Team 201 Progress Report: Future Energy Needs at a Glance

Sean Crowley Ravi Prayaga Ndubusi Anyanwu Hyung Park Trystan Knowles Christopher Baer

I. Introduction

As global energy demand continues to rise, the need for accurate and adaptive load forecasting has become increasingly critical for ensuring energy security, economic stability, and sustainability. Effective forecasting enables grid operators and policymakers to anticipate fluctuations in energy consumption, optimize resource allocation, and integrate renewable energy sources more efficiently. This study provides a comprehensive analysis of key factors influencing future energy demand, including technological advancements, socio-economic trends, and policy interventions. By examining emerging challenges and forecasting methodologies, we aim to develop load forecasting methodologies that will promote sustainability, affordability, and reliability in the evolving energy landscape.

II. Problem Definition

We are generating a heat map visualization of 10 year energy demand forecasts to better understand potential grid strain and failure likelihood in the near-future.

III. Literature Review

Load forecasting in the energy sector is a critical task that enables efficient grid management, resource allocation, and operational planning. The literature on forecasting methodologies highlights a range of techniques, from traditional statistical models to modern machine learning and hybrid approaches, each contributing unique advantages and addressing different forecasting challenges.

A. Market Factors

The evolving energy industry is shaped by technological, socio-economic, and environmental factors that significantly impact energy demand forecasting. A key challenge is the increasing complexity of energy systems due to flexible demand technologies like smart meters and distributed generation, alongside growing electrification and intermittent renewable supply [1]. These factors necessitate adaptive forecasting models, though current research lacks detailed exploration of model implementation. Smart grids also play a crucial role in demand shifts, as the study by Tuomela et al. examines their integration in Finnish households, revealing significant variability in energy savings across different regions [2]. This underscores the need for region-specific forecasting models that account for infrastructural and regulatory differences. Similarly, the building sector remains a major contributor to energy consumption, with regulatory, technological, and behavioral factors influencing long-term demand trends [3]. However, reliance on short-term data limits the ability to predict long-term shifts in energy use. Renewable energy integration further complicates forecasting by impacting grid stability. With increasing distributed energy systems, maintaining grid reliability becomes a challenge, as consumers also become producers [4]. This shift requires models that incorporate both energy consumption and production patterns, particularly in residential sectors. On a broader scale, energy demand drivers differ between developed and developing nations. Economic growth, infrastructure, and climate drive demand in developing countries, whereas institutional factors like innovation and governance play a larger role in developed economies [5]. Additionally, long-term studies show that energy consumption in wealthier nations grows more slowly than GDP, highlighting the role of technological advancements [6]. Socio-economic policies, such as energy taxes and demographic shifts, also influence demand, emphasizing the need for policy-driven forecasting approaches [7]. These studies illustrate the complexity of energy demand forecasting in a rapidly changing landscape. As flexible demand, renewable integration, and regional variations reshape consumption patterns, traditional models must evolve to incorporate these emerging variables effectively.

B. Machine Learning and Deep Learning Approaches

Several studies demonstrate the effectiveness of machine learning (ML) and deep learning (DL) methods in improving load forecasting accuracy. Advanced architectures such as long short-term memory (LSTM) networks and convolutional neural networks (CNNs) capture temporal dependencies and nonlinear patterns in energy consumption data [8]–[11]. These approaches are particularly useful for handling large-scale datasets and dynamic load variations, which are crucial for improving demand predictions in real-world applications. Additionally, hybrid models combining deep learning with traditional statistical methods show promise in reducing forecasting errors by leveraging the strengths of both approaches [12], [13].

C. Statistical and Time Series Models

Traditional statistical techniques, such as autoregressive integrated moving average (ARIMA) and exponential smoothing methods, remain foundational in load forecasting [14], [15]. These models provide interpretable and robust predictions, particularly for short-term forecasting where historical patterns are relatively stable. Despite their limitations in capturing complex nonlinear relationships, their integration with ML-based models can enhance overall forecasting performance by providing baseline predictions that ML models can refine [12], [16].

D. Feature Engineering and Data Preprocessing

Several studies emphasize the importance of feature selection and data preprocessing in forecasting accuracy. Methods such as principal component analysis (PCA) and wavelet transformations help in extracting relevant features from raw energy consumption data, thereby reducing noise and improving model efficiency [3], [17]. Additionally, handling missing data and seasonality adjustments play a crucial role in ensuring model reliability [1], [16]. These insights are valuable for optimizing input features in our forecasting framework, enhancing the predictive power of ML-based models.

