26PESGM1229

Spatiotemporal Probabilistic Forecasting of EV Charging Load with Conditional Diffusion Model

Linbo Zhang
School of Electrical Engineering
Xi'an Jiaotong University
Xi'an, China
978072449@stu.xjtu.edu.cn

Zaibin Jiao
School of Electrical Engineering
Xi'an Jiaotong University
Xi'an, China
jiaozaibin@mail.xjtu.edu.cn

Chenhao Zhao
School of Electrical Engineering
Xi'an Jiaotong University
Xi'an, China
kobe2488@stu.xjtu.edu.cn

Zongbo Li
School of Electrical Engineering
Xi'an Jiaotong University
Xi'an, China
lizongbo@xjtu.edu.cn

Jiawei Yuan
School of Electrical Engineering
Xi'an Jiaotong University
Xi'an, China
yuanjiawei@xjtu.edu.cn

Hanwen Gu
School of Electrical Engineering
Xi'an Jiaotong University
Xi'an, China
guhanwen0104@xjtu.edu.cn

***Abstract*—The rapid growth in the number of electric vehicles has led to a continuous increase in charging demand. However, the randomness and volatility of charging behavior pose new challenges to power grid operation. Therefore, accurate prediction and reliable quantification of charging load uncertainty have become particularly important. However, many existing methods focus only on the time dimension or rely on deterministic prediction, often neglecting the spatial correlation between charging stations. To overcome these limitations, this paper proposes a Spatiotemporal Conditional Diffusion Model (STCDM) for probabilistic charging load prediction of charging station clusters. This model introduces a spatial graph structure as a diffusion condition and designs a spatiotemporal denoising network to fully extract the spatiotemporal correlation features of the load. Experimental results based on real charging data show that STCDM significantly outperforms existing benchmark models in terms of prediction accuracy and uncertainty estimation.**

***Index Terms*—Electric vehicle charging load forecasting, spatiotemporal modeling, probabilistic forecasting, diffusion model.**

I. Introduction

In recent years, electric vehicles (EVs) have become increasingly popular worldwide, and the demand for charging has risen sharply[1]. However, the randomness and volatility of charging behavior bring significant uncertainty to the power distribution network, which can lead to load fluctuations, transformer overloads and voltage deviations. Therefore, accurate forecasting of electric vehicle charging load is crucial for the safe operation of the power grid, demand-side management and smart grid planning[2].

Existing studies on EV charging load forecasting can be broadly grouped into three categories. Model-based methods, such as stochastic queuing models, Monte Carlo simulations, and Markov processes, offer strong interpretability and clear physical meaning [3]. Ref. [4] models EV arrivals and charging service at fast-charging stations using queuing theory and traffic flow prediction. However, these approaches often rely on simplified behavioral and traffic assumptions, which limits accuracy in large-scale systems with strong spatiotemporal coupling.

Data-driven methods, including statistical models and classical machine learning, can capture nonlinear patterns in charging data to some extent. Ref. [5] employs Random Forest and neural-network-based predictors, and Ref. [6] compares SVM, Random Forest, and Gradient Boosting for load forecasting. Despite their effectiveness under certain conditions, these methods typically require substantial feature engineering and struggle to represent long-range temporal dependencies and cross-site correlations.

Deep learning methods further enhance modeling capacity for nonlinear and high-dimensional dynamics. Recurrent architectures such as RNNs, LSTM, and GRU are widely used in short-term load forecasting and generally achieve higher accuracy. Ref. [7] proposes an LSTM-based EV load forecasting model with a gating mechanism that remains robust under degraded data quality. Ref. [8] proposed the LSTM-Transformer hybrid model and verified that RNN/Transformer can better model time dependence. Meanwhile, convolutional neural networks (CNN) have also been used to extract local time features from charging time series [9]. Despite these advances, most existing deep learning models only focus on the time dimension and ignore the spatial correlation between multiple charging stations [10].

In station clusters, charging demand at each site is influenced by nearby stations due to user mobility, proximity, and price sensitivity, so modeling spatial dependence is essential. Graph-based spatiotemporal models improve accuracy, but most of them provide deterministic forecasts and do not explicitly quantify uncertainty, limiting their value for operational planning. Recent studies have used graph convolutional networks (GCNs) [11] and graph attention networks (GATs) [12] to capture the correlation between stations, which has significantly improved the spatiotemporal

979-8-3315-8138-1

prediction performance. However, these methods are usually deterministic point predictions, ignoring the spatial uncertainty of electric vehicle charging behavior, which limits their application value in grid operation [13-15].

