

TrustEnergy: A Unified Framework for Accurate and Reliable User-level Energy Usage Prediction

Dahai Yu¹, Rongchao Xu¹, Dingyi Zhuang², Yuheng Bu³, Shenhao Wang⁴, Guang Wang^{1*}

¹Florida State University, Tallahassee, Florida

²Massachusetts Institute of Technology, Cambridge, Massachusetts

³University of California, Santa Barbara, Santa Barbara, California

⁴University of Florida, Gainesville, Florida

dahai.yu@fsu.edu, rxu@fsu.edu, dingyi@mit.edu, buyuheng@ucsb.edu, shenhaowang@ufl.edu, guang@cs.fsu.edu

Abstract

Energy usage prediction is important for various real-world applications, including grid management, infrastructure planning, and disaster response. Although a plethora of deep learning approaches have been proposed to perform this task, most of them either overlook the essential spatial correlations across households or fail to scale to individualized prediction, making them less effective for accurate fine-grained user-level prediction. In addition, due to the dynamic and uncertain nature of energy usage caused by various factors such as extreme weather events, quantifying uncertainty for reliable prediction is also significant, but it has not been fully explored in existing work. In this paper, we propose a unified framework called TrustEnergy for accurate and reliable user-level energy usage prediction. There are two key technical components in TrustEnergy, (i) a Hierarchical Spatiotemporal Representation module to efficiently capture both macro and micro energy usage patterns with a novel memory-augmented spatiotemporal graph neural network, and (ii) an innovative Sequential Conformalized Quantile Regression module to dynamically adjust uncertainty bounds to ensure valid prediction intervals over time, without making strong assumptions about the underlying data distribution. We implement and evaluate our TrustEnergy framework by working with an electricity provider in Florida, and the results show our TrustEnergy can achieve a 5.4% increase in prediction accuracy and 5.7% improvement in uncertainty quantification compared to state-of-the-art baselines.

Code — <https://github.com/UFOdestiny/TrustEnergy>

Introduction

Energy usage prediction has attracted significant interest from both industry and academia due to its pivotal role in supporting a wide range of practical applications for societal impacts. For example, accurate energy usage prediction is essential for efficient grid management (Jha et al. 2021), outage prevention, and energy distribution optimization. It also contributes to emergency preparedness (Zhong and Sun 2010) by helping stakeholders respond to potential surges

in electricity usage during extreme weather events or crises, ensuring the stability and resilience of the power grid.

Although many approaches have been proposed for energy usage prediction by leveraging sophisticated deep learning architectures (Luo et al. 2024; Xu et al. 2024; Balachander et al. 2024; Buratto et al. 2024; Aguilar Madrid and Antonio 2021; Khan et al. 2022; Abumohsen, Owda, and Owda 2023; Yazici, Beyca, and Delen 2022; Bashir et al. 2022), most of them primarily focus on time series prediction while overlooking the intricate spatial correlations between different households, which can provide useful information for accurate prediction, e.g., households in close proximity may share similar usage patterns. Also, most existing works (Yu, Yin, and Zhu 2018; Tu et al. 2024) predict energy usage at a coarse-grained level (e.g., in a city or region level) and fail to capture fine-grained individual household usage patterns for user-level prediction, which potentially reduces their applicability. Furthermore, quantifying prediction uncertainty is also important, as it provides prediction intervals and confidence levels, enabling stakeholders to mitigate the risks associated with over- or under-predictions. Although there are some methods for spatiotemporal uncertainty quantification (Zhuang et al. 2022; Zhou et al. 2021; Wen et al. 2023; Yu et al. 2025), they fail to capture intrinsic time-varying uncertainty and potential distribution shifts to ensure valid prediction intervals over time.

In this project, by collaborating with a utility provider to improve its energy supply services, we are fortunate to have access to a long-term fine-grained household-level electricity usage dataset, which provides us with a good opportunity to explore user-level energy usage prediction. However, we found that there are two practical challenges. (i) Modeling energy usage at the individual user level is computationally costly when there are hundreds of thousands of users. For example, there are over 60,000 users in our dataset, each exhibiting unique temporal usage patterns and spatial correlations with others. Modeling spatiotemporal usage of each user explicitly would not only result in an intractable computational burden but also pose serious challenges in terms of data sparsity and overfitting. (ii) Spatiotemporal usage patterns are intricately dynamic and time-varying, which are impacted by different factors like seasonal trends, weather conditions, and personal habits. This dynamics potentially

*Prof. Guang Wang is the corresponding author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

causes data distribution shifts and high prediction uncertainty. Effectively modeling such uncertainty remains non-trivial but is essential for delivering reliable predictions.

To address the above two challenges and advance existing works, we propose a unified framework named TrustEnergy for accurate and reliable energy usage prediction. TrustEnergy consists of two key novel components: (i) a Hierarchical Spatiotemporal Representation module (HSTR), which is capable of efficiently capturing both micro user-level and macro region-level patterns based on a new memory-augmented spatiotemporal graph neural network (MASTGNN), and (ii) a Sequential Conformalized Quantile Regression module (SCQR), which is a distribution-agnostic uncertainty quantification method that dynamically adjusts uncertainty bounds to provide valid prediction intervals over time for reliable prediction without making strong assumptions about the underlying data distributions. The key contributions of this paper are as follows:

- Conceptually, we frame household-level energy usage prediction as a spatiotemporal prediction problem rather than a traditional time-series prediction task. We aim to improve not only the accuracy but also the reliability of energy usage predictions for real-world societal impact by incorporating both macro-level regional patterns and micro-level user behaviors.
- Technically, we design a unified framework called TrustEnergy for accurate and reliable energy usage prediction, which includes two key novel modules: (i) a Hierarchical Spatiotemporal Representation Learning module with an innovative MASTGNN to efficiently and effectively capture comprehensive energy usage patterns from both micro level and macro level, and (ii) a distribution-agnostic uncertainty quantification module called SCQR to dynamically adjust uncertainty bounds to ensure valid prediction intervals over time.
- Empirically, we extensively evaluate our TrustEnergy by collaborating with one of the largest municipal utilities in Florida. We compare TrustEnergy with 13 state-of-the-art baselines across six metrics, and the experimental results show the superiority of TrustEnergy, e.g., it increases prediction accuracy by 5.4% and uncertainty quantification by 5.7% over the best baseline on the real-world data from our collaborator. We also verify the generalizability of our TrustEnergy using two other datasets from New York and California.

Problem Formulation

In this section, we formally formulate the energy usage prediction problem as a spatiotemporal graph learning problem since a graph can capture comprehensive dependencies between different entities (e.g., users) in the energy system.

