Comparative Analysis of Machine Learning Models for ERCOT's Short-Term Load Forecasting

Gurkirat Singh

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Masters of Science
in
Computer Science and Applications

Hoda Eldardiry, Chair
Shamar L. Stewart, Co-chair
Sally Hamouda
Hongjie Chen

December 3, 2024
Blacksburg, Virginia

Keywords: Load Forecasting, Machine Learning, Forecast Evaluation, Comparative Analysis, Commodities

Copyright 2025, Gurkirat Singh

Comparative Analysis of Machine Learning Models for ERCOT's Short-Term Load Forecasting

Gurkirat Singh

(ABSTRACT)

This study investigates the efficacy of various machine learning (ML) and deep learning (DL) models for short-term load forecasting (STLF) in the Electric Reliability Council of Texas (ERCOT) grid. A dual comparative approach is employed, evaluating models based on temporal features alone as well as in combination with actual and forecasted weather variables. The research emphasizes region-specific forecasting by capturing heterogeneous load patterns for ERCOT's individual weather zones and aggregating them to predict total load. Model evaluation is conducted using accuracy and bias metrics, with particular attention to high-demand months and peak load hours. The findings reveal that Generalized Additive Models (GAM) consistently outperform other models, most importantly during summer months and peak load hours.

Comparative Analysis of Machine Learning Models for ERCOT's Short-Term Load Forecasting

Gurkirat Singh

(GENERAL AUDIENCE ABSTRACT)

This study explores the effectiveness of various machine learning (ML) and deep learning (DL) models in forecasting short-term electricity demand on the Electric Reliability Council of Texas (ERCOT) grid. A dual approach is used to compare models that rely solely on time-based data versus those that incorporate both historical and forecasted weather information. The research focuses on forecasting demand for specific regions within ERCOT, capturing unique consumption patterns in each area and then combining them for an overall load prediction. Model evaluation is conducted using accuracy and bias metrics, with particular attention to high-demand months and peak load hours. The findings reveal that Generalized Additive Models (GAM) consistently outperform other models, most importantly during summer months and peak load hours.

Dedication

I dedicate this thesis to my parents Gurtej Singh and Paramjeet Kaur, my loving brother - Kulshan Singh and my Guru - Dr. Shamar L. Stewart, all of whom have supported, nurtured and motivated me throughout my life.

Acknowledgments

I would like to express my sincere gratitude to Dr. Hoda Eldardiry and Dr. Shamar Stewart for tremendous support and guidelines throughout my graduate career. I also want to thank the rest of my committee members: Dr. Sally Hamouda, and Dr. Hongjie Chen for providing invaluable feedback. Lastly, I want to thank my friends: Prapti Singh, Sahith Kancharla and all of my mentors for constant support.

Contents

List of Figures viii

List of Tables x

1 Introduction 1

2 Review of Literature 6

3 Data 12

3.1 ERCOT's Electricity Demand Time-series 12

3.2 Seasonality Analysis of ERCOT's Electricity Demand 15

3.3 Temperature Time-series for ERCOT's Weather Zones 18

4 Forecasting Models 21

4.1 Random Forests 22

4.1.1 Decision Trees for Forecasting 22

4.1.2 Random Forest for Forecasting 23

4.2 XGBoost 25

4.2.1 XGBoost for Forecasting 25

4.3 Generalized Additive Models 29

4.3.1 GAM for Forecasting 29

4.4 Long-Short Term Memory 31

5 Evaluation Metrics **34**

5.1 Accuracy 34

5.2 Bias 36

5.3 Test for Differences Between Models 36

6 Results **38**

6.1 Accuracy 40

6.2 Bias 43

6.3 Test for Differences Between Models 46

6.4 Performance by Weather Zones 47

7 Summary and Conclusions **49**

Appendices **51**

Appendix A Supplemental Figures **52**

Bibliography **56**

List of Figures

1.1 ERCOT's Weather Zones 3

3.1 Electricity Demand across ERCOT Weather Zone 13

3.2 Autoplot, ACF, and PACF of ERCOT's Hourly Electricity Demand 15

3.3 Daily Seasonality of ERCOT's Electricity Demand 16

3.4 Monthly Seasonality of ERCOT's Electricity Demand 18

3.5 Impact of Temperature on Electricity Demand across Weather Zones 19

4.1 An Illustration of Decision Tree for Electricity Load Forecasting 23

4.2 Random Forest Architecture 24

4.3 XGBoost Architecture 26

4.4 Training Features Interpretation using GAM 31

4.5 LSTM Memory Cell Architecture 32

6.1 Radar Chart displaying all Error Metrics for All Models 39

6.2 MAE Line Plot by Month for All Models 41

6.3 MAPE Line Plot by Month for All Models 41

6.4 MAE Line Plot by Hour for All Models 42

6.5 MAPE Line Plot by Hour for All Models 42

6.6 MPE Heatmap by Hour for All Models 44

6.7 MPE Heatmap by Month for All Models 45

6.8 Diebold Mariano Test 46

6.9 MAPE by Zone for All Models 47

A.1 MSE Line Plot by Hour for All Models 53

A.2 MSE Line Plot by Month for All Models 53

A.3 RMSE Heatmap by Hour 54

A.4 RMSE Heatmap by Month 55

List of Tables

2.1 Previous Studies on short-term load forecasting for ERCOT 10

3.1 Summary Statistics of Hourly Electricity Demand Time-series 13

3.2 Dummy Regression to test for Seasonality at Hourly and Monthly levels 17

3.3 Summary Statistics of Hourly Temperatures across Weather Zones 20

4.1 Model Names and Input Features 21

6.1 Model Evaluation Metrics Summary 38

Chapter 1

Introduction

Power, as a commodity requiring the supply always to match demand without natural storage, poses a challenge for accurate short-term load forecasting (STLF). For effective grid operations and improved reliability, it is crucial to accurately predict electric load. Doing so allows for pre-positioning the system to acquire services, generation capacity, and reliably meet all load demands in a cost-effective manner. Accurate load forecasting can help minimize risks associated with power unbalancing, frequency deviations arising from high demand, power plant disconnections, and load shedding (Yang et al., 2024).

This challenge of proper forecasting is heightened by increasing climatic uncertainties, complex multi-seasonality, rapid demand fluctuations, changing population dynamics, evolving renewable resources and market volatility within an expansive grid such as the Electric Reliability Council of Texas (ERCOT) (Fan et al., 2021). Accurate forecasting is crucial for ERCOT because it helps maintain grid stability, manage supply-demand balance, and minimize costs associated with over- or under-scheduling energy resources. ERCOT manages the flow of electric power to more than 27 million Texas customers. Although the ERCOT grid is largely isolated from others and has very limited capacity for importing or exporting electricity, it does connect to surrounding regions through direct current interconnections, referred to as DC ties. These connections link ERCOT with the Southwest Power Pool (SPP), certain areas of the Eastern Interconnection, and Mexico. However, these DC ties enable only about 1% of ERCOT's total electric load to be imported or exported. This isolation restricts ERCOT's ability to balance supply and demand by accessing power from outside sources. This makes it crucial for ERCOT to manage electric reliability, especially under tight grid conditions, such as periods of high netload. To maintain stability, ERCOT

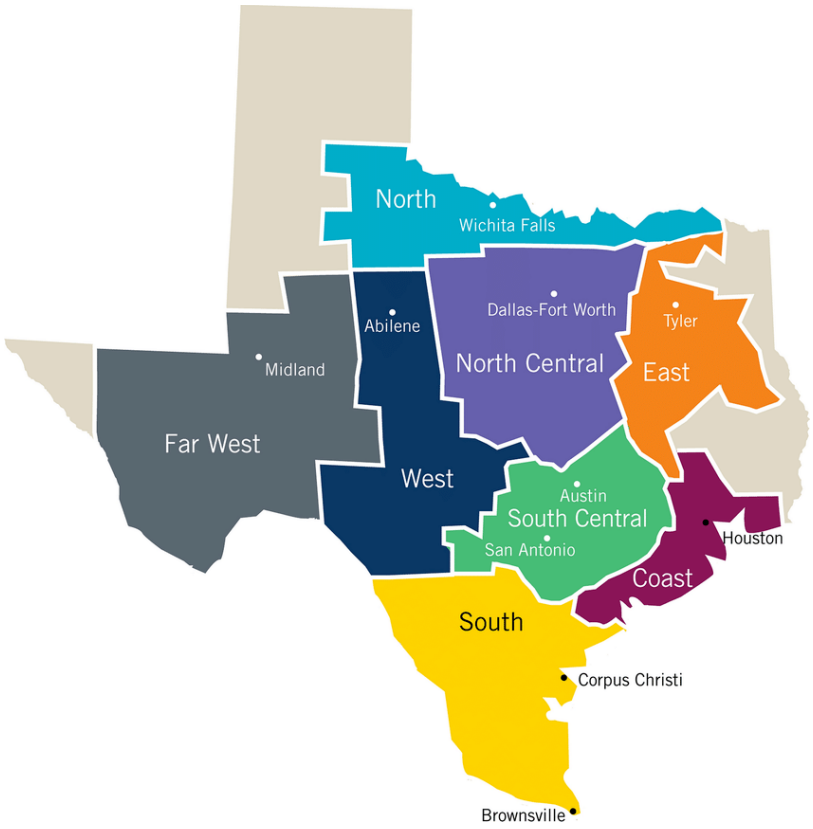
deploys ancillary services (AS) that help balance electricity flow and support grid integrity. These services ensure a stable 60 Hz frequency and provide rapid responses to unexpected shifts in supply or demand, mitigating risks that could otherwise impact the grid's stability. These real-time measures help maintain adequate capacity and supply, preventing or minimizing power blackouts. AS play a key role in maintaining ERCOT's grid stability, complementing short-term load forecasting by addressing immediate imbalances and providing backup power during sudden load fluctuations and generation losses at different locations. Ancillary services are relatively expensive than energy offerings, which makes it crucial for ERCOT to reserve and allocate them in an economical way. More ancillary services are reserved and allocated during high electricity consumption periods, hence accurately forecasting short term load is important for ERCOT to allocate ancillary resources efficiently, keeping the grid balanced in real-time.

While a number of studies (Yang et al., 2024; Rice et al., 2022; Hossain and Mahmood, 2020; Nguyen and Hansen, 2017; Aguilar Madrid and Antonio, 2021) have focused on forecasting demand in the ERCOT grid, there is a general paucity of work focused on applying and comparing Machine Learning (ML) and Deep Learning (DL) models on ERCOT's data, despite their potential to improve forecasting accuracy. ML and DL models can process large volumes of complex data, such as multi-seasonality, weather effects, and non-linear electricity consumption trends, to capture intricate relationships that traditional methods may overlook. Some studies fail to account for the role of weather (Aguilar Madrid and Antonio, 2021; Rice et al., 2022; Tarmanini et al., 2023; Gökçe and Duman, 2022), other use the historical/realized temperatures and weather variables and often ignore additional information available to the practitioner over the out of sample period. In contrast, we use full set of information available at the time of forecast such as forecasted temperature variables by location which is available 24 hours ahead of the forecasting period (Yang et al., 2024; Hossain and Mahmood, 2020; Cordeiro-Costas et al., 2023; Zhang and Jánošík, 2024; Aguilar Madrid and Antonio, 2021).

Given the isolated nature of the ERCOT grid, which limits energy exchange with other regions and makes it more susceptible to internal demand fluctuations and supply constraints, it is essential

to assess the application, performance, and robustness of state-of-the-art ML and DL forecasting models for accurate short-term load forecasting, using available information at the time, specifically designed for ERCOT's unique conditions.

Figure 1.1: ERCOT's Weather Zones



Notes:

ERCOT is divided into several weather zones: Coast, East, North, West, Far West, South, North Central (NORTHCEN) and South Central (SOUTHCEN) 1.1. Weather zones are spatially separated regions with heterogeneous demand, observed in Table 3.1, Figure 3.1), and weather characteristics as observed in Table 3.3. The relationship between electricity demand and temperature varies across different weather zones as observed in Figure 3.5. To address the complexity of forecasting in ERCOT's expansive grid, where climate and demand patterns vary significantly across regions, our study evaluates the performance of several ML (XGBoost (XGB), Random Forest (RF) and Generalized Additive Models (GAM)) and DL (Long Short-Term Memory (LSTM)) models on

ERCOT's hourly load data. GAM models, in particular, are assessed for their ability to capture local demand fluctuations by incorporating forecasted temperatures and specific weather conditions across ERCOT's various zones. By systematically comparing these models, our study highlights the importance of weather zone-based forecasting in ERCOT's context. This approach is vital, as ERCOT's demand often peaks in different zones depending on localized weather, making zone-based forecasts a key factor for accurately predicting demand across the grid.

The primary goal of this research is to compare and identify models that offer the highest robustness and accuracy for ERCOT's unique load patterns, especially during peak load hours and seasons. We adopt two comparative approaches to evaluate and identify the best models: one using only temporal features and the other combining temporal and temperature features. In both approaches, we first forecast the next hour's load for each individual weather zone and then aggregate these predictions to estimate ERCOT's total hourly load. Our findings show that GAM models, which account for both overall trends and local weather zone-specific variations, achieve superior accuracy. Specifically, the GAM (weather) model recorded the lowest error rates across metrics such as Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), with a MAPE of 0.81%. This accuracy is particularly evident during peak summer and winter months, when localized, weather-driven demand fluctuations are most pronounced. In contrast, models such as XGB and RF had higher error rates (MAPE up to 2.70% during summer months), indicating challenges in capturing ERCOT's seasonal and regional peak loads. Although the LSTM model demonstrated strong temporal pattern recognition with an MAPE of 1.03%, it did not outperform the GAM models in terms of zone-specific accuracy. When comparing models using only temporal variables with those incorporating both temporal and temperature variables, we found an overall improvement in accuracy across hours and months. However, this improvement also came with an increase in mean percentage error (MPE), indicating that the models developed a higher tendency to under-predict.

We also examine systematic biases, defined as the consistent tendency to overestimate or underestimate the actual load. Our analysis reveals that the GAM (seasonal) model achieves the most

balanced and least biased results across both hours and months, as illustrated in Figures 6.6 and 6.7, respectively. In contrast, RF and XGB models consistently show relatively large positive MPE indicating bias to underestimate the actual load for all hours and months. Lastly, we evaluated the differences in predictive accuracy between the models using the Diebold-Mariano (DM) test. The results indicate that the predictive accuracies of the models do not differ significantly from one another.

Overall, the results indicate that GAM models, with their ability to capture both local and seasonal trends using detailed weather zone data, offer ERCOT a practical solution for improving load forecasting accuracy and enhancing grid reliability during critical demand periods.

Chapter 2

Review of Literature

Electric load forecasting is categorized into several classifications based on the duration of the forecast. This classification encompasses Long-term load forecasting, which anticipates electricity demand over several years (Lindberg et al., 2019); Mid-term load forecasting, which addresses periods ranging from weeks to months (Feng and Qian, 2021); and Short-term load forecasting, which focuses on shorter intervals, typically spanning minutes to hours or days (Guo et al., 2021). This study will concentrate on a comparative analysis of machine learning techniques specifically applied to hourly short-term load forecasting.

Different theoretical approaches exist for STLF. In electric load forecasting literature, there are two main categories of forecasting methods. The first category comprises traditional statistical and econometric techniques such as auto-regressive moving average (ARMA) (Hinman and Hickey, 2009), auto-regressive integrated moving average (ARIMA) (Chodakowska et al., 2021), seasonal auto-regressive integrated moving average (SARIMA), Holt-Winters exponential smoothing (Chapagain et al., 2020), and multiple linear regression (MLR). Traditional statistical and econometric techniques are usually only able to capture linear features in the time series data and often fail to capture non-linear dependencies between the input features and target variables. These approaches often have limited flexibility in handling complex exogenous variables, such as temperature, especially when non-linear relationships exist between input variables and the target load. To overcome these limitations, ML and DL models are used. In recent years, the rapid advancements in ML and DL technologies have significantly improved STLF, a critical area in energy management and grid stability. Machine learning techniques such as XGBoost (Shabbir et al., 2023; Gökçe and Duman, 2022) and Random Forest (Fan et al., 2021; Gökçe and Duman, 2022; Nallathambi and Ramasamy,

2017) have become popular for their robustness and accuracy in predicting electricity demand, especially when handling structured data with complex seasonality patterns in the electric load. GAM provide an interpretable approach to modeling non-linear relationships in load forecasting Obst et al. (2021). Meanwhile, tools such as Prophet, developed by Facebook, have been effective in capturing seasonal trends and holidays, offering simplicity and flexibility in energy demand forecasting Guo et al. (2020). Furthermore, deep learning approaches, such as LSTM networks, Gated Recurrent Unit (GRU), and Recurrent Neural Network (RNN), have proven to be highly effective in capturing complex temporal dependencies and fluctuations in electricity load Abumohsen et al. (2023).

Numerous studies have explored the application of ML and deep learning DL techniques for STLF across various datasets. These datasets include household and building load data, regional consumption figures, and grid demand measurements. For instance, Shabbir et al. (2023) demonstrated the effectiveness of XGBoost for STLF using household data from Estonia, revealing that XGBoost outperformed both Random Forest and CatBoost algorithms. Similarly, Tarmanini et al. (2023) forecasted electric load for 709 randomly selected households in Ireland, utilizing ARIMA and Artificial Neural Networks (ANN). Their findings revealed that ANN outperformed ARIMA when dealing with non-linear load data.

Fan et al. (2021) conducted a study that hybridized grey catastrophe theory with random forests and support vector regression (SVR) to enhance STLF, focusing on data from the Australian Energy Market Operator. Furthermore, Kamoona et al. (2023) highlighted that simple machine learning methods, particularly when leveraging temperature and energy consumption as inputs to polynomial regression models, yielded more accurate predictions compared to existing baseline models in Australia's national grid demand response market.