Therefore, this paper proposes a Spatiotemporal Conditional Diffusion Model (STCDM) for probabilistic prediction of electric vehicle charging load at multiple charging stations. This model is an improvement on the diffusion framework, constructing the conditional diffusion model by introducing a spatial graph structure and designing a denoising network to extract multi-scale spatiotemporal features, thereby capturing complex spatiotemporal dependencies and providing charging station clusters with prediction results that balance accuracy and uncertainty.

The remainder of this paper is organized as follows: Section II details the conditional diffusion model and the design of the denoising network; Section III tests the model prediction results and conducts ablation experiments; Section IV summarizes the entire paper.

II. Probabilistic Forecasting Framework

A. Denoising Diffusion Model

DDPM learns a data distribution by progressively adding Gaussian noise and then reversing the process via a learned denoiser, enabling uncertainty-aware probabilistic forecasting, as illustrated in Fig. 1. In this framework, two stochastic processes are defined: the forward diffusion and the reverse denoising processes.

1) Forward Process

The forward process gradually converts a clean sample x_0 into noise by sequentially adding Gaussian perturbations. The model for each step is as follows:

$$q(x_n \mid x_{n-1}) = \mathrm{N}\left(x_n; \sqrt{1-\beta_n}\, x_{n-1}, \beta_n I\right) \tag{1}$$

where $\beta_n \in (0,1)$ controls the noise intensity at step n, and I denotes the identity matrix. By chaining all steps, we obtain:

$$q(x_{1:N} \mid x_0) = \prod_{n=1}^{N} q(x_n \mid x_{n-1}) \tag{2}$$

Defining $\alpha_n = 1 - \beta_n$ and $\hat{\alpha}_n = \prod_{i=1}^{n} \alpha_n$, the closed-form expression for any step can be derived as:

$$x_n = \sqrt{\bar{\alpha}_n}\, x_0 + \sqrt{1-\bar{\alpha}_n}\, \epsilon, \quad \epsilon \sim \mathrm{N}(0, I) \tag{3}$$

This formulation allows direct sampling of x_n at any noise level without iterative computation of intermediate states.

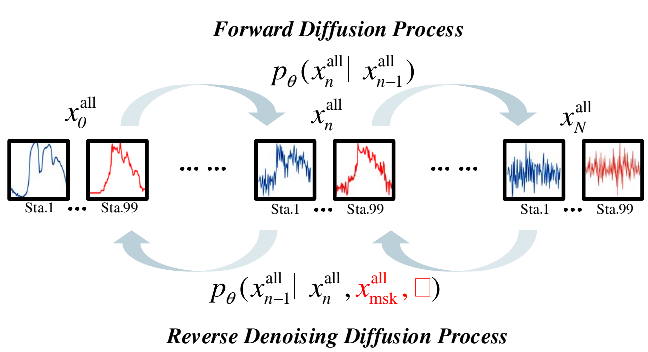

Fig. 1. Forward diffusion and reverse diffusion process.

2) Reverse Process

The reverse process learns to recover data from noise by inverting the forward diffusion through a parameterized Markov chain. Starting from pure Gaussian noise $x_N \sim \mathrm{N}(0, I)$, the process is defined as:

$$p_\theta(x_{0:N}) = p(x_N) \prod_{n=1}^{N} p_\theta(x_{n-1} \mid x_n) \tag{4}$$

Given sufficiently small β_n, the reverse transitions share the same Gaussian form as the forward process:

$$p_\theta(x_{n-1} \mid x_n) = \mathrm{N}\left(x_{n-1}; \mu_\theta(x_n, n), \sigma_\theta(x_n, n)\right) \tag{5}$$

Where, the mean μ_θ and variance σ_θ are predicted by a neural network parameterized by θ. The mean can be expressed as:

$$\mu_\theta(x_n, n) = \frac{1}{\sqrt{\alpha_n}}\left(x_n - \frac{\beta_n}{\sqrt{1-\bar{\alpha}_n}}\epsilon_\theta(x_n, n)\right) \tag{6}$$

where $\epsilon_\theta(x_n, n)$ denotes the predicted noise at step n.

