Assessment 3:

Group Project Report (Team Echo)

Enhancing Electricity Demand Forecasting in New South Wales

Submitted by

Chuang, Keith – z5449930 (Data Analyst)

Gandhi, Rupesh – z5368767 (Lead Researcher)

Submitted on

20th April 2024

The University of New South Wales (UNSW)

Master of Data Science Course

(ZZSC9020 – Data Science Project)

Hexamester 2, 2024

Personal Declaration

During our analysis, Team Echo encountered several challenges that affected our team composition and contributions. Initially comprising four members, our team consisted of **Joe Melbin** as the Team Leader, **Rupesh Gandhi** as the Lead Researcher, **Keith Chuang** and **Yiu Tong Chiu (Davis)** as Data Analyst.

Regrettably, Joe Melbin, our Team Leader, encountered unforeseen circumstances and had to apply for special consideration, ultimately withdrawing from the course after two weeks. Additionally, Yiu Tong Chiu has been unresponsive since Week 1, failing to attend meetings or contribute to Assessment 3 (Report) and the associated code.

Considering these challenges, Keith Chuang and Rupesh Gandhi decided to continue with the project independently, focusing on time series models and hybrid variations. It's important to note that Team Echo should now be considered a two-member team, and **all submissions and grades should be attributed solely to the efforts of Keith Chuang** (z5449930) **and Rupesh Gandhi** (z5368767).

To ensure transparency, we have communicated this situation in advance to Dr. Rohitash Chandra (Course Convenor), Dr. Wei Tan (Online Lecturer), and Dr. Sonit Singh (Online Lecturer).

From Team Echo:

Keith Chuang (Data Analyst)
Rupesh Gandhi (Lead Researcher)

Table of Content

1 Abstract 5
2 Introduction and Motivation 6
3 Literature review 7
3.1 Methodology for Literature review 7
3.2 Conventional Models 7
3.2.1 Time Series models 7
3.2.2 Regression models (RMs) 8
3.2.3 Gray Models (GM) 9
3.3 AI-based models 9
3.3.1 ANN-based models 9
3.3.2 DL models 11
3.3.3 SVR-based models 12
3.4 Conclusion commentary on Models 12
4 Exploratory Data Analysis 14
4.1 Objective 14
4.2 Data Sources 14
4.3 Data Loading 14
4.4 Data Cleaning 14
4.5 Data Transformation 15
4.6 Exploratory Analysis 15
4.6.1 Descriptive Statistics 15
4.7 Univariate Analysis – Demand 16
4.7.1 Weekly-of-the-month demand pattern 16
4.7.2 Seasonal demand 17
4.7.3 Average demand during a 24-hour period 18
4.8 Univariate Analysis - Temperature 19
4.9 Bivariate Analysis - Demand vs Temperature 19
4.10 Bivariate Analysis – Holiday on Demand 21
4.11 Actual Demand vs Forecast Demand 21
4.12 Daily Demand vs Solar Exposure 23
4.13 Daily Temperature vs Solar Exposure 24
4.14 Correlation of Different Time Frequencies and Demand 25
4.14.1 Correlation of Demand and Daily Frequencies Variables 26
4.15 Autocorrelations 26
4.15.1 Stationarity 27
5 Modelling 28
5.1 Models Considered 28
5.1.1 Statistical Metrics 28
5.1.2 Forecast accuracy comparison for the models 28
5.2 Selection of the Final Model 29
5.3 Model Performance and Benchmarking 29
5.3.1 Model Performance 29
6 Discussion 32
6.1 Complementary insights and strategic enhancements 32
6.2 Strategic enhancements and further integration proposals 32
7 Conclusion and Further Recommendations 34

7.1 Exogenous variables analysed during EDA 35
8 Bibliography 36

Table of Figures

Figure 1: Process flow for literature review 7
Figure 2: Fluctuation of electricity in NSW 16
Figure 3: Weekly demand patterns of historic data. 17
Figure 4: Seasonal demand of historic data 18
Figure 5: Mean demand plotted against day-of-week 18
Figure 6: Average daily demand 18
Figure 7: Temperature variation over time 19
Figure 8: Temperature versus electricity demand relationship. 20
Figure *9*: Demand during holidays. 21
Figure 10: Forecast vs Total Demand (2017) 22
Figure 11: Time Shifted Forecast vs Demand (2017) 22
Figure 12: Solar exposure vs daily demand. 23
Figure 13: Solar Exposure vs Temperature. 24
Figure 14: Correlation – Time vs Demand 25
Figure 15: Correlation - Demand vs *v*ariables 26
Figure 16: Statistical metric comparison of models. 28
Figure 17: Accuracy for models considered. 29
Figure 18: Model performance for 24 hours (48 intervals) 30
Figure 19: Benchmark forecast accuracy. 30
Figure 20: Model Forecast and benchmark with actual data. 31
Figure 21: Mean solar exposure 35
Figure 22: Rainfall impact 35

