WattCast: Spatiotemporal Forecasting of Low-Voltage Feeder Energy Consumption

Vaibhav Sourirajan, vs2787

May 11, 2025

Abstract

Accurately forecasting energy consumption is essential for efficient grid management, demand response optimization, and the integration of renewable energy sources. In this work, we present WattCast, a spatiotemporal forecasting framework designed to predict future energy consumption at the low-voltage feeder level. WattCast leverages the Weaver Energy Dataset, which provides half-hourly aggregated energy consumption data from the United Kingdom. Our approach augments the time series data with learned substation embeddings derived from a graph representation of substations, capturing spatial relationships based on their physical proximity and consumption patterns. We compare the performance of this hybrid model—combining LSTM with graph-augmented embeddings—against an LSTM without graph augmentation and several baseline models, including SARIMA, Prophet, and other naive approaches. The hybrid model achieves a mean $R^2 = 0.72$, the highest among all models evaluated, offering a robust solution for accurate feeder-level energy forecasting and highlighting the benefits of hybrid spatiotemporal modeling.

Introduction

Reliable energy consumption forecasting is a cornerstone of modern energy management. In the electrical grid, substations function as critical nodes, stepping down high-voltage electricity from transmission lines to lower voltages suitable for local distribution. Feeders, which extend outward from substations, serve as the final major link in the distribution network, making accurate consumption forecasting crucial for maintaining local grid stability—especially as demand becomes increasingly volatile due to factors such as electric vehicle (EV) charging and the proliferation of distributed solar generation.

This growing demand volatility imposes substantial stress on low-voltage distribution networks, which have traditionally relied on static, coarse-grained models, often based on annual averages or conservative peak assumptions. Such models fail to capture the dynamic, localized fluctuations introduced by modern electrification, leading to inefficiencies and rendering them inadequate for contemporary grid management. Traditional time-series methods including ARIMA and Prophet, while able to capture temporal patterns in stationary time series, are limited when faced with the non-linear dependencies of energy forecasting. To address these limitations, learning-based approaches such as Long Short-Term Memory (LSTM) networks are promising as they can capture complex relationships in sequential data. However, a conventional LSTM method applied to distributed feeder-level energy consumption forecasting problems models each feeder independently, neglecting spatial dependencies between neighboring feeders that may exhibit correlated consumption patterns. This can result in poor peak prediction and reduced robustness under system-wide changes.

In response, we introduce WattCast, a spatiotemporal forecasting framework designed to accurately predict energy consumption at the low-voltage feeder level. Our approach combines the strengths of LSTM-based temporal modeling with graph-based spatial embeddings, effectively capturing both local consumption dynamics and cross-feeder interactions. This study makes the following contributions:

1. We establish strong baseline forecasting performance using traditional models, including ARIMA, persistence, and boosting methods, providing a foundation for comparative analysis.
2. We develop a robust forecasting framework that combines LSTM-based temporal modeling with graph-based spatial embeddings, effectively capturing both local and cross-feeder interactions.
3. We perform a comprehensive evaluation on the Weaver Energy Dataset, demonstrating that the hybrid model, outperforms baseline approaches and provides insights into the benefits of spatiotemporal modeling.

Related Work

Energy consumption forecasting has been extensively studied, with approaches ranging from statistical models to advanced machine learning techniques. Traditional energy forecasting methods in industry often rely on static modeling, where tools generate single expansion plans that cannot adapt to changing conditions, such as the increasing penetration of distributed energy resources or evolving market dynamics (1). These methods typically depend on static assumptions, making them rigid and unable to capture the dynamic nature of energy consumption in modern grids. For instance, in the European Union, 86% of grid investments are still directed toward passive infrastructure upgrades (e.g., physical expansions), rather than active, flexible management measures that can dynamically respond to changing demand (2).

Among classical approaches, AutoRegressive Integrated Moving Average (ARIMA) and its seasonal variant (SARIMA) have been widely used for time series forecasting in the energy sector due to their interpretability and strong theoretical foundation (3; 4). Fathin et al. demonstrated the effectiveness of ARIMA in energy forecasting, where an ARIMA (8,2,0) model achieved a mean percentage error of 5.3%. However, ARIMA's reliance on stationarity and linearity limits its ability to model the complex, non-linear patterns often observed in energy consumption. Seasonal ARIMA (SARIMA) extends ARIMA by incorporating seasonal components, making it better suited for data with periodic patterns. Studies generally use SARIMA in conjunction with other machine learning methods to improve performance. Zhao demonstrated the utility of SARIMA for urban electricity consumption forecasting, showing that a hybrid SARIMA-Random Forest model reduced the Mean Absolute Error (MAE) from 944.903 (SARIMA alone) to 562.767 (5). This hybrid approach highlights how SARIMA's limitations can be mitigated by combining it with more flexible machine learning methods.