The model is trained by minimizing the expected mean squared error between the true and predicted noise:

$$\min_\theta \mathrm{L}(\theta) = \min_\theta \mathrm{E}\left\|\epsilon - \epsilon_\theta\left(\sqrt{\alpha_n}\, x_0 + \sqrt{1-\alpha_n}\, \epsilon, n\right)\right\|_2^2 \tag{7}$$

After training, new samples are generated by iteratively denoising from $x_N \sim \mathrm{N}(0, I)$ to x_0.

B. Conditional Diffusion Model

Deterministic spatiotemporal forecasters capture spatial–temporal dependencies but cannot quantify predictive uncertainty, whereas diffusion-based probabilistic models characterize uncertainty well yet often underuse spatial structure. To address these limitations, we extend conditional diffusion to graph-structured spatiotemporal forecasting and propose the Spatiotemporal Conditional Diffusion Model (STCDM), which jointly learns spatiotemporal representations and uncertainty, improving accuracy and distributional robustness.

1) Conditional Diffusion Process

Since charging behavior is influenced by both spatial correlation between charging stations and external factors such as electricity price, weather, and time-of-day patterns, we concatenate time features, meteorological features, and price features with the load sequence to form a node feature tensor $x_t \in \mathbb{R}^{F \times V \times T}$. All exogenous features are resampled/aligned to the 5-min grid and restricted to information available at forecast issue time. We then incorporate spatiotemporal context information into the model to make the diffusion process more suitable for charging load prediction scenarios, where F represents the number of feature dimensions, V represents the number of charging stations, and T represents the time horizon. A graph structure $\mathrm{G} = \{\mathrm{V}, \mathrm{E}, A\}$ is introduced to represent spatial dependencies, where $A \in \mathbb{R}^{V \times V}$ is a weighted adjacency matrix based on the geographical distances between stations.

The model aims to predict the probabilistic distribution of future charging loads x^p conditioned on historical observations x^h, time steps T^h, and graph structure G. Extending the DDPM framework, the conditional reverse diffusion process is formulated as:

$$p_\theta(x^p_{0:N} \mid x^h, \mathrm{G}) = p(x^p_N) \prod_{n=1}^{N} p_\theta(x^p_{n-1} \mid x^p_n, x^h, \mathrm{G}) \tag{8}$$

The transition distribution follows:

$$\begin{aligned} & p_\theta(x^p_{n-1} \mid x^p_n, x^h, \mathrm{G}) \\ & = \mathrm{N}\left(x^p_{n-1}; \mu_\theta(x^p_n, n \mid x^h, \mathrm{G}), \sigma_\theta(x^p_n, n \mid x^h, \mathrm{G})\right) \end{aligned} \tag{9}$$

where μ_θ and σ_θ are predicted by a neural network conditioned on both x^h and G. The conditional training objective is given by:

$$\min_\theta \mathrm{L}(\theta) = \min_\theta \mathrm{E}_{x^p_0, \epsilon} \left\| \epsilon - \epsilon_\theta(x^p_n, n \mid x^h, \mathrm{G}) \right\|_2^2 \tag{10}$$

This formulation enables the diffusion model to learn complex spatiotemporal and contextual dependencies while maintaining uncertainty-aware probabilistic forecasting.

2) Generalized Conditional Diffusion Model

The historical sequence x^h and prediction target x^p belong to separate sample spaces, which limits temporal information flow between consecutive intervals. To overcome this, we define a unified latent variable $x^{all}=[x^h, x^p] \in R^{F \times V \times T}$, where $T=T_h+T_p$ denotes the total time horizon. The unobserved future part is masked as $x_{msk}{}^{all}$, enabling both historical and predictive data to share the same spatiotemporal space.

Under this unified representation, the reverse diffusion process is reformulated as a masked conditional diffusion:

$$p_\theta(x^{\mathrm{all}}_{0:N} \mid x^{\mathrm{all}}_{\mathrm{msk}}, \mathrm{G}) = p(x^{\mathrm{all}}_N) \prod_{n=1}^{N} p_\theta(x^{\mathrm{all}}_{n-1} \mid x^{\mathrm{all}}_n, x^{\mathrm{all}}_{\mathrm{msk}}, \mathrm{G}) \tag{11}$$

with learnable parameters θ. The corresponding training objective follows the conditional denoising score matching formulation:

$$\min_\theta \mathrm{L}(\theta) = \min_\theta \mathrm{E}_{x^{\mathrm{all}}_0, \epsilon} \left\| \epsilon - \epsilon_\theta(x^{\mathrm{all}}_n, n \mid x^{\mathrm{all}}_{\mathrm{msk}}, \mathrm{G}) \right\|_2^2 \tag{12}$$

Compared with the original conditional setting, this generalized form integrates x^h and x^p into a single latent space for joint reconstruction and prediction, and enhances spatiotemporal continuity across historical–future boundaries, improving the model's ability to learn holistic dynamic distributions.