Reference to Tables

Table 1: referenced papers for time series models 8
Table 2: referenced papers for regression model 9
Table 3: referenced papers for gray models 9
Table 4: referenced papers for FFNN and BPNN 10
Table 5: referenced papers for ANFIS 11
Table 6: referenced papers for DL model 11
Table 7: referenced papers for SVR models 12
Table 8: SARIMA variants considered for modelling. 28
Table 9: Accuracy comparison of various models 28
Table 10: SARIMA with temperature and Fourier defining factors. 29
Table 11: Mathematical model 29

1 Abstract

Endgame Economics, our client, highlights the critical need for accurate and responsive electricity demand forecasting in New South Wales (NSW). Precise forecasting is essential for policy development and grid stability in a region impacted by complex factors including weather variations, economic shifts, and demographic changes. This report utilises comprehensive datasets spanning January 2010 to March 2021, which include half-hourly records of electricity demand alongside correlating variables such as temperature, solar exposure, and public holidays.

Our exploratory data analysis (EDA) underscored the complex seasonality of electricity demand, subject to large variations between the different seasons and within the single days, especially when facing extreme climatic events and holidays. This complex seasonality necessitated the adoption of forecasting models capable of capturing both high-frequency daily cycles and broader seasonal trends.

In our study, we evaluated various forecasting models to optimise electricity demand prediction in New South Wales. We explored an Autoregressive Integrated Moving Average (ARIMA) model incorporating both daily and weekly Fourier terms, as well as a Seasonal ARIMA (SARIMA) model with only weekly Fourier terms. The comparison revealed that the SARIMA model with solely weekly Fourier terms outperformed the ARIMA model. This suggests that the SARIMA model's inherent seasonal adjustment capabilities effectively capture the daily seasonality more efficiently than the Fourier terms modelled in the ARIMA framework. Building upon these insights, our final model—a SARIMA with weekly Fourier terms and temperature data—was developed to further refine forecast accuracy. This model leverages the periodicity captured by the weekly Fourier components alongside the temperature variables, providing a robust tool for predicting electricity demand fluctuations while accommodating seasonal and temperature influences.

Team Echo's project source code is in the below GitHub public repository Team-Echo---Enhancing-Electricity-Demand-Forecasting-in-New-South-Wales/src at main · keithzrc/Team-Echo---Enhancing-Electricity-Demand-Forecasting-in-New-South-Wales (github.com)

The full project repository is in the below link https://github.com/keithzrc/Team-Echo---Enhancing-Electricity-Demand-Forecasting-in-New-South-Wales

The insights generated from this study provide advanced tools for improving demand response strategies and optimising grid operations. Recommendations for ongoing enhancements and potential explorations into additional predictive variables are discussed to keep pace with the rapidly evolving energy sector.

This comprehensive analysis not only advances the methodology behind electricity demand forecasting but also aligns strategically with Endgame Economics' objectives, thereby enhancing decision-making processes and contributing to the operational stability and economic efficiency of the energy infrastructure in NSW.

2 Introduction and Motivation

In the current dynamic landscape marked by technological progress and societal shifts, there's a growing need for innovative solutions to address complex challenges. Team Echo responded to this call by taking on the task of forecasting electricity consumption in New South Wales (NSW). The goal is to confront rising energy costs and empower consumers through accurate predictions, fostering informed decision-making and enhancing energy affordability and accessibility in the region. Leveraging historical data and diverse variables, Team Echo is developing tailored forecast models to navigate the intricacies of NSW's energy landscape. Their commitment reflects a dedication to driving positive change and shaping a more sustainable future for the region.

Team Echo’s motivation stems from a deep-seated conviction that every challenge presents an opportunity for growth and transformation. The team is inspired by the potential to bridge gaps, break barriers, and pave the way for a brighter future for NSW residents, energy providers, environmental groups, and policy makers. Whether it's advancing scientific understanding, fostering social inclusion, or driving sustainable development, we believe that our collective efforts can affect positive change on a global scale. The team is driven by a conviction that challenges offer pathways for growth and transformation, inspiring their pursuit of innovative solutions in energy consumption forecasting. Drawing on extensive research categorising models into conventional and AI-based approaches, the paper synthesises insights from notable studies to develop tailored forecasting models for New South Wales (NSW).

