Energy Grid Forecasting and Anomaly Detection

Cody Hill
University of Colorado Boulder
Boulder, USA
cody.hill-1@colorado.edu

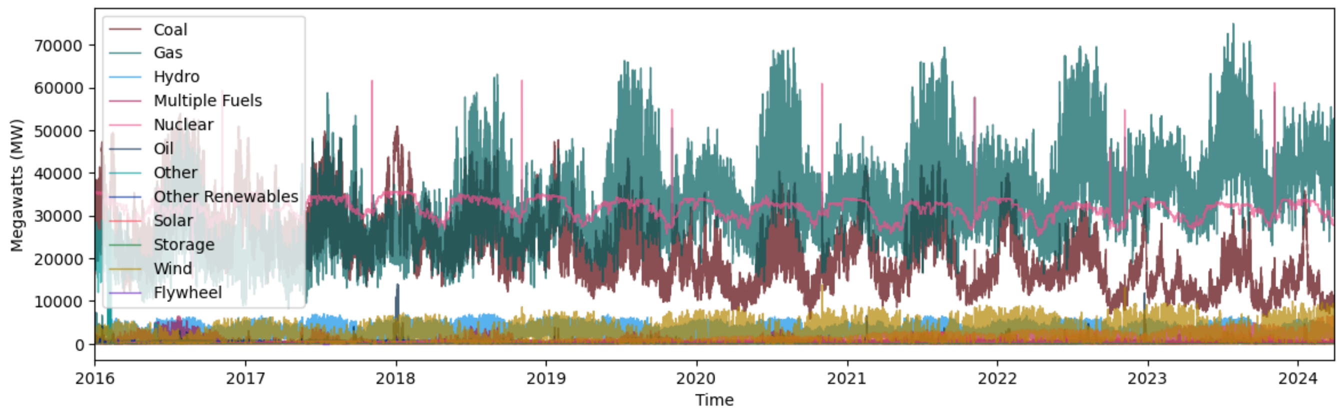

Figure 1: Energy Generated in Megawatts (MW) by Fuel Type.

ABSTRACT

Energy forecasting is a vital part of today's energy markets and operations. It is used to set the cost of electricity, determine when power generation should be increased, and decide when power needs to be directed to or from other regions of the electrical grid. Electrical energy demand continues to rise, especially with the recent widespread adoption of electric vehicles and heat pumps. As a result, the electrical grid will need to become more robust and adaptive as this increase in demand comes in tandem with a transition away from fossil fuels and into a more diversified energy market and with it, more complex cycles. These circumstances make accurate forecasting and periodicity identification of electrical loads and generation that much more important, and it is necessary to research the differences in forecasting applications and relevance in various short-term, medium-term, and long-term settings.

Many of the statistical time series forecasting techniques of the past are still in use today, each with their own strengths and weaknesses, often trading compute efficiency with lesser accuracy in long-term forecast horizons or ability to capture multiple seasonal periods. More modern machine learning models and neural network-based architectures tend to generalize through forecasting horizons better, but often at great cost of compute resources making them inefficient for short forecast horizon forecasts. Though because of the potential for complexity of energy load patterns, both categories see benefits in task-specific tuning. By evaluating each forecasting model's accuracy across various context windows and forecasting horizons, this project presents a thorough comparison on which techniques are most relevant in the task-specific energy grid context and forecast horizons.

Data Mining Project '24, April 2024, Boulder, CO
2024. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

CCS CONCEPTS

• **Mathematics of computing → Time series analysis**; • **Computing methodologies → Supervised learning by regression**; *Classification and regression trees*; *Support vector machines*; *Neural networks*; • **Applied computing → Forecasting**.

KEYWORDS

time series forecasting, anomaly detection, energy load, renewables, energy grid

ACM Reference Format:
Cody Hill. 2024. Energy Grid Forecasting and Anomaly Detection. In *Proceedings of Data Mining Project (Data Mining Project '24).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 AUTHOR'S NOTE

This document is to be used as a project proposal for the beginning planning stages of the project as well as a living-document as the project progresses. Up until project completion when a new version will be created to fill in the results.