Hierarchical Graphs Construction Formally, a graph can be defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V} = (v_1, v_2, \dots, v_N)$ is a set of $|\mathcal{V}| := N$ nodes, $\mathcal{E} = (e_1, e_2, \dots)$ is a set of edges, and \mathcal{A} is an adjacency matrix. The adjacency matrix derived from this graph can be denoted as $\mathcal{A} \in \mathbb{R}^{N \times N}$, where \mathcal{A}_{v_i, v_j} denotes the edge weight between node v_i and node v_j , which

can be denoted as:

$$\mathcal{A}_{v_i, v_j} = \begin{cases} \exp(-\frac{d_{ij}^2}{\sigma^2}), & i \neq j \text{ and } \exp(-\frac{d_{ij}^2}{\sigma^2}) > r \\ 0, & \text{otherwise.} \end{cases}, \quad (1)$$

where d_{ij} represents the distance of the centroids between node v_i and node v_j . σ^2 and r are thresholds to control the distribution and sparsity of the matrix, where a large r value can be used to accelerate the model training process.

Based on our data analysis, a user's energy usage can be predicted not only from their historical data but also by leveraging the usage patterns of nearby users or those with similar usage behaviors. Therefore, instead of constructing a single graph, we construct two different graphs to comprehensively learn hierarchical patterns from the user aspect and region aspect, which capture energy usage patterns at the micro level and macro level, respectively. We first construct a **micro-level graph** $\mathcal{G}_u = (\mathcal{V}_u, \mathcal{E}_u, \mathcal{A}_u)$ as defined before to model the correlation between all users, where each user is a node, and the edge is the geographical distance between households. Similarly, a **macro-level graph** is defined as $\mathcal{G}_r = (\mathcal{V}_r, \mathcal{E}_r, \mathcal{A}_r)$, where each node denotes a geospatial region, and each edge means the geographical distance between two regions.

Energy Usage Prediction

Given the above-constructed hierarchical graphs, we use \mathcal{X}^u and \mathcal{X}^r to denote the micro-level and macro-level graph signals, respectively. Suppose we use $x_t^i \in \mathbb{R}$ to denote the value of node i at time t and $X_t = (x_t^1, x_t^2, \dots, x_t^N) \in \mathbb{R}^N$ denotes the values of all the N nodes at time t . Then, historical data of τ time steps can be defined $\mathcal{X}_{1:\tau}^u = (X_1^u, X_2^u, \dots, X_\tau^u) \in \mathbb{R}^{N^u \times \tau}$ for micro-level graph, and $\mathcal{X}_{1:\tau}^r = (X_1^r, X_2^r, \dots, X_\tau^r) \in \mathbb{R}^{N^r \times \tau}$ for macro-level graph. Conventional energy usage prediction aims to learn a mapping function f from historical data to predict values at the following T time steps, which can be expressed as:

$$[\mathcal{X}_{1:\tau}^u, \mathcal{X}_{1:\tau}^r] \xrightarrow{f} \hat{\mathcal{X}}_{\tau+1:\tau+T}^u, \quad (2)$$

where $\hat{\mathcal{X}}_{\tau+1:\tau+T}^u \in \mathbb{R}^{N^u \times T}$.

Instead of solely considering prediction accuracy by outputting a single expected value, we also aim to enhance prediction reliability by providing a prediction interval distribution. A common approach to achieve this is the quantile regression, which outputs different quantiles (e.g., 90th percentile) of the predicted variable under the miscoverage rate of α , where we define $q_{\alpha_{lo}}$, $q_{\alpha_{up}}$, and $q_{\alpha_{mi}}$ as lower quantile, upper quantile, and median, respectively. The uncertainty-aware (i.e., probabilistic) prediction can be expressed as:

$$[\mathcal{X}_{1:\tau}^u, \mathcal{X}_{1:\tau}^r] \xrightarrow{\mathcal{F}} [\hat{L}_{\tau+1:\tau+T}, \hat{U}_{\tau+1:\tau+T}, \hat{M}_{\tau+1:\tau+T}], \quad (3)$$

where $\hat{L}_{\tau+1:\tau+T} = \hat{q}_{\alpha_{lo}}(\mathcal{X}_{1:\tau}^u) \in \mathbb{R}^{N^u \times T}$ is the lower quantile prediction, $\hat{U}_{\tau+1:\tau+T} = \hat{q}_{\alpha_{up}}(\mathcal{X}_{1:\tau}^u) \in \mathbb{R}^{N^u \times T}$ is the upper quantile prediction, and $\hat{M}_{\tau+1:\tau+T} = \hat{q}_{\alpha_{mi}}(\mathcal{X}_{1:\tau}^u) \in \mathbb{R}^{N^u \times T}$ is the median prediction, equivalent to $\hat{\mathcal{X}}_{\tau+1:\tau+T}^u$ in Eq. 2.

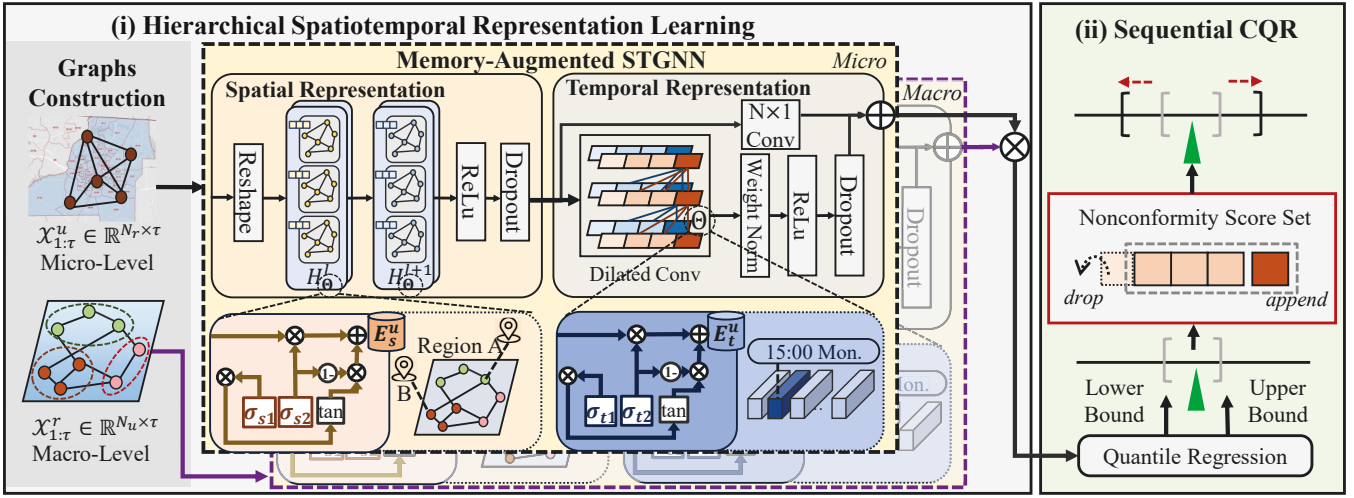


Figure 1: An overall framework of TrustEnergy, which consists of (i) a Hierarchical Spatiotemporal Representation Learning module to efficiently capture both macro-level and micro-level patterns, and (ii) a Sequential Conformalized Quantile Regression module to dynamically adjust uncertainty bounds to ensure valid prediction intervals over time.