Finally, Krstonijević (2022) proposed an adaptive load forecasting methodology based on GAM with big data estimation, tested on data from the New York Independent System Operator (NYISO). This approach dynamically selected the most significant input features for specific load series, demonstrating its adaptability and accuracy.

A comparative analysis of ML and DL algorithms is essential for understanding their respective strengths, weaknesses, and suitability for STLF. Such an analysis enables the identification of the vulnerabilities and limitations of each model, including their capacity to handle various data types, sensitivity to outliers, ability to incorporate exogenous variables, and overall performance across diverse datasets and time periods.

Numerous studies have conducted comparative analyses of traditional statistical methods alongside ML and DL algorithms to ascertain which approaches yield the "best" results in the context of STLF. This underscores the significance of ongoing research aimed at improving forecasting accuracy and reliability. For instance, Guo et al. (2021) compared the performance of three commonly used machine learning methods for STLF– support vector machine (SVM), RF, and LSTM– using real load-based simulated data. Gokce and Dunman Gökçe and Duman (2022) evaluated the performance of simple regression, RF, and XGBoost on hourly load consumption data for Turkey, concluding that the RF provided the most accurate forecasts in this context.

Solyali (2020) analyzed the performance of ANN, adaptive neural fuzzy inference system (ANFIS), MLR, and SVM, focusing on the various features in the Cyprus dataset, including temperature, humidity, Gross National Income (GNI) per Capita, population, electricity price, and generation (MW). Chaturvedi et al. (2022) conducted a comparative analysis of SARIMA, LSTM, and FB Prophet for predicting peak monthly energy demand for India and its five electrical zones. Cordeiro-Costas et al. (2023) compared the characteristics and performance of ML (SVR, RF, XGBoost) and DL (LSTM, multi-layered perceptron (MLP)) algorithms for forecasting building hourly load, employing box plots to analyze variability and biases in model forecasts. Zhang and Jánošík (2024) proposed an enhanced STLF hybrid machine learning model, integrating CatBoost and XGBoost as base models, and assessed their performance through various statistical error metrics such as RMSE, MAPE, and R-squared. Finally, Aguilar Madrid and Antonio (2021) compared the performance of k-nearest neighbors (KNN), RF, Predispatch, MLR, SVR, XGB, algorithms for weekly (168-hour ahead) load forecasting in the building sector utilizing lagged load, temporal, and weather variables (temperature and humidity) as input features for David, Santiago and Panama City, their results

highlighted XGBoost as the best-performing model and lagged load, hour of day, and temperature being most important features.

Every dataset possesses unique characteristics and load profiles that necessitate customized forecasting approaches. In the case of ERCOT, an isolated grid with a rapidly evolving mix of renewable energy sources and battery storage, load profiles can vary considerably across different weather and load zones. While some research has been conducted on applying and comparing ML and deep learning DL models to ERCOT's data, as summarized in Table 2.1, the scope remains relatively limited, and several shortcomings in the prevailing modeling approaches have been identified.

For instance, certain studies fail to adequately incorporate the impact of weather on electricity demand (for e.g., Aguilar Madrid and Antonio, 2021; Rice et al., 2022; Tarmanini et al., 2023; Gökçe and Duman, 2022). Others often use the historical/realized temperatures, overlooking additional information, such as forecasted temperatures, available during the out-of-sample period ((for e.g., Yang et al., 2024; Hossain and Mahmood, 2020; Cordeiro-Costas et al., 2023; Zhang and Jánošík, 2024; Aguilar Madrid and Antonio, 2021). Given ERCOT's distinctive characteristics and the isolated nature of its grid, it is essential to rigorously evaluate the application, performance, and robustness of advanced ML and DL forecasting models explicitly designed for ERCOT's unique load profile context.

This analysis highlights the advantages of specific models concerning parsimony– the balance between model simplicity and predictive power– as well as their, robustness in the presence of noisy or incomplete data. Additionally, certain models demonstrate a greater ability to capture inherent non-linear relationships and complex seasonal patterns within the load data. In short, different models exhibit distinct and varying capabilities in handling evolving load profiles, influenced by factors such as weather conditions, seasonal demand patterns, and population dynamics. Understanding how these algorithms perform under varying scenarios, including their adaptability to new data and changing patterns, is crucial for selecting the most "appropriate" forecasting technique.

In this work, two different comparative analysis approaches are used to determine the performance and robustness of different ML (XGBoost, RF, GAM) and DL (LSTM) algorithms. The first

Table 2.1: Previous Studies on short-term load forecasting for ERCOT

	Author	Input(s)	Forecasting Methods	Findings/Results	Period
1.	Yang et al. (2024)	Previous Load, Time, Weather (Temperature, Wind, SWRad, LWRad)	Comparative analysis of Support Vector Machines (SVR), Long-Short-Term Memory & Long-Term Recurrent Convolutional Network (LRCN) with combinations of different time/weather variables. Forecasting 1 hour ahead.	1. SVM is the best-performing model, with LSTM second-best 2. LSTM achieves the best result using Load, Hour, Month, and Temp. 3. Short-term features such as Hour and Temp provide unnecessary redundancy and may even decrease accuracy in models.	2011-2021 (Hourly)
2.	Hossain and Mahmood (2020)†	Previous Load, Time, Holiday, Weather (Temperature, Relative Humidity, Pressure)	Two forecasting models using LSTM: 1. The first model predicts a single step ahead load 2. Other model predicts multi-step intraday rolling horizons (12, 24 hours etc)	1. LSTM outperforms Generalized Regression Neural Networks (GRNNs) and Extreme Learning Machines (ELMs) 2. It is shown that the highest accuracy can be achieved at a certain sequence length, where increasing the size of the look-back window further reduces accuracy (window size =24)	2012-2015 (Hourly)
3.	Nguyen and Hansen (2017)	Time	Statistical approach in modeling load, used ARIMA and SARIMA. Forecasting 1 hour ahead	1. ARIMA outperforms SARIMA when the smaller dataset is used as it captures the weekly seasonality more effectively, but SARIMA outperforms ARIMA with a bigger dataset i.e. the full dataset. 2. ACF and PACF suggest that a correlation recorded at the same hour is still not accounted for.	2002-2016 (Hourly)
4.	Rice et al. (2022)	Previous Load, Time	Ensemble approach using LASSO and ridge regressions for 1-41 hour ahead load forecasting. Experimented with 3 forecasting horizons: real-time forecasting: t+1-t+17, Day-Ahead Markets (DAM) forecasting t+18-t+41, and total t+1-t+41. Models were updated every month when used for forecasting the next month.	1. Models outperformed ERCOT's models over all 3 forecasting horizons for test data (calendar year 2020). 2. Tested for alternative forecasting approaches, i.e., ARIMA & Prophet against ERCOT, yielded poorer performance.	2017 - 2021 (Hourly)

Notes: † indicates that this study was conducted for ERCOT West. All others were conducted at the total grid level. SWRad and LWRad refer to Short-Wave Radiation and Long-Wave Radiation, respectively. SWRad and LWRad are important meteorological variables that impact weather and temperature patterns.

approach compared the models using only temporal features as inputs. In the second approach, the models are trained on temporal variables, and actual temperatures and historical forecasted temperatures are added, extracted from the Weather Assumptions report from ERCOT, into the feature input set to test the ability of the models to generalize the load forecasts, and to avoid look-ahead bias. Lastly, both approaches were weather zone based forecasting techniques, meaning first the electric load of each zone was forecasted and then aggregated to find total electric load for ERCOT. This approach enabled the models to capture region based patterns and relationships.

Chapter 3

Data

We utilized two primary datasets for our analysis, the first being ERCOT's hourly electricity demand data, which is publicly accessible on the U.S. Energy Information Administration (EIA) website. This dataset includes detailed hourly data from January 1, 2019, to January 31, 2024 (44,544 timestamps), capturing both total electricity demand for the ERCOT region and demand segmented by various regions in Texas, including the Coast, South, West, North, East, North Central, South Central, and Far West areas. This comprehensive data allows for an in-depth examination of electricity usage patterns across the state. The second dataset comprises of hourly temperature data, which is obtained from the publicly available "Weather Assumptions" report provided by ERCOT. This dataset includes both historical hourly realized temperatures and 24-hour-ahead forecasted temperatures for all ERCOT weather zones, spanning from January 1, 2019, to January 31, 2024. This data allows us to obtain the relationships between temperature and electricity demand across weather zones.

3.1 ERCOT's Electricity Demand Time-series

The summary statistics presented for ERCOT and its weather zones in Table 3.1 provide a detailed understanding of the distribution of electricity demand across various regions. The mean values highlight the average electricity demand over the given period, with ERCOT's overall hourly average being 46,512.98 MW. Among the individual weather zones, the NORTHCEN region exhibits the highest average demand (13,911.43 MW), followed by the COAST zone (12,709.53 MW), while the NORTH zone has the lowest average demand (1,142.95 MW). This indicates that the magni-

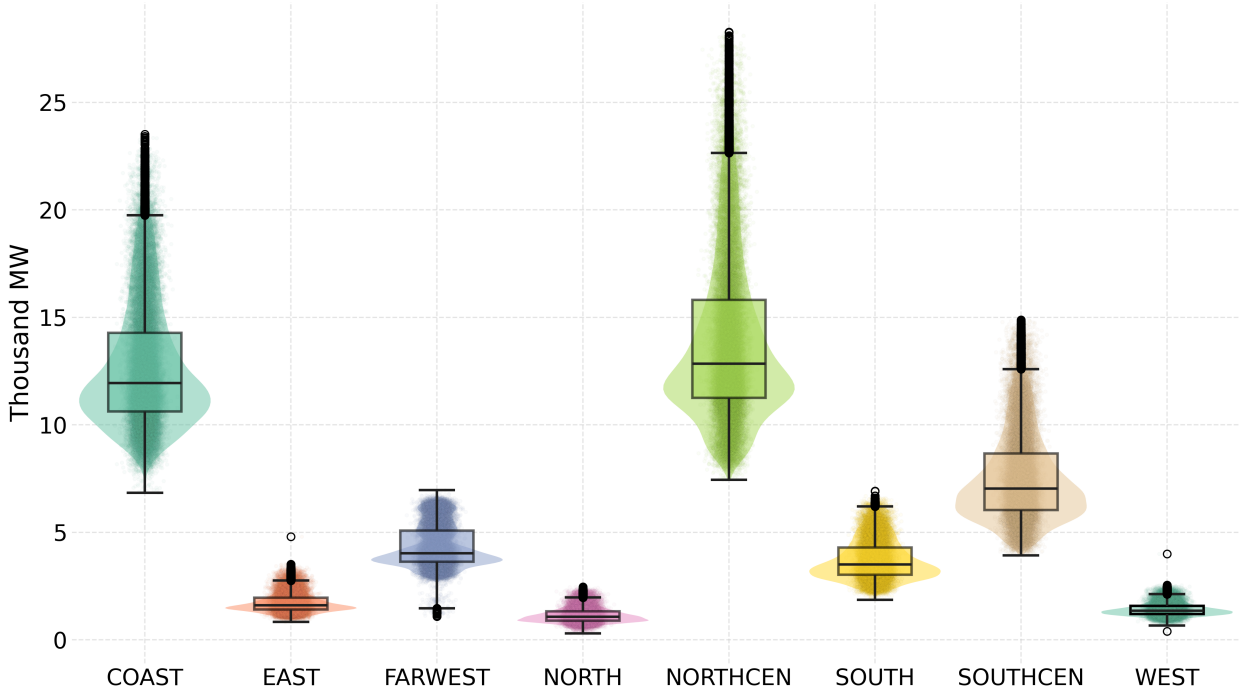
Table 3.1: Summary Statistics of Hourly Electricity Demand Time-series

Statistic	ERCOT	COAST	EAST	FARWEST	NORTH	NORTHCEN	SOUTH	SOUTHCEN	WEST
Mean	46512.98	12709.53	1720.62	4346.73	1142.95	13911.43	3718.14	7547.79	1417.41
Minimum	27449.00	6844.00	829.00	1087.00	300.00	7439.00	1860.00	3924.00	400.00
Maximum	85432.00	23516.00	4800.00	6963.00	2467.00	28269.00	6924.00	14887.00	4000.00
Standard Deviation	10636.19	2883.16	431.98	971.59	332.20	3841.90	911.52	2071.25	295.61
Coefficient of Variation (CV)	0.23	0.23	0.25	0.22	0.29	0.28	0.25	0.27	0.21
Skewness	0.99	0.94	0.98	0.60	0.91	1.06	0.72	0.95	0.84
Kurtosis	0.57	0.32	0.65	-0.57	0.65	0.68	-0.19	0.44	0.54

Notes: The table above provides summary statistics for ERCOT and its various Weather Zones, detailing measures such as the mean, minimum, maximum, standard deviation, coefficient of variation, skewness, and kurtosis. Here, NORTHCEN and SOUTHCEN denote the North Central and South Central zones, respectively. The CV is calculated as the ratio of the standard deviation to the mean, allowing for a relative comparison of variability across different zones. This metric effectively illustrates the extent of variability in electricity demand relative to the mean, making it useful for understanding demand consistency within each weather zone.

tude of electricity demand varies significantly across zones, influenced by factors such as population density, industrial activity, and climate conditions specific to each zone. The minimum, maximum,

Figure 3.1: Electricity Demand across ERCOT Weather Zone



Notes: This figure presents violin and box plots depicting the distribution of electricity demand across various ERCOT weather zones, with distributions showing a right-skewed pattern in all zones. The labels NORTHCEN and SOUTHCEN represent the North Central and South Central weather zones, respectively.

and standard deviation values collectively highlight the variability in electricity demand across ERCOT and its weather zones, as depicted in the box plot shown in Figure 3.1. ERCOT's hourly demand ranges from 27,449 MW at its lowest to 85,432 MW at its peak, showcasing substantial variability in magnitude. Among the weather zones, NORTHCEN and COAST experience the highest maximum demands (28,269 MW and 23,516 MW, respectively), which can be attributed to high population densities or industrial activities in these areas. In contrast, zones such as NORTH and WEST exhibit relatively low maximum demands (2,467 MW and 4,000 MW, respectively), reflecting smaller grid requirements during peak periods. This variability in magnitude, as illustrated in Figure 3.1, suggests that different zones experience unique patterns of electricity demand, likely influenced by factors such as regional demographics, industrialization, and weather patterns.

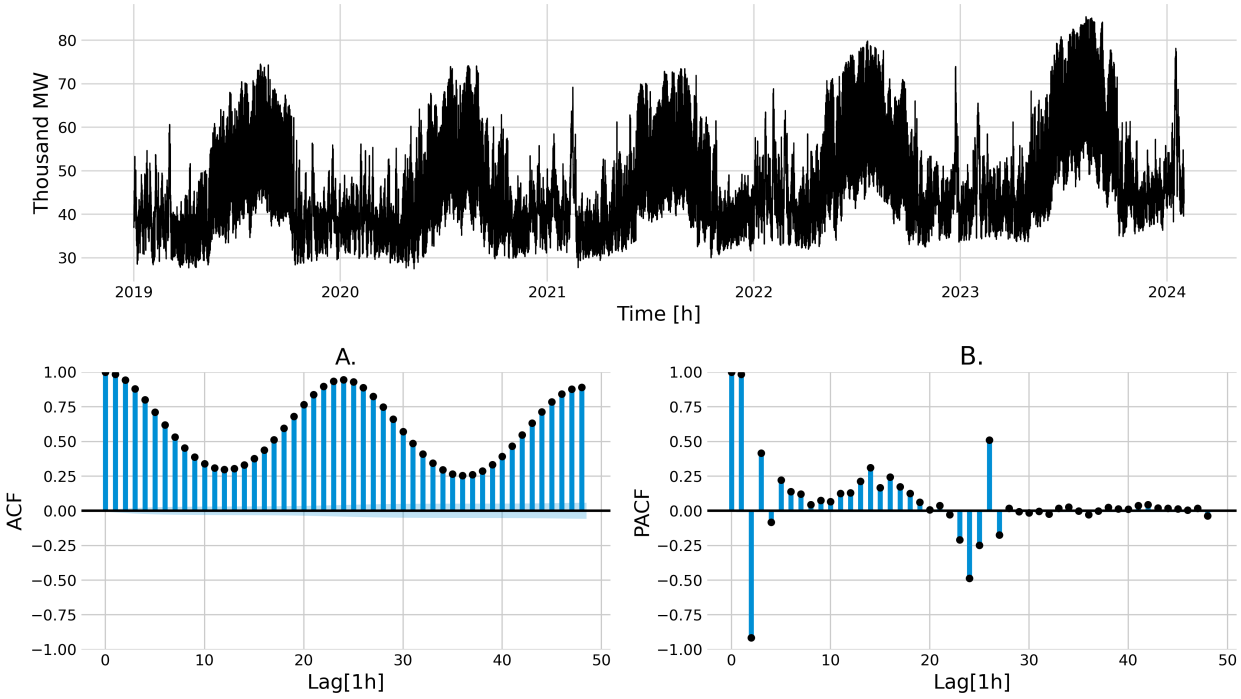
The Coefficient of Variation (CV) provides a standardized measure of demand variability relative to the mean across zones. In essence, zones with higher CV values, such as NORTH and NORTHCEN, reflect a larger degree of demand volatility, reinforcing the importance of capturing these fluctuations through zone-specific load forecasting. The lower CVs in FARWEST and WEST suggest these regions experience less relative variability in demand, which might be influenced by renewable energy smoothing demand peaks. The combination of these metrics and the box plot representation illustrates both the magnitude of variability and the dynamic nature of electricity demand across different zones, reinforcing the importance of zone-based load forecasting to account for these regional differences and capture non-linear patterns in demand.