2 INTRODUCTION

Energy forecasting is an important task to not only determine the expected load on the energy grid and the requirements of energy providers to meet that demand by either ramping up production or diverting sources altogether [8]. Identifying hourly, daily, seasonal, or even yearly energy demands accurately has substantial impact on not only the logistics of the infrastructure of our energy grid, but also the complex energy marketplace. These intervals, or forecast horizons, in the context of predictive models are defined in Table 1. Short-term forecasts can be useful for players in power markets where prices dynamically shift based on generation supply, customer demand, and financial spot and derivative contracts which are all based on energy forecasting [7]. Very short-term forecasts can be important for power distributors and operators that want to meet customer demands more efficiently. Long-term forecasts are important in infrastructure planning and energy provider and distributers long-term planning. An example of a long-term forecast can be seen in Figure 2.

In the past this predictive role was largely upheld with statistical point-estimate models, but with the growth of our predictive models along with the diversification of the energy grid, more modern probabilistic interval forecasting techniques have begun to be used [8, 13, 16]. This diversification of the energy grid not only comes from energy reforms by way of deregulating the national electric market in the 1990's, but emerging cultural desires and improved technology in renewable energies, driving a need for a better electrical grid [11].

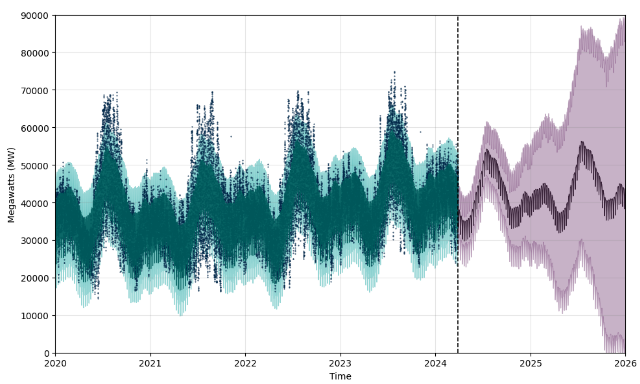

Figure 2: Long-Term Forecast of Natural Gas Energy Generation Using Meta's Prophet Model. Actual observations are represented by the blue points, forecasted fit and predictions are in teal and purple respectively.

Electricity generation has always been mostly an on-demand industry and largely continues to be, where the energy we generate needs to be transmitted and used within a short period of time because we lack efficient systems for energy storage at scale [12]. This fact presents additional complications when many renewable energies only produce energy in certain conditions, making its effective use more reliant on forecasting to navigate a more diverse and complex power grid. Additionally, with the cultural and regulatory push towards end-users adopting the use of electric appliances, electric vehicles, and heat pumps, energy demand will continue to change [6]. These circumstances mean accurate forecasting is becoming even more important with improvements having the potential to better reduce reliance on fossil-fuel energies by more efficiently filling in gaps of energy demands with a more diversified and efficient energy grid.

There is little consensus among researchers and professionals about what exemplifies state-of-the-art in time series forecasting.

Table 1: Defining prediction forecast horizon intervals. The forecast horizon determines how far into the future a forecast model predicts.

No.	Forecast Horizon Intervals	Duration
1	Very Short-Term	$X < 1$ Hour
2	Short-Term	1 Hour $\leq X <$ 1 Week
3	Medium-Term	1 Week $< X \leq$ 1 Year
4	Long-Term	$X >$ 1 Year

This seems to be due to many things including that time series forecasting is heavily task-specific, meaning one model may perform well with one type of data and forecast horizon, but might perform poorly when compared to other techniques with a different horizon or data type [17]. This is why it is still important to experiment and compare forecasting techniques on domain-specific time series data.

