations on the graph of the PEMS03 and PEMS04 datasets. Specifically, during the training phase, we randomly add edges between nodes

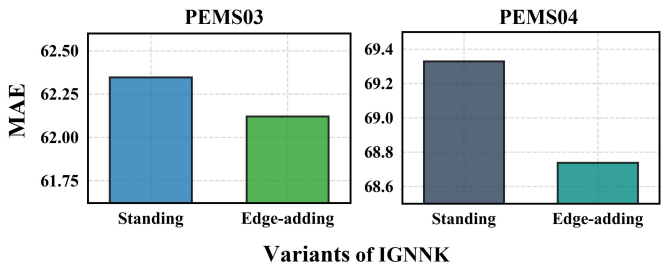

Figure 1: MAE comparison of standing IGNNK and edge-adding IGNNK. After adding edges, the MAE of IGNNK decreases.

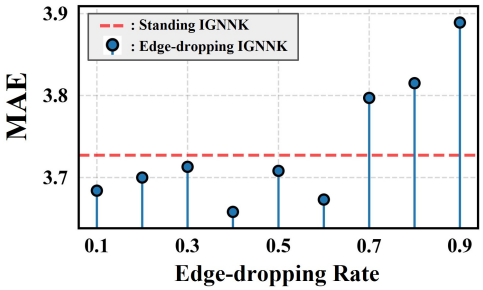

Figure 2: MAE comparison of IGNNK with different edge-dropping rate on PEMS-BAY dataset. Within a certain range of edge-dropping rates, IGNNK's MAE decreases.

and other nodes within their Biconnected Components (BCCs). This strategy aims to provide IGNNK with additional topological information, as BCCs being special subgraphs capture local connectivity. The empirical results are shown in Figure 1. After adding edges, the Mean Absolute Error (MAE) of IGNNK decreases, indicating improved performance.

Observation I: The PCGs are constrained by sparsity, which hinders their performance. However, these limitations can be alleviated by incorporating additional topological prompts into the ISK model.

- **Preliminary Experiment II**: To verify **Conjecture II**, we perform edge-dropping[1] operations on the graph constructed from the PEMS-BAY dataset. Specifically, we randomly remove edges with varying probabilities during training. The empirical results are shown in Figure 2. As seen, when the edge-dropping rate is between 0.1 and 0.6, the MAE of IGNNK decreases.

  Observation II: The PCGs are subject to a density limitation: excessive edges introduce noise into the model. Thus, designing noise-reduction strategies can help mitigate this issue and enhance performance.

A.2.2 Noise Edges in SPGs In this section, we analyze the issues caused by noise edges in SPGs. Through preliminary experiments, we observe that discarding some edges can mitigate the limitations introduced by noise edges.

[1] In this section, edge-dropping is implemented by replacing the edges with extremely low intensity values.

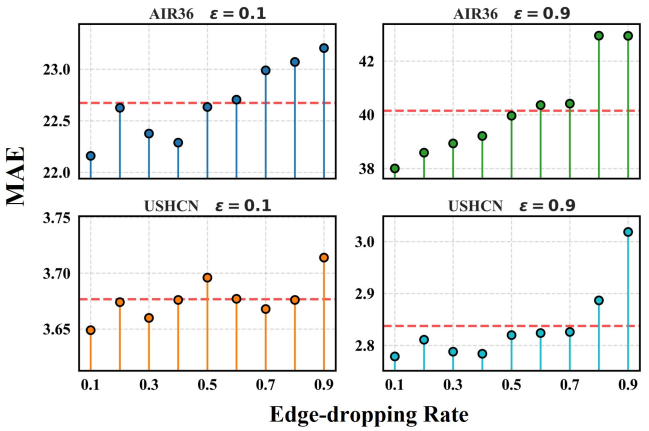

Figure 3: MAE comparison of IGNNK with different edge-dropping rates and ε on AIR36 and USHCN. The red horizontal line represents the MAE of the standing IGNNK. As observed, no matter how ε is selected, edge-dropping decreases MAE.

Table 2: Statistical information of SPGs on different datasets.