E. Hybrid and Ensemble Methods

The combination of different modeling techniques has been explored to improve forecasting precision. Hybrid models that integrate statistical methods with deep learning architectures or ensemble approaches, such as boosting and bagging, have been found to outperform individual models by leveraging multiple predictive strengths [9], [13], [16], [18]. These methods provide robustness against fluctuations and uncertainties in energy demand, making them particularly relevant for our research, where high accuracy and adaptability are required.

F. Applications in the Energy Sector

The application of these models to real-world energy demand forecasting has demonstrated their potential in grid optimization, renewable energy integration, and demand-side management. Studies analyzing regional and national-scale energy consumption highlight the importance of incorporating external factors such as weather conditions, economic indicators, and policy changes into forecasting models [2], [3], [17], [19]. These insights will be crucial for refining our approach by integrating external variables that influence energy demand dynamics.

G. Conclusion

The reviewed literature underscores the necessity of adopting a multi-faceted approach to load forecasting, leveraging both statistical and ML-driven methods. While deep learning techniques enhance the ability to model complex patterns, traditional methods provide a solid foundation for interpretable and stable predictions. Hybrid and ensemble techniques further improve forecasting reliability, making them ideal candidates for our research into energy demand prediction. Incorporating feature engineering strategies and external influencing factors will be critical in optimizing our forecasting framework for practical implementation in the energy sector.

IV. Proposed Method

In Progress. This study aims to enhance energy demand forecasting by leveraging machine learning techniques to analyze US historical data and produce dynamic, geographically mapped projections of future consumption. Unlike traditional forecasting methods that primarily rely on statistical models with near-term forecasts, our approach integrates interactive heat maps to illustrate predicted demand shifts over a 10 year time period, aligning with existing grid infrastructure. This heat map will allow for regional outlook of consumption not based on political boundaries in the US, but by energy provider infrastructure layout. Combining both current abilities to meet demand with future consumer energy needs will allow providers to optimize resource allocation, policymakers to make steps towards preventing grid failures, and consumers to better comprehend the near-future load expectations.

A. Data Preprocessing

The data preprocessing phase involved several systematic steps to prepare the Electricity Generation dataset for model development. The dataset comprises historical electricity generation data collected between 2015 and 2024, with each source file containing monthly net generation figures for the corresponding year. Below is a detailed outline of the preprocessing procedures:

- **Data Acquisition and Collection:** Electricity Generation data was acquired for each year from 2015 to 2024. Each dataset contained detailed records of monthly net generation, which served as the foundational element for further analysis.
- **Data Integration:** To facilitate comprehensive temporal analysis, the monthly net generation data from each annual dataset were integrated. This process involved aligning and merging the data across all years to construct a unified dataset. The result was a consolidated time series that represents the monthly net generation values from 2015 through 2024, ensuring that the dataset is chronologically complete and consistent.
- **Data Cleaning and Missing Value Treatment:** The initial integrated dataset was examined for inconsistencies, errors, and missing values. Since the net generation values are inherently numeric, any missing or null entries identified during this inspection were systematically replaced with a value of 0. This imputation strategy is particularly suitable for the domain, as it preserves the numeric integrity of the data and mitigates potential issues during subsequent statistical analysis and model training.
- **Final Dataset Verification:** After the integration and cleaning steps, the final consolidated dataset underwent verification procedures. This ensured that all records adhered to the required format and that no gaps existed in the time series data from 2015 to 2024. The verification step is crucial for maintaining data quality and ensuring that the dataset is ready for advanced analytics and model development.

B. Time-Series Modeling

One of the modeling methodologies to predict the electricity demand is to consider the historic data as a time series dataset. Of the various data elements available in the dataset chosen, electricity generation and storage are key indicators to determine demand. Of course, the assumption is that the electricity generation and storage are keeping up with demand over time.

One of the cleaned datasets consists of generation data for each state in the United States and from various sources the electricity is generated for each month across years. This data is now aggregated and transformed to reflect the total electricity generated per state for each month across years. This transformed data is then ready to be analyzed and modeled as a time series dataset. Fig. 1 shows the sample view of the constructed time series dataset.