3 RELATED WORK

Seasonal Autoregressive Integrated Moving Average (SARIMA) and Exponential Smoothing models tend to be useful in short and medium-term forecasts but can experience limitations modeling complex periodicity and interactions [9, 14]. Newer, but certainly not state-of-the-art, machine learning techniques such as Support Vector Machines, XGBoost, and a variety of Long Short-Term Memory (LSTM) network architectures tend to generalize better and excel across multiple forecast horizons [14, 15].

Hybrid models and ensemble methods, utilizing a variety of preprocessing techniques and combining models into a unified framework, are becoming more popular as of late. The process typically involves splitting the job of extracting different types of features from the data, say the temporal and spatial relationships of the time series, and feeding those outputs into yet another separate model to perform the forecasting [10]. This technique combines the strengths of multiple models resulting in improved results across all forecast horizons [9, 10]. Many of these still utilize LSTM layers as their attention mechanism [15].

Transformer-based models and other frameworks developed and popularized in the Large Language Model (LLM) space have found a foothold in time series forecasting in the last few years as well, but their use in time series forecasting is still being researched [18]. Similar to the LSTM, transformer-based models capture complex context and relationships in data, but across greater distances of the training corpus with increased efficiency because of their ability to be parallelized effectively [19]. This type of model benefits from pre-training on a large corpus of diverse data which simply does not exist in the public domain for time series data [1]. With more time and data, researchers believe transformer-based models can be effective at time series forecasting [1, 19].

4 METHODS AND PROPOSED WORK

Several models stemming from different statistical and machine-learning families will be trained to perform electricity load forecasting across the different forecast horizon intervals listed in Table 1.

Each model's amount of compute required for training and evaluation scores will be compared across different forecast horizons. See evaluation section for more details on the specific metrics used.

4.1 Data

Electric load data was collected by the Pennsylvania-New Jersey-Maryland Interconnection (PJM). PJM is a regional electrical transmission organization which coordinates electricity transmission in all or parts of Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, and the District of Columbia [2]. A map of this can bee seen in Figure 3. Weather data was retrieved from the National Oceanic and Atmospheric Administration's (NOAA) Regional Climate Center (RCC) Applied Climate Information System (ACIS) [3]. The data was manually downloaded by selecting a weather station and choosing relevant dates and features with the daily data listings form.

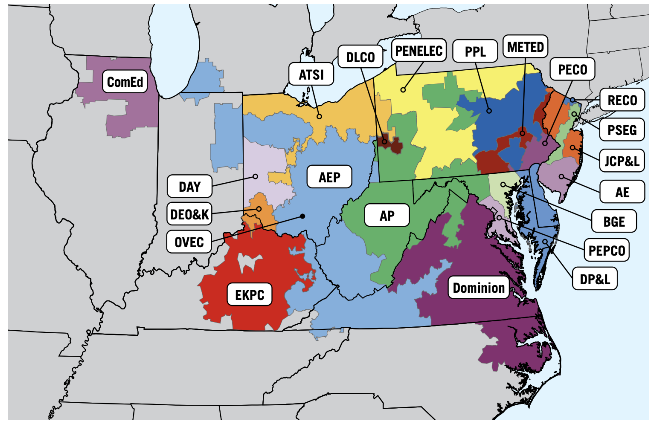

Figure 3: Regional Transmission Zones Labeled by Distributor IDs [5].

Table 2: Dataset size and date ranges.

Dataset	Size (Rows)	Date Range
Hourly Load	5,446,243	1993/01/01 - 2024/03/27
Daily Weather	32,400	1993/01/01 - 2024/04/09

The electric load data sourced from PJM is a set of hourly load in megawatt-hours verified by the individual electric distribution companies [4]. The hourly load data is flagged with geographical location data which will be leveraged to include regional weather data as a covariate in the models. The weather data has a daily sample frequency which was the only data available at the time. Hourly data would have been preferable to match the frequency of the load data, but even at the mismatched sample rate, will be a valuable exogenous covariate. Features included in this dataset are the maximum and minimum recorded temperature, average temperature, precipitation in inches. After data discovery, 5 zones were chosen for their sample size, variety in signal shape, and data continuity. The weather data was selected to match the zones of interest from the load dataset and along with the zone map in Figure 3, weather stations were chosen for each zone. The weather stations chosen to represent each zone were chosen under the following criteria: 1.) Weather station is geographically central to the zone. 2.) Weather station uptime and reliability (airport weather stations often chosen for this quality). An overview of the datasets can be seen in Table 2.