Skewness and kurtosis offer insights into the distribution shapes of the demand data. Most zones exhibit positive skewness, meaning that demand is skewed towards higher values, with NORTHCEN (1.06) being the most skewed. The positive skewness indicates that although lower demand values are more common, there are instances of exceptionally high demand. Kurtosis values are low across the zones, indicating that the demand distributions are relatively flat compared to a normal distribution, with FARWEST showing negative kurtosis (-0.57), suggesting fewer extreme values than expected, likely due to significant increases in solar and wind build-outs, which could smooth out overall demand patterns and lead to fewer extreme fluctuations in the short term.

In summary, the statistical analysis reveals significant heterogeneity in electricity demand across ERCOT's zones, with the COAST and NORTHCEN zones being dominant in terms of both average demand and variability. Meanwhile, zones such as FARWEST and SOUTH are exhibiting different patterns due to renewable energy build-outs. This variability underscores the importance of region-specific electricity demand forecasting and ensuring grid stability across the ERCOT system.

3.2 Seasonality Analysis of ERCOT's Electricity Demand

Figure 3.2: Autoplot, ACF, and PACF of ERCOT's Hourly Electricity Demand

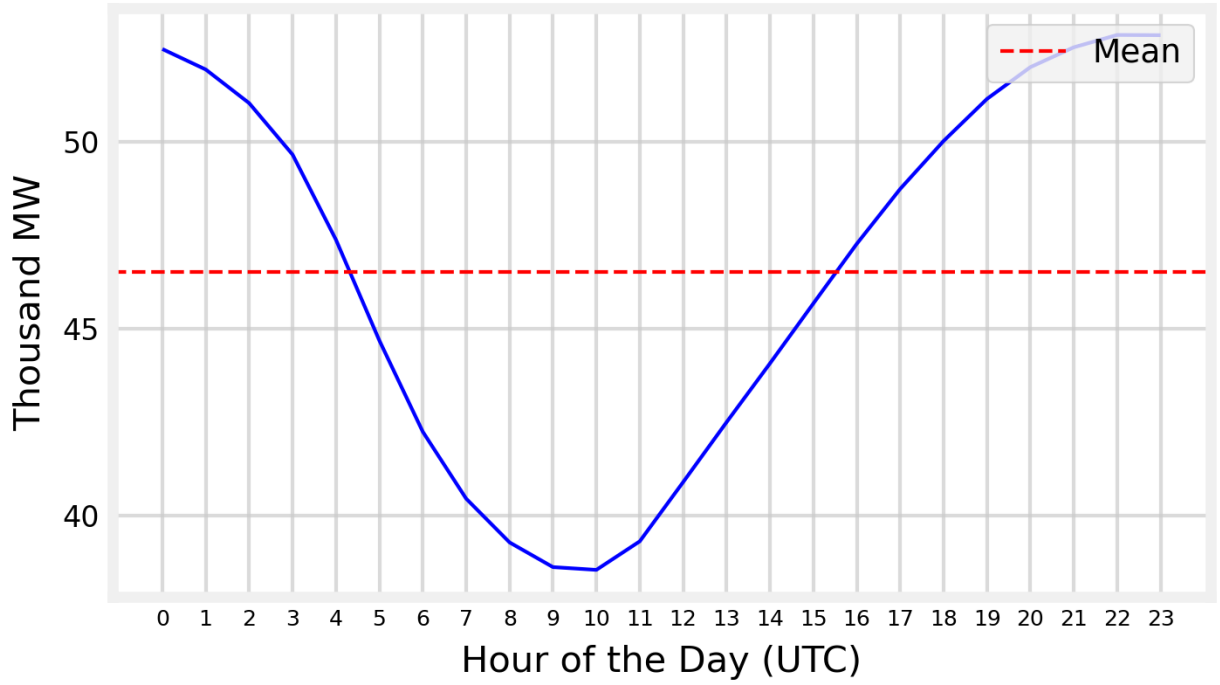


Notes: This figure displays ERCOT's electricity demand time series along with its ACF and PACF plots. The time series plot reveals an annual seasonal pattern, while the ACF plot shows a repeating pattern every 24 lags, indicating daily seasonality. This daily seasonality is further supported by the PACF plot, where the first 24 lags are significant, reinforcing a 24-lag seasonal structure.

The time series, ACF, and PACF plots in Figure 3.2 further support the findings from the summary statistics, offering a more detailed view of the temporal structure of ERCOT's electricity demand. The time series plot clearly shows strong seasonal trends, with higher demand during summer peaks.

The autocorrelation function (ACF) plot confirms the presence of a daily seasonality, showing strong positive correlations at 24-hour intervals, indicating repeated demand patterns across days. This observation is further supported by the PACF plot, which shows that the first 24 lags are significant. Beyond lag 24, there is a noticeable decay and all subsequent lags are statistically insignificant. Furthermore, we ranked the significance of lags using "Feature Importance" in tree-based models (XGBoost and RF), and found that Lag 1, Lag 2 and Lag 24 capture the majority of the importance, emphasizing that recent demand values and daily patterns (24-hour intervals) are the most critical predictors of electricity demand.

Figure 3.3: Daily Seasonality of ERCOT's Electricity Demand



Notes: This figure illustrates the daily profile of electricity demand, highlighting peak usage during the evening hours. Specifically, demand reaches its highest at hour ending 22, 23, 0, and 1 UTC, which align with 5 PM, 6 PM, 7 PM, and 8 PM Central Time, respectively.

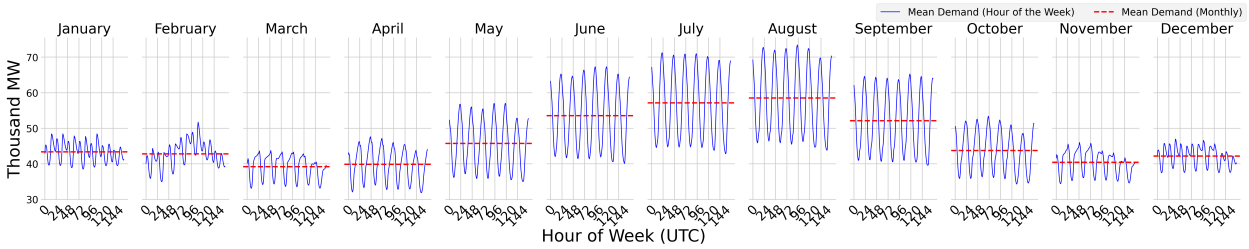
Figures 3.3 and 3.4 further highlight these seasonal patterns at multiple time scales. The hourly electricity demand plot captures the daily oscillations, with the lowest demand occurring during mid-morning hours in UTC time zone and the peak demand rising in the evening. The monthly

Table 3.2: Dummy Regression to test for Seasonality at Hourly and Monthly levels

	ERCOT	COAST	EAST	FARWEST	NORTH	NORTHCEN	SOUTHCEN	SOUTH	WEST
const	49298.83*** (180.49)	12441.01*** (43.38)	1927.51*** (8.24)	4486.27*** (25.66)	1281.38*** (7.16)	15669.60*** (67.46)	8182.30*** (36.92)	3760.73*** (14.99)	1560.80*** (6.30)
Hour_1	−547.40** (217.57)	−225.05*** (52.29)	−17.41* (9.94)	5.36 (30.93)	−14.04 (8.64)	−165.07** (81.32)	−49.93 (44.50)	−66.63*** (18.08)	−3.42 (7.59)
Hour_2	−1446.82*** (217.57)	−467.02*** (52.29)	−52.06*** (9.94)	9.93 (30.93)	−35.90*** (8.64)	−492.15*** (81.32)	−229.87*** (44.50)	−137.40*** (18.08)	−25.81*** (7.59)
Hour_3	−2822.27*** (217.57)	−834.67*** (52.29)	−108.36*** (9.94)	8.76 (30.93)	−66.52*** (8.64)	−961.28*** (81.32)	−513.44*** (44.50)	−250.16*** (18.08)	−56.43*** (7.59)
Hour_4	−5100.18*** (217.57)	−1404.27*** (52.29)	−204.60*** (9.94)	−18.62 (30.93)	−121.89*** (8.64)	−1764.50*** (81.32)	−971.76*** (44.50)	−439.76*** (18.08)	−114.66*** (7.59)
Hour_5	−7794.86*** (217.57)	−2062.34*** (52.29)	−318.33*** (9.94)	−60.33* (30.93)	−185.62*** (8.64)	−2718.40*** (81.32)	−1512.96*** (44.50)	−665.98*** (18.08)	−187.25*** (7.59)
Hour_6	−10233.09*** (217.57)	−2672.94*** (52.29)	−418.90*** (9.94)	−101.58*** (30.93)	−241.24*** (8.64)	−3582.67*** (81.32)	−2016.13*** (44.50)	−888.10*** (18.08)	−254.24*** (7.59)
Hour_7	−12023.67*** (217.57)	−3131.01*** (52.29)	−488.36*** (9.94)	−134.03*** (30.93)	−279.33*** (8.64)	−4209.22*** (81.32)	−2389.04*** (44.50)	−1061.49*** (18.08)	−300.72*** (7.59)
Hour_8	−13193.50*** (217.57)	−3433.45*** (52.29)	−531.15*** (9.94)	−153.92*** (30.93)	−302.39*** (8.64)	−4615.65*** (81.32)	−2630.46*** (44.50)	−1181.39*** (18.08)	−330.10*** (7.59)
Hour_9	−13853.86*** (217.57)	−3599.49*** (52.29)	−553.87*** (9.94)	−169.22*** (30.93)	−313.09*** (8.64)	−4839.98*** (81.32)	−2769.07*** (44.50)	−1259.50*** (18.08)	−348.55*** (7.59)
Hour_10	−13927.80*** (217.57)	−3610.81*** (52.29)	−552.41*** (9.94)	−176.98*** (30.93)	−313.62*** (8.64)	−4849.07*** (81.32)	−2794.70*** (44.50)	−1291.61*** (18.08)	−355.58*** (7.59)
Hour_11	−13164.27*** (217.57)	−3412.86*** (52.29)	−515.98*** (9.94)	−174.26*** (30.93)	−294.08*** (8.64)	−4550.64*** (81.32)	−2663.21*** (44.50)	−1254.19*** (18.08)	−342.74*** (7.59)
Hour_12	−11579.78*** (217.57)	−3057.04*** (52.29)	−435.93*** (9.94)	−152.09*** (30.93)	−251.86*** (8.64)	−3944.74*** (81.32)	−2358.17*** (44.50)	−1138.03*** (18.08)	−300.03*** (7.59)
Hour_13	−9978.28*** (217.57)	−2709.95*** (52.29)	−358.25*** (9.94)	−131.61*** (30.93)	−209.70*** (8.64)	−3309.48*** (81.32)	−2038.96*** (44.50)	−1024.04*** (18.08)	−252.33*** (7.59)
Hour_14	−8406.61*** (217.57)	−2280.15*** (52.29)	−287.27*** (9.94)	−116.88*** (30.93)	−166.45*** (8.64)	−2731.54*** (81.32)	−1764.92*** (44.50)	−900.52*** (18.08)	−213.63*** (7.59)
Hour_15	−6803.05*** (217.57)	−1747.44*** (52.29)	−226.18*** (9.94)	−91.39*** (30.93)	−124.48*** (8.64)	−2221.09*** (81.32)	−1525.36*** (44.50)	−736.99*** (18.08)	−186.07*** (7.59)
Hour_16	−5209.02*** (217.57)	−1191.86*** (52.29)	−168.52*** (9.94)	−66.29** (30.93)	−92.33*** (8.64)	−1734.11*** (81.32)	−1290.56*** (44.50)	−555.95*** (18.08)	−163.69*** (7.59)
Hour_17	−3741.18*** (217.57)	−688.21*** (52.29)	−120.64*** (9.94)	−37.89 (30.93)	−67.86*** (8.64)	−1296.71*** (81.32)	−1053.60*** (44.50)	−381.70*** (18.08)	−143.92*** (7.59)
Hour_18	−2459.22*** (217.57)	−283.80*** (52.29)	−78.41*** (9.94)	−14.32 (30.93)	−46.10*** (8.64)	−913.90*** (81.32)	−821.53*** (44.50)	−225.05*** (18.08)	−120.06*** (7.59)
Hour_19	−1333.94*** (217.57)	53.21 (52.29)	−37.86*** (9.94)	3.19 (30.93)	−25.18*** (8.64)	−580.66*** (81.32)	−598.82*** (44.50)	−92.04*** (18.08)	−93.29*** (7.59)
Hour_20	−483.91** (217.57)	283.51*** (52.29)	−12.53 (9.94)	16.11 (30.93)	−7.82 (8.64)	−331.31*** (81.32)	−407.46*** (44.50)	10.51 (18.08)	−65.11*** (7.59)
Hour_21	46.52 (217.57)	392.66*** (52.29)	1.16 (9.94)	17.41 (30.93)	2.31 (8.64)	−169.92** (81.32)	−257.49*** (44.50)	76.39*** (18.08)	−38.75*** (7.59)
Hour_22	377.19* (217.57)	396.21*** (52.29)	6.79 (9.94)	21.45 (30.93)	8.95 (8.64)	−37.85 (81.32)	−124.16*** (44.50)	105.34*** (18.08)	−16.00** (7.59)
Hour_23	371.02* (217.57)	257.25*** (52.29)	5.67 (9.94)	14.44 (30.93)	9.02 (8.64)	30.34 (81.32)	−25.07 (44.50)	79.22*** (18.08)	−4.05 (7.59)
Month_2	−553.98*** (151.25)	−151.40*** (36.35)	−3.46 (6.91)	−498.51*** (21.50)	−58.11*** (6.00)	34.01 (56.53)	−7.25 (30.93)	76.24*** (12.57)	53.30*** (5.28)
Month_3	−4122.66*** (147.32)	−80.28** (35.41)	−253.79*** (6.73)	−409.87*** (20.94)	−202.96*** (5.85)	−2239.79*** (55.07)	−769.35*** (30.13)	−24.35** (12.24)	−142.80*** (5.14)
Month_4	−3489.87*** (148.65)	384.06*** (35.73)	−264.40*** (6.79)	−357.37*** (21.13)	−231.70*** (5.90)	−2476.21*** (55.56)	−600.36*** (30.40)	198.83*** (12.35)	−142.27*** (5.19)
Month_5	2379.98*** (147.32)	2368.22*** (35.41)	−76.33*** (6.73)	−286.70*** (20.94)	−129.63*** (5.85)	−677.63*** (55.07)	524.15*** (30.13)	688.52*** (12.24)	−30.47*** (5.14)
Month_6	10184.25*** (148.65)	4130.86*** (35.73)	240.34*** (6.79)	−64.76*** (21.13)	135.56*** (5.90)	2522.50*** (55.56)	1926.87*** (30.40)	1126.86*** (12.35)	167.89*** (5.19)
Month_7	13785.12*** (147.32)	4581.69*** (35.41)	396.33*** (6.73)	141.91*** (20.94)	266.02*** (5.85)	4313.95*** (55.07)	2576.16*** (30.13)	1278.66*** (12.24)	231.24*** (5.14)
Month_8	15168.37*** (147.32)	4983.83*** (35.41)	456.49*** (6.73)	178.88*** (20.94)	295.16*** (5.85)	4717.01*** (55.07)	2861.55*** (30.13)	1402.87*** (12.24)	273.19*** (5.14)
Month_9	8763.03*** (148.65)	3440.74*** (35.73)	174.11*** (6.79)	102.74*** (21.13)	121.51*** (5.90)	1976.98*** (55.56)	1827.09*** (30.40)	1018.49*** (12.35)	102.93*** (5.19)
Month_10	419.21*** (147.32)	1320.80*** (35.41)	−141.75*** (6.73)	17.49 (20.94)	−114.72*** (5.85)	−1247.34*** (55.07)	223.06*** (30.13)	469.89*** (12.24)	−107.44*** (5.14)
Month_11	−2897.19*** (148.65)	−20.47 (35.73)	−188.95*** (6.79)	43.77** (21.13)	−116.39*** (5.90)	−1971.00*** (55.56)	−514.36*** (30.40)	−20.73* (12.35)	−109.60*** (5.19)
Month_12	−1139.51*** (147.32)	176.26*** (35.41)	−86.64*** (6.73)	157.46*** (20.94)	−64.27*** (5.85)	−1059.39*** (55.07)	−190.60*** (30.13)	−18.22 (12.24)	−53.61*** (5.14)
R-squared	0.61	0.69	0.51	0.06	0.37	0.58	0.57	0.64	0.39
R-squared Adj.	0.61	0.69	0.51	0.06	0.37	0.58	0.57	0.64	0.39
F-statistic	2064.69	2982.57	1358.09	84.23	779.77	1841.54	1749.26	2280.92	832.01
Log-Likelihood	-455132.46	-391626.33	-317663.39	-368231.13	-311407.76	-411297.62	-384439.74	-344308.49	-305659.77
AIC	910334.93	783322.67	635396.77	736532.27	622885.52	822665.24	768949.49	688686.99	611389.53
BIC	910639.58	783627.31	635701.42	736836.92	623190.17	822969.89	769254.13	688991.64	611694.18

Notes: The table displays the results of our seasonality tests conducted through dummy variable regression. The hourly indicators, denoted as $Hour\_\#$, reflect deviations from the reference hour 0, while the monthly indicators, $Month\_\#$, signify deviations from the baseline month of January (month 1). The respective standard errors are presented in parentheses. *, **, and *** reflect statistical significance at the 10%, 5%, and 1% levels, respectively. NORTHCEN and SOUTHCEN represent North Central and South Central zones, respectively.