Date <date>	State <chr>	Total.Generation <dbl>
2015-01-01	AK	327312.64
2015-02-01	AK	295036.21
2015-03-01	AK	295942.31
2015-04-01	AK	222601.15
2015-05-01	AK	251415.94
2015-06-01	AK	212635.74
2015-07-01	AK	233481.72
2015-08-01	AK	239849.36
2015-09-01	AK	230783.07
2015-10-01	AK	191489.91

Fig. 1. Sample view of time-series dataset

Analysis of the time-series data was performed by plotting to observe trends, seasonality, or cyclical patterns. Fig. 2 shows this for California

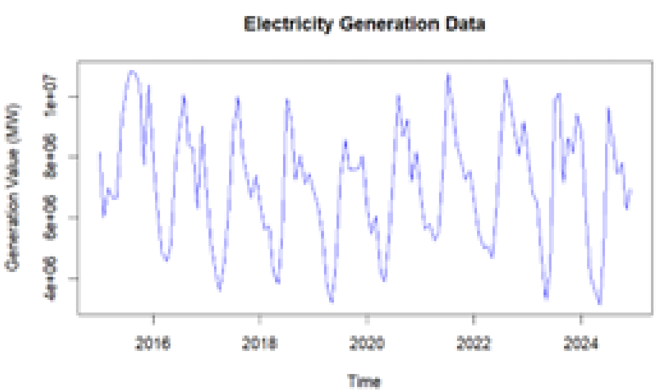

Fig. 2. Electricity Generation Data from 2015 to 2024

An Autocorrelation Function (ACF) plot was then constructed to observe the stationarity behavior of each individual dataset as seen in Fig. 3.

Date <date>	State <chr>	Total.Generation <dbl>
2015-01-01	AK	327312.64
2015-02-01	AK	295036.21
2015-03-01	AK	295942.31
2015-04-01	AK	222601.15
2015-05-01	AK	251415.94
2015-06-01	AK	212635.74
2015-07-01	AK	233481.72
2015-08-01	AK	239849.36
2015-09-01	AK	230783.07
2015-10-01	AK	191489.91

Fig. 3. ACF Plot of Electricity Generation Data

After splitting into training and test datasets, the following models were chosen:

- ARIMA: Previously listed time-series data is univariate
- ARIMA-GARCH: Previously listed time-series data is univariate
- VAR: when considering both electricity generation, storage, and other endogenous/exogeneuos time-seires it is multivariate.

C. CNN-LSTM Hybrid Architecture

The energy demand data, sourced from U.S. power plants, was first cleaned by converting net generation values to numeric types, handling non-numeric values as NaN. The data was then aggregated by state and year to enable state-level analysis, transforming it from wide to long format using the melt function. The month names were extracted, converted to datetime objects, and the dataset was sorted by state and date for temporal analysis.

To prepare the data for the CNN-LSTM model, the net generation values were normalized using MinMaxScaler, scaling the values between 0 and 1. Sequences of 12 months of past energy generation data were created, where each input sequence consisted of the last 12 months of generation, and the output was the next month's energy generation. The dataset was split into training (70%), validation (15%), and test (15%) sets, and saved in .npz format for easy manipulation during model training.

The model architecture combined Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) layers to capture both local patterns and temporal dependencies in the data. The input sequences were passed through several 1D convolutional layers, starting with 48 filters, followed by 32 and 16 filters in subsequent blocks. Each convolutional block was followed by a flattening operation, with dropout layers (rate: 0.25) applied to prevent overfitting. The CNN layers extracted high-level features from the data.

After the convolutional layers, the data was reshaped into the required format for the LSTM layers, which consisted of three layers, each with 20 units. The first two LSTM layers returned sequences, retaining temporal features across time steps, while the final LSTM layer returned only the last output, representing the model's prediction for the next time step. Another dropout layer was applied after the LSTM layers to further reduce overfitting. The final output was produced by passing the concatenated CNN and LSTM features through a dense layer with a tanh activation function, which was appropriate due to the scaled energy generation values.

The model was trained using the Adam optimizer and Mean Squared Error (MSE) loss function. The model was evaluated using Mean Absolute Error (MAE) as the performance metric. After training for 50 epochs, the model achieved a Root Mean Squared Error (RMSE) of 0.022 and an R-squared value of 0.979. These results indicate that the model accurately forecasts energy demand, explaining approximately 97.9% of the variance in the test data.