Scenarios	Air Quality		Precipitation	
Datasets	AIR36		USHCN	
# of Nodes	36		1218	
Threshold ε	0.1	0.9	0.1	0.9
Avg Degree	18.17±8.71	2.15±2.55	559.59±158.61	44.90±18.18

As discussed in Section A.1, unlike PCGs, which benefit from expert intervention, SPGs face greater limitations due to the lack of such expert knowledge. A crucial step in constructing SPGs is calculating the geographical distance between pairs of nodes. However, geographical distance often introduces noise, as proximity does not always guarantee reachability between two nodes. Table 2 summarizes the statistics of SPGs for the air quality and precipitation scenarios, using threshold values ϵ of 0.1 and 0.9. When $\epsilon = 0.1$, the average degree increases to 18.17 for AIR36 and 559.59 for USHCN, resulting in a denser graph. In contrast, when $\epsilon = 0.9$, the average degree of SPGs for AIR36 is 2.15. For USHCN, the average degree of SPGs drops to 44.90 compared to its 559.59 at $\epsilon = 0.1$, indicating a relatively sparse graph. **Conjecture III:** Using a large ϵ can reduce edge noise to some extent, but it does not fully mitigate the noise issue.

- **Preliminary Experiment III:** To validate **Conjecture III**, we conduct experiments on AIR36 and USHCN datasets with various edge-dropping rates and threshold values ϵ using IGNNK. The results are presented in Figure 3. When $\epsilon = 0.1$, we observe a decrease in MAE under certain edge-dropping rates, suggesting the presence

of noise. Notably, when $\epsilon = 0.9$, the MAE of IGNNK decreases as the edge-drop rate ranges from 0.1 to 0.5.

Observation III: Although increasing ϵ to make the SPGs sparser can help, noise caused by edges constructed from geographical distance still persists. In other words, the noise edges inherent to SPGs remain, regardless of the selected value for ϵ.

B. Experimental Configurations

Table 3: The overall information for datasets.

Datasets	Traffic Flow		Traffic Speed	Air Quality	Precipitation
	PEMS03	PEMS04	PEMS-BAY	AIR36	USHCN
Region	California		San Francisco	Beijing	United States
Graph type	Pairwise Connectivity Graphs			Spatial Proximity Graphs	
Nodes	358	307	325	36	1,218
Timesteps	26,208	16,992	52,128	8,759	1,440
Interval	5min	5min	5min	1hour	1month

Datasets. We evaluate DARKFARSEER on five benchmark datasets spanning four application scenarios: traffic flow, traffic speed, air quality, and precipitation. Specifically, PEMS03 and PEMS04 represent sparse PCGs, PEMS-BAY is a dense PCG, and AIR36 and USHCN are SPGs. Detailed dataset statistics are provided in Table 3.

- **PEMS03** (Song et al. 2020): The dataset includes data from multiple traffic sensors located along the I-880 highway in California.
- **PEMS04** (Song et al. 2020): The dataset includes data from multiple traffic monitoring sensors located on specific highway networks in California, primarily along the I-105 and SR-73 highways. We only consider the traffic flow output from PEMS03 and PEMS04.
- **PEMS-BAY** (Li et al. 2018): The dataset comprises data from numerous traffic sensors strategically positioned across the San Francisco Bay Area's extensive highway network.
- **AIR36** (Cini, Marisca, and Alippi 2022): The dataset is a specialized subset focused on air quality research. We consider only the PM2.5 pollutant.
- **USHCN** (Wu et al. 2021): The dataset provides monthly climate data for 1218 stations across the Contiguous United States. We utilize the precipitation data from this dataset, covering the period from 1899 to 2019.

Baselines. We select two traditional non-learning methods and five SOTA deep learning-based baselines for comparison:

- **MEAN**: At fixed timestamps, we use the mean of each observed node in the test set to fill virtual nodes.
- **KNN**: K-Nearest Neighbors fills virtual nodes by using the average of the K nearest observed nodes. We set $K = 10$.
- **IGNNK** (Wu et al. 2021) (AAAI'21): A graph-based ISK model employs a three-layer Graph Diffusion Layer with residual connections.
- **DualSTN** (Hu et al. 2023) (TNNLS'23): In both long-term and short-term modes, pattern learning is performed using graph and attention.
- **INCREASE** (Zheng et al. 2023) (WWW'23): INCREASE aggregates the heterogeneous spatial information of K-nearest neighbors to perform ISK.
- **KCP** (Li et al. 2024a) (AISTATS'24): A graph-based CL framework for ISK. It constructs spatial contrast and temporal heterogeneity contrast through neighbor comparisons and prototype heads.
- **KITS** (Xu et al. 2025) (AAAI'25): A model that introduces an Increment Training Strategy to mitigate the "graph gap" by adding virtual nodes to the training graph.