To overcome the limitations of classical statistical methods, deep learning techniques—particularly Recurrent Neural Networks (RNNs) such as LSTM and Gated Recurrent Unit (GRU) networks—have gained popularity in energy forecasting (6; 7). LSTM networks are specifically designed to capture long-range dependencies in sequential data, making them well-suited for time series prediction tasks. However, despite their flexibility, LSTM models also suffer from the notable challenge of modeling cross-feeder interactions, modeling feeders independently. A comparative study between ARIMA and LSTM for electricity consumption prediction further emphasized these challenges, showing that while LSTM outperformed ARIMA for complex, non-linear datasets, it struggled to generalize in scenarios with sparse training data (7). This has motivated the exploration of hybrid approaches that combine LSTM with spatial modeling techniques.

Graph Neural Networks (GNNs) have emerged as a powerful tool for capturing spatial dependencies in complex systems, making them highly suitable for energy forecasting in grid networks. A comprehensive survey by researchers categorized GNNs into several types, including recurrent graph neural networks, convolutional graph neural networks, graph autoencoders, and spatiotemporal graph neural networks (ST-GNNs) (8). These models excel at capturing complex interactions between interconnected entities, making them ideal for spatiotemporal forecasting tasks. In energy forecasting, GNNs can model the influence of neighboring feeders on a target feeder's consumption, improving the model's robustness and predictive accuracy. Campagne et.al proposed a graph-based energy forecasting model where each node represents a region's combined load, and edges encode interactions between these regions (9). This model significantly improved forecasting accuracy by leveraging the interconnected nature of energy consumption across regions.

The strengths of GNNs in capturing cross-feeder interactions make them an ideal complement to temporal models like LSTM, forming the basis for hybrid spatiotemporal frameworks like WattCast. In this work, we build on these by exploring a modularized approach by integrating LSTM-based temporal modeling with GNN-based spatial embeddings, providing a unified solution for low-voltage feeder-level energy forecasting.

Methodology

Dataset

In this study, we leverage the Weave Energy Dataset, a comprehensive energy consumption dataset that provides half-hourly aggregated energy consumption data across the United Kingdom. The dataset encompasses over 100,000 low-voltage feeders, resulting in approximately 2 billion rows of consumption records for the year 2024. This extensive coverage makes it an ideal testbed for developing and evaluating spatiotemporal forecasting methods. Given computational constraints, we focus specifically on forecasting for the feeders in West London in December 2024. The region of interest is defined by a bounding box of geographic coordinates:

$$\text{West London Bounding Box: } (-0.6°\mathrm{E}, 51.4°\mathrm{N}) \text{ to } (0.4°\mathrm{E}, 51.7°\mathrm{N})$$

This focus allows for a more manageable subset of the dataset, while still maintaining sufficient spatial and temporal diversity for robust model evaluation. Our analysis centers on December 2024, providing insights into energy consumption patterns during a period of seasonal demand variability. Figure 1 provides an overview of energy consumption across the United Kingdom at 8:00 PM on July 14, 2024, showcasing the spatial distribution of electricity usage. A more localized view of consumption within the city of Southampton is presented in Figure 2, capturing the detailed consumption patterns at the feeder level. Finally, Figure 3 illustrates the temporal dynamics of energy consumption across all feeders connected to a single substation in December 2024, highlighting the variability of demand over time.

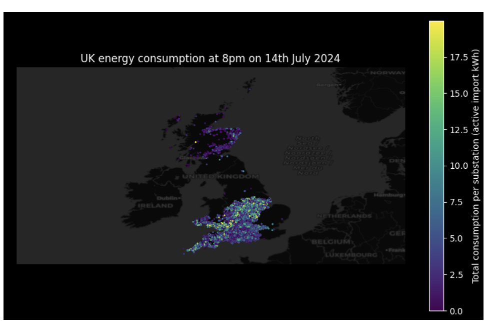


Figure 1: Energy consumption across the UK at 8:00 PM on July 14, 2024.