Figure 3.4: Monthly Seasonality of ERCOT's Electricity Demand



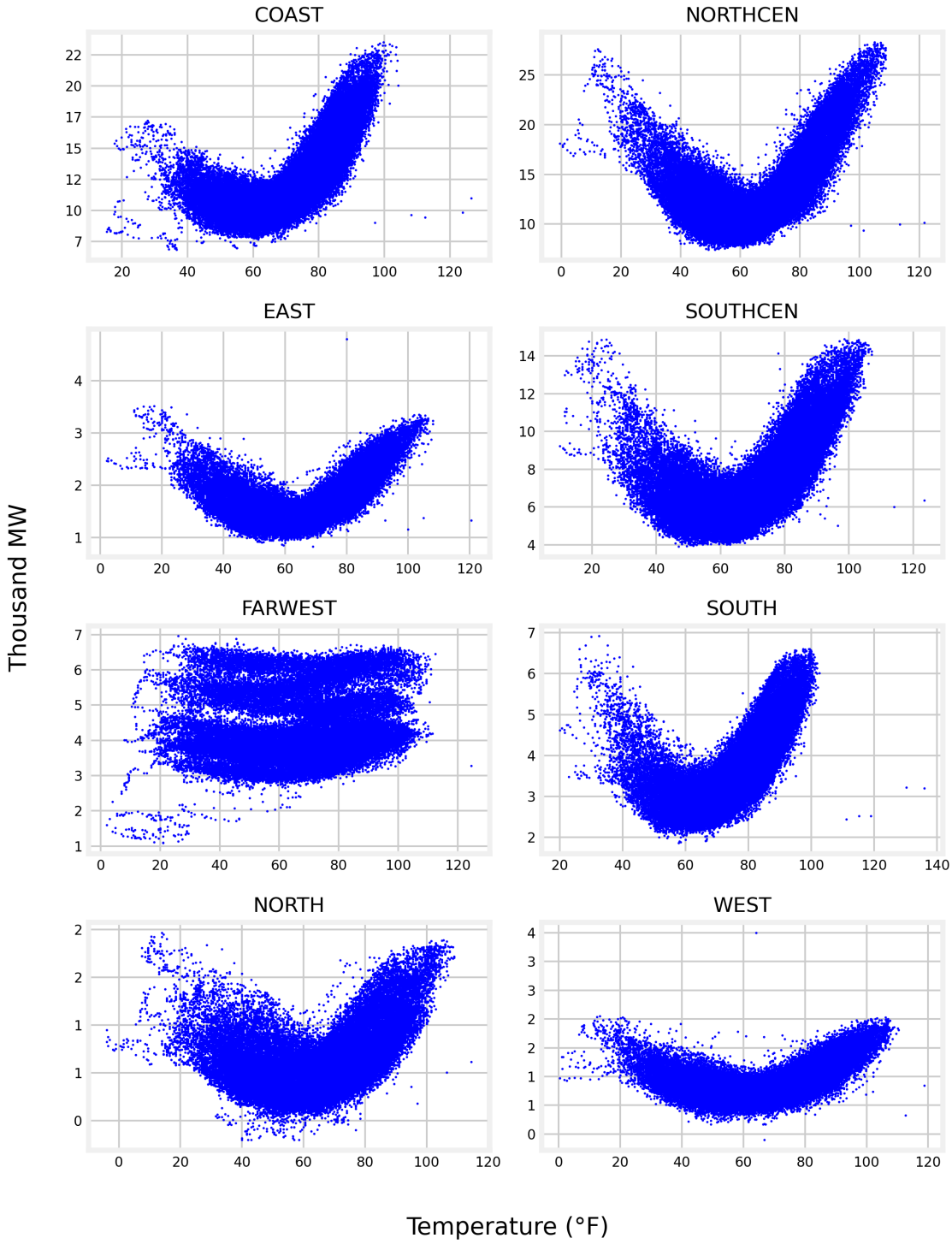
Notes: This figure shows the weekly variations in electricity demand by month, with peak demand occurring prominently in summer. Notably, winter months (December, January, and February) exhibit dual daily peaks, driven by increased demand for heating during both early morning and late evening hours. Conversely, summer months display a single daily peak, reflecting the demand for cooling concentrated in the afternoon and evening. These daily and weekly load profiles vary month-to-month in ERCOT, driven by shifts in weather patterns.

profiles reveal how demand fluctuates across months, with July and August showing the largest peaks, reflecting the summer heat and increased air conditioning usage. Conversely, the demand in winter months has two peaks of load demand during the day (in the morning and evening) due to electricity demand from heating, though with noticeable fluctuations in January and February due to holiday-related factors and weather anomalies. To formally test the seasonality, a dummy regression was conducted, and results are presented in Table 3.2. Significant coefficients for Hourly and Monthly dummies at 10%, 5%, and 1% significance levels indicate a strong presence of multi-seasonality at different time intervals across all zones. Thus, the various demand time series are non-stationary, as they exhibit pronounced seasonal effects that vary across different time scales. These combined insights from Figures 3.2, 3.3 and 3.4 and dummy regression results from Table 3.2 emphasize the pronounced seasonal dependencies in ERCOT's electricity demand, driven by both daily and monthly seasonal cycles across different weather zones.

3.3 Temperature Time-series for ERCOT's Weather Zones

Temperature variability across zones plays a critical role in influencing electricity demand, as fluctuations in temperature drive changes in heating and cooling requirements, leading to significant variations in load patterns and grid management challenges. From the Table 3.3, FARWEST (0.29),

Figure 3.5: Impact of Temperature on Electricity Demand across Weather Zones



Notes: This figure illustrates the impact of temperature on electricity demand across weather zones. Most zones exhibit a parabolic pattern, reflecting increased electricity demand driven by cooling needs at high temperatures and heating requirements at low temperatures. NORTHCEN and SOUTHCEN refer to North Central and South Central weather zones respectively.

Table 3.3: Summary Statistics of Hourly Temperatures across Weather Zones

Statistic	COAST	EAST	FARWEST	NORTH	NORTHCEN	SOUTHCEN	SOUTH	WEST
Mean	71.36	67.27	65.58	63.71	66.80	69.46	74.31	66.63
Minimum	15.30	2.00	2.00	-4.00	-0.50	10.00	20.00	0.40
Maximum	126.50	120.50	124.50	114.50	121.75	123.50	136.00	118.80
Standard Deviation	14.04	16.47	19.31	18.31	17.98	16.22	13.21	18.48
Coefficient of Variation (CV)	0.20	0.24	0.29	0.29	0.27	0.23	0.18	0.28
Skewness	-0.55	-0.37	-0.20	-0.23	-0.29	-0.41	-0.65	-0.27
Kurtosis	-0.21	-0.41	-0.66	-0.57	-0.53	-0.31	0.16	-0.52

NORTH (0.29) and WEST (0.28) exhibit the highest CVs, indicating relatively greater temperature variability compared to their mean values. In contrast, COAST (0.20) and SOUTH (0.18) show the lowest CVs, reflecting more stable temperatures. The remaining zones, such as EAST (0.24), NORTHCEN (0.27), and SOUTHCEN (0.23), fall in the mid-range of variability. This suggests that while temperature variability exists across all zones, FARWEST, NORTH, and WEST stand out for their significantly higher relative variability, whereas COAST and SOUTH demonstrate more consistent temperature patterns.

The relationship between temperature and electricity demand across ERCOT's weather zones is critical for understanding the demand behavior. Figure 3.5 illustrates the parabolic relationship between electricity demand and temperature across eight zones. In most zones, there is a distinct parabolic pattern, highlighting the significant role that weather conditions play in driving demand. Demand increases sharply at both lower and higher temperature extremes. This trend is particularly evident in zones such as COAST, NORTHCEN, and SOUTHCEN, where peak demand occurs at very high temperatures during the summer months due to cooling needs and at low temperatures during the winter months due to heating requirements. These zones experience significant demand fluctuations because of their large populations. Zones such as FARWEST show a more varied and damped pattern, likely reflecting long-term demand growth trend due to more renewable generation being installed. Strong correlation between temperature and demand emphasizes the necessity of incorporating detailed weather forecasts into demand forecasting models, as failing to account for weather fluctuations could result in inaccurate load predictions.

Chapter 4

Forecasting Models

Table 4.1: Model Names and Input Features

Model	Algorithm	Input Features	Annotation
Model 1	RF	Hour, Weekday, Hour of the Week, Week of the Year, Month, Prev. Load (Lag 1, Lag 2, Lag 24), Trend, Hourly Temperature, Max Temp., and Min Temp.	RF (weather)
Model 2	XGB	Hour, Weekday, Hour of the Week, Week of the Year, Month, Prev. Load (Lag 1, Lag 2, Lag 24), Trend, Hourly Temperature, Max Temp., and Min Temp.	XGB (weather)
Model 3	GAM	Hour, Weekday, Hour of the Week, Week of the Year, Month, Prev. Load (Lag 1, Lag 2, Lag 24), Trend, Hourly Temperature, Max Temp, and Min Temp.	GAM (weather)
Model 4	LSTM	Timestamp and Prev. Load ($n = 72$)	LSTM
Model 5	GAM	Hour, Weekday, Hour of the Week, Week of the Year, Month, Prev. Load (Lag 1, Lag 2, Lag 24), Trend	GAM (seasonal)
Model 6	RF	Hour, Weekday, Hour of the Week, Week of the Year, Month, Prev. Load (Lag 1, Lag 2, Lag 24), Trend	RF (seasonal)
Model 7	XGB	Hour, Weekday, Hour of the Week, Week of the Year, Month, Prev. Load (Lag 1, Lag 2, Lag 24), Trend	XGB (seasonal)

Notes: The table presents a mapping of model names, annotations, and the corresponding input features. "Prev. Load" refers to the realized electricity demand at the specified lagged timestamp. "Max Temp." and "Min Temp." represent the day's maximum and minimum temperatures (°F), respectively. The parameter n indicates the lookback window size for the LSTM model. The LSTM model considers the realized electricity demand over the past 72 hours to predict the demand for the next timestamp.

In this section, we describe the forecasting methods used to predict electricity demand. To ensure a robust analysis, we implemented a diverse set of models, including machine learning approaches such as RF, XGBoost, and GAM, and a deep learning model (LSTM). These models produced accurate results in previous studies as presented in the literature section and were selected for

their distinct capabilities in handling seasonality, and time dependencies in the data. Each model was trained and evaluated using same set of input features designed to capture temporal, weather-related, and historical demand patterns. Table 4.1 provides a detailed mapping of the models, their respective input features, and annotations. The input features were engineered based on findings from the literature, as summarized in Table 2.1, alongside insights from the autocorrelation function (ACF) and partial autocorrelation function (PACF) analyses shown in Figure 3.2 (to determine the number of lags of previous load). Feature importance graphs were also utilized to ensure that only the most significant features were included in the models.

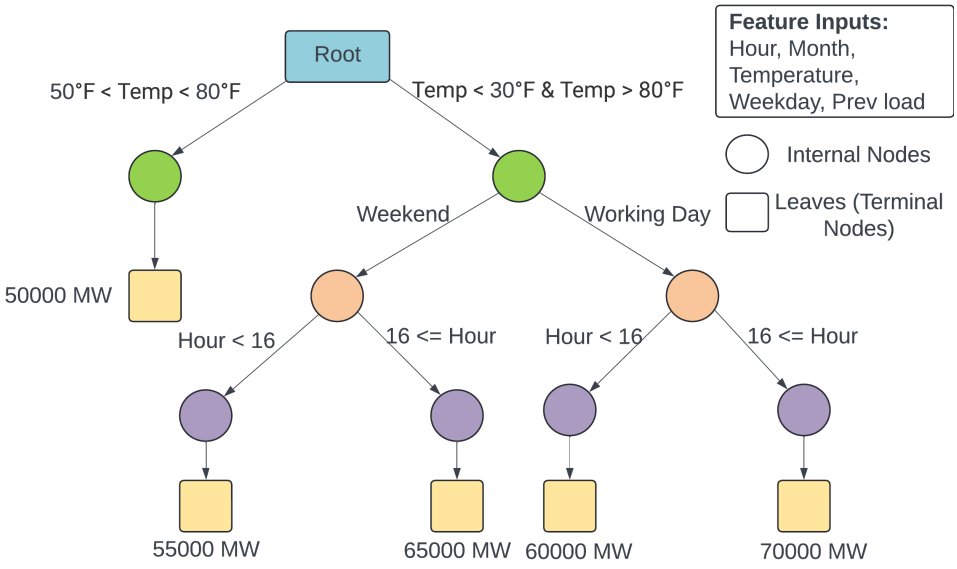
4.1 Random Forests

Random Forest (RF) regression is an ensemble learning technique that combines multiple decision trees to develop more accurate and generalized model. It is widely used for tasks such as regression, classification, and forecasting due to its ability to handle high-dimensional data, resistant to overfitting, and produce robust generalized predictions. The method extends the basic concept of decision trees and applies it to a large forest of trees, allowing the model to generalize better over unseen data.

4.1.1 Decision Trees for Forecasting

A decision tree makes predictions by splitting data into smaller subsets based on a specific feature. It works as a flowchart where each node represents a decision based on features and each branch represents the outcome of the decision. The goal of each split is to improve the homogeneity of the data in the resulting branches, making the target variable more predictable. The model starts with the entire dataset at the root node and decides which feature to split on. It chooses the feature such as temperature, hour, or weekday, that best separates the data (for regression, by minimizing Mean Squared Error (MSE)). Then the data is split into two groups based on the selected feature. Each split tries to make the groups as homogeneous as possible in terms of the target variable. For

Figure 4.1: An Illustration of Decision Tree for Electricity Load Forecasting



Notes: This figure illustrates an example of how the decision tree algorithm predicts hourly load by traversing through the features.

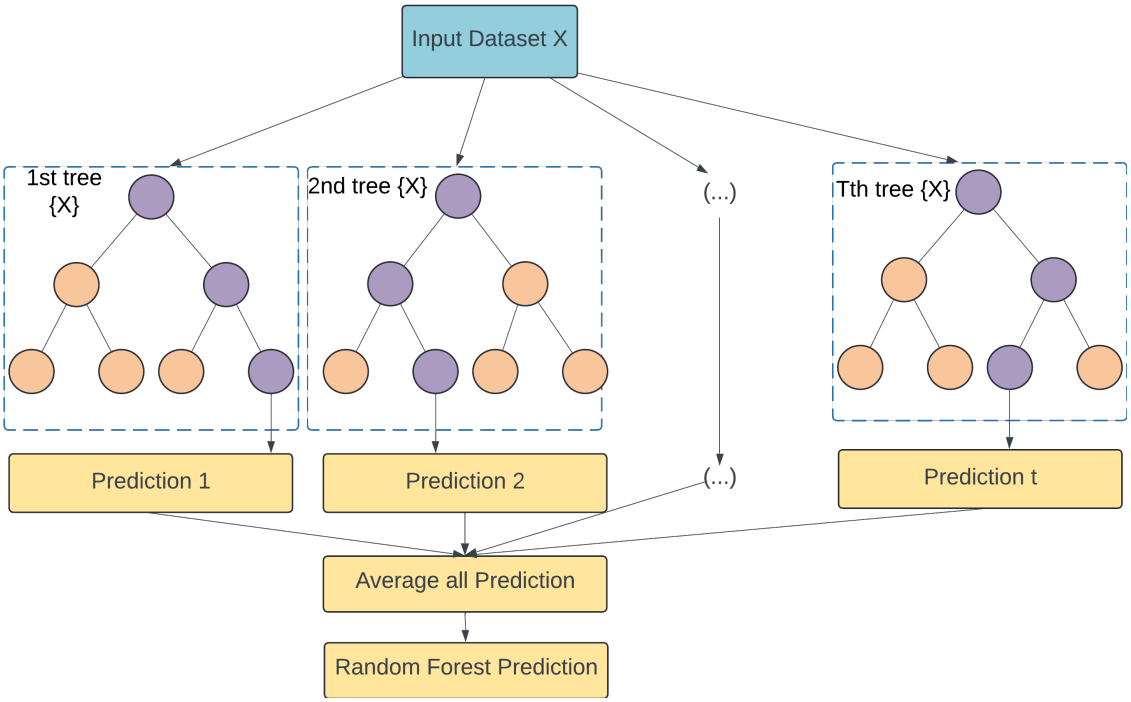
example, the Figure 4.1 is illustrating working of a random decision tree where the tree first splits based on temperature—if it's between 50°F and 80°F, the load is predicted to be lower (50,000 MW). The tree recursively partitions the feature space at internal nodes by choosing the optimal feature j and split-point s that minimize MSE. This splitting process continues for each group, creating branches, until a stopping condition is met (e.g., minimum node size or maximum depth). At the leaf nodes (end of the branches), the model makes a prediction, which is the average of the target values in that node.

While decision trees are simple and interpretable, they tend to overfit, especially on noisy datasets. This is where Random Forest comes into play, leveraging an ensemble of trees to reduce variance and improve predictive accuracy.

4.1.2 Random Forest for Forecasting

Random Forest Regression builds multiple decision trees (often hundreds or thousands), as shown in Figure 4.2 and aggregates their predictions by averaging to reduce overfitting and improve

Figure 4.2: Random Forest Architecture



Notes: This figure illustrates the architecture and functioning of a RF. An RF is composed of T decision trees. The input dataset X is fed into each tree, which independently generates a prediction. These individual predictions are then averaged to produce the final Random Forest prediction.

generalization. A key advantage of this approach, known as bootstrap aggregation or bagging, is its ability to mitigate noise by creating multiple uncorrelated trees. Each tree is trained on a different bootstrapped sample of the data, which introduces diversity. Additionally, at each split in a tree, instead of evaluating all features, only a random subset of features is considered, and the best split among these is selected based on minimizing the MSE. This process reduces correlation between the trees. While a single decision tree is prone to overfitting and may be sensitive to noise, the averaging effect of many uncorrelated trees significantly reduces variance, leading to a more robust and generalized model.