By leveraging historical data and diverse variables, Team Echo aims to contribute to improved energy management and affordability in the region, aligning their efforts with a vision of fostering sustainable development and positive societal impact. Notable studies by (Daut, et al., 2017), (Bourdeau, Zhai, Nefzaoui, Guo, & Chatellier, 2019), (Connor, Chan, & Laforge, 2012)1, (Wei N. , Li, Duan, Liu, & Zeng, 2019) and (Ahmad, Chen, Yabin, & Jiangyu, 2018) provide insights into the strengths, weaknesses, and applications of these conventional, AI-based and hybrid methods. These analyses encompass a wide range of factors, including prediction scopes, data properties, method applications, input data characteristics, pre-processing methods, building typologies, targeted energy end-uses, and accuracy assessments. By synthesizing this research, Team Echo aims to develop effective forecasting models tailored to NSW's energy needs, thereby contributing to improved energy management and affordability. These models establish explicit relationships between consumption and influencing factors such as temperature, GDP, and population (Debnath & Mourshed, 2018). In contrast, AI-based models such as artificial neural networks (ANN) and support vector regression (SVR) do not require explicit relationships but learn from historical data, excelling in handling nonlinear problems and short-term forecasting. Post literature review and deep analysis of historic data, Team Echo chose a hybrid approach using SARIMA and Fourier to develop the forecast for NSW.[2]

Motivated by the belief that collective efforts can affect meaningful change on a global scale, Team Echo's dedication to bridging gaps and breaking barriers propels them forward in their mission. Through their commitment to developing effective forecasting models, they aspire to empower consumers, drive informed decision-making, and pave the way for a brighter energy future in NSW. Through this paper, we aspire to embark on a journey of exploration and discovery, delving into models which could open avenues for further research using our hybrid model as basis to assist in controlling cost and providing optimised outcomes to providers, users and policymakers.

[1] Short-term load forecasting (STLF)
[2] Further details of the process and choice of model is provided in the Discussion section

3 Literature review

3.1 Methodology for Literature review

Team followed the below process for literature review and modelling:

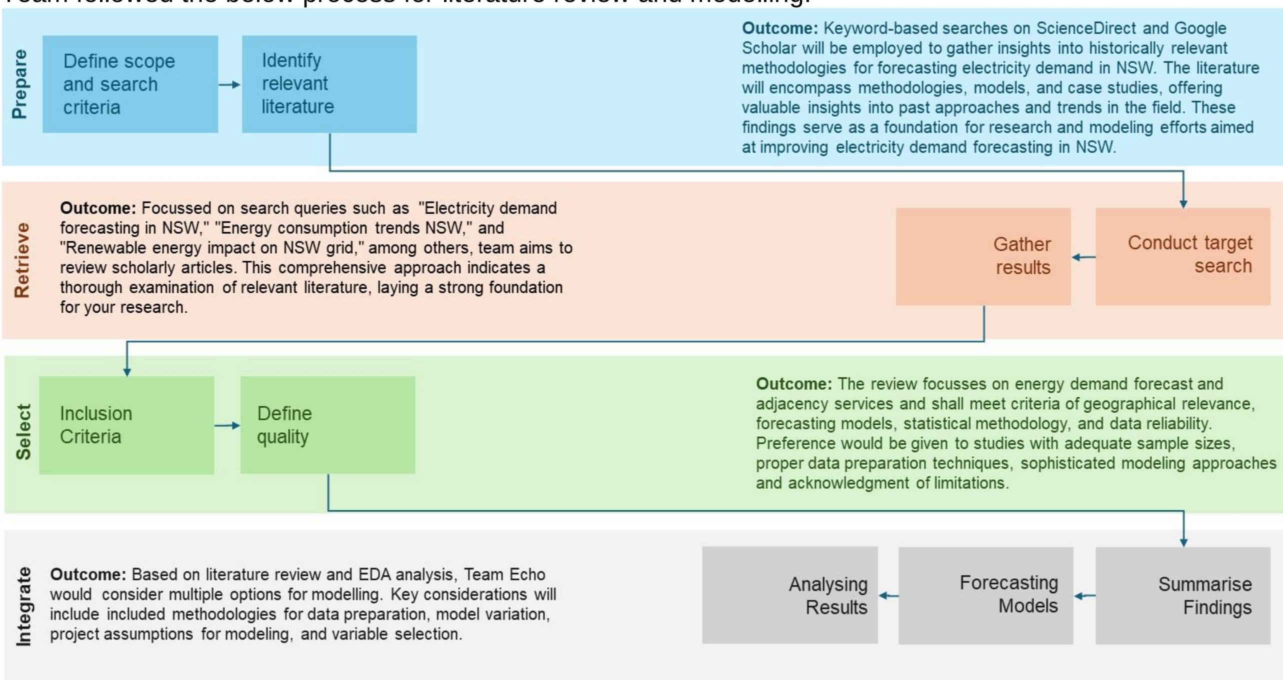


Figure 1: Process flow for literature review

3.2 Conventional Models

In energy consumption forecasting, conventional models such as time series, regression, and gray models are standard. Researchers have sought to improve their accuracy by refining their structures or integrating them with advanced methods. This has led to the creation of various enhanced conventional models and hybrid combinations.

3.2.1 Time Series models

Time series (TS) models are forecasting methods that utilize historical demand data without requiring external variables, as highlighted by (A. Azadeh, 2015) . These models, such as the autoregressive (AR) and moving-average (MA) models, are commonly employed in energy consumption forecasting due to their simplicity and effectiveness. They require only a small amount of historical data for construction, many TS approaches represent regression models since the predicted value is estimated based on one or more previous values.

(Xu & Wang, 2010) focussed on improving the MA model with a second order polynomial curve, producing the polynomial curve and MA combination projection model (**PCMACP**). Results demonstrate the PCMACP model's superior reliability, with a Mean Absolute Percentage Error (MAPE) lower than previous methods such as back propagation neural network (BPNN) and GM.

This category includes univariate time series models such as autoregressive moving average (ARMA) models, Popular other techniques are autoregressive integrated moving average (ARIMA)[3] (Wiki Article, 2024) models for non-stationary time series (by introducing a lag), seasonal autoregressive integrated moving average (SARIMA) models for seasonality and ARMA models for exogenous variables (ARMAX). A typical multivariate TS method is vector auto-regression and includes smoothing models and ARCH

[3] The parameters in the ARIMA model can be determined by optimisation algorithms, such as Box-Jenkins, GA and particle swarm optimisation (PSO) (Meer van der, Shepero, Svensson, Widen, & Munkhammar, 2018)

techniques. To achieve accurate forecasting results, external variables should be considered and introduced into conventional models.

Researcher/Scientist	Models developed
(A. Azadeh, 2015)	Forecasted electricity consumption in Iran with moving average (MA) to make data trend-free to train the ANN
(Abdel-Aal & Al-Garni, 1997)	forecasted the monthly electricity consumption of Saudi Arabia with univariate Box-Jenkins analysis and ARIMA.
(Ervural, Beyca, & Zaim, 2016)	proposed a forecasting method integrating genetic algorithm (GA) and ARMA for the natural gas consumption of Istanbul (Turkey)
(Hussain, Rahman, & Memon, 2016)	forecasted the monthly electricity consumption of Pakistan using the univariate Box-Jenkins analysis and ARIMA.
(Pappas, et al., 210)	used ARMA model to forecast the electricity load forecasting of the Hellenic power system.
(He & Lin, 2018)	integrated mixed data sampling model with autoregressive distributed lag to forecast China's long-term energy demand.
(Crompton & Wu, 2005)	forecasted energy consumptions in China using the BVAR model and considered the effect of macroeconomic variables.
(Zhu, Wang, Zhao, & Wang, 2011)	integrated the MA-C-WH procedure with SARIMA with weight coefficients, and thereafter optimised the coefficients using an hybrid PSO algorithm for China. This model exhibited a better prediction accuracy.

Table 1: referenced papers for time series models

3.2.2 Regression models (RMs)

Regression model is a statistical approach used to understand the relationship between one or more independent variables (predictors) and a dependent variable (outcome). It aims to estimate how changes in the independent variables are associated with changes in the dependent variable. In essence, a regression model quantifies the effect of the independent variables on the dependent variable, allowing for prediction, inference, and understanding of relationships within data. LogR was one the first models applied in energy forecasting (Forouzanfar, Doustmohammadi, Menhaj, & Hasanzadeh, 2010).