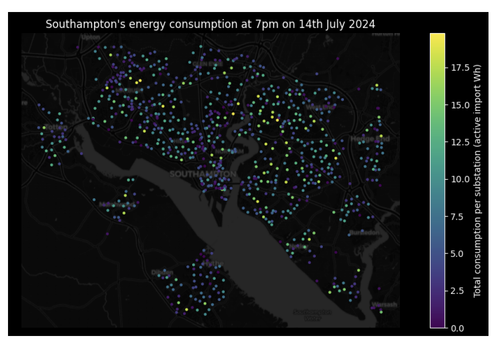


Figure 2: Energy consumption in Southampton at 7:00 PM on July 14, 2024.

Problem Formulation

The objective of this study is to develop a spatiotemporal forecasting model capable of accurately predicting energy consumption at the low-voltage feeder level in West London. Formally, the problem can be defined as follows:

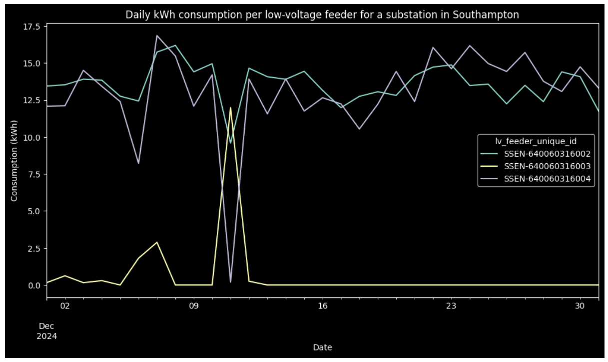


Figure 3: Energy consumption across all feeders of a single substation in December 2024, highlighting daily demand patterns.

Given a set of N feeders $\{F_1, F_2, \ldots, F_N\}$, each characterized by a time series of historical energy consumption values, our goal is to forecast the future energy consumption of each feeder over a specified prediction horizon. Specifically, let:

1. $X_{i,t} \in \mathbb{R}$ represent the energy consumption of feeder F_i at time t.
2. $\mathbf{X}_t = [X_{1,t}, X_{2,t}, \ldots, X_{N,t}]^\top \in \mathbb{R}^N$ denote the vector of energy consumption values for all N substations at time t.
3. $s = 48$ represent the sequence length (one full day of half-hourly measurements).

Our goal is to learn a mapping function $f : \mathbb{R}^s \rightarrow \mathbb{R}$ for each feeder independently, such that:

$$\hat{X}_{i,t+s+1} = f(X_{i,t:t+s})$$

where:

1. $\hat{X}_{i,t+s+1}$ is the predicted energy consumption of feeder F_i for the next time step following the input sequence.
2. $X_{i,t:t+s}$ represents the historical consumption values for feeder F_i used as input.

Spatial Dependencies and Graph Representation

Although the forecasting is performed at the feeder level, the Weave Dataset does not include the individual feeder locations, releasing only substation-level coordinates. We thus formulate our graph at the substation level where each substation is represented as a node in an undirected graph $G = (V, E)$. The exact edge and node features for this problem is described in the experimental setup below.

Model Architectures

Our proposed forecasting framework, WattCast, is designed as a hybrid spatiotemporal model that combines a LSTM network for temporal modeling with a Graph Convolutional Network (GCN) for spatial modeling. This architecture is specifically tailored to capture both the temporal dependencies of energy consumption at each substation and the spatial dependencies between interconnected substations.

WattCast utilizes a LSTM network to model the temporal dynamics of energy consumption at each feeder. The LSTM processes a sequence of historical consumption values for each feeder, learning to capture complex temporal patterns.

GCN for Spatial Modeling

To capture spatial dependencies between substations, WattCast employs a GCN. The GCN propagates information between connected substations using the following update rule:

$$\mathbf{H}^{(l+1)} = \sigma(\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}\mathbf{H}^{(l)}\mathbf{W}^{(l)})$$

where:

1. $\mathbf{H}^{(l)}$ is the feature matrix at layer l.
2. $\mathbf{D}$ is the degree matrix of the graph.
3. $\mathbf{W}^{(l)}$ is the learnable weight matrix for layer l.
4. $\sigma(\cdot)$ is an activation function (e.g., ReLU).

The final output of the GCN is a set of learned substation embeddings $\mathbf{E} \in \mathbb{R}^{S \times d}$, capturing spatial relationships between substations.