Metrics. Following previous work (Zheng et al. 2023; Li et al. 2024a; Xu et al. 2025), we adopt three commonly used evaluation metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Relative Error (MRE) (Cao et al. 2018; Cini, Marisca, and Alippi 2022). Lower values for these metrics indicate better model performance. The formulas are defined as follows:

$$\text{MAE} = \frac{1}{|\mathcal{V}^u|} \sum_{i \in \mathcal{V}^u} |\mathbf{X}_i - \hat{\mathbf{X}}_i|, \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{1}{|\mathcal{V}^u|} \sum_{i \in \mathcal{V}^u} (\mathbf{X}_i - \hat{\mathbf{X}}_i)^2}, \tag{4}$$

$$\text{MRE} = \frac{\sum_{i \in \mathcal{V}^u} |\mathbf{X}_i - \hat{\mathbf{X}}_i|}{\sum_{i \in \mathcal{V}^u} |\mathbf{X}_i|}. \tag{5}$$

Implementation Details. Unless otherwise specified, we set the random seed to 0 (this seed is used for the selection of virtual nodes) for selecting kriging nodes with an observed-to-virtual nodes ratio of 3:1. The dataset is split into training, validation, and test sets in a ratio of 6:2:2. Experiments are conducted four times, and the averaged results are reported. The time window p is set to 24. We normalize the training, validation, and test sets using the global statistics of the observed nodes from the training set. For optimization, the Adam optimizer is employed. The batch size is set to 32, the number of training epochs to 300, and gradient clipping is 5.0. The adjacency matrices constructed for all datasets are normalized using Equation 2, where ε is set to $+\infty$ for PEMS03, PEMS04 and PEMS-BAY, and to 0.1 for AIR36 and USHCN. In addition, our code is implemented using Python 3.10.9 and PyTorch 2.2.1. All experiments are conducted on a system equipped with an NVIDIA GeForce RTX 4090 GPU (with CUDA 11.1), a 12th Gen Intel(R) Core(TM) i9-12900K CPU, and 64GB of RAM, running Ubuntu 20.04.6 LTS. For detailed model parameters, please refer to the code.

C. Definition of Transductive SK (TSK)

In TSK, virtual sensors and the complete spatial relationships $\mathcal{R}$ are accessible during training. Let $\mathcal{F}$ denote a kriging model. The kriging during training and inference for TSK can be formulated as:

$$\hat{\mathbf{X}}^u_{t:t+p} = \mathcal{F}(\mathbf{X}^o_{t:t+p}, \mathcal{R}), \tag{6}$$

where p represents the time window size.

D. Detailed Related Work

Spatio-temporal Kriging

Kriging is a prominent geostatistical method used for spatial interpolation. Spatio-temporal kriging extends this technique to include the temporal dimension, enabling the estimation of values at unobserved locations at different times (Xu et al. 2025). Spatio-temporal kriging is broadly classified into transductive and inductive approaches. Transductive models require all nodes to be available during the training phase and are unable to naturally generate representations for new (virtual) nodes without retraining. Early Spatio-temporal kriging methods were often represented by transductive approaches, such as those based on matrix factorization (Zhou et al. 2012; Bahadori, Yu, and Liu 2014; Takeuchi, Kashima, and Ueda 2017) and graph-based approaches (Xu et al. 2020).

Although these works have achieved certain successes, TSK is difficult to extend to new nodes. Compared to TSK, inductive models are capable of handling new nodes introduced after training, making them more practical. ISK has recently sparked a wave of research (Appleby, Liu, and Liu 2020; Wu et al. 2021; Zhang et al. 2022; Hu et al. 2023; Xu et al. 2025; Li et al. 2024a). The key to ISK lies in transferring temporal patterns from observed nodes to virtual nodes. Deep learning-based ISK methods can be broadly categorized into two paradigms according to the scale of transfer: **partial transfer** and **global transfer**. Global transfer aims to migrate the temporal patterns of all nodes to virtual nodes. The most representative approaches employ GNNs. For example, IGNNK (Wu et al. 2021) uses Graph Diffusion Layers (Li et al. 2018), and KCP (Li et al. 2024a) utilizes GraphSAGE (Hamilton, Ying, and Leskovec 2017). Currently, the vast majority of methods rely on global transfer. Partial transfer generally involves migrating based on the k-nearest neighbors of virtual nodes. For instance, INCREASE (Zheng et al. 2023) aggregates and transfers information from heterogeneous k-neighbors specific to a given virtual node.

The quality of the graph structure significantly affects the effectiveness of graph-based ISK, yet few works have addressed the issue of graph structure quality. In contrast, our work, DARKFARSEER, is specifically designed to focus on graph modeling tailored for virtual nodes. This strategic focus aims to integrate and leverage the advantages offered by both partial transfer and global transfer, moving beyond the single-paradigm approaches dominant in the field.