4.2 XGBoost

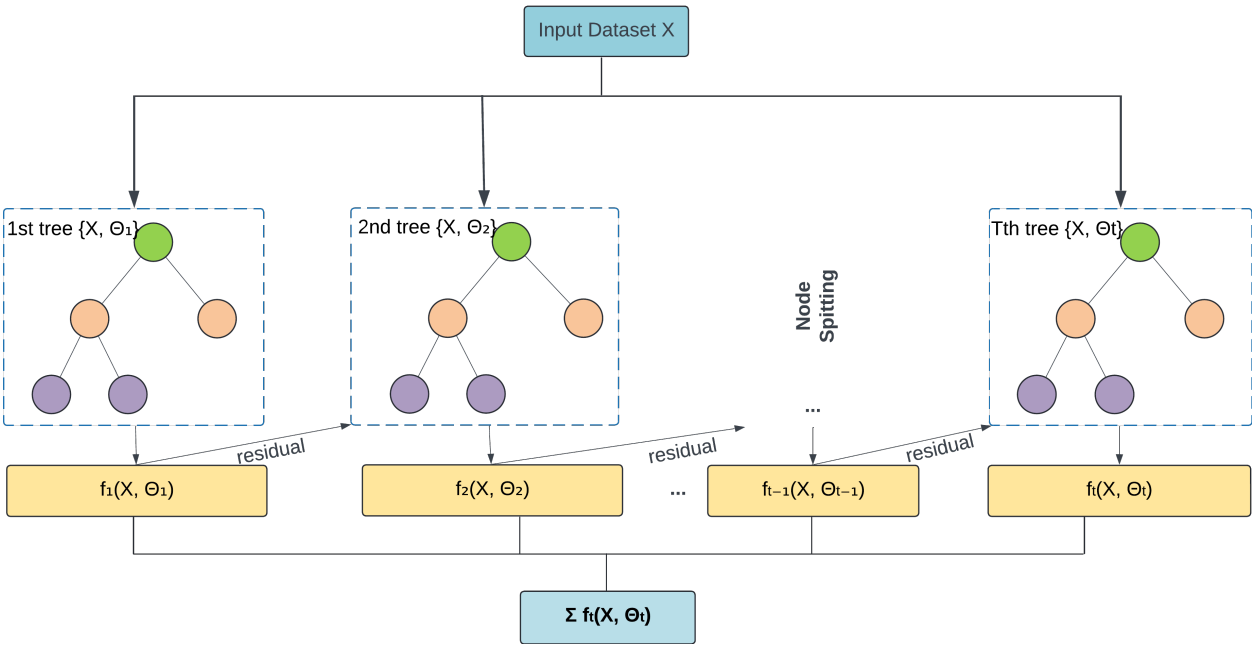
XGBoost (Extreme Gradient Boosting) is a powerful and efficient implementation of gradient-boosted decision trees designed for high performance in both classification and regression tasks. It gained wide popularity due to its scalability, speed, and ability to handle large datasets with complex patterns. Since ERCOT load dataset contains complex patterns such as multi-seasonality and non-linear relationships between load and temperature, XGBoost becomes an ideal algorithm to solve the problem of STLF. XGBoost improves upon traditional gradient boosting by incorporating regularization techniques, efficient tree construction, and advanced optimization strategies, making it ideal for high-dimensional, noisy data.

4.2.1 XGBoost for Forecasting

XGBoost follows a boost mechanism, which means it builds trees sequentially, as demonstrated in Figure 4.3 where each new tree attempts to correct the errors of the previous trees. The fundamental principle behind XGBoost is that the model combines many weak learners (decision trees) to form a strong predictive model.

A matrix $X \in \mathbb{R}^{n \times d}$, where n is the number of samples and d is the number of features, is fed into the XGBoost algorithm as input to predict the target variable y, which is defined as $y \in \mathbb{R}^n$, i.e.

Figure 4.3: XGBoost Architecture



Notes: This figure illustrates the architecture of an XGBoost model. XGBoost builds T decision trees sequentially, where each tree (f_t) represents a function mapping inputs (X) to predictions based on split conditions (Θ_t). The final prediction is the sum of outputs from all trees, minimizing residual errors at each step.

hourly electricity load. Before the training starts, XGBoost allows for several data preprocessing steps such as handling missing data and random column sub-sampling, to create diversity among the trees, making the model more generalized and reducing overfitting. Upon starting the training process, XGBoost initializes the model with a constant prediction for all data points. This initial prediction is the mean of the target values:

$$\hat{y_i}^{(0)} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{4.1}$$

This forms the base model, and the subsequent trees are built to minimize the residuals between this prediction and the actual target values using the objective function that balances loss minimization and model complexity.

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{t=1}^{T} \Omega(f_t) \tag{4.2}$$

where $l(y_i, \hat{y}_i^{(t)})$ is the loss function which is MSE in this case, and $\Omega(f_t)$ is the regularization term that penalizes the complexity of the trees.

In each iteration t, XGBoost constructs a new decision tree f_t, defined as $f_t(X, \theta_t)$ where X is the input dataset and θ_t represents the parameters of t^{th} decision tree, that aims to correct the errors made by the previous model. The difference between the true target y_i and the prediction $\hat{y_i}^{(t-1)}$ is called the residual: $r_i^{(t)} = y_i - \hat{y_i}^{(t-1)}$. The new tree f_t is trained to predict residuals, not actual target values. This tree tries to learn the errors that the previous model made. To update predictions effectively, gradient boosting computes two key terms: the gradient $g_i^{(t)}$, and the Hessian $h_i^{(t)}$. The gradient represents how much the model's predictions need to be adjusted (in the direction that minimizes the error), and the Hessian is the second-order derivative that captures the curvature of the error surface. These values are computed for each data point and

help the model optimize the loss function.

$$g_i^{(t)} = \frac{\partial l\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial \hat{y}_i^{(t-1)}} \tag{4.3}$$

$$h_i^{(t)} = \frac{\partial^2 l\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial \left(\hat{y}_i^{(t-1)}\right)^2} \tag{4.4}$$

When a decision tree splits the data into two parts, the model calculates the **gain** to measure how much this split improves the prediction and is given by:

$$\text{Gain} = \frac{1}{2}\left(\frac{\left(\sum_{i\in L} g_i^{(t)}\right)^2}{\sum_{i\in L} h_i^{(t)} + \lambda} + \frac{\left(\sum_{i\in R} g_i^{(t)}\right)^2}{\sum_{i\in R} h_i^{(t)} + \lambda} - \frac{\left(\sum_{i\in L\cup R} g_i^{(t)}\right)^2}{\sum_{i\in L\cup R} h_i^{(t)} + \lambda}\right) - \gamma \tag{4.5}$$

where L, R are left and right splits of the data, λ is the regularization term for controlling the leaf size and γ is the penalty for creating new leaf nodes. The gain depends on both the gradient and Hessian, along with regularization parameters that prevent overfitting. The higher the gain, the better the split improves the model's performance.

After splitting, each leaf in a tree contains predictions. The predicted value at each leaf node is computed using the ratio of the sum of gradients to the sum of Hessians, adjusted by a regularization term λ.

$$\hat{y}_{\text{leaf}} = -\frac{\sum_{i\in\text{leaf}} g_i^{(t)}}{\sum_{i\in\text{leaf}} h_i^{(t)} + \lambda} \tag{4.6}$$

This ensures that the predicted values are balanced between improving accuracy and preventing overfitting. Following the training of each tree, its predictions are added to the previous predictions from earlier trees. This update is scaled by a learning rate η, which controls how much impact the

new tree has on the final prediction.

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta \cdot f_t(X_i) \tag{4.7}$$

Once all trees have been trained, the final prediction for each data point is obtained by summing up the predictions from all trees, each scaled by the learning rate. This combination of many trees, each correcting the mistakes of the previous ones, results in a powerful predictive model.

$$\hat{y}_i = \sum_{t=1}^{T} \eta \cdot f_t(X_i) \tag{4.8}$$

4.3 Generalized Additive Models

Generalized Additive Models (GAM) are a flexible extension of generalized linear models (GLMs) that allow for modeling non-linear relationships between the dependent variable and each of the independent variables. Unlike traditional linear models that assume a strict linear relationship between input and output variables, GAM allow for non-linear smoothing functions to be applied to individual predictors while maintaining the additivity of the model. This makes GAM particularly useful for problems where the relationship between predictors and the target variable is more complex, such as temperature and electricity load.

4.3.1 GAM for Forecasting

The input data for a GAM includes a matrix $X \in \mathbb{R}^{n \times d}$, where n is the number of samples and d is the number of features. Each row represents a data instance, and each column represents a different feature i.e. $X_1, X_2, ..., X_d$ which include temporal variables, previous load, and temperature.

GAM has a model structure expressed as:

$$g(\mathbb{E}[Y]) = \beta_0 + f_1(X_1) + f_2(X_2) + \cdots + f_d(X_d) \tag{4.9}$$

where $g(\mathbb{E}[Y])$, $\log \mathbb{E}[Y]$, is the link function, which transforms the expected value of the response variable into a form compatible with linear regression. The link function connects the non-linear model output to the linear predictors, allowing for flexible modeling of different types of data. β_0 is the intercept term, and $f_1(X_1) + f_2(X_2) + \cdots + f_d(X_d)$ are smooth, non-linear functions applied to the features, capturing more complex relationships.

Each smooth function in the model is represented using a basis function expansion:

$$f_i(X) = \sum_{k=1}^{K} \alpha_k B_k(X) \tag{4.10}$$

where $f_i(X)$ is the smooth function, α_k are the coefficients, and $B_k(X)$ are the basis functions, we are using **cubic splines** as the basis functions. A cubic spline is a piecewise polynomial function of degree three that is smooth at the connections between the segments (called knots). These splines offer a balance between flexibility and smoothness, allowing the model to capture complex, non-linear relationships in the data while avoiding overfitting. K is the number of basis functions that control the smoothness of the function. More basis functions allow the curve to capture more detail, while fewer basis functions result in a smoother curve.

To avoid overfitting, GAM introduces a penalty term in the objective function, controlling the roughness of the smooth function:

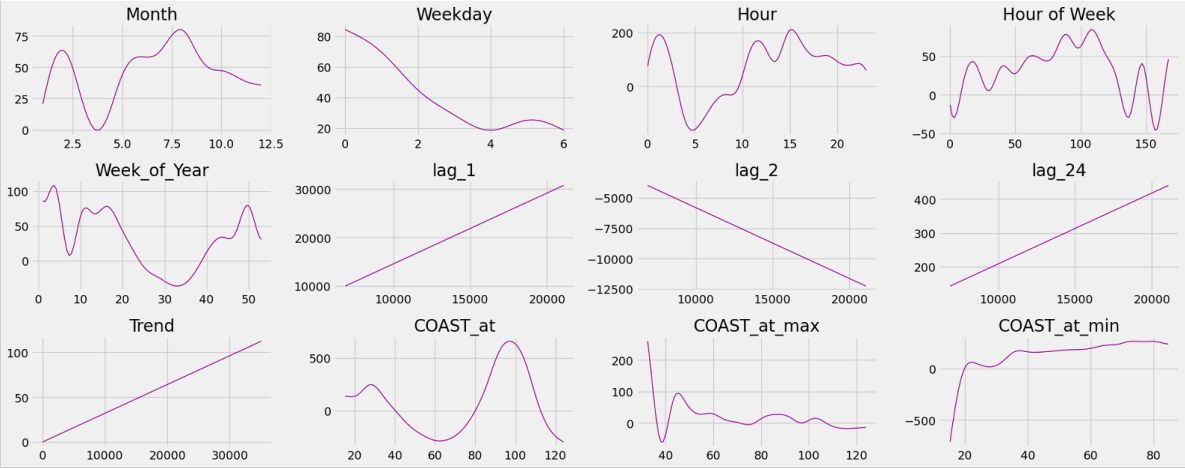
$$\text{Penalized Likelihood} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \int \left(\frac{d^2 f(x)}{dx^2} \right)^2 dx \tag{4.11}$$

The first term represents the squared error, while the second term penalizes large changes in the curvature of the smooth function, with λ controlling the level of smoothness.

The model is fit by estimating the smooth functions and coefficients for the linear terms. We use penalized least squares to optimize these terms and substitute them into the model structure in Equation (4.9). $\tilde{c}$

One of the main advantages of GAM is their interpretability. Unlike more complex machine learning

Figure 4.4: Training Features Interpretation using GAM



Notes: This figure illustrates the interpretability of the structure of the GAM model for COAST weather zone and the dependence of the electric load on each input feature. COAST_at, COAST_at_max, and COAST_at_min denote the actual hourly temperature, the maximum daily temperature, and the minimum daily temperature in the COAST weather zone, respectively. For out-of-sample testing and predictions, actual temperatures are replaced with forecasted temperatures, denoted by COAST_ft, COAST_ft_max, and COAST_ft_min, respectively.

models, GAM allow us to visualize and understand the individual contributions of each predictor to the overall outcome. This ability to capture non-linear relationships, as demonstrated by the smooth curves for variables such as Hour and COAST_at in Figure 4.4, clearly illustrates how each feature contributes to the model's output. This provides both transparency and valuable insights into the model's predictive process.

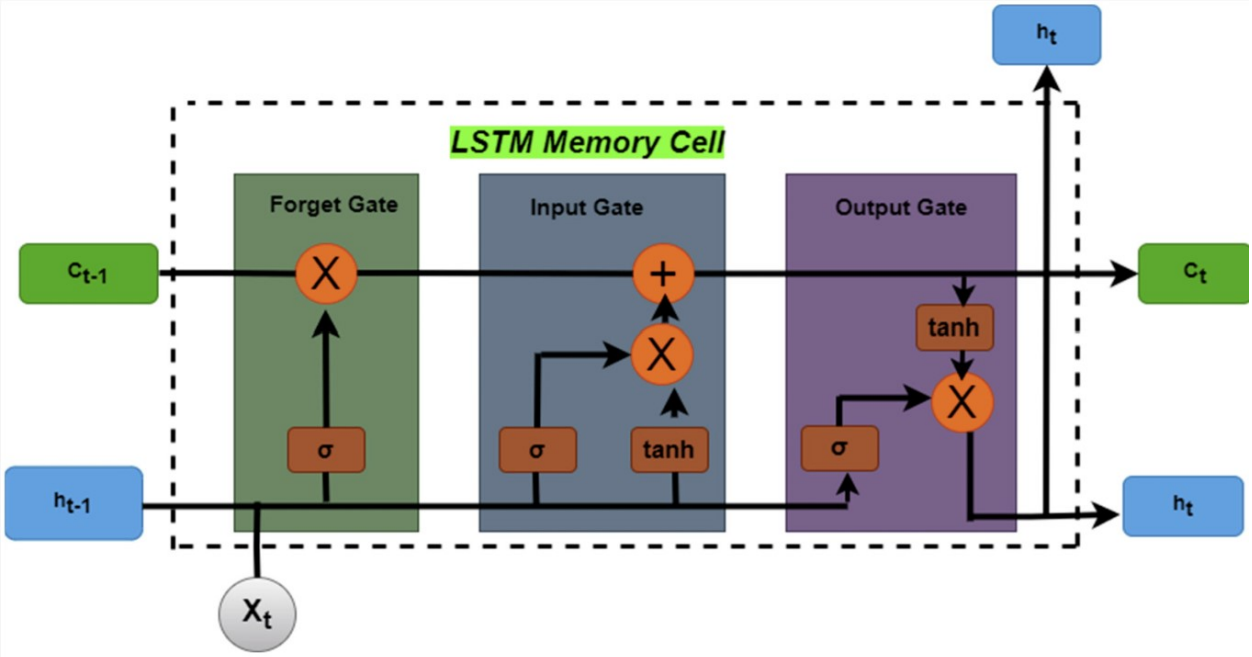
4.4 Long-Short Term Memory

Long Short-Term Memory (LSTM) networks are a special type of RNN designed to handle sequential data and solve the problem of vanishing gradients that traditional RNNs suffer from. LSTMs are widely used in tasks that involve time-series forecasting, natural language processing (NLP), and other problems where long-term dependencies are crucial.

The architecture of LSTM is built to maintain long-term memory, preserve information over many time steps, and selectively forget irrelevant information through their gating mechanism. This

makes LSTM useful in time-series forecasting applications such as electricity load forecasting, where different time periods pertain to different seasonal patterns, so retaining the seasonality information of different time periods plays a crucial role in determining the accuracy of the load forecasts.

Figure 4.5: LSTM Memory Cell Architecture



Notes: This figure illustrates the structure and operation of an LSTM memory cell. The LSTM memory cell consists of three primary gates: the forget gate, input gate, and output gate, which regulate the flow of information. At each time step, the forget gate decides what information from the previous cell state (c_{t-1}) should be discarded. It applies a sigmoid activation function (σ) to the previous hidden state (h_{t-1}) and current input (x_t). The input gate, consisting of a sigmoid function (σ) and a *tanh* function, determines what new information to add to the current cell state (c_t). Finally, the output gate computes the next hidden state (h_t) by applying both a sigmoid function (σ) and a *tanh* activation to the current cell state. The cell states (c_{t-1} and c_t) and hidden states (h_{t-1} and h_t) represent the memory and output of the cell, with the input at the current time step denoted as x_t. Together, these gates allow the LSTM to manage long-term dependencies and adapt to new inputs efficiently.