Methodology

In this part, we introduce the detailed design of our TrustEnergy. The overview framework is shown in Figure 1.

Hierarchical Spatiotemporal Representation

In this work, we focus on user-level energy usage prediction, which is computationally expensive for existing spatiotemporal graph learning frameworks (Tu et al. 2024; Qian et al. 2023; Wen et al. 2023; Wu et al. 2019; Dong et al. 2024) to model fine-grained spatial and temporal patterns due to large-scale users and huge timesteps. For example, in our problem, there are $T = 17,520$ 30-minute time slots in a year and over $N = 60,000$ households, so the parameter space is at a billion scale. Directly optimizing parameters in such a large parameter space is computationally expensive. To address this issue, inspired by HimNet (Dong et al. 2024), we propose a memory-augmented spatiotemporal graph neural network (MASTGNN) to efficiently and effectively capture hierarchical spatiotemporal patterns.

MASTGNN includes a memory-augmented mechanism based on a well-designed shared parameter pool to update internal states, which significantly reduces model computational complexity. Specifically, the memory denotes a shared parameter pool that stores a fixed number of trainable parameter blocks. At each update step, task-specific parameters are dynamically constructed by retrieving and aggregating relevant blocks from this pool. This memory-augmented mechanism can be adapted to any STGNN framework. Without loss of generality, we employ an effective backbone combining Diffusion Graph Convolutional Networks (Li et al. 2017) and Temporal Convolutional Networks (Zhuang et al. 2022) to model the spatial and temporal patterns, and our memory-augmented mechanism will be utilized to efficiently update model parameters.

Parameter Pool Construction We construct parameter pools from both spatial and temporal dimensions to model user behaviors. We build a spatial embedding matrix $E_s^u \in \mathbb{R}^{N^u \times d_s}$, where N^u denotes the total number of users in the dataset, d_s is the dimension of the embedding vector assigned to each location. Similarly, the spatial embedding for the macro-level graph is $E_s^r \in \mathbb{R}^{N^r \times d_s}$, where N^r denotes the total number of regions. From the temporal dimension, we found a significant energy usage behavior shift for different hours, days, and months. Therefore, we construct a time-of-day embedding $D_{\text{tod}} \in \mathbb{R}^{N_{\text{tod}} \times d_{\text{tod}}}$, a day-of-week embedding $D_{\text{dow}} \in \mathbb{R}^{N_{\text{dow}} \times d_{\text{dow}}}$, and a month-of-year embedding $D_{\text{moy}} \in \mathbb{R}^{N_{\text{moy}} \times d_{\text{moy}}}$, where d_{tod} , d_{dow} and d_{moy} denotes the number of time steps per day, the number of days in a week and the number of months in a year, respectively. Typically, d_{tod} is dependent on the original time interval, $d_{\text{dow}} = 7$, and $d_{\text{moy}} = 12$. The last sample's time information (tod, dow, moy) of each batch is used as the index to extract the corresponding embedding $E_{\text{tod}} \in \mathbb{R}^{B \times d_{\text{tod}}}$, $E_{\text{dow}} \in \mathbb{R}^{B \times d_{\text{dow}}}$, and $E_{\text{moy}} \in \mathbb{R}^{B \times d_{\text{moy}}}$. The three embeddings are concatenated to obtain the overall temporal embedding matrix $E_t^u = E_t^r = E_{\text{tod}} \parallel E_{\text{dow}} \parallel E_{\text{moy}} \in \mathbb{R}^{B \times d_t}$, where B is the batch size and $d_t = d_{\text{tod}} + d_{\text{dow}} + d_{\text{moy}}$.

For the spatial dimension, we define a parameter pool $P_s \in \mathbb{R}^{d_s \times M}$ and generate spatial parameters using the spatial embedding $E_s \in \mathbb{R}^{N \times d_s}$:

$$\theta_s = E_s \cdot P_s, \quad (4)$$

where $\theta_s \in \mathbb{R}^{N \times M}$. Similarly, the temporal parameter pool is defined as $P_t \in \mathbb{R}^{d_t \times M}$, containing d_t candidate vectors. Given the temporal embedding $E_t \in \mathbb{R}^{B \times d_t}$ (from B samples), the temporal parameters are generated by:

$$\theta_t = E_t \cdot P_t, \quad (5)$$

where $\theta_t \in \mathbb{R}^{B \times M}$.

Furthermore, to efficiently generate parameters for large-scale user sets, we partition the full parameter sets P_s and

P_t into multiple smaller blocks. Take P_s as an example:

$$P_s = [P_{s_1}, P_{s_2}, \dots, P_{s_I}], \quad P_{s_i} \in \mathbb{R}^{d_s \times m_{s_i}}, \quad \sum_{i=1}^I m_{s_i} = M. \quad (6)$$

Each block P_{s_i} is represented by a centroid vector $\mu_i \in \mathbb{R}^k$, which summarizes the block's parameter distribution. Given a query E_s , we first perform a coarse selection step by measuring the similarity between E_s and each centroid μ_i . The index of the most similar block is obtained by:

$$i = \arg \max_i \text{sim}(E_s, \mu_i), \quad (7)$$

where $\text{sim}(\cdot, \cdot)$ denotes a similarity metric such as cosine similarity or dot product. Finally, the parameter θ_s is generated by multiplying the query vector E_t with the selected block P_{s_i} :

$$\theta_s = E_s \cdot P_{s_i}. \quad (8)$$

By reusing shared parameter pools and querying them adaptively via embeddings, TrustEnergy effectively reduces the computational complexity from $\mathcal{O}(TN)$ to $\mathcal{O}(k)$, where $k = B + N$ is the size of the parameter pool. This enables efficient and adaptive learning across highly heterogeneous user-level energy usage patterns.