C. Spatiotemporal Denoising Network

To model spatiotemporal dependencies in graph-structured time series, we design a spatiotemporal denoising graph encoder–decoder. As shown in Fig. 2, the model performs reverse diffusion using local graph convolution and gated temporal filtering, enabling the extraction of both temporal continuity and spatial coherence. The encoder and decoder each consist of multiple Spatiotemporal Blocks (ST-Blocks) that integrate temporal, spatial, and noise embeddings. This symmetric architecture jointly captures temporal dynamics, spatial structure, and noise evolution throughout the diffusion process.

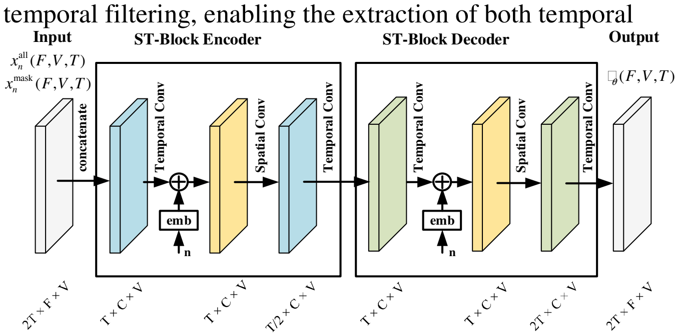

Fig. 2. Structure of the spatiotemporal denoising graph block.

1) Temporal Layers

The temporal layer encodes short-term and periodic patterns using gated 1D convolutions. Given input sequence x and kernel $\Omega \in R^K$, temporal convolution aggregates neighbor observations within a sliding window:

$$z_t = \mathrm{GLU}(x) = (\Omega * x) \odot \sigma(\Omega * x) \tag{13}$$

where $*$ denotes convolution, $\sigma(\cdot)$ is the sigmoid gate, and $\odot$ is element-wise multiplication. This gating mechanism adaptively regulates temporal information flow, enhancing physically consistent evolution while suppressing fluctuations. Deconvolution layers mirror this structure to recover the original temporal resolution after denoising.

2) Noise-Level Embedding

To maintain spatiotemporal consistency with the diffusion process, each ST-Block introduces a noise-level embedding layer. The noise level $n \in [1, N]$ is encoded using sinusoidal position encoding:

$$\mathbf{e}(n) = \left[\ldots, \cos\left(n / r^{-\frac{2d}{D}}\right), \sin\left(n / r^{-\frac{2d}{D}}\right), \ldots \right]^{\mathrm{T}} \tag{14}$$

where d indexes embedding dimensions and r=10,000.

The temporal output is modulated via affine transformations:

$$\tilde{Z}_t = Z_t \odot \left(1 + \phi(e_t)\right) + \psi(e_t) \tag{15}$$

where ϕ and ψ are linear mappings. This modulation adaptively rescales and shifts temporal features according to the denoising stage, enhancing generative consistency and stability.

3) Spatial Layers

The spatial layer explicitly models graph topology and spatial correlations using spectral graph convolution over the adjacency matrix A:

$$H_t = \sum_{k=0}^{K} \theta_k T_k(\tilde{L}) \tilde{Z}_t \tag{16}$$

$$\tilde{L} = \frac{2L}{\lambda_{\max}} - I \tag{17}$$

where $L=I-D^{-1/2}AD^{-1/2}$ is the normalized Laplacian, and $T_k(\cdot)$ denotes the Chebyshev polynomial of order k.

This operation aggregates contextual information from K-hop neighbors, embedding both local smoothness and global structural coherence. Empirically, K=3 achieves a balance between accuracy and computational efficiency.

By stacking ST-Blocks in a hierarchical encoder–decoder architecture, STCDM progressively refines the noise representations, enabling multi-scale temporal compression, diffusion-aware transformation, and robust spatial correlation inference.