Graph Modeling for Time Series

Since the variable dimensions of multivariate time series (MTS) can often be represented as graphs, the application of graph representation learning techniques to MTS analysis has become quite prevalent. Different types of graphs applied to MTS can facilitate the learning of varied inductive biases. For instance, the introduction of dynamic graphs allows models to learn time-varying spatial relationships (Ye et al. 2022; Zhao et al. 2023; Li et al. 2023; Liu et al. 2024). Incorporating hypergraphs enables the learning of more complex and intertwined many-to-many spatial relationships (Wu et al. 2023; Yin et al. 2023; Shang and Chen 2024). Thanks to its simple and effective design paradigm, some studies employ contrastive learning (CL). to enhance graph learning for time series (Qin et al. 2023; Wang et al. 2023).

CL has emerged as a powerful self-supervised paradigm for this task. It learns meaningful representations by maximizing the agreement between different augmented "views" of the same data, thereby reducing the dependency on labeled datasets. A critical component is the design of advanced augmentation strategies that preserve intrinsic data properties. For instance, SFA (Zhang et al. 2023) rebalances spectral features to enhance robustness, while other approaches introduce regularization terms to explicitly preserve the graph's topological structure. In specific applications like sequential recommendation, methods such as TGCL4SR (Zhang et al. 2024) employ dual augmentations like neighbor sampling and temporal perturbations to handle sparsity and noise. Notably, these innovations typically focus on the contrastive mechanism itself, rather than requiring specific modifications to the underlying GNN architecture. Furthermore, the utility of CL extends beyond learning robust embeddings to helping the model better understand and learn the graph structure itself. The S3-CL (Ding et al. 2023) improves node discrimination by contrasting representations derived from local versus higher-order neighborhoods, thus leveraging multi-scale structural information.

While these advanced methods provide a strong foundation, their application to the specific demands of ISK reveals challenges, particularly in unifying information transfer strategies for virtual nodes and ensuring robustness against imperfect graph structures. To address these very issues, the key modules of DARKFARSEER, namely RSCL and SGDs, have been meticulously developed. They are designed to significantly enhance the model's robustness, specifically to handle and mitigate challenges associated with sparse and noisy graph structures in the context of ISK.

References

Appleby, G.; Liu, L.; and Liu, L.-P. 2020. Kriging convolutional networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 3187–3194.

Bahadori, M. T.; Yu, Q. R.; and Liu, Y. 2014. Fast multivariate spatio-temporal analysis via low rank tensor learning. *Advances in neural information processing systems*, 27.

Cao, W.; Wang, D.; Li, J.; Zhou, H.; Li, L.; and Li, Y. 2018. Brits: Bidirectional recurrent imputation for time series. *Advances in neural information processing systems*, 31.

Cini, A.; Mandic, D.; and Alippi, C. ???? Graph-based Time Series Clustering for End-to-End Hierarchical Forecasting. In *Forty-first International Conference on Machine Learning*.

Cini, A.; Marisca, I.; and Alippi, C. 2022. Filling the G_ap_s: Multivariate Time Series Imputation by Graph Neural Networks. In *International Conference on Learning Representations*.

Cui, Y.; Zheng, K.; Cui, D.; Xie, J.; Deng, L.; Huang, F.; and Zhou, X. 2021. METRO: a generic graph neural network

framework for multivariate time series forecasting. *Proceedings of the VLDB Endowment*, 15(2): 224–236.

Ding, K.; Wang, Y.; Yang, Y.; and Liu, H. 2023. Eliciting structural and semantic global knowledge in unsupervised graph contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 7378–7386.

Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.

Hu, J.; Liang, Y.; Fan, Z.; Liu, L.; Yin, Y.; and Zimmermann, R. 2023. Decoupling long-and short-term patterns in spatiotemporal inference. *IEEE Transactions on Neural Networks and Learning Systems*.

Jin, M.; Koh, H. Y.; Wen, Q.; Zambon, D.; Alippi, C.; Webb, G. I.; King, I.; and Pan, S. 2024. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *International Conference on Learning Representations*.

Li, Z.; Nie, Y.; Li, Z.; Bai, L.; Lv, Y.; and Zhao, R. 2024a. Non-Neighbors Also Matter to Kriging: A New Contrastive-Prototypical Learning. In *International Conference on Artificial Intelligence and Statistics*, 46–54. PMLR.

Li, Z.; Xia, L.; Xu, Y.; and Huang, C. 2024b. GPT-ST: generative pre-training of spatio-temporal graph neural networks. *Advances in Neural Information Processing Systems*, 36.

Li, Z. L.; Zhang, G. W.; Yu, J.; and Xu, L. Y. 2023. Dynamic graph structure learning for multivariate time series forecasting. *Pattern Recognition*, 138: 109423.