Parameter Updating Mechanism Our method builds upon a memory-augmented recurrent unit as the fundamental component, where both spatial and temporal dynamics are captured via structured gating and message passing. The fundamental building block is formulated by:

$$\begin{aligned} r_t &= \sigma(\Theta_r \star_G [X_t, H_{t-1}] + b_r) \\ u_t &= \sigma(\Theta_u \star_G [X_t, H_{t-1}] + b_u) \\ c_t &= \tanh(\Theta_c \star_G [X_t, (r_t \odot H_{t-1})] + b_c) \\ H_t &= u_t \odot H_{t-1} + (1 - u_t) \odot c_t \end{aligned} \quad (9)$$

where $X_t \in \mathbb{R}^N$ and $H_t \in \mathbb{R}^{N \times h}$ denote the input and output at timestep t . h is the hidden layer size. r_t and u_t are the reset and update gates. $\Theta_r, \Theta_u, \Theta_c$ are parameters for the corresponding filters in the graph convolution operation \star_G , which is defined as:

$$Z = U \star_G = \sum_{k=0}^K \tilde{A}^k U W_k, \quad (10)$$

where $U \in \mathbb{R}^{N \times C}$ is the input, $Z \in \mathbb{R}^{N \times h}$ is the output, and C denotes the input channels. \tilde{A} represents the topology of graph \mathcal{G} , and $W \in \mathbb{R}^{K \times C \times h}$ is the kernel parameter Θ . We use temporal embeddings $E_t \in \mathbb{R}^{B \times d_t}$ and spatial embeddings $E_s \in \mathbb{R}^{N \times d_s}$ to dynamically generate context-dependent meta-parameters via low-rank linear attention:

$$W_t = E_t \cdot P_t, \quad W_s = E_s \cdot P_s, \quad (11)$$

where $P_t \in \mathbb{R}^{d_t \times K \times C \times h}$ and $P_s \in \mathbb{R}^{d_s \times K \times C \times h}$ are meta-parameter pools for the temporal and spatial dimensions, respectively. $W_t \in \mathbb{R}^{B \times K \times C \times h}$ and $W_s \in \mathbb{R}^{N \times K \times C \times h}$ are the corresponding meta-parameters. Given input sequences $X \in \mathbb{R}^{B \times T \times N}$, the model computes hidden states $H \in \mathbb{R}^{B \times T \times N \times h}$ using the generated W_t and W_s . These hidden states are directly passed to a lightweight prediction head (e.g., linear layer or MLP) to obtain the final prediction output.

Loss Function Since our target is to output the lower quantile, upper quantile, and median of the prediction, we design a hybrid loss function by combining both the pinball loss $\rho_\alpha(Y, \hat{X})$ for quantile regression and mean absolute error (MAE) loss for median prediction, which is denoted as:

$$\begin{aligned} \mathcal{L}(Y, \hat{X}) &= \rho_{\alpha_{\text{lo}}}(Y, \hat{X}) + \rho_{\alpha_{\text{up}}}(Y, \hat{X}) \\ &\quad + \text{MAE}(Y, \hat{q}_{\alpha_{\text{mi}}}(X)), \end{aligned} \quad (12)$$

where $\rho_\alpha(Y, \hat{X})$ is calculated:

$$\rho_\alpha(Y, \hat{X}) = \begin{cases} \alpha(Y - \hat{X}) & \text{if } Y - \hat{X} > 0 \\ (1 - \alpha)(Y - \hat{X}) & \text{otherwise} \end{cases}. \quad (13)$$

Sequential Conformalized Quantile Regression

Prediction intervals from traditional quantile regression may be poorly calibrated in energy usage prediction scenarios due to distribution shifts caused by factors such as seasonal changes, weather fluctuations, and behavioral changes, which can lead to miscoverage and the failure to provide reliable uncertainty estimates. To address this challenge, we propose Sequential Conformalized Quantile Regression (SCQR), a distribution-free uncertainty quantification method that ensures valid prediction intervals with guaranteed coverage. SCQR dynamically adjusts uncertainty bounds over time, enhancing quantile regression and maintaining a valid coverage rate without strong assumptions about the underlying data distribution.

Given a dataset from previous n timesteps $\{(X_i, Y_i)\}_{i=1}^n$, we first split it into a training set \mathcal{D}_{tr} , a calibration set \mathcal{D}_{ca} , and a test set \mathcal{D}_{te} . We then obtain the nonconformity scores set $\mathcal{E} = \{\hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_t\}$, where $t = |\mathcal{D}_{\text{ca}}|$ and $\hat{\epsilon}_i$ is computed by nonconformity measure S :

$$\hat{\epsilon}_i = S\{(X_i, Y_i)\} = \max\{\hat{q}_{\alpha_{\text{lo}}}(X_i) - Y_i, Y_i - \hat{q}_{\alpha_{\text{up}}}(X_i)\}, \quad (14)$$

where $i \in \mathcal{D}_{\text{ca}}$. To conformalize the prediction interval, we compute $Q_{1-\alpha}(\mathcal{E}, \mathcal{D}_{\text{ca}})$ on the set \mathcal{E} :

$$Q_{1-\alpha}(\mathcal{E}, \mathcal{D}_{\text{ca}}) = \text{quantile}((1 - \alpha)(1 + 1/|\mathcal{D}_{\text{ca}}|)) \text{ of set } \mathcal{E}. \quad (15)$$

Finally, given new input data X_{new} , we set low = $\hat{q}_{\alpha_{\text{lo}}}(X_{\text{new}})$ and high = $\hat{q}_{\alpha_{\text{up}}}(X_{\text{new}})$. We then construct the prediction interval for Y_{new} as:

$$\hat{C}(X_{\text{new}}) = [\text{low} - Q_{1-\alpha}(\mathcal{E}, \mathcal{D}_{\text{ca}}), \text{high} + Q_{1-\alpha}(\mathcal{E}, \mathcal{D}_{\text{ca}})]. \quad (16)$$

To comprehensively capture the underlying temporal dependencies inherent in sequential data, SCQR explicitly incorporates the historical behavior of residuals (i.e., nonconformity scores) when constructing valid prediction intervals. This approach is different from traditional conformal prediction paradigms that assume exchangeability and independence among samples by sequentially adapting the conformal score set to reflect the temporal structure of the data stream. Formally, consider a test-time scenario where a sequence of test samples $\{(X_{t+1}, Y_{t+1}), (X_{t+2}, Y_{t+2}), \dots\}$ arrives in a streaming or temporally ordered fashion. Let $\mathcal{E} = \{\hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_t\}$ denote the current set of nonconformity scores derived from a held-out calibration set \mathcal{D}_{ca} . When the first test sample (X_{t+1}, Y_{t+1}) is observed, SCQR computes a new nonconformity score $\hat{\epsilon}_{t+1} = S(X_{t+1}, Y_{t+1})$ using a predefined scoring function $S(\cdot)$, typically based on

Algorithm 1: Sequential Conformalized Quantile Regression

Input: Training set \mathcal{I}_{tr} , calibration set \mathcal{I}_{ca} , test set \mathcal{I}_{te} ; a quantile regressor \mathcal{F} ; target error rate α ; nonconformity measure S .