In Figure 4.5, the structure of an LSTM memory cell is shown. The working of an LSTM cell proceeds step by step as follows: First, the input x_t, which represents the current electricity demand at hour t, is combined with the previous hidden state h_{t-1}, summarizing the information of the demand from previous hours, and both are passed into the LSTM cell. The first operation is the forget gate, which determines how much of the previous cell state C_{t-1}(the memory of past electricity demand patterns) should be retained or discarded. This is done by applying a sigmoid

function on h_{t-1} and x_t, producing a forget factor f_t. Next, the input gate determines how much new information from x_t (current electricity demand) should be added to the cell state. A new candidate cell state $\tilde{C}_t$ is generated by passing the current input and hidden state through a $tanh$ function, and the input gate i_t decides the portion of this information to add. The cell state is then updated by combining the retained old cell state and the new candidate state using the forget and input gates. Lastly, the output gate decides what part of the updated cell state should influence the next hidden state. The new hidden state $h_t$is calculated by passing the updated cell state through a $tanh$ function, modulated by the output gate value o_t , and is sent to the next time step, completing the process. This step-by-step process enables LSTM cells to selectively remember or forget information and handle long-term dependencies in sequence data such as hourly electricity demand, where past demand patterns influence future predictions.

Chapter 5

Evaluation Metrics

In evaluating the performance of the forecasting models, we apply several widely used error metrics. These metrics provide a comprehensive assessment of the model's forecasting accuracy, bias, and overall performance. In this section, we outline the metrics employed and describe their significance in evaluating out-of-sample forecasts.

5.1 Accuracy

Accuracy measures provide insights into the magnitude and scale of forecast errors, which are essential for evaluating how close the predictions are to the actual values. We use several well-established error metrics to quantify the forecasting accuracy of the models. These metrics allow us to compare and rank models based on their out-of-sample performance.

The MAE measures the average absolute difference between the actual and predicted values. It indicates the magnitude of the forecast errors, regardless of their direction (i.e., whether they are underestimates or overestimates). A lower MAE value indicates that, on average, the model produces predictions that are closer to the observed values.

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{5.1}$$

where y_t is the actual load at timestamp t, $\hat{y}_t$ is the predicted load at timestamp t, and n is the total number of observations.

The Mean Squared Error (MSE) captures the average of the squared differences between actual and

predicted values. Squaring the errors penalizes larger deviations more severely, making this metric sensitive to outliers. A smaller MSE indicates a model that makes smaller and more consistent errors.

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \tag{5.2}$$

The Root Mean Squared Error (RMSE) is the square root of the MSE, providing a measure of error in the same units as the data, which makes it easier to interpret. Similar to the MSE, it emphasizes larger errors, making it particularly useful when large errors are undesirable. A smaller RMSE signifies better predictive accuracy.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{5.3}$$

The Mean Absolute Percentage Error (MAPE) expresses the average absolute error as a percentage of the actual values, providing a scale-independent measure of forecasting accuracy. This is useful for comparing the performance across different scales or units of measurement. However, it is sensitive to small values in the denominator.

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{5.4}$$

A lower MAPE indicates that the forecast is more accurate in percentage terms, making it easier to assess performance across different models and datasets.

Coefficient of Determination (R^2) is a commonly used measure of goodness-of-fit in regression models. It quantifies the proportion of the variance in the actual data that is explained by the model. A higher R^2 value, with 1 being the maximum, signifies that the model captures a larger portion of the variance in the actual data, implying better predictive performance.

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{5.5}$$

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

5.2 Bias

Accuracy measures such as MAE, RMSE, and MAPE describe the overall magnitude of errors but do not indicate whether the errors exhibit any systematic bias. Bias occurs when a model consistently overestimates or underestimates the actual values. Detecting bias is crucial for understanding whether a model's errors are due to random fluctuations or if there is a consistent pattern of misestimation.

The Mean Percentage Error (MPE) assesses the average direction of the forecast errors. A positive MPE suggests that the model tends to underestimate the actual values, while a negative MPE indicates a tendency to overestimate. This metric is particularly useful for identifying systematic bias in the model's predictions, especially in contexts such as load forecasting where different seasons or hours may have distinct forecasting characteristics.

$$\text{MPE} = \frac{100}{n} \sum_{t=1}^{n} \frac{y_t - \hat{y}_t}{y_t} \tag{5.6}$$

5.3 Test for Differences Between Models

In addition to measuring the accuracy and bias of individual models, it is important to determine whether the differences in predictive accuracy between models are statistically significant. Comparing models allows us to assess whether an alternative forecasting approach offers a meaningful improvement over the benchmark model.

We use the Diebold-Mariano (DM) test developed by Diebold and Mariano (2002) to compare the predictive accuracy of two forecasting models by analyzing the differences in their forecast errors. Specifically, it tests whether the forecast errors from two models are significantly different from each other. This is useful for comparing the performance of competing models, especially when deciding

whether an alternative approach offers a statistically significant improvement over a benchmark model. It is defined as:

$$DM = \frac{\bar{d}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(d_t - \bar{d}\right)^2}} \tag{5.7}$$

where $d_t = e_{1,t}^2 - e_{2,t}^2$, is the difference between the squared forecast errors of the two models, and $\bar{d}$ is the mean of d_t. The test evaluates whether the forecast errors from two models are significantly different from each other. The null hypothesis (H_0) assumes that the two models have equal predictive accuracy, meaning the difference between their forecast errors is zero. In contrast, the alternative hypothesis (H_1) suggests that the models have unequal predictive accuracy, indicating a non-zero difference in their forecast errors. If the DM statistic is significantly different from zero, the null hypothesis is rejected, implying that one model outperforms the other in terms of forecasting accuracy.

The Diebold-Mariano test provides a formal statistical framework to compare the performance of different models, helping to ensure that any observed differences in accuracy are not due to random chance but reflect genuine performance improvements.

Chapter 6

Results

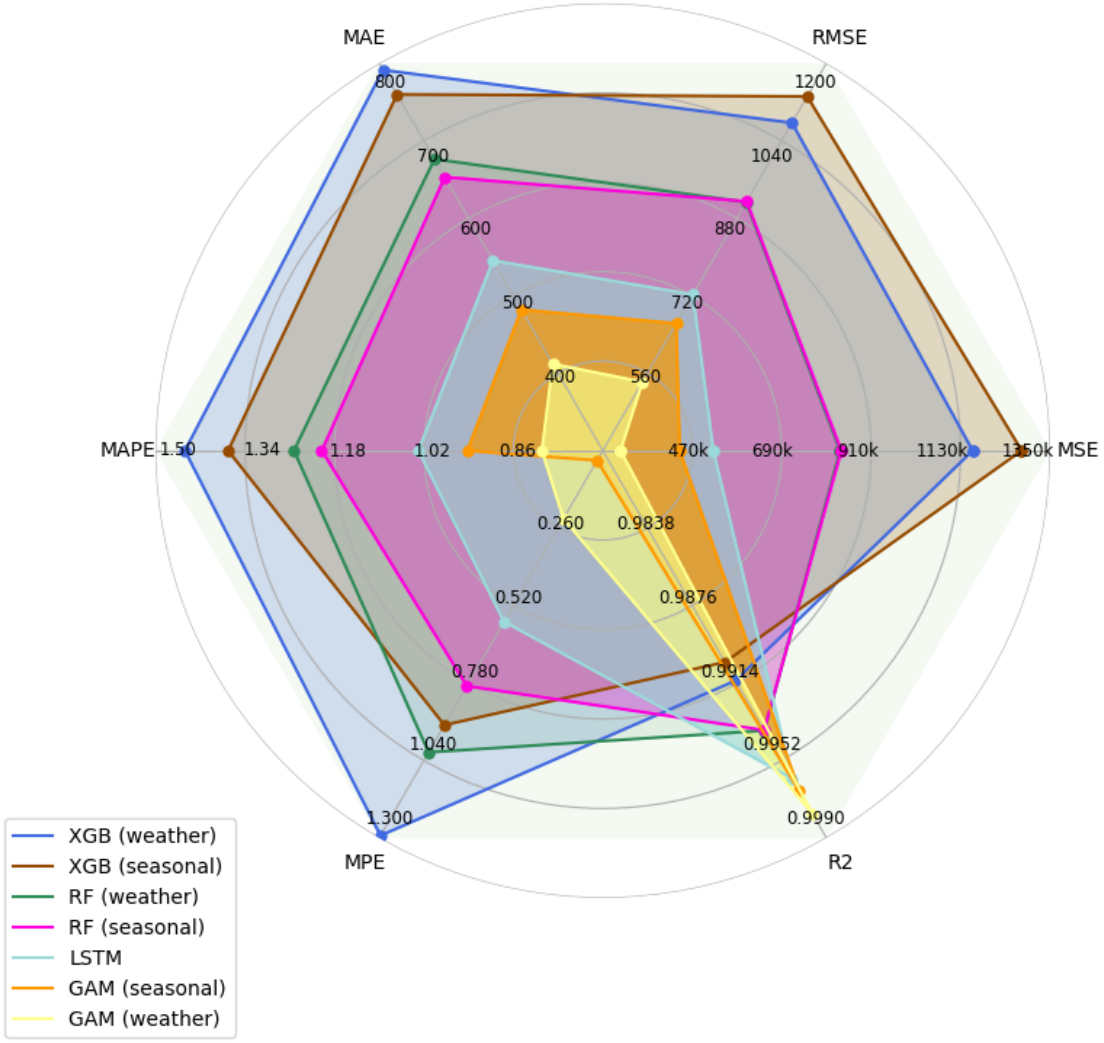
To evaluate the out-of-sample performance of the models, we begin by splitting the data into a training and test set. The full dataset spans from **January 1, 2019, to January 31, 2024**, with the training period covering **January 1, 2019, to December 31, 2022**. This results in a training sample of four years, allowing the models to capture seasonality, trends, and exogenous variables such as temperature. We then evaluate the models on an out-of-sample test set from **January 1, 2023, to January 31, 2024**, where instead of using actual historical temperatures we used historical temperature forecasts which were made on a daily basis for the next 24 hours, preventing any look-ahead bias in the hourly demand forecasts. These hourly demand forecasts are then compared against actual electricity load data for accuracy metrics such as MAE, MSE, RMSE, MAPE, and R^2.

Table 6.1: Model Evaluation Metrics Summary

Models	MSE	RMSE	MAE	MAPE	MPE	R^2
RF (weather)	833706.40	913.08	676.30	1.25	1.01	0.99
XGB (weather)	1159725.42	1076.91	791.33	1.45	1.29	0.99
GAM (weather)	293701.32	541.94	411.88	0.81	0.23	1.00
LSTM	521971.56	722.48	545.34	1.03	0.57	1.00
GAM (seasonal)	439287.81	662.79	481.87	0.94	0.03	1.00
RF (seasonal)	836843.46	914.79	653.10	1.20	0.79	0.99
XGB (seasonal)	1280221.56	1131.47	759.80	1.37	0.92	0.99

Table 6.1 and Figure 6.1 provides a detailed comparison of all models for forecasting hourly electricity demand using key evaluation metrics: MSE, RMSE, MAE, MAPE, MPE, and R^2. The

Figure 6.1: Radar Chart displaying all Error Metrics for All Models



GAM (weather) model stands out with the lowest error metrics across most categories, including MAPE of 0.81%, RMSE of 541.94, and a very low MAE of 411.88, indicating superior accuracy and minimal forecast error. On the other hand, XGB (seasonal) and XGB (weather) show the highest error rates, with an MAPE of 1.37% and 1.45% respectively and RMSE of 1,131.47, and 1,076.91 respectively, reflecting the model's struggle with accuracy, especially in capturing growth trend variations. While LSTM performs well with an RMSE of 722.48, it does not outperform the GAM models, particularly in terms of bias (as indicated by MPE).

6.1 Accuracy

To evaluate the out-of-sample performance of our forecasting models, we utilize a set of commonly applied accuracy metrics, including MAE, MSE, RMSE, MAPE, and R^2. Figures 6.2, A.2 and 6.3 present line charts of these statistics, segmented by month, to examine the errors in hourly load forecasts. Our analysis reveals that both variations of the GAM models—one incorporating temperature variables and one without, consistently outperform alternative models in terms of accuracy, with LSTM following closely behind.

Given that ERCOT is an isolated grid, the precision of forecasts during peak load seasons and high-demand hours is particularly critical. To compare the models' performance during these periods, we examine Figures 6.2, A.2 and 6.3, which illustrate that most models, including RF, XGB, and LSTM, exhibit large errors during the peak summer months (July and August) and winter months (November, December, January). The RMSE heatmap A.4 further supports this finding, highlighting significant forecasting errors during these periods. The underlying reasons for these errors, particularly in summers, stem from both fundamental factors and limitations of certain models. ERCOT's load consistently sets new peak records during summer, and decision-tree-based models such as XGBoost and RF struggle to predict load values outside the range of their training data. As these models were trained on data from 2019 to 2023, they fail to capture the higher peak loads observed in the 2024 data. In contrast, LSTM and GAM models demonstrate superior

Figure 6.2: MAE Line Plot by Month for All Models

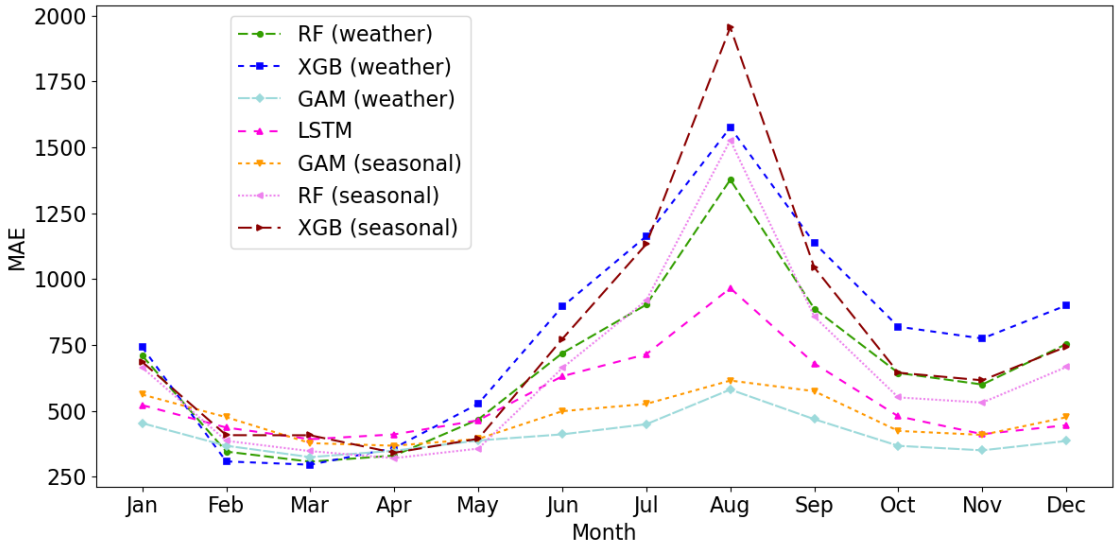
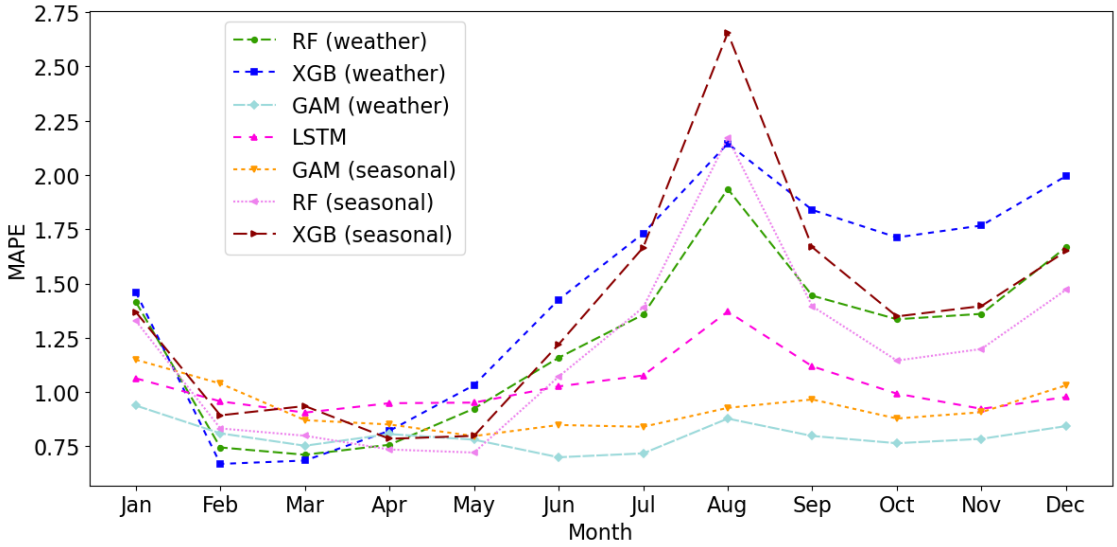


Figure 6.3: MAPE Line Plot by Month for All Models



Notes: The sub-figures depict the resulted line plots of MAE and MAPE error metrics for all models grouped by months.