D. Training and Inference Process

1) Training Process

During training, the masked tensor $x_{msk}{}^{all}$ is constructed based on observed entries. Given the complete ground-truth signal $x_0{}^{all}$ and graph G, noisy samples are generated as $x_n^{\mathrm{all}} = \sqrt{\bar{\alpha}_n}\, x_0^{\mathrm{all}} + \sqrt{1-\bar{\alpha}_n}\, \epsilon$,The denoising network ϵ_θ is optimized by minimizing the conditional objective in Eq. (12), learning to recover noise components conditioned on both the masked input and graph topology. This enables the model to capture spatially coherent and temporally consistent structures under stochastic diffusion noise.

2) Inference Process

During inference, the trained model iteratively samples $x_{n-1}{}^{all}$ from Eq. (11) through the reverse diffusion process, guided by $x_{msk}{}^{all}$ and graph G. The proposed STCDM framework generates all future steps within a single reverse sequence, ensuring multi-step temporal consistency, spatial coherence, and computational efficiency in probabilistic load forecasting.

III. Case Study

The proposed STCDM method was evaluated on real-world datasets, compared with representative probabilistic baselines, and validated through a series of ablation studies.

A. *Dataset and Experimental Setup*

1) Dataset

We used a real-world dataset from public electric vehicle charging stations in Shenzhen, China, covering a six-month period from September 1, 2022 to February 28, 2023. This dataset records 1,682 charging stations and 24,798 charging piles at a time resolution of 5 minutes. Each record includes a matrix of charging load, electricity price, weather conditions, and station spacing, enabling the model to capture the spatial and seasonal variations in charging behavior.

2) Evaluation Metrics

We adopt the Continuous Ranked Probability Score (CRPS) to evaluate probabilistic forecasting performance, which measures the discrepancy between the predicted cumulative distribution F and the observed value x:

$$\mathrm{CRPS}(F,x) = \int_{\mathbb{R}} \left(F(z) - \mathrm{I}\{x \le z\}\right)^2 \mathrm{d}z \tag{18}$$

For deterministic evaluation, we also report Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE):

$$\mathrm{MAE}(Y,\hat{Y}) = \frac{1}{|Y|}\sum_{i=1}^{|Y|} |Y_i - \hat{Y}_i| \tag{19}$$

$$\mathrm{RMSE}(Y,\hat{Y}) = \sqrt{\frac{1}{|Y|}\sum_{i=1}^{|Y|} (Y_i - \hat{Y}_i)^2} \tag{20}$$

Lower values of these metrics denote higher forecasting accuracy, while CRPS primarily measures the reliability of probabilistic outputs.

3) Parameter Settings

The dataset was divided chronologically in a 6:2:2 ratio into training, validation, and test sets. The model was trained using the Adam optimizer with a learning rate of 0.002 for 200 epochs. The batch size is 16, the diffusion steps were $T = 200$, the latent embedding dimension was set to 32, and the historical prediction timeframes were $T_h = 12$ and $T_h = \{3,6,9,12\}$. During training, we optimize the denoiser using the conditional denoising objective in Eq. (12) without backpropagating through all diffusion steps, so the per-iteration computational cost is comparable to that of a standard spatiotemporal forecasting network. At inference time, the runtime increases approximately linearly with the number of reverse diffusion steps N. To control spatial computation, we adopt a low-order Chebyshev approximation (K=3) in the graph convolution. We report parameter count and empirical end-to-end inference latency as practical indicators of computational cost. On a single NVIDIA RTX 4070, STCDM with 3.76M parameters forecasts all 1,682 stations in 15.4 s using 50 reverse steps and 38.6 s using 200 reverse steps, corresponding to 9.1 ms and 22.9 ms per station.

B. *Predictive Performance Evaluation*

To evaluate the predictive performance and generalization ability of the proposed STCDM model, we adopted a cross-time and cross-site testing method, using data from different observation stations on the same day and data from the same observation station on different dates. This design covers both spatial variability between observation stations and temporal variability between seasons, thereby validating the model's performance.