Output: Prediction intervals $\hat{C}(X_t)$ for all $t \in \mathcal{I}_{te}$.

- 1: Train the quantile regressor to obtain $\{\hat{q}_{\alpha_{lo}}, \hat{q}_{\alpha_{up}}\} \leftarrow \mathcal{F}(\{(X_i, Y_i)\}_{i \in \mathcal{I}_{tr}})$.
 - 2: Compute nonconformity scores $\mathcal{E} = \{\hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_t\}$, where $t = |\mathcal{I}_{ca}|$ and $\hat{\epsilon}_i = S((X_i, Y_i))$ for $i \in \mathcal{I}_{ca}$, as defined in Eq. 14.
 - 3: **for** each time step $t \in \mathcal{I}_{te}$ **do**
 - 4: Compute the quantile $Q_{1-\alpha}(\mathcal{E}, \mathcal{D}_{ca})$, as in Eq. 15.
 - 5: Construct the prediction interval $\hat{C}(X_t)$, as in Eq. 16.
 - 6: Obtain new nonconformity score $\hat{\epsilon}_t \leftarrow S((X_t, Y_t))$.
 - 7: Update the new set \mathcal{E} by removing the oldest element and adding the newest $\hat{\epsilon}_t$.
 - 8: **end for**
-

the deviation between the ground truth Y_{t+1} and the model’s predictive quantile estimate at X_{t+1} .

To maintain a rolling calibration window and preserve the sequential dependency, the earliest score $\hat{\epsilon}_1$ in \mathcal{E} is discarded, and the newly computed score $\hat{\epsilon}_{t+1}$ is appended to the set, yielding an updated nonconformity set: $\mathcal{E} = \{\hat{\epsilon}_2, \hat{\epsilon}_3, \dots, \hat{\epsilon}_{t+1}\}$. This dynamic update mechanism ensures that the conformal quantile $Q_{1-\alpha}(\mathcal{E}, \mathcal{D}_{ca})$, which determines the prediction interval width for a given coverage level $1 - \alpha$, is recalibrated in real time to accommodate evolving patterns in the residual distribution. Subsequently, the updated quantile estimate is utilized to compute the prediction interval $\hat{C}(X_{t+2})$ for the next incoming test sample (X_{t+2}, Y_{t+2}) , maintaining validity under the sequential regime. This recursive procedure continues iteratively, such that the conformal set \mathcal{E} is refreshed at each time step to reflect the most recent residual dynamics. The process is illustrated in Algorithm 1, ensuring that each prediction interval is adaptively informed by the most temporally relevant uncertainty information, thereby enhancing calibration fidelity in non-stationary or temporally correlated settings.

Evaluation

In this project, we collaborate with an electricity provider in Florida to evaluate our TrustEnergy framework. Specifically, we try to answer the following questions:

- **RQ 1:** How does TrustEnergy perform compared to state-of-the-art models across different metrics?
- **RQ 2:** How does TrustEnergy perform under abnormal scenarios such as extreme weather events?
- **RQ 3:** What is the individual contribution of each module to TrustEnergy’s performance?
- **RQ 4:** Can TrustEnergy generalize to other regions?

Experimental Setup

Datasets We evaluate the performance of our TrustEnergy using fine-grained 30-minute interval electricity consumption data for 61,672 households across 253 regions in

Models	Metrics					
	MAE	RMSE	MAPE	MPIW	WINK	Cov
STGCN	49.72	59.65	1.37%	99.75	132.17	✗
GWNET	63.52	70.36	1.58%	115.37	138.62	✗
ASTGCN	56.25	64.17	1.39%	105.28	130.49	✗
AGCRN	<u>48.35</u>	<u>58.31</u>	1.40%	96.23	120.85	✓
StemGNN	51.66	63.28	1.32%	94.05	113.43	✓
DSTAGNN	58.71	68.79	1.47%	107.67	125.81	✗
STZINB	55.31	68.43	1.28%	96.70	120.63	✗
PDFormer	53.71	66.07	1.42%	97.31	117.20	✗
DiffSTG	53.28	67.91	1.30%	95.83	125.70	✗
DeepSTUQ	55.75	68.05	1.32%	94.25	122.10	✓
PowerPM	50.25	60.11	<u>1.29%</u>	99.52	135.41	✗
Chronos	48.50	59.02	1.35%	<u>90.94</u>	<u>108.57</u>	✓
Moment	53.20	64.75	1.38%	95.42	118.94	✗
TrustEnergy	45.72	55.65	1.28%	85.72	102.31	✓

Table 1: Comparison with 13 state-of-the-art baselines using Florida data on six metrics. The best results are presented in bold, and the second-best results are underlined. ✓ means that the method reaches the target coverage (i.e., coverage $\geq 90\%$) while ✗ means that it fails to achieve it.

Florida (FL) in 2018. In addition, we also use two other real-world datasets from New York (NY) (NYISO New York State) and California (CA) (CAISO California State) to verify the generalizability of our design.

Baselines We compare TrustEnergy with 13 state-of-the-art baselines, including Spatial-Temporal Graph Convolution Network (STGCN) (Yu, Yin, and Zhu 2018), Graph WaveNet (GWNET) (Wu et al. 2019), (ASTGCN) (Guo et al. 2019), Adaptive Graph Convolutional Recurrent Network (AGCRN) (Bai et al. 2020), Spectral Temporal GNN (StemGNN) (Cao et al. 2020), Dynamic Spatial-Temporal Aware Graph Neural Network (DSTAGNN) (Lan et al. 2022), STZINB (Zhuang et al. 2022), PDFormer (Jiang et al. 2023), DiffSTG (Wen et al. 2023), DeepSTUQ (Qian et al. 2023), PowerPM (Tu et al. 2024), Chronos (Ansari, Stella, and Turkmen 2024), and Moment (Goswami et al. 2024).

Evaluation Metrics We utilize three metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), to evaluate the deterministic performance, and three other metrics, including Mean Prediction Interval Width (MPIW) (Zhuang et al. 2022), Winkler score (WINK) (Han 2023), and coverage (COV) (Zhuang et al. 2024), to evaluate the performance of uncertainty quantification.