Figure 6.4: MAE Line Plot by Hour for All Models

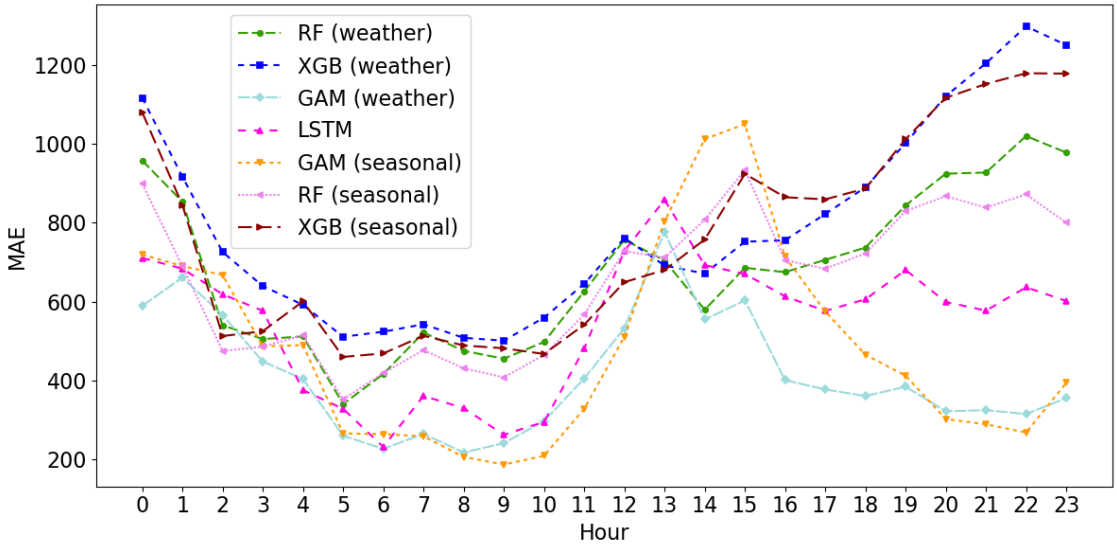
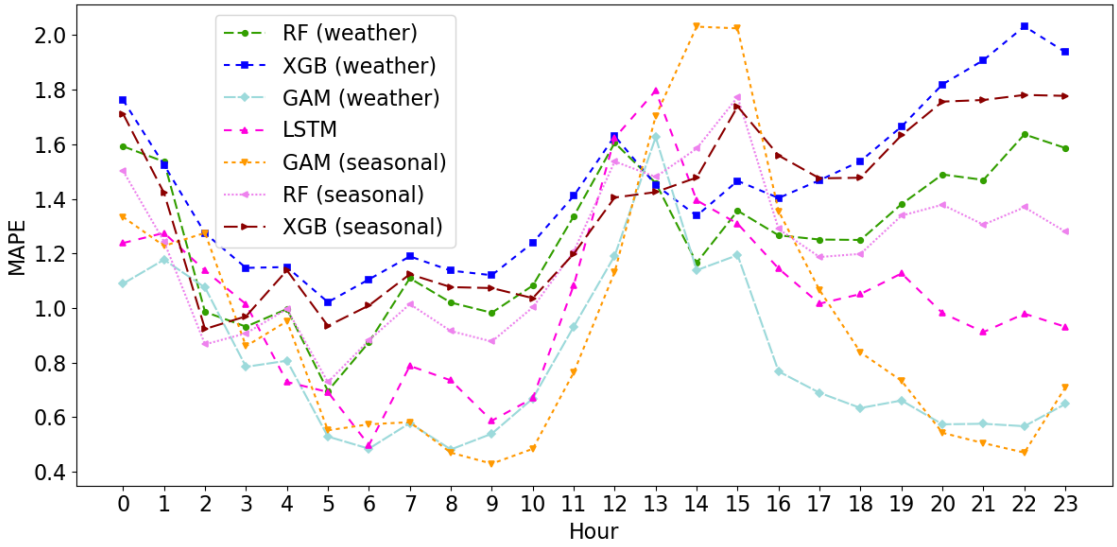


Figure 6.5: MAPE Line Plot by Hour for All Models



Notes: The sub-figures depict the resulted line plots of MAE and MAPE error metrics for all models grouped by hours (UTC time).

performance during the summer due to their ability to capture both short- and long-term seasonal trends, including out-of-range values, in their forecasts.

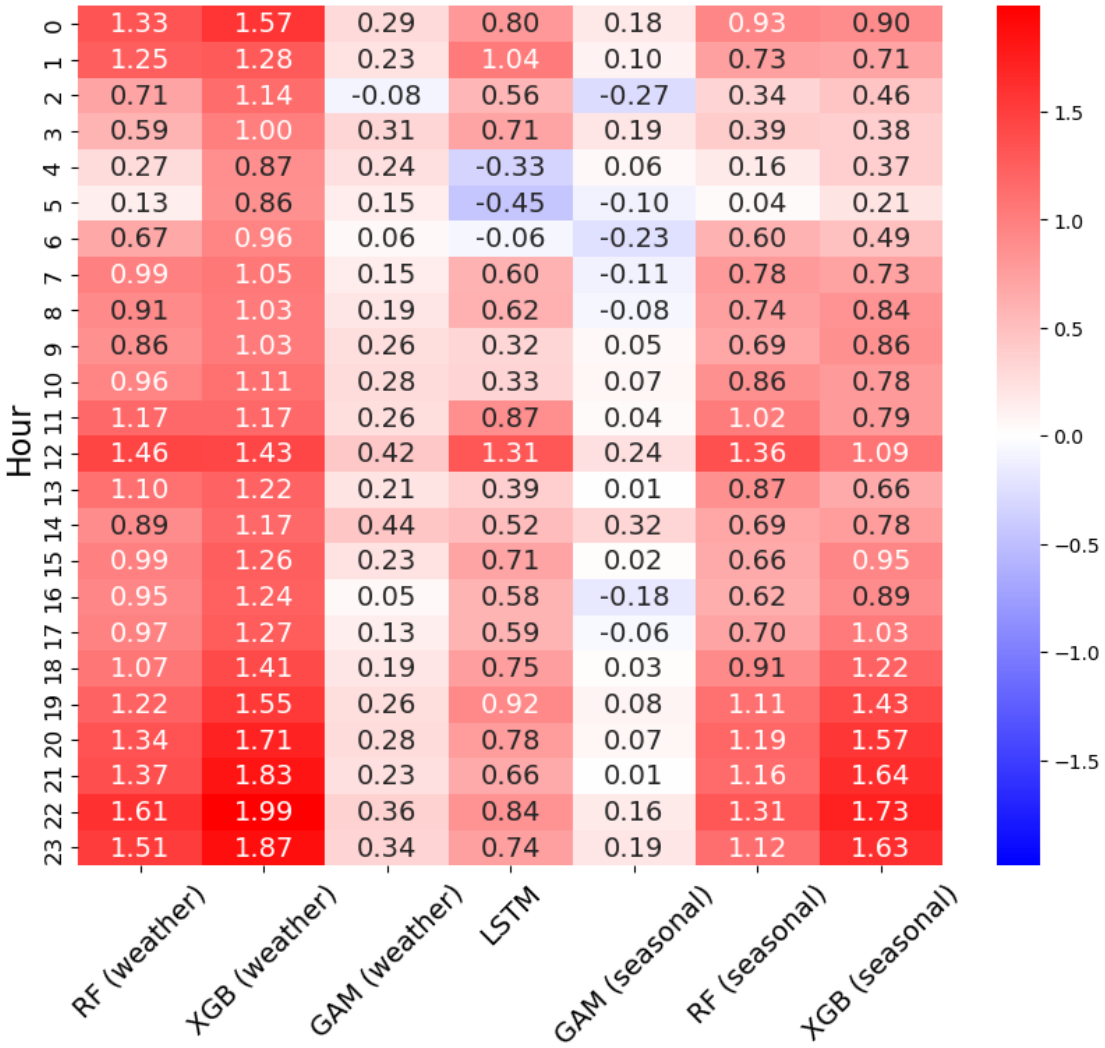
Figures 6.4, A.1 and 6.5 detail the performance of the models based on the hour of the day. The largest MAPE and MAE errors are observed during the 6-9 AM CT (12-15 UTC) and 5-9 PM CT (22-2 UTC) time ranges. ERCOT's daily load consumption patterns vary by season: in summer, load peaks occur during 5-9 PM CT, while winter sees dual peaks during 6-9 AM CT and 5-9 PM CT. This variation in daily consumption patterns often leads models to learn a generalized hourly shape, contributing to large errors during these specific periods. The winter forecasts, in particular, exhibit larger MAPE values, which align with changes in hourly consumption patterns, as demonstrated in Figures 6.5 and 6.3.

In summary, GAM models consistently outperform the other models across most accuracy error metrics, particularly during peak load months and high-demand hours. Their superior ability to capture both local and global seasonal trends, alongside exogenous factors such as temperature, underscores their effectiveness in forecasting hourly electricity demand.

6.2 Bias

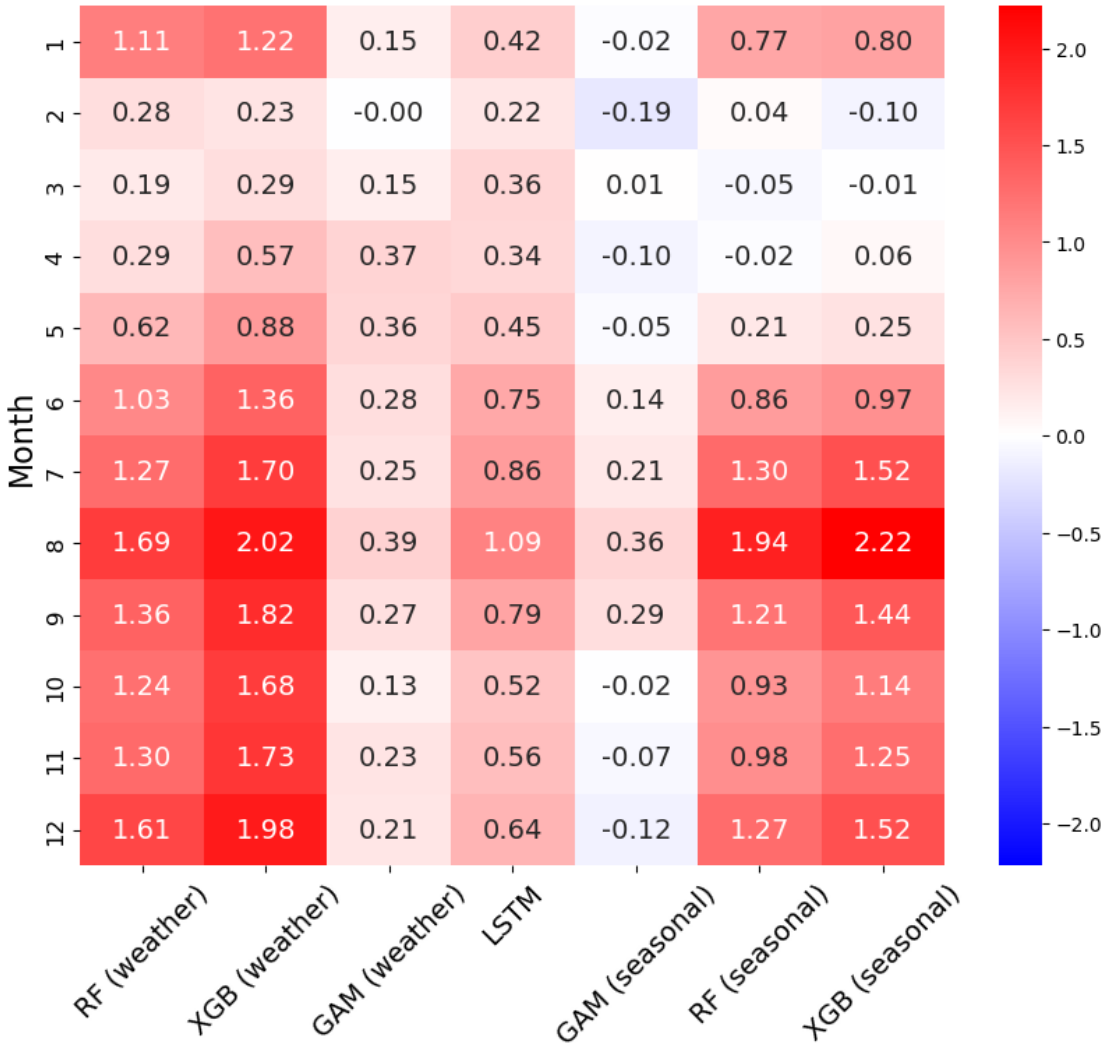
Bias analysis, based on the MPE statistic shown in Figures 6.6 and 6.7, reveals that the RF and XGBoost models exhibit positive forecast errors on a monthly scale, particularly in January and from June through December, indicating a consistent tendency to under-predict hourly electricity demand. On an hourly basis, these models continue to under-predict load across most hours, with the exception of hours ending 10 and 11 PM CT (4 and 5 UTC), where the errors are more balanced. This persistent under-prediction bias suggests that RF and XGB models are highly prone to underestimating demand during both monthly and hourly intervals. Similarly, LSTM models also produce positive forecast errors, but these are smaller in magnitude compared to RF and XGB models, indicating a relatively lower bias. In some instances, particularly during 10-11 PM CT (4-5 UTC), LSTM models generate negative errors, leading to over-predicted load forecasts. In contrast,

Figure 6.6: MPE Heatmap by Hour for All Models



Notes: The figure highlights the bias present in the models. Red indicates a tendency to under-estimate, while blue indicates a tendency to over-estimate. The hours are shown in UTC time.

Figure 6.7: MPE Heatmap by Month for All Models

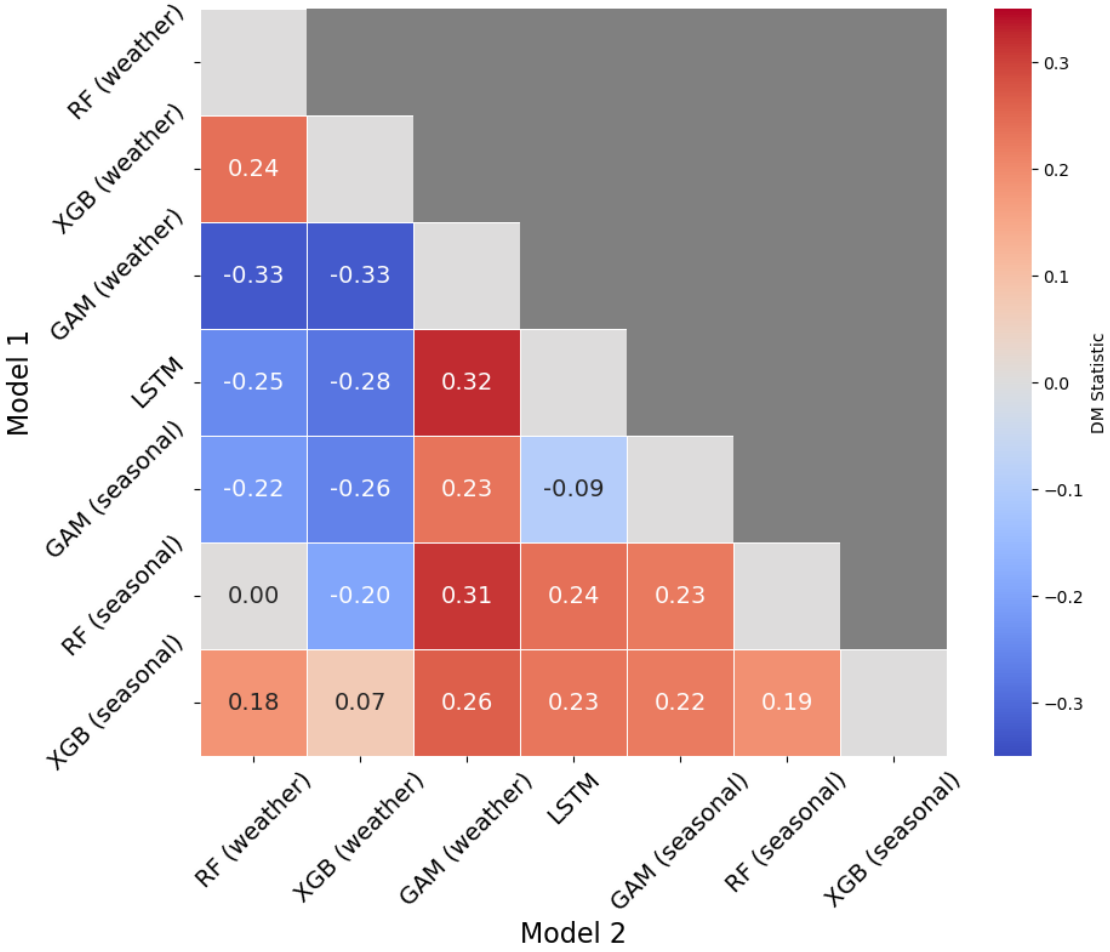


Notes: The figure highlights the bias present in the models. Red indicates a tendency to under-estimate, while blue indicates a tendency to over-estimate.

GAM models produce the smallest errors among all models, displaying minimal bias. The GAM (weather) model's errors are too small to indicate any consistent directional bias, while the GAM (seasonal) model displays a well-distributed error profile, leading to unbiased forecasts of hourly electricity load. This demonstrates the capacity of GAM models to generate reliable, unbiased load forecasts across different seasons and hours.