Researcher/Scientist	Models developed
(Forouzanfar, Doustmohammadi, Menhaj, & Hasanzadeh, 2010)	offers two parameter estimation methods: nonlinear programming (NLP) and genetic algorithm (GA) to forecast natural gas consumption in Iran underlying LogR. The forecasting considers both yearly and seasonal consumption patterns, providing insights into long-term and short-term trends.
(Thatcher, 2007)	introduces a linear regression model that correlates regional electricity demand with climate variables to capture intraday demand variability. This model enables prediction of regional load duration curves (LDCs) for four Australian states participating in National Electricity Market (NEM).
(H. Fan and I.F. MacGill and A.B. Sproul, 2015)	introduced linear regression model incorporating factors such as household demographics, behaviour, building and appliances, and climate impact, using a dataset that included half-hour interval electricity readings and survey data from 9903 households from Australia's Smart Grid Smart City (SGSC) project, providing good accuracy in forecasting average daily electricity demand.
(Radharani Panigrahi, 2022)	developed a Machine Learning Categorical Boosting (ML CatBoost) Regressor model to predict hourly electricity demand using data from the Electricity Reliability Council of Texas and introduced two hybrid nonlinear models blending a linear model with an artificial neural network (ANN). These highly accurate hybrid models effectively addressed heteroscedasticity in input data, resulting in reduced errors for multi-step-ahead forecasting.
(Limanond, Jomnonkwao, & Srikaew, 2011)	introduced two novel hybrid nonlinear models blending a linear model with an artificial neural network (ANN), specifically designed to enhance forecast accuracy. These models are adept at handling heteroscedasticity in input data, resulting in reduced errors for multi-step-ahead forecasting. Higher forecast accuracy was achieved considering the model captured complex

	non-linear relationships effectively. Integration of heteroscedastic variations into the input layer of the hybrid univariate model significantly enhances modelling accuracy for multi-step-ahead forecasting

Table 2: referenced papers for regression model

3.2.3 Gray Models (GM)

Gray forecasting, a non-statistical method suitable for nonlinear time series, is valuable when data are limited. GM is commonly used in energy consumption forecasting due to its simplicity and accuracy, integrating with MLR and seasonal factors for forecasting precision. Conventional GM's limitations in handling nonhomogeneous exponential sequences led to development of self-adapting GM. GM optimized by ordinary least squares (OLS) method and ant lion optimizer (ALO) algorithm was used to predict China's natural gas demands.

Nonlinear Gray Bernoulli model (NGBM) and Grey Verhulst Model (GVM) (Shaikh, Ji, Shaikh, Mirjat, & Uqaili, 2017) are models derived from gray models to deal with the nonlinear and fluctuant data patterns. The models utilise the specific power exponent function to manifest the nonlinear characteristics of the data pattern and determine the characteristics of the model curve.

Researcher/Scientist	Models developed
(Bianco, Manca, Nardini, & Mineaa, 2010)	adopted a trigonometric GM with rolling mechanism (TGMRM) to estimate electricity consumption. The two models lead to similar results, with an average deviation of less than 5%
(Kumar & Jain, 2010)	predicted the consumption of coal and electricity (in utilities) consumption in India using a GM with rolling mechanism and singular spectrum analysis (SSA)
(Wang, Du, Lu, & Yang, 2018)	indicated that traditional rolling GM (1,1) cannot always achieve satisfactory forecasting results because of the different trends or characteristics of samples
(Liu, Moreno, & García Ana Salomé, 2016)	designed a hybrid model that combined the GM and BPNN to predict the primary energy consumption, this model is called as Grey Neural Network, and Input-Output Combining Forecasting Model (GNF-IO model)

Table 3: referenced papers for gray models

3.3 AI-based models

In energy consumption forecasting, Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Random Forest (RF) are prominent AI-based models. However, ANN and SVR are particularly popular and extensively applied for solving a wide range of forecasting problems in this domain.

3.3.1 ANN-based models

Artificial Neural Networks (ANN) provide a framework for various machine learning algorithms to collaborate and process intricate data inputs. In the realm of energy consumption forecasting, ANN-based models encompass a variety of architectures, including feedforward neural networks (FFNN), backpropagation neural networks (BPNN), adaptive network-based fuzzy inference systems (WNN), Echo State Networks (ESN), and deep learning (DL) models.

3.3.1.1 FFNN and BPNN

Feedforward Neural Networks (FFNN) transmit information in a unidirectional manner, from input nodes through hidden layers to output nodes, while Backpropagation Neural Networks (BPNN) additionally propagate error differences backward through hidden layers (Kermanshahi, 1998). These conventional Artificial Neural Network (ANN) models have been extensively used in energy consumption forecasting, supported by studies such as. Optimisation algorithms applied to FFNN parameters have been found to reduce forecasting errors, enhancing their effectiveness in energy consumption prediction tasks.