Overall Performance Comparison (RQ 1)

An overall comparison of our TrustEnergy and other baseline models is presented in Table 1. We found that our TrustEnergy consistently achieves the best performance across all metrics for both prediction accuracy and reliability. Specifically, our framework reduces MAE by approximately 5.4% compared to the best baseline AGCRN. In addition, TrustEnergy also demonstrates superior performance on prediction reliability, with a 5.7% improvement in WINK

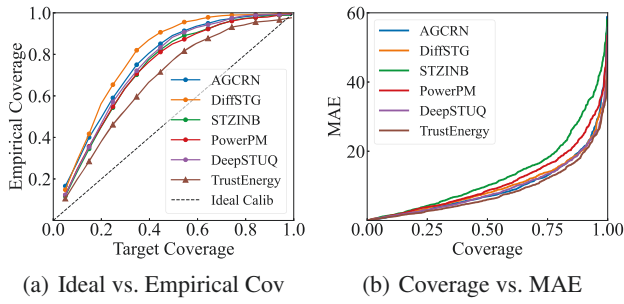


Figure 2: Uncertainty calibration results.

and reaching the target coverage. Figure 2(a) compares nominal coverage levels with empirical coverage, which shows TrustEnergy achieves the most reliable calibration as its curve is closest to the diagonal line. Figure 2(b) also demonstrates the reliability by comparing the coverage and MAE. It is observed that TrustEnergy achieves a higher coverage with the same MAE level, indicating its better performance for uncertainty quantification.

Case Studies for Abnormal Situations (RQ 2)

One of the most challenging aspects of energy usage prediction lies in handling abnormal or extreme scenarios, such as hurricanes and heat waves, that often lead to sudden and significant deviations in energy usage. Accurate and reliable energy usage prediction under these extreme weather events is critical for ensuring grid stability, optimizing energy resource allocation, and enabling timely decision-making for both utilities and emergency response teams.

We utilize three representative extreme weather events: a major hurricane, a tornado, and an extreme heat wave in Florida, to show the effectiveness of our method. As shown in Figure 3, TrustEnergy maintained accurate predictions during the hurricane, with ground truth values mostly falling within the prediction intervals, while AGCRN failed to capture abrupt energy usage changes. Similarly, under the heat wave scenario, where energy demand surged, TrustEnergy outperformed DeepSTUQ with tighter intervals. These results demonstrate TrustEnergy’s superior adaptability and robustness under extreme conditions.

Ablation Study (RQ 3)

We conduct an ablation study to evaluate the contribution of each key component in TrustEnergy. Specifically, we compare TrustEnergy with its following three variants:

- **w/o HSTR**: Replace hierarchical spatiotemporal representation module with only macro-level embeddings.
- **w/o MASTGNN**: Remove the memory-augmented parameter update mechanism and use the conventional spatiotemporal learning module with DGCN and TCN.
- **w/o SCQR**: Replace SCQR with traditional CQR, eliminating sequential dependency modeling.

From Table 2, we found that the model performance without the HSTR module significantly decreases (e.g.,

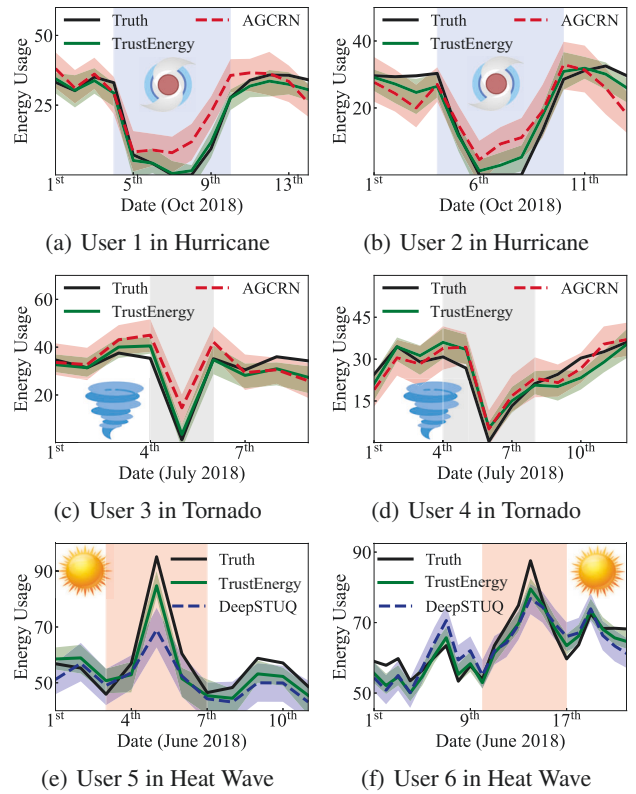


Figure 3: Prediction results under diverse extreme weather conditions.

Methods	Metrics					
	MAE	RMSE	MAPE	MPIW	WINK	COV
w/o HSTR	49.35	47.35	1.48%	91.27	114.78	✗
w/o MASTGNN	47.54	57.88	1.33%	87.26	108.51	✗
w/o SCQR	-	-	-	87.05	104.70	✗
TrustEnergy	45.72	55.65	1.28%	85.72	102.31	✓

Table 2: Ablation study of key design components.

more than 7.3% reduction in MAE on the Florida dataset), which indicates the importance of capturing comprehensive micro-level and macro-level patterns. Similarly, by removing the memory-augmented parameter update mechanism, we found that the performance also decreased in all metrics (e.g., 3.8% reduction in MAE). In addition, we found that the uncertainty quantification performance decreased without the SCQR module since the prediction results cannot reach the target coverage.

Generalizability to Other Areas (RQ 4)

To verify the generalizability of our TrustEnergy to different geographical areas, we also evaluate it on the NY and CA datasets. As shown in Table 3, TrustEnergy consistently outperforms state-of-the-art baselines on most metrics, demonstrating its effectiveness and generalizability to broader areas. Also, the ablation study in Table 4 validates the utility of each component in our TrustEnergy since the performance