4.2 Preprocessing

The first step in preprocessing the datasets will be checking for any missing dates. Any missing rows will added by inserting the average value of the two surrounding observed data. Then the data will be analyzed for outliers, using a tumbling window method, replacing any identified outliers with the median of the window size. The two datasets will then be date aligned and merged.

4.3 Feature Engineering

Feature engineering is a large part of time series forecasting, typically past values of the target variable are highly correlated to future values. Several methods will be used to generate covariates that either have a signal smoothing effect, identifies seasonal periods, or encodes exogenous information.

1. **Date features** will be encoded in categorical features representing Hours, Days, Months, Years, and Day of Year - this will help the models identify recurring temporal relationships, or seasonality, in the target variable.
2. **Rolling mean** of the target variable of the previous 1 month of data points, adding the rolling mean helps smooth out the signal, increasing model performance.
3. **Lagged features** at specific intervals are useful if an effort is made to match known seasonal periods within the data. Lagged features of the target variable at 1 hour, 6 hours, 24 hours, 1 week, and 1 month will be used. *Note, a 1 year lagged feature would also be effective, but would have null values for the first year which can be problematic for some models and those rows would need to be removed.*
4. **Fourier transform spectral density**, similar to lagged features, helps identify correlated seasonal periods. By performing the Fast Fourier Transform on the target variable and taking the absolute value of it, the most relevant seasonal frequencies will have the greatest magnitudes. An example of the resulting spectral density plot can be seen in Figure 4.

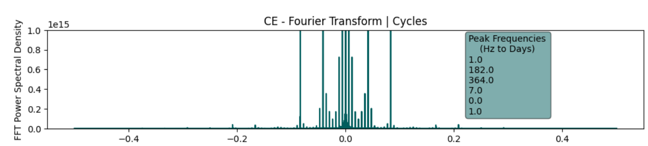

Figure 4: Spectral density plot of the ComEd zone. To cast frequency units to days: $\frac{1}{freq*24}$.

5. **First and Second-Order Derivatives/Differencing** finds the instantaneous rate of change between two observed data points and the acceleration of the rate of change respectively. This encodes direction, magnitude, and change of direction

in time series data. An example plot of both orders of derivatives can be seen in Figure 7 in the appendix.

Specific intervals of features such as lag, rolling mean, derivatives need to be considered carefully when used as exogenous features calculated from the target variable. This can be a great source of data leakage into the testing set if these are not data you would have in a prediction setting. One method to overcome this is to take a naive approach and assign last season's values. Another is to create a hybrid model which forecasts the exogenous features to include in the forecast. Finally, a rolling forecast can be used predict new datapoints and retrain the model with those predictions up until the intended forecast horizon is achieved. Here, the shorter length interval features will be excluded to prevent data leakage, but will remain implemented but not included as an exercise in time series feature engineering.

4.4 Tools and Techniques

The load data will then be analyzed and plotted for historical trends and to identify periodicity (cycles) at different context window intervals which will help set model expectations for the various forecast horizons.

Models of varying complexity and methodology will be used to better represent the broad spectrum of techniques used today, in an attempt to capture a good representation of which contemporary models work best for different forecast horizons.

Proposed models:

Baseline

- Seasonal Naive Forecast Model
- Mean

Statistical

- Seasonal Autoregressive Integrated Moving Average (SARIMA)
- Multiple Seasonal-Trend decomposition using LOESS (MSTL)

Machine Learning

- Support Vector Machine (SVM)
- XGBoost

Deep Learning

- Temporal Fusion Transformers (TFT)

Some models may require data transformation or normalization to perform effectively. For example, the SVM and TFT increase performance with scaled feature matrices.

$$Z = \frac{(X - \bar{X})}{S}$$

Centers inputs around 0 with the sample mean, and scales them with the sample standard deviation.