Researcher/Scientist	Models developed
(Szoplik, 2015)	examined gas demand forecasting using artificial neural networks (ANN), specifically multilayer perceptron (MLP) models, trained on actual natural gas consumption data from Szczecin, Poland. Time and temperature are considered in the model to capture the impact on gas consumption. The

	study identifies the MLP 22-36-1[4] model as effective for predicting gas consumption throughout the year and at any time of day.
(Pao, 2009)	proposed a hybrid model that combine Feedforward Neural Network (FFNN) with Linear Regression (LR) to improve forecasting accuracy compared to conventional FFNN models. Results showed that MAPE for the hybrid models were consistently below 5%. The flexibility of hybrid ANNs to capture complex nonlinear relationships contributes to their superior performance and higher accuracy.
(Chae, Horesh, Hwang, & Lee, 2016)	introduced a short-term building energy usage forecasting model, utilising a Feedforward Neural Network (FFNN) optimised by the Bayesian regularisation algorithm. The model aims to forecast day-ahead electricity usage in 15-minute intervals. Key predictors such as day type indicator, time-of-day, HVAC set temperature schedule, outdoor air dry-bulb temperature, and outdoor humidity are identified through variable importance analysis

Table 4: referenced papers for FFNN and BPNN

3.3.1.2 ANFIS

ANFIS, or Adaptive Neuro-Fuzzy Inference System, merges the adaptive capabilities of neural networks with the interpretability of fuzzy logic systems, making it a powerful tool for tackling problems with non-linear relationships between input and output variables. Particularly relevant in energy demand forecasting due to its applicability to complex, multi-variable systems, ANFIS models offer advantages such as transparency, interpretability, and adaptability across diverse data types. Although computationally intensive and requiring meticulous parameter tuning, ANFIS-based models have demonstrated superior accuracy compared to traditional methods like ARIMA and standalone artificial neural network (ANN) models, as evidenced by lower Mean Absolute Percentage Error (MAPE) values.

A notable advancement in energy consumption forecasting integrates ANFIS with stochastic frontier analysis (SFA) and fuzzy data envelopment analysis (FDEA) algorithms, enhancing long-term natural gas consumption predictions. Further refinements involve incorporating computer simulation's random number generation to capture intricate input data behaviours, thus elevating forecast accuracy. Studies emphasize ANFIS's efficacy as a component in hybrid models, significantly enhancing prediction precision.

Researcher/Scientist	Models developed
(Yang, Chen, Wang, Li, & Lian, 2016)	proposed a combined forecasting approach integrating Back Propagation (BP) neural network, Adaptive Network-based Fuzzy Inference System (ANFIS), and Difference Seasonal Autoregressive Integrated Moving Average (diff-SARIMA). Initially, each method (BP, ANFIS, and diff-SARIMA) is employed separately to generate forecasts. Subsequently, optimal weight coefficients are determined for each forecasting result, and these results are aggregated by multiplying the coefficients with the respective forecasts and summing them. This process yields the final forecasting results. BP and ANFIS are chosen for their ability to handle nonlinear data, while diff-SARIMA is selected for its proficiency in addressing linear and seasonal data patterns. The combined method outperforms individual methods, demonstrating enhanced accuracy in forecasting by effectively reducing errors between actual and forecasted values.
(Azadeh, Ghaderi, & Sohrabkhani, 2008)	The study presents an integrated algorithm for forecasting monthly electrical energy consumption in Iran, combining artificial neural network (ANN), computer simulation, and design of experiments with stochastic procedures. Initially, an ANN approach is utilized for forecasting, which is compared to a time series model. Computer simulation generates random variables to simulate the effects of probabilistic distribution on monthly electricity consumption, leading to the development of a simulated-based ANN model. Analysis of variance (ANOVA) is conducted on four treatments: actual data, time series, ANN, and simulated-based ANN, to determine the best model. The algorithm

[4] MLP 22-36-1 refers to a specific configuration of a multilayer perceptron (MLP) neural network. In this notation, "22" indicates the number of neurons in the input layer, "36" represents the number of neurons in the hidden layer, and "1" signifies the number of neurons in the output layer.

	offers flexibility in model selection based on ANOVA and mean absolute percentage error (MAPE) results. It recognizes conventional time series as a potential optimal model for forecasting, challenging the notion that ANN consistently provides the best estimates.