Dataset	Models	Metrics					
		MAE	RMSE	MAPE	MPIW	WINK	COV
NY	STGCN	0.492	49.58	0.99%	3.328	3.331	✗
	GWNET	0.199	8.364	0.60%	0.934	1.234	✗
	ASTGCN	0.149	3.311	0.58%	0.558	0.851	✗
	AGCRN	0.147	3.361	0.57%	0.522	0.746	✗
	StemGNN	0.275	7.828	0.94%	0.888	1.437	✓
	DSTAGNN	0.181	4.738	0.74%	0.774	1.045	✗
	STZINB	0.149	3.825	0.54%	0.588	0.781	✓
	PDFormer	0.243	8.352	0.86%	0.760	1.378	✗
	DiffSTG	0.281	4.863	0.94%	1.108	1.751	✓
	DeepSTUQ	0.310	5.132	0.98%	1.205	1.868	✗
	PowerPM	0.298	4.953	0.98%	0.854	1.361	✗
	Chronos	0.158	4.365	0.59%	0.574	1.353	✓
	Moment	0.192	5.174	0.63%	0.623	1.268	✗
TrustEnergy	0.137	3.321	0.57%	0.540	0.727	✓	
CA	STGCN	0.607	33.27	1.36%	3.500	3.776	✓
	GWNET	0.353	5.565	1.15%	1.243	2.155	✗
	ASTGCN	0.344	3.142	1.28%	0.973	2.353	✗
	AGCRN	0.328	3.256	1.24%	0.854	2.303	✗
	StemGNN	0.416	5.090	1.45%	1.575	2.475	✓
	DSTAGNN	0.327	4.014	1.23%	0.848	2.345	✗
	STZINB	0.350	3.606	1.41%	0.905	2.471	✗
	PDFormer	0.381	4.128	1.39%	1.274	2.484	✗
	DiffSTG	0.358	4.034	1.29%	1.101	2.254	✓
	DeepSTUQ	0.384	4.152	1.36%	1.268	2.225	✗
	PowerPM	0.337	4.128	1.28%	0.957	2.261	✗
	Chronos	0.348	3.816	1.28%	0.944	2.235	✓
	Moment	0.352	4.013	1.29%	1.035	2.274	✓
TrustEnergy	0.327	3.012	1.14%	0.896	2.325	✓	

Table 3: Comparison with 13 state-of-the-art baselines on two other datasets from New York and California.

of the proposed framework will decrease without any key module.

Related Work

Energy Usage Prediction

As an important real-world problem, energy usage prediction has attracted much interest from both academia and industry. Traditional machine learning methods have long been employed, such as linear regression (Hong, Wang, and Willis 2011), support vector regression (Sapankevych and Sankar 2009), decision tree-based models (Lloyd 2014), and random forest regression (Wu et al. 2015). However, these

Dataset	Methods	Metrics					
		MAE	RMSE	MAPE	MPIW	WINK	COV
NY	w/o HSTR	0.431	7.583	1.26%	1.459	1.952	✗
	w/o MASTGNN	0.358	5.752	1.15%	1.357	1.864	✗
	w/o SCQR	-	-	-	0.875	1.054	✓
	TrustEnergy	0.137	3.321	0.57%	0.540	0.727	✓
CA	w/o HSTR	0.471	7.713	1.43%	1.368	2.675	✗
	w/o MASTGNN	0.405	5.257	1.39%	1.257	2.568	✗
	w/o SCQR	-	-	-	1.125	2.463	✓
	TrustEnergy	0.327	3.012	1.14%	0.896	2.325	✓

Table 4: Ablation study on NY and CA datasets.

methods often struggle to capture the complex, nonlinear relationships inherent in energy usage data. In recent years, more advanced deep learning methods (Luo et al. 2024; Balachander et al. 2024; Buratto et al. 2024) have been leveraged to predict energy usage by leveraging sophisticated deep learning architectures, but most of them focus on aggregated time series prediction while ignoring rich user-level and spatial information, which undermines their capabilities for making accurate fine-grained predictions. In addition, energy usage prediction reliability will directly impact real-world decision-making, such as resource allocation or restoration under extreme weather conditions. However, this topic has not been fully explored by existing work.

Spatiotemporal Uncertainty Quantification

Recent efforts on spatiotemporal uncertainty quantification provide a good opportunity to quantify energy usage prediction reliability. Wen *et al.* (Wen et al. 2023) extend denoising diffusion probabilistic models to spatiotemporal graphs via DiffSTG for capturing intrinsic uncertainties. Kexin *et al.* (Huang et al. 2024) adapt conformal prediction to graph-based models to provide valid uncertainty guarantees. Zhuang *et al.* (Zhuang et al. 2022) propose STZINB, a zero-inflated negative binomial GNN designed for modeling uncertainty in sparse urban data. Qian *et al.* (Qian et al. 2023) present DeepSTUQ, which leverages dual neural sub-networks to estimate uncertainty in traffic prediction tasks.

Nevertheless, most existing methods (Xu et al. 2023; Faustine and Pereira 2022) for spatiotemporal uncertainty quantification either impose restrictive assumptions on spatial and temporal dependencies or fail to jointly model the interplay between them, which may cause the prediction intervals to be poorly calibrated, and they cannot dynamically adjust uncertainty bounds over time to adapt to the distribution shifts caused by factors like extreme weather events.

Conclusion

In this work, we design a unified framework called TrustEnergy to improve both the accuracy and reliability of user-level energy usage prediction. There are two key novel designs in TrustEnergy, i.e., (i) a hierarchical spatiotemporal representation module to efficiently capture both macro and micro patterns with a new MASTGNN, and (ii) an innovative distribution-agnostic uncertainty quantification method called SCQR to dynamically adjust uncertainty bounds to ensure valid prediction intervals over time. By collaborating with an electricity provider in Florida, we extensively evaluate TrustEnergy by comparing it against 13 state-of-the-art baselines across six metrics. The results demonstrate TrustEnergy effectively outperforms baselines and improves prediction accuracy by around 5.4% and prediction reliability by around 5.7%. Notably, TrustEnergy proves to be particularly effective under extreme weather conditions such as hurricanes, highlighting its potential for real-world societal impact. We also validate its generalizability using real-world datasets from New York and California, which also indicate the effectiveness of TrustEnergy.

Acknowledgments

We sincerely thank all anonymous reviewers for their insightful comments and valuable suggestions. This work is partially supported by the FSU Sustainability & Climate Solutions Grant Program, FSU/AWS Computer Support Seed Fund, and FSU Startup Fund. Shenhao Wang acknowledges the support from UFL's 2023 Research Opportunity Seed Fund (ROSF2023).