Hybrid Spatiotemporal Architecture

WattCast integrates the LSTM and GCN into a unified spatiotemporal architecture for each feeder:

1. The LSTM processes the historical consumption values for each feeder, producing a hidden state $\mathbf{h}_t$ that encodes temporal dependencies.
2. The GCN generates spatial embeddings $\mathbf{E}$ for each substation, capturing spatial relationships.
3. The LSTM hidden state $\mathbf{h}_t$ and the GCN embedding $\mathbf{e}$ corresponding to the substation of the feeder are concatenated:
$$\mathbf{F} = [\mathbf{h}_t \,||\, \mathbf{e}]$$
4. The concatenated representation $\mathbf{F}$ is passed through a fully connected (linear) layer to obtain the final prediction

This hybrid architecture allows WattCast to effectively leverage both temporal and spatial information, providing a comprehensive understanding of energy consumption patterns across substations.

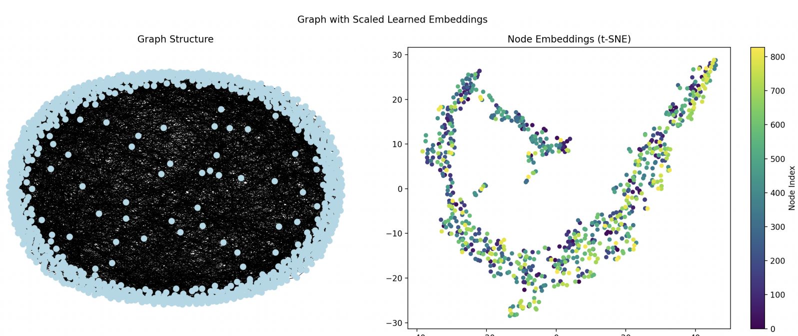


Figure 4: Graph structure representing substations as nodes and their spatial relationships as edges, used in the GCN of WattCast.

Experimental Setup

Data Preprocessing and Substation Profile Generation

We begin by extracting the energy consumption data for December 2024 from the Weave Energy Dataset, focusing on low-voltage feeders within the West London region.

Each substation's energy consumption is represented as a 48-dimensional vector, corresponding to the first 48 timestamps of December, summed across all feeders in the substation. This aggregated consumption is then scaled and log-transformed. Formally, the aggregated and transformed consumption for substation S_i at time t is defined as:

$$X_{i,t}^{scaled} = \log\left(\frac{\sum_{f=1}^{F_i} X_{f,t}}{\max\left(\sum_{f=1}^{F_i} X_{f,t}\right)} + 1\right)$$

where:

1. $X_{f,t}$ is the energy consumption of feeder f at time t.
2. F_i is the total number of feeders associated with substation S_i.

Each substation's 48-dimensional vector captures both the temporal distribution of energy consumption across a typical day (our assumed daily energy consumption cycle) and the magnitude of consumption for each substation. Importantly, we do not normalize the data, as our goal is to maintain the relative differences in consumption magnitudes between substations.

Graph Construction

The spatial relationships between substations are captured through a graph structure, where each substation is represented as a node. The edges between nodes are determined using a nearest-neighbor approach. Each substation is connected to its 5 nearest neighbors, calculated based on the Euclidean distance between the substations. An edge is created between two substations if they are among each other's nearest neighbors. The resulting graph is represented by an adjacency matrix $\mathbf{A}$, where $A_{ij} > 0$ indicates a connection between substations S_i and S_j. The edge weights are set to the inverse of the Euclidean distance, providing greater influence to closer neighbors.

Training and Evaluation Setup

We train the GCN first on the generated substation profiles. Figure 4 contains a representation of the graph structure and a low-dimensional visual of the learned substation embeddings using t-SNE.

The LSTM is trained with the following hyperparameters: sequence length (48), optimizer (Adam, learning rate = 0.001), batch size (32), LSTM hidden dimension (64), GCN hidden dimension (16), number of GCN layers (2), activation function (ReLU), loss function (Mean Squared Error), training epochs (100). Both training and evaluation are performed using teacher forcing, where the true consumption values are provided as input at each step during the prediction horizon. This ensures that the model is trained and evaluated with accurate historical context, avoiding the accumulation of errors that can occur in autoregressive forecasting. The dataset is split into training (first 80%), testing (last 20%) with performance evaluated across three standard metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 Score. WattCast is implemented entirely in PyTorch, with training on either cuda or Mac mps backend.