Table 5: referenced papers for ANFIS

3.3.2 DL models

Recent advancements in deep learning (DL) models have led to their ability to produce forecasting predictions comparable to or superior to human experts, especially in domains like stock market index forecasting. Recurrent Neural Network (RNN) stands out as a key DL model for energy consumption forecasting, utilising its internal state to process input sequences and improve prediction accuracy. (He W. , 2017) combined RNN with Convolutional Neural Network (CNN) to successfully forecast electricity load in a northern Chinese city, with CNN extracting crucial historical data features and RNN capturing dynamic load patterns.

Long Short-Term Memory (LSTM), a specialized form of RNN, (Graves, 2012), (Wei N. , Li, Duan, Liu, & Zeng, 2019) has been shown to outperform traditional models, addressing issues such as vanishing or exploding gradients in the back-propagation process. Additionally, integrating sequence-to-sequence architecture into LSTM models enables the estimation of loads for future time steps, enhancing prediction performance compared to other models such as Artificial Neural Networks (ANN), Support Vector Regression (SVR), and traditional RNNs.

Researcher/Scientist	Models developed
(Miljanovic, 2012)	The purpose of this paper is to perform evaluation of two different neural network architectures used for solving temporal problems, i.e. time series prediction. The data sets in this project include Mackey-Glass, Sunspots, and Standard & Poor's 500, the stock market index. The study also presents a comparison study on the two networks and their performance
(Kermanshahi, 1998)	investigates long-term load forecasting using two artificial neural networks: a recurrent neural network (RNN) for one-year ahead predictions and a three-layer feed-forward back-propagation (BP) for forecasts spanning 5 and 10 years. Trained on data from 1975 to 1994 and tested on subsequent years, the models demonstrate reasonable accuracy both within and beyond the training period. To address the absence of weather data beyond two weeks, a sensitivity program is devised to estimate future temperatures. The study concludes by conducting load forecasts for nine utilities in Japan
(He W. , 2017)	presents a novel Deep Neural Network (DNN) architecture for short-term load forecasting, leveraging multiple input features. It employs Convolutional Neural Network (CNN) components to extract features from historical load sequences and Recurrent Components to model implicit dynamics. Dense layers are used to process other types of features. Experimental results on a large dataset from a city in North China demonstrate the effectiveness of the proposed method, which is also adaptable to other time series prediction tasks.
(Wei N. , Li, Duan, Liu, & Zeng, 2019)	introduces a hybrid model for daily natural gas load forecasting, combining Principal Component Correlation Analysis (PCCA) with Long Short-Term Memory (LSTM) networks. PCCA extracts relevant factors while removing redundancies from the dataset. LSTM is utilised for load prediction. Using recent data from Xi'an (China) and Athens (Greece), incorporating 14 weather factors, the proposed PCCA-LSTM model outperforms other methods such as PCA-LSTM, BPNN, and SVR, demonstrating the lowest mean absolute percentage errors.
(Qing & Niu, 2018)	proposes a model for day-ahead solar irradiance prediction using weather forecasting data, addressing the challenge of obtaining historical irradiance data for on-site photovoltaic generation. The model employs long short-term memory (LSTM) networks to capture temporal dependencies. This model achieves 18.34% improvement in root mean square error (RMSE) compared to feedforward neural networks. The model demonstrates better generalisation capability and reduced overfitting, improving accuracy.

Table 6: referenced papers for DL model

3.3.3 SVR-based models

A superior performing model, in energy consumption forecasting as highlighted in (Shine, Murphy, & Scully, 2018), tackles challenges like local minima and overfitting by employing Support Vector Regression (SVR) with time series reconstruction properties. Additionally, the Least Square Support Vector Machine (LSSVM), an improved version of SVM, transforms the quadratic programming problem into a linear equation via a new quadratic loss function. This transformation enhances accuracy while reducing computational burden. (Shayeghi, Ghasemi, Moradzadeh, & Nooshyar, 2015) further enhanced this concept, developing a hybrid LSSVM model for estimating electricity load.