5 EVALUATION

5.1 Training, Validation, and Testing Data Split

The data will be split into a training, validation, testing split. A validation set is necessary here because several of the proposed forecasting models will require hyperparameter tuning. Once the data has been aggregated, cleaned, and finalized the exact ratio may be adjusted to ensure the test set is large enough to span the largest forecast horizon at least one or several times, depending on what is feasible with the data on hand.

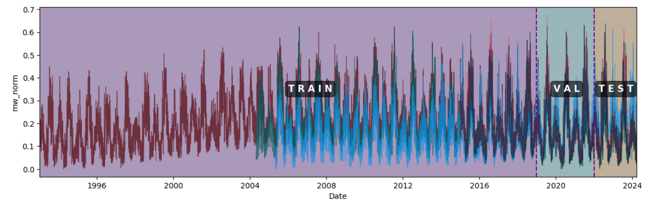

Figure 5: Plot of the training, validation, and testing split.

To avoid data leakage and maintaining temporal relationships, the test set will always be assigned the most recent dates with zero overlap in the training or validation sets. Also, it is ensured that each set contains a reasonable amount of each periodicity. Details about the resulting split can be seen in Table 3 and Figure 5.

Table 3: Dataset size and date ranges.

Split Set	Date Range
Training	1993/01/01 - 2018/12/31
Validation	2019/01/01 - 2021/12/31
Testing	2022/01/01 - 2024/03/27

5.2 Metrics

Symmetric Mean Absolute Percentage Error (sMAPE) will be the evaluation metric used to compare all models. This will ensure comparisons of error across different models and forecasting horizons are normalized.

$$\text{sMAPE} = \frac{200}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|} \tag{1}$$

where y_i and $\hat{y}_i$ are the actual values and forecasted values respectively, and n is the number of points in the set.

One possible limitation of sMAPE is if a predicted value or actual value equals zero, the metric balloons to the maximum error value, and if both equals zero, the metric is undefined. Chosen evaluation metrics often have strengths and weaknesses, and it is yet unclear if this limitation will occur here.

Root Mean Squared Error (RMSE) will also be used in model evaluation to complement sMAPE. The value of including RMSE is that it is an easily interpretable quantitative metric resulting in a value that is in the same unit as the response. It also heavily penalizes predictions the further they are from the actual value.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{2}$$

where y_i and $\hat{y}_i$ are the actual values and forecasted values respectively, and n is the number of points in the set.

In conjunction with the RMSE, a Normalized RMSE (NRMSE) will also be used to compare models between zones, since each zone have varying scales of the target variable.

$$\text{NRMSE} = \frac{RMSE}{\bar{Y}} \tag{3}$$

where $\bar{Y}$ is the mean of the actual values.

6 DISCUSSION

More discussions and analysis will be added here when the project has been completed.

6.1 Project Timeline

Table 4 outlines the specific tasks and their projected schedule.

Table 4: A general framework onto which the project will progress. This timeline is subject to change.

Timeline	Agenda Item	Progress	Done
Days 1 - 5	Data Mining	–	X
	Hourly Load Data	–	X
	Hourly Generation Data	–	X
	Weather Data	–	X
	Cleaning/Preprocessing	–	X
	Datetime Alignment	–	X
	Aggregate by Zones	–	X
	Outlier Inspection	–	X
	Corrupted Data Removal	–	X
	Feature Engineering	–	X
	Exploratory Data Analysis	–	X
	Train/Val/Test Split	–	X
Days 6-8	Anomaly Detection		
	Forecast Model Setup	–	X
	Training	–	X
	Val. Optimizations	–	X
	Build Eval. Functions	X	
Days 9-10	Test Results		
	Discussion		
	Conclusion		

6.2 Potential Challenges

A few challenges may arise impeding progress or causing the project to fall behind schedule are expected to be the following. Generally, properly fitting the data to the number of proposed forecast models within the timeline may pose a challenge. Specifically, the neural-network architectures can take considerable time to build from the ground up without intimate knowledge and experience of their specific architectures. Also, completely fleshing out the three goals of the project on schedule may be cause for concern and need reevaluation – anomaly detection, hourly load forecasting, and hourly energy generation forecasting.