We compare WattCast to a vanilla LSTM implementation and several baselines including last value prediction (Naive), last day mean value prediction (Mean (1 day)), replicating full previous day cycle (Persistence), Seasonal ARIMA with 48 step cycle (SARIMA), Light Gradient Boosting Machine (Boosting), and Prophet.

Results

Table 1 and Table 2 report the mean/standard deviation and median/inter-quartile range, respectively across all models and metrics. We note that there were several feeders that had anomalous consumption and skewed several metrics, motivating the reporting of median and IQR to better gauge the average performance of models. The median RMSE, MAE are lowest for the LSTM without augmented graph embeddings, though performance is very similar to WattCast. The introduced complexity of learned embeddings likely requires further optimization. WattCast achieves the highest mean and median R^2, 0.72 and 0.74 respectively, confirming that it better captures the variability in energy consumption across feeders, making it a more robust model for energy forecasting in spatially distributed networks. The deep learning approaches drastically outperform all the baselines. Sample predictions can be seen in Figure 5.

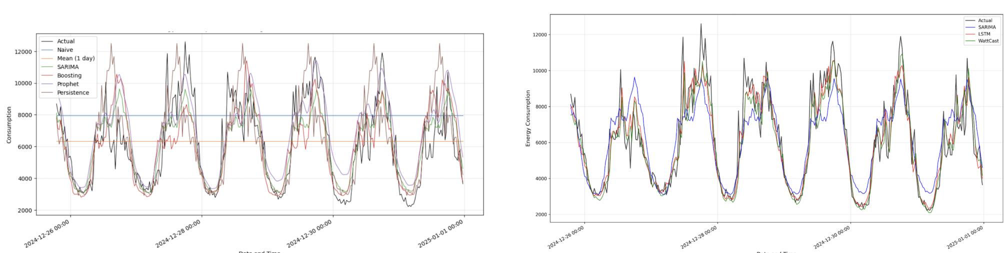
These findings highlight the trade-offs associated with spatiotemporal modeling. The LSTM model is highly effective at capturing temporal patterns but is limited in its ability to account for interactions between substations. By contrast, WattCast leverages both temporal and spatial information, leading to a more nuanced understanding of consumption dynamics.

Model	RMSE (Mean ± Std)	MAE (Mean ± Std)	R^2 (Mean ± Std)
Naive	1634195.13 ± 59185882.30	577900.84 ± 21744710.85	-0.98 ± 1.64
Mean (1 day)	1633584.81 ± 59185954.12	577406.12 ± 21744713.12	-0.15 ± 0.27
Persistence	**1633395.32 ± 59185984.11**	**577034.27 ± 21744715.01**	-0.14 ± 0.71
SARIMA	1639061.59 ± 59186005.94	582986.56 ± 21745301.56	-5908.90 ± 189150.55
Boosting	11037914.41 ± 162417789.93	9581136.42 ± 149913745.30	-5530924920.54 ± 145458084668.64
Prophet	2722850.52 ± 64236974.20	1516106.96 ± 31519259.42	-52944315.02 ± 1705892275.07
LSTM	1853368.82 ± 60051296.79	739421.60 ± 22268124.19	0.67 ± 0.15
WattCast	2164677.62 ± 64948851.69	901593.55 ± 24245035.34	**0.72 ± 0.12**

Table 1: Baseline Model Performance (Mean ± Std)

Model	RMSE (Median, IQR)	MAE (Median, IQR)	R^2 (Median, IQR)
Naive	2749.29 (1752.37 - 4104.62)	2282.66 (1449.06 - 3428.39)	-0.38 (-1.15 - -0.09)
Mean (1 day)	2294.91 (1521.94 - 3294.65)	1890.45 (1237.41 - 2723.76)	-0.06 (-0.19 - -0.01)
Persistence	2118.98 (1500.47 - 2913.93)	1594.32 (1113.39 - 2221.70)	0.03 (-0.37 - 0.31)
SARIMA	1620.59 (1159.45 - 2271.95)	1308.33 (912.26 - 1848.58)	0.44 (0.19 - 0.60)
Boosting	1959.99 (1363.64 - 2816.24)	1535.10 (1049.63 - 2228.62)	0.16 (-0.20 - 0.44)
Prophet	1688.43 (1176.74 - 2367.24)	1361.28 (919.70 - 1938.18)	0.38 (0.14 - 0.55)
LSTM	**1189.29 (877.04 - 1603.14)**	**898.35 (645.05 - 1210.70)**	0.70 (0.59 - 0.78)
WattCast	1200.88 (914.19 - 1554.55)	903.03 (679.97 - 1186.17)	**0.74 (0.65 - 0.81)**