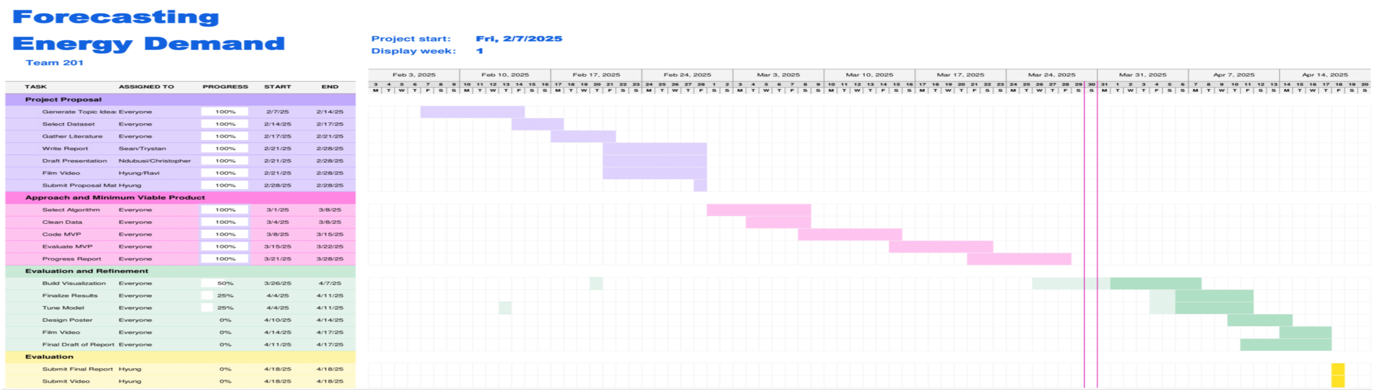

Fig. 4. Updated Gantt Chart of Team 201 Project Schedule.

This CNN-LSTM hybrid model successfully integrates feature extraction and sequence learning, making it well-suited for forecasting future energy generation based on historical data.

D. Visualization

These models will be used to generate energy predictions that, averaged across all models will be show in a heat map across the United States to better understand future states needs as seen in Fig. 6. These were constructed using D3.

E. Progress and Planning

V. Evaluation

In Progress. Proposed Evaluations of Model currently involve using a validation subset of our data for testing.

A. Time-Series Modeling

In progress.

B. CNN-LSTM Hybrid Architecture

In Progress.

VI. Conclusion and Discussion

In Progress

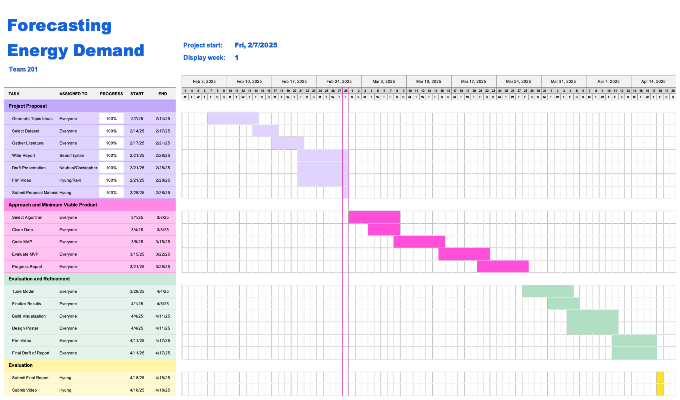

Fig. 5. Original Gantt Chart of Team 201 Project Schedule.

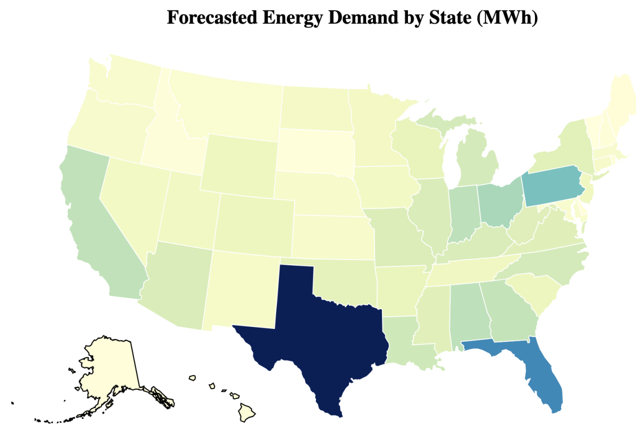

Fig. 6. 10 year outlook of US Energy Demand

A. Upcoming Experiments

Given feedback from proposal with regards to providing innovative visualizations, a different schematic of how to display risk of blackouts and with a more granular regional approach. In terms of further experimental design the focus of our efforts are to be set forth creating more innovative visualizations, as our models are currently able to predict energy demand on a state by state basis with passing validation criteria.