Liu, H.; Yang, D.; Liu, X.; Chen, X.; Liang, Z.; Wang, H.; Cui, Y.; and Gu, J. 2024. Todynet: temporal dynamic graph neural network for multivariate time series classification. *Information Sciences*, 120914.

Liu, Y.; Liu, Q.; Zhang, J.-W.; Feng, H.; Wang, Z.; Zhou, Z.; and Chen, W. 2022. Multivariate time-series forecasting with temporal polynomial graph neural networks. *Advances in neural information processing systems*, 35: 19414–19426.

Qin, S.; Chen, L.; Luo, Y.; and Tao, G. 2023. Multi-view graph contrastive learning for multivariate time series anomaly detection in IoT. *IEEE Internet of Things Journal*.

Shang, C.; Chen, J.; and Bi, J. ???? Discrete Graph Structure Learning for Forecasting Multiple Time Series. In *International Conference on Learning Representations*.

Shang, Z.; and Chen, L. 2024. Mshyper: Multi-scale hypergraph transformer for long-range time series forecasting. *arXiv preprint arXiv:2401.09261*.

Song, C.; Lin, Y.; Guo, S.; and Wan, H. 2020. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 914–921.

Takeuchi, K.; Kashima, H.; and Ueda, N. 2017. Autoregressive tensor factorization for spatio-temporal predictions. In *2017 IEEE international conference on data mining (ICDM)*, 1105–1110. IEEE.

Wang, Y.; Xu, Y.; Yang, J.; Wu, M.; Li, X.; Xie, L.; and Chen, Z. 2023. Graph Contextual Contrasting for Multivariate Time Series Classification. *arXiv preprint arXiv:2309.05202*.

Wu, J.; Qi, Q.; Wang, J.; Sun, H.; Wu, Z.; Zhuang, Z.; and Liao, J. 2023. Not Only Pairwise Relationships: Fine-Grained Relational Modeling for Multivariate Time Series Forecasting. In *IJCAI*, 4416–4423.

Wu, Y.; Zhuang, D.; Labbe, A.; and Sun, L. 2021. Inductive graph neural networks for spatiotemporal kriging. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 4478–4485.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; and Zhang, C. 2019. Graph wavenet for deep spatial-temporal graph modeling. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 1907–1913.

Xu, D.; Wei, C.; Peng, P.; Xuan, Q.; and Guo, H. 2020. GE-GAN: A novel deep learning framework for road traffic state estimation. *Transportation Research Part C: Emerging Technologies*, 117: 102635.

Xu, Q.; Long, C.; Li, Z.; Ruan, S.; Zhao, R.; and Li, Z. 2025. Kits: Inductive spatio-temporal kriging with increment training strategy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 12945–12953.

Ye, J.; Liu, Z.; Du, B.; Sun, L.; Li, W.; Fu, Y.; and Xiong, H. 2022. Learning the evolutionary and multi-scale graph structure for multivariate time series forecasting. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, 2296–2306.

Yin, N.; Shen, L.; Xiong, H.; Gu, B.; Chen, C.; Hua, X.-S.; Liu, S.; and Luo, X. 2023. Messages are never propagated alone: Collaborative hypergraph neural network for time-series forecasting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhang, S.; Chen, L.; Wang, C.; Li, S.; and Xiong, H. 2024. Temporal graph contrastive learning for sequential recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 9359–9367.

Zhang, Y.; Li, A.; Li, J.; Han, D.; Li, T.; Zhang, R.; and Zhang, Y. 2022. SpecKriging: GNN-based secure cooperative spectrum sensing. *IEEE Transactions on Wireless Communications*, 21(11): 9936–9946.

Zhang, Y.; Zhu, H.; Song, Z.; Koniusz, P.; and King, I. 2023. Spectral feature augmentation for graph contrastive learning and beyond. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 11289–11297.

Zhao, K.; Guo, C.; Cheng, Y.; Han, P.; Zhang, M.; and Yang, B. 2023. Multiple time series forecasting with dynamic graph modeling. *Proceedings of the VLDB Endowment*, 17(4): 753–765.

Zheng, C.; Fan, X.; Wang, C.; Qi, J.; Chen, C.; and Chen, L. 2023. Increase: Inductive graph representation learning for spatio-temporal kriging. In *Proceedings of the ACM Web Conference 2023*, 673–683.

Zhou, T.; Shan, H.; Banerjee, A.; and Sapiro, G. 2012. Kernelized probabilistic matrix factorization: Exploiting graphs and side information. In *Proceedings of the 2012 SIAM international Conference on Data mining*, 403–414. SIAM.