The granularity of the data may post issues in accurately forecasting the shorter forecast horizons. Since all collected data has a time-step (sample rate) granularity of one hour, predictions less than one hour have little basis in reality. Data with smaller a time-step would be required for accurate shorter forecast horizons.

6.3 Alternate Approaches

If the project is running behind the timeline a few different approaches can be considered to make the target deadlines.

(1) The number of models can be reduced but leaving at least one remaining from each model type.
(2) In the case of some of the neural-network based architectures and machine-learning approaches, complexity can be reduced by utilizing more out-of-the-box approaches/models.
(3) Instead of forecasting both hourly load usage and generation by fuel type, the scope of the project can be reduced to focus on just one of these domains.
(4) Further reduce the scope to solely focus on anomaly detection.

7 RESULTS

As the other models are still being worked on here is a preliminary look at the results for the baseline models and XGBoost which has been tuned and finalized.

Preliminary Results:

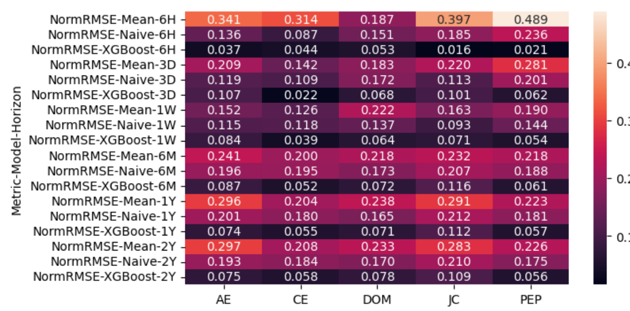

Figure 6: Results of the models that are completed.

8 CONCLUSION

By analyzing and modeling energy time series data we can efficiently and accurately forecast into the future different interval forecast horizons, which correspond to real-world decisions. Depending on the time-scale in which those decisions rely, analysts must appropriately identify what side of the accuracy and efficiency tradeoff is more important in their use case. Modern techniques in time series forecasting have reduced the errors in all forecast horizon intervals and continue to generalize those prediction abilities across many domains, but many of the most accurate models take a considerable amount of time or hardware to train. This type of research and technology is necessary for uninterrupted energy generation and distribution due to the ever-increasing demands and complexity of our energy grids.

8.1 Key Findings

XGBoost performs surprisingly well in this domain at many forecast horizons, greatly improving over the baseline methods. On top of its performance the computational demands are very low and training takes negligible time compared to other machine learning and neural network-based methods.

Once the remaining models are completed, a further comparative analysis across all horizons and methods can be completed.

8.2 Future Work

To be completed upon project completion.

9 APPENDICES

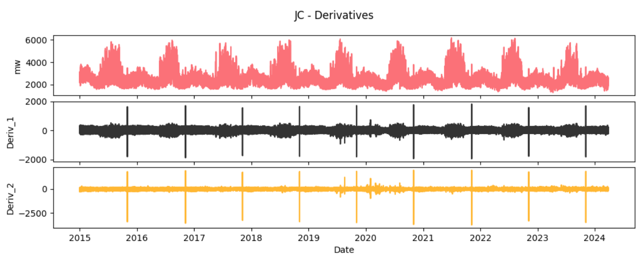

Figure 7: First and second-order derivatives/differencing of Megawatts in black and yellow respectively.

ACKNOWLEDGMENTS

Sqyd, for always brewing the morning coffee – I know it's 8 cups of water, now.