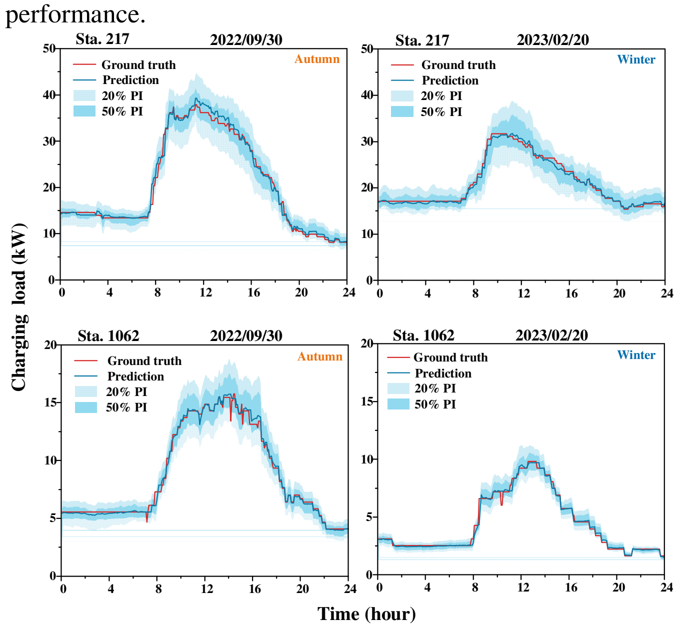

Fig. 3. Predictive performance across different sites and seasons

As shown in Fig. 3, the proposed STCDM model accurately captures charging load trajectories across different sites and seasons. For sites 217 and 1062, the predicted curves closely match actual measurements during both peak and off-peak periods, demonstrating the excellent nonlinear time series characterization capabilities of model. The 20% and 50% prediction intervals are well-calibrated, with most observed data points falling within the expected uncertainty range, confirming the ability to stably quantify uncertainty.

Cross-seasonal experimental results demonstrate that the model successfully captures variations in load amplitude and peak duration from autumn to winter, proving its prediction accuracy under different time conditions. Similarly, the consistent performance across different sites confirms that STCDM effectively captures spatial dependence, thus

979-8-3315-8138-1

maintaining robust prediction accuracy in regions with different load characteristics.

TABLE I

COMPARED TO THE PROBABILITY BASELINE

Metric / Method	CRPS				
	15	30	45	60	Average
DeepAR	0.39	0.45	0.51	0.57	0.48
CSDI	0.33	0.38	0.42	0.49	0.41
TimeGrad	0.31	0.35	0.40	0.45	0.38
STCDM	0.23	0.27	0.35	0.39	0.31

We compared STCDM with three probabilistic baseline models: DeepAR, CSDI, and TimeGrad. All methods were trained under the same settings and evaluated using the CRPS metric. As shown in Table I, STCDM achieved the lowest CRPS values across all prediction periods, indicating its ability to accurately capture the spatiotemporal characteristics between charging stations while maintaining the accuracy of uncertainty measurements in the diffusion model. STCDM also achieves the lowest Brier scores across all horizons, ranging from 0.078 to 0.105, indicating improved calibration compared with the baselines.

Overall, the results confirm that STCDM has high prediction accuracy and can quantify the uncertainty of charging behavior. Furthermore, this model is applicable to real-world charging station cluster scenarios with different spatial layouts and seasonal characteristics.

C. *Ablation Study*

We conducted elimination studies to evaluate the contributions of key components in STCDM, including the spatial module, the temporal module, and the proposed spatiotemporal coding-decoding module (retaining only the TCN for feature extraction). All variants were evaluated using mean squared error (MAE), root mean square error (RMSE), and CRPS under the same settings.

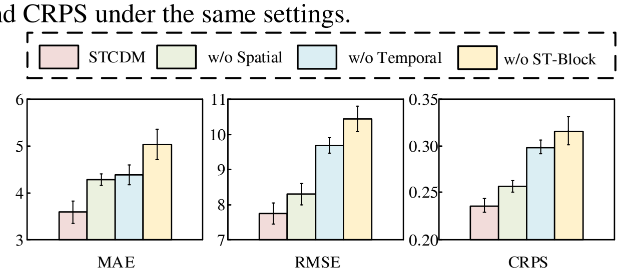

Fig. 4. Ablation study

Fig. 4 shows that removing any component leads to a degraded performance across all metrics. Removing the spatial module severely impacts accuracy, validating the importance of modeling spatial correlations between nodes. Removing the temporal module causes degradation across all metrics. Removing the spatiotemporal coding-decoding module results in the largest performance degradation, confirming its central role in jointly capturing spatiotemporal dependencies. These results demonstrate the necessity of each component.