6.3 Test for Differences Between Models

Figure 6.8: Diebold Mariano Test

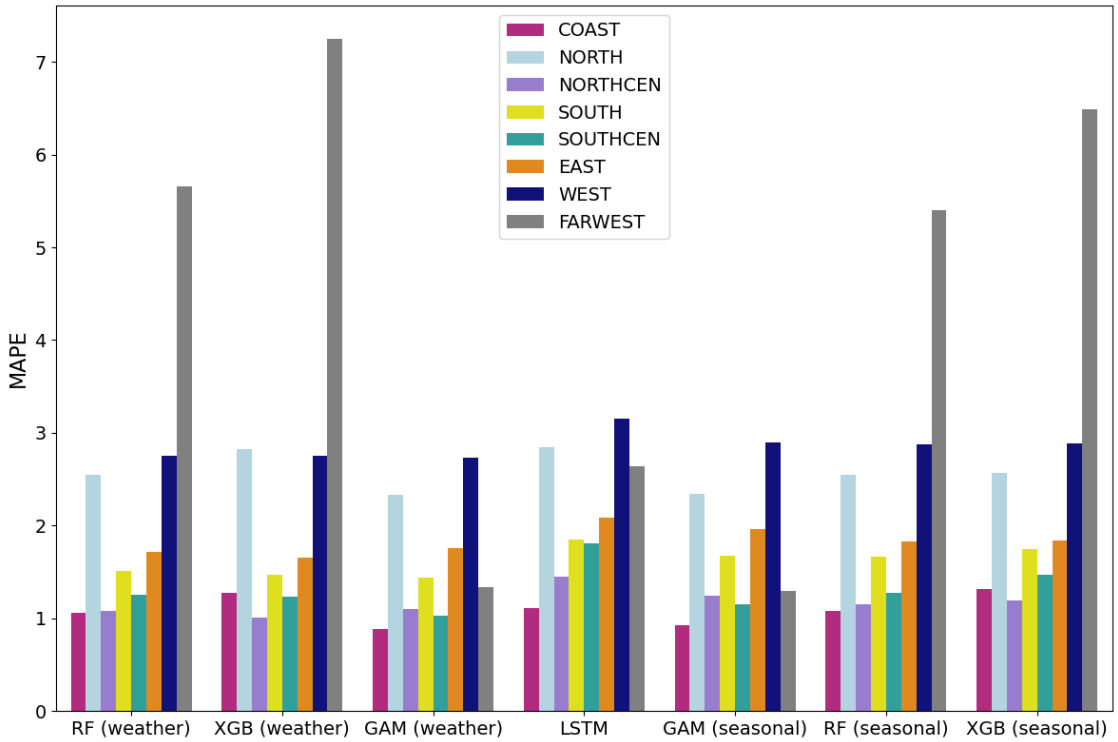


The DM test results in Figure 6.8 indicate that the forecasts from all the models are statistically

indistinguishable. This is because the DM statistics and p-values obtained when comparing each model pair were insignificant at 1%, 5% and 10% significance levels, indicating a failure to reject the null hypothesis, which states that both models have the same predictive accuracy. However, we anticipate that if the DM statistic is computed for peak electricity demand months (July and August) and daily peak load hours, it could lead to a different results where models differ in predictive accuracy, as shown by accuracy analysis, over monthly and hourly periods depicted in Figures 6.5, 6.3, 6.4 and 6.2.

6.4 Performance by Weather Zones

Figure 6.9: MAPE by Zone for All Models



Notes: The figure shows the error composition of the models across ERCOT's weather zones. RF and XGB demonstrate the poorest performance in the FARWEST weather zone, where their errors are the largest. NORTHCEN and SOUTHCEN refer to the North Central and South Central weather zones, respectively.

Figure 6.9 highlights MAPE for different models across ERCOT zones. Tree-based models, such

as RF and XGB, exhibit notably higher errors in the FARWEST zone. This is attributed to the unique positive trend in electricity demand in FARWEST, driven by significant renewable energy buildout over the years. Unlike other zones, FARWEST's rapid and sustained growth in demand creates challenges for these models, as they primarily rely on historical patterns and fail to extrapolate effectively for unseen scenarios. This limitation causes RF and XGB to underperform when forecasting above certain thresholds, reflecting their inherent difficulty in capturing strong, zone-specific growth trends. In contrast, models such as GAM and LSTM, which can better incorporate temporal trends and predict outside of the training data range, perform relatively well in this scenario.

Chapter 7

Summary and Conclusions

Accurate short-term load forecasting is critical for ERCOT's operations and planning. This study examines and compares several ML methods, including XGBoost, RF, and GAM, alongside a deep learning approach, LSTM, to identify the best-performing model during periods of high electricity demand. A dual comparative approach is adopted, evaluating models using temporal features alone and in combination with actual weather data for training and forecasted weather data for out-of-sample testing. The study emphasizes region-specific forecasting by capturing diverse load patterns across ERCOT's individual weather zones and aggregating them to predict total load.

Model performance is assessed using accuracy metrics such as MAPE, MSE, MAE, and RMSE, as well as bias metrics such as MPE, with particular focus on high-demand months and peak load hours. Variants of the GAM model consistently outperform other models across most error metrics, especially during months with the highest loads and during peak-load hours. GAM model that incorporates both temporal and weather variables demonstrates the best accuracy with the lowest error metrics, while the GAM model using only temporal features is the least biased. Including temperature variables improves accuracy but increases model bias, leading to under-prediction of electric load.

An analysis of errors by zone reveals that tree-based models, such as XGBoost and Random Forest, struggle to predict values outside the range of their training data, particularly in the FARWEST zone, which has experienced significant growth in electricity demand over the years. These models also perform poorly during summer months and peak-load hours, largely due to the same limitation.

The DM test is applied to all possible model pairs to assess differences in predictive accuracy. On

the full range of test dataset, the DM test indicates that the predictive accuracy of the models is statistically indistinguishable. However, it remains unclear whether this finding holds true during peak-load months and hours, indicating further investigation.

There are several avenues for future research that include applying the DM test to specifically compare model performance during peak-load hours and high-demand months, providing more refined statistical insights into differences in predictive accuracy of models under these critical conditions. Secondly, application and evaluation of models such as Localized Random Forest and Random Forest combined with Linear Regression, to better capture predictions beyond the training range. Developing a hybrid model that integrates tree-based methods with GAM could also be a promising approach, aiming to reduce errors during peak hours and summer months when accuracy is most crucial. Finally, extrapolating the models to predict electricity demand 24 hours ahead and comparing their results using the same evaluation criteria could offer valuable insights for day-ahead operational planning.

Appendices

Appendix A

Supplemental Figures

Figure A.1: MSE Line Plot by Hour for All Models

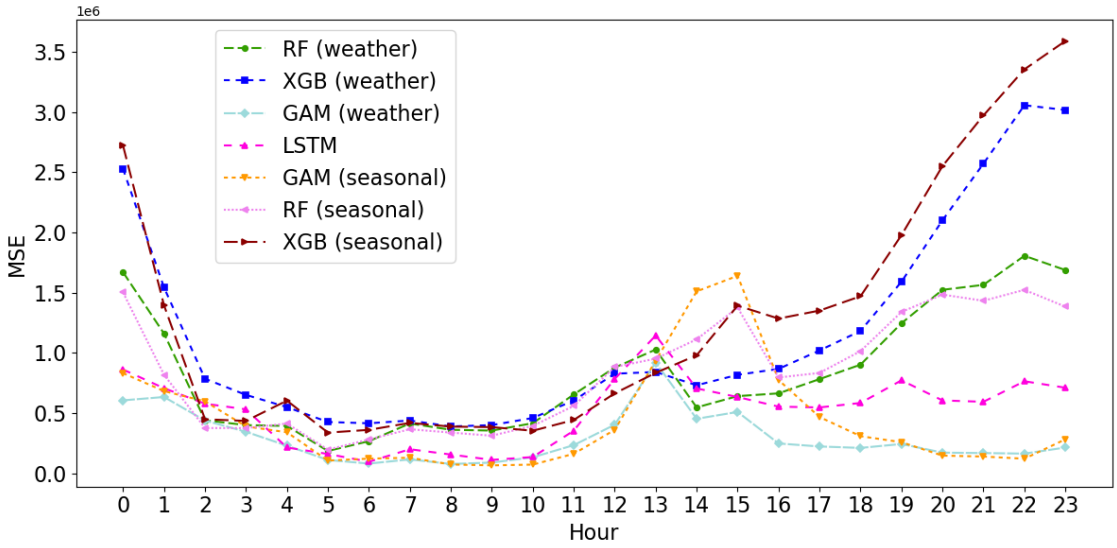


Figure A.2: MSE Line Plot by Month for All Models

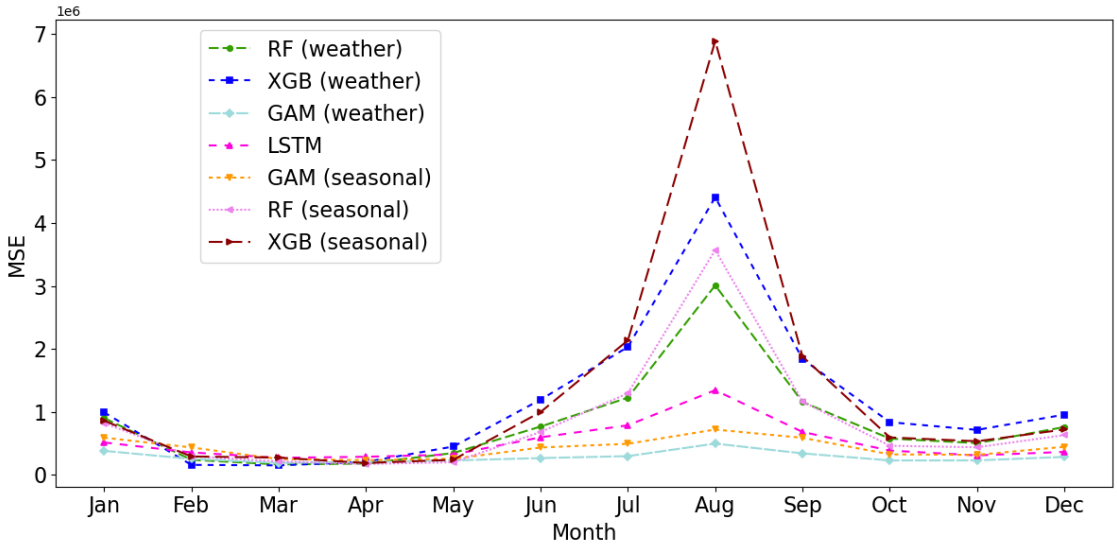


Figure A.3: RMSE Heatmap by Hour

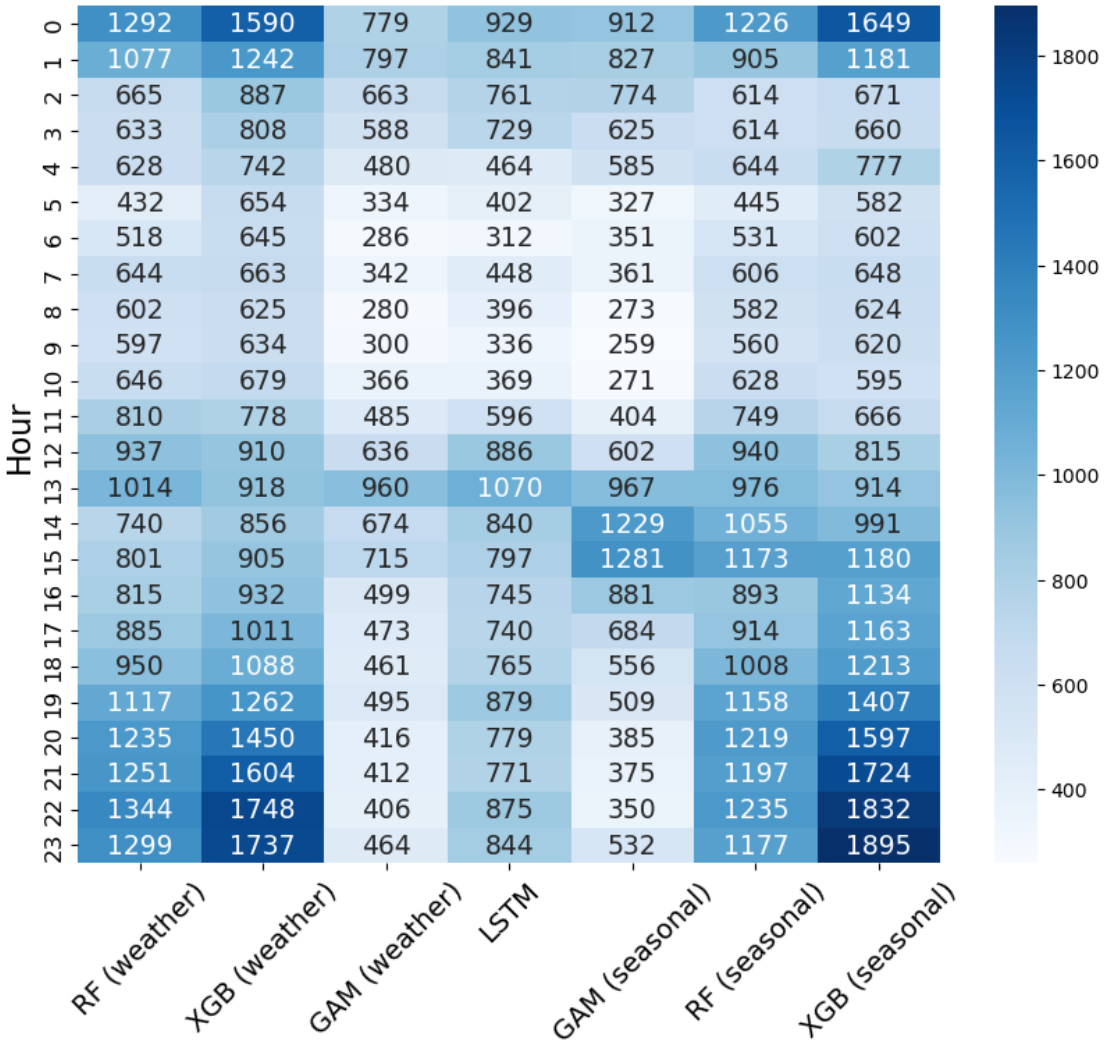


Figure A.4: RMSE Heatmap by Month

Bibliography

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.

Chapagain, K., Kittipiyakul, S., and Kulthanavit, P. (2020). Short-term electricity demand forecasting: Impact analysis of temperature for Thailand. *Energies*, 13(10):2498.

Chaturvedi, S., Rajasekar, E., Natarajan, S., and McCullen, N. (2022). A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India. *Energy Policy*, 168:113097.

Chodakowska, E., Nazarko, J., and Nazarko, Ł. (2021). Arima models in electrical load forecasting and their robustness to noise. *Energies*, 14(23):7952.

Cordeiro-Costas, M., Villanueva, D., Eguía-Oller, P., Martínez-Comesaña, M., and Ramos, S. (2023). Load forecasting with machine learning and deep learning methods. *Applied Sciences*, 13(13):7933.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144.

Fan, G.-F., Yu, M., Dong, S.-Q., Yeh, Y.-H., and Hong, W.-C. (2021). Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. *Utilities Policy*, 73:101294.

Feng, Q. and Qian, S. (2021). Research on power load forecasting model of economic development zone based on neural network. *Energy Reports*, 7:1447–1452.

Guo, C., Ge, Q., Jiang, H., Yao, G., and Hua, Q. (2020). Maximum Power Demand Prediction Using Fbprophet With Adaptive Kalman Filtering. *IEEE Access*, 8:19236–19247.

Guo, W., Che, L., Shahidehpour, M., and Wan, X. (2021). Machine-Learning based methods in short-term load forecasting. *The Electricity Journal*, 34(1):106884.

Gökçe, M. M. and Duman, E. (2022). Performance Comparison of Simple Regression, Random Forest and XGBoost Algorithms for Forecasting Electricity Demand. In *2022 3rd International Informatics and Software Engineering Conference (IISEC)*, pages 1–6.

Hinman, J. and Hickey, E. (2009). Modeling and forecasting short-term electricity load using regression analysis. *Journal of Institute for Regulatory Policy Studies*, pages 1–51.

Hossain, M. S. and Mahmood, H. (2020). Short-Term Load Forecasting Using an LSTM Neural Network. In *2020 IEEE Power and Energy Conference at Illinois (PECI)*, pages 1–6.

Kamoona, A., Song, H., Keshavarzian, K., Levy, K., Jalili, M., Wilkinson, R., Yu, X., McGrath, B., and Meegahapola, L. (2023). Machine learning based energy demand prediction. *Energy Reports*, 9:171–176.

Krstonijević, S. (2022). Adaptive Load Forecasting Methodology Based on Generalized Additive Model with Automatic Variable Selection. *Sensors*, 22(19):7247.

Lindberg, K. B., Seljom, P., Madsen, H., Fischer, D., and Korpås, M. (2019). Long-term electricity load forecasting: Current and future trends. *Utilities Policy*, 58:102–119.

Nallathambi, S. and Ramasamy, K. (2017). Prediction of electricity consumption based on DT and RF: An application on USA country power consumption. In *2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE)*, pages 1–7.

Nguyen, H. and Hansen, C. K. (2017). Short-term electricity load forecasting with Time Series Analysis. In *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pages 214–221.

Obst, D., De Vilmarest, J., and Goude, Y. (2021). Adaptive methods for short-term electricity load forecasting during COVID-19 lockdown in France. *IEEE transactions on power systems*, 36(5):4754–4763.

Rice, R., North, K., Hansen, G., Pearson, D., Schaer, O., Sherman, T., and Vassallo, D. (2022). Time-Series Forecasting Energy Loads: A Case Study in Texas. In *2022 Systems and Information Engineering Design Symposium (SIEDS)*, pages 196–201.

Shabbir, N., Ahmadiahangar, R., Rosin, A., Jawad, M., Kilter, J., and Martins, J. (2023). XgBoost based Short-term Electrical Load Forecasting Considering Trends and Periodicity in Historical Data. In *2023 IEEE International Conference on Energy Technologies for Future Grids (ETFG)*, pages 1–6.

Solyali, D. (2020). A comparative analysis of machine learning approaches for short-/long-term electricity load forecasting in Cyprus. *Sustainability*, 12(9):3612.

Tarmanini, C., Sarma, N., Gezegin, C., and Ozgonenel, O. (2023). Short term load forecasting based on ARIMA and ANN approaches. *Energy Reports*, 9:550–557.

Yang, J., Tuo, M., Lu, J., and Li, X. (2024). Analysis of Weather and Time Features in Machine Learning-aided ERCOT Load Forecasting. In *2024 IEEE Texas Power and Energy Conference (TPEC)*, pages 1–6.

Zhang, L. and Jánošík, D. (2024). Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches. *Expert Systems with Applications*, 241:122686.