Luna, for breaking up the hyper-focused monotony with mind-clearing afternoon walks.

REFERENCES

[1] 2024. https://huggingface.co/blog/autoformer
[2] 2024. https://www.pjm.com/about-pjm/who-we-are/territory-served
[3] 2024. https://scacis.rcc-acis.org/#
[4] 2024. https://dataminer2.pjm.com/feed/hrl_load_metered/definition
[5] 2024. https://www.pjm.com/library/maps.aspx
[6] James Archsmith, Erich Muehlegger, and David S. Rapson. 2022. Future Paths of Electric Vehicle Adoption in the United States: Predictable Determinants, Obstacles, and Opportunities. *Environmental and Energy Policy and the Economy* 3 (2022), 71–110. https://doi.org/10.1086/717219 arXiv:https://doi.org/10.1086/717219
[7] Derek W. Bunn. 2004. *Modelling Prices in Competitive Electricity Markets*. Wiley.
[8] Tao Hong and Shu Fan. 2016. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting* 32, 3 (2016), 914–938. https://doi.org/10.1016/j.ijforecast.2015.11.011
[9] Konstantin Hopf, Hannah Hartstang, and Thorsten Staake. 2023. Meta-Regression Analysis of Errors in Short-Term Electricity Load Forecasting. In *Companion Proceedings of the 14th ACM International Conference on Future Energy Systems* (Orlando, FL, USA) *(e-Energy '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 32–39. https://doi.org/10.1145/3599733.3600248
[10] Heng Hua, Mingping Liu, Yuqin Li, Suhui Deng, and Qingnian Wang. 2023. An ensemble framework for short-term load forecasting based on parallel CNN and GRU with improved ResNet. *Electric Power Systems Research* 216 (2023), 109057. https://doi.org/10.1016/j.epsr.2022.109057
[11] Paul Joskow. 2001. U.S. Energy Policy During the 1990s. *National Bureau of Economic Research, Inc, NBER Working Papers* (01 2001).
[12] Chun Sing Lai, Giorgio Locatelli, Andrew Pimm, Xiaomei Wu, and Loi Lei Lai. 2021. A review on long-term electrical power system modeling with energy storage. *Journal of Cleaner Production* 280 (2021), 124298. https://doi.org/10.1016/j.jclepro.2020.124298
[13] Eunjung Lee and Wonjong Rhee. 2021. Individualized Short-Term Electric Load Forecasting With Deep Neural Network Based Transfer Learning and Meta Learning. *IEEE Access* PP (01 2021), 1–1. https://doi.org/10.1109/ACCESS.2021.3053317
[14] Shahzad Muzaffar and Afshin Afshari. 2019. Short-Term Load Forecasts Using LSTM Networks. *Energy Procedia* 158 (2019), 2922–2927. https://doi.org/10.1016/j.egypro.2019.01.952 Innovative Solutions for Energy Transitions.
[15] Lichao Sun, Hang Qin, Krzysztof Przystupa, Michał Majka, and Orest Kochan. 2022. Individualized Short-Term Electric Load Forecasting Using Data-Driven Meta-Heuristic Method Based on LSTM Network. *Sensors* 22 (10 2022), 7900. https://doi.org/10.3390/s22207900
[16] Rafał Weron. 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting* 30, 4 (2014), 1030–1081. https://doi.org/10.1016/j.ijforecast.2014.08.008
[17] Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yu Tong, and Bixiong Xu. 2021. TS2Vec: Towards Universal Representation of Time Series. In *AAAI Conference on Artificial Intelligence*. https://api.semanticscholar.org/CorpusID:237497421
[18] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. 2022. Are Transformers Effective for Time Series Forecasting? arXiv:2205.13504 [cs.AI]
[19] Vincent Zhihao Zheng and Lijun Sun. 2024. Multivariate Probabilistic Time Series Forecasting with Correlated Errors. arXiv:2402.01000 [stat.ML]