Table 2: Baseline Model Performance (Median, IQR)



(a) Baseline Model Performance

(b) LSTM Model Performance

Figure 5: Comparison of Baseline and LSTM Model Predictions for Feeder SSEN-2803005160

Conclusion

In this work, we presented WattCast, a hybrid spatiotemporal forecasting model designed to accurately predict energy consumption at the feeder level. WattCast combines an LSTM network for temporal modeling with a GCN for spatial modeling, capturing both the temporal evolution of energy consumption at each substation and the spatial relationships between interconnected substations. Our approach was evaluated using the Weave Energy Dataset, with performance assessed against a range of baseline models, including traditional time-series methods (Naive, Mean, Persistence, ARIMA, Boosting, and Prophet) and a standalone LSTM model.

The results demonstrate that WattCast achieves competitive performance, performing slightly worse than the LSTM model on RMSE and MAE but achieving a higher R^2 score. This suggests that while the spatial component introduces additional complexity, it also enhances the model's ability to capture variance in energy consumption. These findings validate the importance of spatiotemporal modeling in energy forecasting, particularly for spatially distributed networks such as substations.

Future work can build on this foundation by exploring more advanced graph neural network architectures, such as Graph Attention Networks (GAT), which may better capture complex spatial relationships. Additionally, the robustness of WattCast could be enhanced through regularization techniques, improved graph construction methods, or adaptive graph structures that dynamically adjust connections based on changing consumption patterns. This work demonstrates the potential of hybrid spatiotemporal modeling for energy forecasting, paving the way for more sophisticated approaches in this domain.

References

[1] A. Degleris, A. El Gamal, and R. Rajagopal, "Scalable and interactive electricity grid expansion planning," arXiv, Preprint, October 2024, accessed: 2025-05-11. [Online]. Available: https://arxiv.org/abs/2410.13055

[2] R. Rana and I. B. Sperstad, "How do we integrate local energy communities into grid planning?" SINTEF Energy Research, Norway, Report, September 2023, accessed: 2025-05-11. [Online]. Available: https://blog.sintef.com/energy/how-do-we-integrate-local-energy-communities-into-grid-planning/

[3] J. Polprasert, V. A. H. Nguyen, and S. N. Charoensook, "Forecasting models for hydropower production using arima method," in *2021 9th International Electrical Engineering Congress (iEECON)*. IEEE, 2021, accessed: 2025-05-11. [Online]. Available: https://doi.org/10.1109/IEECON51072.2021.9440293

[4] M. Elsaraiti, G. Ali, H. Musbah, A. Merabet, and T. Little, "Time series analysis of electricity consumption forecasting using arima model," in *2021 IEEE Green Technologies Conference (GreenTech)*. IEEE, 2021, pp. 259–262, accessed: 2025-05-11. [Online]. Available: https://doi.org/10.1109/GreenTech48523.2021.00049

[5] S. Zhao, "Urban electricity consumption forecasting based on sarima and random forest modeling," *Journal of Electrotechnology, Electrical Engineering and Management*, vol. 7, no. 1, pp. 113–118, 2024, accessed: 2025-05-11. [Online]. Available: https://www.clausiuspress.com/article/12194.html

[6] S. Mahjoub, L. Chrifi-Alaoui, B. Marhic, and L. Delahoche, "Predicting energy consumption using lstm, multi-layer gru and drop-gru neural networks," *Sensors*, vol. 22, no. 11, p. 4062, 2022, accessed: 2025-05-11. [Online]. Available: https://www.mdpi.com/1424-8220/22/11/4062

[7] K. Enerud, "Predicting electricity consumption with arima and recurrent neural networks," Uppsala University, Sweden, Thesis, February 2024, accessed: 2025-05-11. [Online]. Available: https://www.diva-portal.org/smash/get/diva2:1834866/FULLTEXT01.pdf

[8] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2020, accessed: 2025-05-11. [Online]. Available: https://arxiv.org/abs/1901.00596

[9] E. Campagne, Y. Amara-Ouali, Y. Goude, and A. Kalogeratos, "Leveraging graph neural networks to forecast electricity consumption," arXiv, Preprint, August 2024, accessed: 2025-05-11. [Online]. Available: https://arxiv.org/abs/2408.17366