IV. CONCLUSIONS

This paper proposes a Spatiotemporal Conditional Diffusion Model (STCDM) for probabilistic prediction of electric vehicle charging load across multiple charging stations. By combining graph-based spatial conditioning with a spatiotemporal denoiser, STCDM captures cross-station dependence and delivers competitive probabilistic accuracy on real data. Ablation results verify the necessity of the spatial, temporal, and joint spatiotemporal components, and the graph-conditioned formulation facilitates transfer by updating the station graph and normalization, with optional lightweight fine-tuning under domain shifts.

REFERENCES

[1] A. Paudel, P. Mandal, and G. Ravikumar, "Forecasting charging demand patterns of electric vehicles using hybrid deep learning approaches," in Proc. IEEE Power & Energy Society General Meeting (PESGM), Seattle, WA, USA, 2024, pp. 1–5.

[2] H. T. Nguyen and D.-H. Choi, "Distributionally robust model predictive control for smart electric vehicle charging station with V2G/V2V capability," IEEE Transactions on Smart Grid, vol. 14, no. 6, pp. 4621–4633, 2023.

[3] K. Liu and Y. Liu, "Stochastic user equilibrium based spatial-temporal distribution prediction of electric vehicle charging load," Applied Energy, vol. 339, no. 120943, 2023.

[4] X. Zhang, K. W. Chan, H. Li, H. Wang, J. Qiu, and G. Wang, “Deep-Learning-Based Probabilistic Forecasting of Electric Vehicle Charging Load with a Novel Queuing Model,” IEEE Trans. Cybern., vol. 51, no. 6, pp. 3157–3170, 2021.

[5] W. Khan, W. Somers, S. Walker, K. de Bont, J. Van der Velden, and W. Zeiler, "Comparison of electric vehicle load forecasting across different spatial levels with incorporated uncertainty estimation," Energy, vol. 283, no. 129213, 2023.

[6] W. U. Rehman, R. Bo, H. Mehdipourpicha, and J. Kimball, "Sizing energy storage system for energy arbitrage in extreme fast charging station," in Proc. IEEE Power & Energy Society General Meeting (PESGM), Washington, DC, USA, 2021, pp. 1–5.

[7] X. Shen, H. Zhao, Y. Xiang, P. Lan, and J. Liu, “Short-term electric vehicles charging load forecasting based on deep learning in low-quality data environments,” Electric Power Systems Research, vol. 212, no. 108247, pp, 0378-7796, 2022.

[8] A. Hussain, V. Eswarakrishnan, A. Aslam, and others, "c," Scientific Reports, vol. 15, no. 36639, 2025.

[9] X. Li, Y. Xu, X. Zhang, W. Shi, Y. Yue, and Q. Li, "Improving short-term bike sharing demand forecast through an irregular convolutional neural network," Transportation Research Part C: Emerging Technologies, vol. 147, no. 103984, 2023.

[10] K. Liu and Y. Liu, "Stochastic user equilibrium based spatial-temporal distribution prediction of electric vehicle charging load," Applied Energy, vol. 339, no. 120943, 2023.

[11] J. Shi, W. Zhang, Y. Bao, D. W. Gao, and Z. Wang, "Load forecasting of electric vehicle charging stations: Attention based spatiotemporal multi–graph convolutional networks," IEEE Transactions on Smart Grid, vol. 15, no. 3, pp. 3016–3027, 2024.

[12] T. Xie, X. Yun, G. Zhang, H. Li, K. Zhang, and R. Wang, "Charging station cluster load prediction: Spatiotemporal multi-graph fusion technology," Renewable and Sustainable Energy Reviews, vol. 206, no. 114855, 2024.

[13] X. Huang, D. Wu, and B. Boulet, "MetaProbformer for charging load probabilistic forecasting of electric vehicle charging stations," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 10, pp. 10445–10455, 2023.

[14] X. Zhang, K. W. Chan, H. Li, H. Wang, J. Qiu, and G. Wang, "Deep-learning-based probabilistic forecasting of electric vehicle charging load with a novel queuing model," IEEE Transactions on Cybernetics, vol. 51, no. 6, pp. 3157–3170, 2020.

[15] K. Zheng, H. Xu, Z. Long, Y. Wang, and Q. Chen, "Coherent hierarchical probabilistic forecasting of electric vehicle charging demand," IEEE Transactions on Industry Applications, vol. 61, no. 1, pp. 1329–1340, Jan.–Feb. 2025.