B. Revised Plans

To reflect the need for more innovative visualizations the timeframe in our planning has been adjusted to reflect this. All Team Members have contributed a similar amount of effort.

References

[1] S. Tuomela, M. de Castro Tomé, N. Iivari, and R. Svento, "Impacts of home energy management systems on electricity consumption," *Applied Energy*, vol. 299, p. 117310, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261921007224

[2] S. Pfenninger, A. Hawkes, and J. Keirstead, "Energy systems modeling for twenty-first century energy challenges," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 74–86, 2014.

[3] M. Santamouris and K. Vasilakopoulou, "Present and future energy consumption of buildings: Challenges and opportunities towards decarbonisation," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 1, p. 100002, 2021.

[4] O. Smith, O. Cattell, E. Farcot, R. D. O'Dea, and K. I. Hopcraft, "The effect of renewable energy incorporation on power grid stability and resilience," *Science Advances*, vol. 8, no. 9, p. eabj6734, 2022. [Online]. Available: https://www.science.org/doi/abs/10.1126/sciadv.abj6734

[5] I. Dokas, M. Panagiotidis, S. Papadamou, and E. Spyromitros, "The determinants of energy and electricity consumption in developed and developing countries: International evidence," *Energies*, vol. 15, no. 7, p. 2558, 2022.

[6] Z. Csereklyei, M. d. M. Rubio-Varas, and D. I. Stern, "Energy and economic growth: the stylized facts," *The Energy Journal*, vol. 37, no. 2, pp. 223–256, 2016.

[7] A. Zaharia, M. C. Diaconeasa, L. Brad, G.-R. Lădaru, and C. Ioanăș, "Factors influencing energy consumption in the context of sustainable development," *Sustainability*, vol. 11, no. 15, p. 4147, 2019.

[8] M. Emami Javanmard and S. Ghaderi, "Energy demand forecasting in seven sectors by an optimization model based on machine learning algorithms," *Sustainable Cities and Society*, vol. 95, p. 104623, 8 2023.

[9] S. S. Barik, S. Nayak, and T. Ranjan, "Electricity consumption and prediction using machine learning models," *ResearchGate*, 6 2020.

[10] R. V. Klyuev, I. D. Morgoev, A. D. Morgoeva, O. A. Gavrina, N. V. Martyushev, E. A. Efremenkov, and Q. Mengxu, "Methods of forecasting electric energy consumption: A literature review," *Energies*, vol. 15, no. 23, p. 8919, 2022.

[11] K. Ullah, M. Ahsan, S. M. Hasanat, M. Haris, H. Yousaf, S. F. Raza, R. Tandon, S. Abid, and Z. Ullah, "Short-term load forecasting: A comprehensive review and simulation study with cnn-lstm hybrids approach," *IEEE Access*, vol. 12, pp. 111 858–111 881, 2024.

[12] Y. Li, "Energy consumption forecasting with deep learning," *ResearchGate*, 2023.

[13] H.-x. Zhao, Y. Lee, J.-S. Chou, A. Ahmad, M. Biswas, I. Khandelwal, W. Zhang, and et al., "Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders," *Energy*, 9 2018.

[14] L. Suganthi and A. A. Samuel, "Energy models for demand forecasting—a review," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 2, p. 1223–1240, 2 2012.

[15] N. Christiansen, M. Kaltschmitt, and F. Dzukowski, "Electrical energy consumption and utilization time analysis of hospital departments and large scale medical equipment," *Energy and Buildings*, vol. 131, pp. 172–183, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778816308234

[16] Q. Qiao, A. Yunusa-Kaltungo, and R. E. Edwards, "Towards developing a systematic knowledge trend for building energy consumption prediction," *Journal of Building Engineering*, vol. 35, p. 101967, 2021.

[17] M. Matos, J. Almeida, P. Gonçalves, F. Baldo, F. J. Braz, and P. C. Bartolomeu, "A machine learning-based electricity consumption forecast and management system for renewable energy communities," *MDPI*, 1 2024.

[18] T. Carriere and G. Kariniotakis, "An integrated approach for value-oriented energy forecasting and data-driven decision-making application to renewable energy trading," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6933–6944, 2019.

[19] E. Zhou, S. Gadzanku, C. Hodge, M. Campton, S. de la Rue du Can, and J. Zhang, *Best practices in electricity load modeling and forecasting for long-term power system planning*. National Renewable Energy Laboratory, 2023.