References

- Abumohsen, M.; Owda, A. Y.; and Owda, M. 2023. Electrical load forecasting using LSTM, GRU, and RNN algorithms. *Energies*, 16(5): 2283.
- Aguilar Madrid, E.; and Antonio, N. 2021. Short-term electricity load forecasting with machine learning. *Information*, 12(2): 50.
- Ansari, A. F.; Stella, L.; and Turkmen. 2024. Chronos: Learning the Language of Time Series. *Transactions on Machine Learning Research*.
- Bai, L.; Yao, L.; Li, C.; Wang, X.; and Wang, C. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems*, 33: 17804–17815.
- Balachander, T.; Khot, S. A.; Huseyn, R.; Garg, S.; Vijay, S.; and Pandey, V. 2024. An Innovative Method for Short Term Electrical Load Forecasting Based on Adaptive CNN-MRMR Model. In *2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC)*, 1–6. IEEE.
- Bashir, T.; Haoyong, C.; Tahir, M. F.; and Liqiang, Z. 2022. Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. *Energy reports*, 8: 1678–1686.
- Buratto, W. G.; Muniz, R. N.; Nied, A.; and Gonzalez, G. V. 2024. Seq2Seq-LSTM with attention for electricity load forecasting in Brazil. *IEEE Access*.
- CAISO. California State.
- Cao, D.; Wang, Y.; Duan, J.; Zhang, C.; Zhu, X.; Huang, C.; Tong, Y.; Xu, B.; Bai, J.; Tong, J.; et al. 2020. Spectral temporal graph neural network for multivariate time-series forecasting. *Advances in neural information processing systems*, 33: 17766–17778.
- Dong, Z.; Jiang, R.; Gao, H.; Liu, H.; Deng, J.; Wen, Q.; and Song, X. 2024. Heterogeneity-informed meta-parameter learning for spatiotemporal time series forecasting. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, 631–641.
- Faustine, A.; and Pereira, L. 2022. FPSeq2Q: Fully parameterized sequence to quantile regression for net-load forecasting with uncertainty estimates. *IEEE Transactions on Smart Grid*, 13(3): 2440–2451.
- Goswami, M.; Szafer, K.; Choudhry, A.; Cai, Y.; Li, S.; and Dubrawski, A. 2024. MOMENT: A Family of Open Time-series Foundation Models. In *International Conference on Machine Learning*.
- Guo, S.; Lin, Y.; Feng, N.; Song, C.; and Wan, H. 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 922–929.
- Han, J. 2023. Probabilistic multivariate time series forecasting and robust uncertainty quantification with applications in electricity price prediction.
- Hong, T.; Wang, P.; and Willis, H. L. 2011. A naïve multiple linear regression benchmark for short term load forecasting. In *2011 IEEE power and energy society general meeting*, 1–6. IEEE.
- Huang, K.; Jin, Y.; Candes, E.; and Leskovec, J. 2024. Uncertainty quantification over graph with conformalized graph neural networks. *Advances in Neural Information Processing Systems*, 36.
- Jha, N.; Prashar, D.; Rashid, M.; Gupta, S. K.; and Saket, R. 2021. Electricity load forecasting and feature extraction in smart grid using neural networks. *Computers & Electrical Engineering*, 96: 107479.
- Jiang, J.; Han, C.; Zhao, W. X.; and Wang, J. 2023. Pdfformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 4365–4373.
- Khan, Z. A.; Ullah, A.; Haq, I. U.; Hamdy, M.; Mauro, G. M.; Muhammad, K.; Hijji, M.; and Baik, S. W. 2022. Efficient short-term electricity load forecasting for effective energy management. *Sustainable Energy Technologies and Assessments*, 53: 102337.
- Lan, S.; Ma, Y.; Huang, W.; Wang, W.; Yang, H.; and Li, P. 2022. DSTAGNN: Dynamic Spatial-Temporal Aware Graph Neural Network for Traffic Flow Forecasting. In Chaudhuri, K.; Jegelka, S.; Song, L.; Szepesvari, C.; Niu, G.; and Sabato, S., eds., *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, 11906–11917. PMLR.
- Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2017. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*.
- Lloyd, J. R. 2014. GEFCom2012 hierarchical load forecasting: Gradient boosting machines and Gaussian processes. *International Journal of Forecasting*, 30(2): 369–374.
- Luo, S.; Wang, B.; Gao, Q.; Wang, Y.; and Pang, X. 2024. Stacking integration algorithm based on CNN-BiLSTM-Attention with XGBoost for short-term electricity load forecasting. *Energy Reports*, 12: 2676–2689.
- NYISO. New York State.
- Qian, W.; Zhang, D.; Zhao, Y.; Zheng, K.; and James, J. 2023. Uncertainty quantification for traffic forecasting: A unified approach. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, 992–1004. IEEE.
- Sapankevych, N. I.; and Sankar, R. 2009. Time series prediction using support vector machines: a survey. *IEEE computational intelligence magazine*, 4(2): 24–38.
- Tu, S.; Zhang, Y.; Zhang, J.; Fu, Z.; Zhang, Y.; and Yang, Y. 2024. PowerPM: Foundation Model for Power Systems.

Advances in Neural Information Processing Systems, 37: 115233–115260.

Wen, H.; Lin, Y.; Xia, Y.; Wan, H.; Wen, Q.; Zimmermann, R.; and Liang, Y. 2023. Diffstg: Probabilistic spatio-temporal graph forecasting with denoising diffusion models. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*, 1–12.

Wu, X.; He, J.; Zhang, P.; and Hu, J. 2015. Power system short-term load forecasting based on improved random forest with grey relation projection. *Automation of Electric Power Systems*, 39(12): 50–55.

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

Xu, C.; Sun, Y.; Du, A.; and Gao, D.-c. 2023. Quantile regression based probabilistic forecasting of renewable energy generation and building electrical load: A state of the art review. *Journal of Building Engineering*, 107772.

Xu, H.; Hu, F.; Liang, X.; Zhao, G.; and Abugunmi, M. 2024. A framework for electricity load forecasting based on attention mechanism time series depthwise separable convolutional neural network. *Energy*, 299: 131258.

Yazici, I.; Beyca, O. F.; and Delen, D. 2022. Deep-learning-based short-term electricity load forecasting: A real case application. *Engineering Applications of Artificial Intelligence*, 109: 104645.

Yu, B.; Yin, H.; and Zhu, Z. 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 3634–3640.

Yu, D.; Zhuang, D.; Jiang, L.; Xu, R.; Ye, X.; Bu, Y.; Wang, S.; and Wang, G. 2025. UQGN: Uncertainty Quantification of Graph Neural Networks for Multivariate Spatiotemporal Prediction. In *33rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2025)*.

Zhong, S.; and Sun, Z. 2010. Challenges and opportunities in emergency management of electric power system black-out. In *2010 International Conference on E-Product E-Service and E-Entertainment*, 1–4. IEEE.

Zhou, Z.; Wang, Y.; Xie, X.; Qiao, L.; and Li, Y. 2021. STU-aNet: Understanding Uncertainty in Spatiotemporal Collective Human Mobility. In *Proceedings of the Web Conference 2021, WWW '21*, 1868–1879. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383127.

Zhuang, D.; Bu, Y.; Wang, G.; Wang, S.; and Zhao, J. 2024. Sauc: Sparsity-aware uncertainty calibration for spatiotemporal prediction with graph neural networks. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, 160–172.

Zhuang, D.; Wang, S.; Koutsopoulos, H.; and Zhao, J. 2022. Uncertainty Quantification of Sparse Travel Demand Prediction with Spatial-Temporal Graph Neural Networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '22*, 4639–4647.

New York, NY, USA: Association for Computing Machinery. ISBN 9781450393850.