Researcher/Scientist	Models developed
(Cao & Wu, 2016)	highlights the limitations of Support Vector Regression (SVR) due to challenges in parameter selection and handling seasonal effects. To address these issues, the model proposes a hybrid approach that integrates SVR with the Fruit Fly Optimization (FOA) Algorithm and seasonal index adjustment. Evaluation using data from China and the United States demonstrates the effectiveness of the hybrid model, indicating its potential for improving electricity consumption forecasting.
(Ding, Zhang, & Yuan, 2017)	Introduced GA-SVR and GA-WD-SVR prediction models for short-term and ultra-short-term cooling load forecasts in office buildings. These forecasts aid in optimizing HVAC system operations for energy efficiency. The GA-SVR model predicts short-term cooling loads for the next day, while the GA-WD-SVR model predicts ultra-short-term loads for the next hour. Utilizing meteorological data and historical cooling load records, the models are trained and evaluated on an office building in Tianjin. Results demonstrate the superior performance of the GA-WD-SVR model for ultra-short-term predictions.
(Yaslan & Bican, 2017)	a hybrid method combining Empirical Mode Decomposition (EMD) and Support Vector Regression (SVR) was proposed. This method does not rely on specific Intrinsic Mode Functions (IMFs) for denoising and model learning. Experimental findings indicate the superiority of this hybrid method over traditional SVR techniques and non-feature-based denoised-SVR methods in forecasting electricity load demand.
(Sujjaviriyasup, 2017)	proposed a hybrid forecasting model that combined MODWT, SVM, and DE optimization. The model combined SSA, SVR, and cuckoo search (CS) algorithm for electric load forecasting. This model outperformed traditional methods and other hybrid models in forecasting annual electricity consumption, offering superior accuracy and precision.

Table 7: referenced papers for SVR models

3.4 Conclusion commentary on Models

Energy consumption forecasting is a pivotal aspect of energy management across various sectors, facilitating informed decision-making, optimised resource allocation, and enhanced energy efficiency. Traditional statistical methods like time series analysis, regression analysis, and Gray forecasting have long served as foundational tools for forecasting energy consumption. While effective, these methods may struggle to capture the intricate nonlinear relationships present in energy consumption data, as we have seen during the EDA and data pre-processing stages.

The emergence of artificial intelligence (AI) and machine learning (ML) has further revolutionised forecasting, offering more sophisticated models capable of handling complex data patterns and nonlinear relationships. AI-based models such as Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Random Forest (RF) have demonstrated promising results in improving forecasting accuracy compared to conventional methods. Leveraging large datasets and advanced computational techniques, these models can discern intricate patterns and make precise predictions.

Hybrid models, which amalgamate multiple forecasting techniques, represent another significant advancement in forecasting accuracy. By combining statistical methods with machine learning algorithms or integrating various machine learning techniques, hybrid models can capture a broader spectrum of data patterns and relationships, leading to more accurate forecasts.

Central to our methodology is the utilisation of time series models, which enable us to analyse historical demand data and identify trends and patterns over time. Given the seasonality inherent in energy consumption data, we have incorporated Fourier transforms to effectively capture periodic fluctuations and ensure robust forecasting. This hybrid approach combines the strengths of SARIMA (Seasonal Autoregressive Integrated Moving Average) in capturing time series patterns with the Fourier transforms' ability to effectively handle seasonality in the data. Furthermore, acknowledging the significant impact of temperature on energy demand, we have chosen to incorporate temperature data as a crucial input variable in our models. By integrating temperature square into our analysis, we aim to capture the nonlinear relationship between temperature and energy consumption, thereby enhancing the accuracy of our forecasts.

Team Echo was inspired by our review of multiple papers on using SARIMA-Fourier model. The model extends traditional SARIMA framework by incorporating Fourier terms to capture seasonal patterns in the data. By decomposing the time series data into seasonal components using Fourier transforms, the model can better account for periodic fluctuations and improve forecasting accuracy. In another approach, Fourier terms are added to the SARIMA model to explicitly model the seasonal patterns in the data. By incorporating Fourier terms as regressors in the SARIMA framework, the model can effectively capture the seasonal variations and produce more accurate forecasts.

These examples highlight how integrating Fourier transforms with SARIMA models can improve the accuracy of energy consumption forecasts by better capturing seasonal patterns and fluctuations in the data. By leveraging the strengths of both approaches, our hybrid Fourier-SARIMA model aims to provide Endgame Economics with reliable and accurate forecasts essential for policy formulation and grid stability in NSW.

4 Exploratory Data Analysis

This section embarks on an in-depth exploration of electricity use in NSW. We integrate data on demand, temperature changes, forecasted electricity demand, and other important external factors. The objective is to summarise these disparate data points to provide a better understanding of the factors shaping electricity demand in NSW.

4.1 Objective

Our analytical journey is guided by several key objectives:

- To uncover key trends and seasonality in electricity demand in NSW, identifying when and how demand fluctuates.
- To investigate how temperature changes affect how much electricity is used.
- To assess the accuracy of electricity demand forecasts against actual consumption data, pinpointing areas of improvement for future predictions.
- To incorporate additional data, recognizing their potential impact on electricity consumption patterns.

4.2 Data Sources

Our examination is supported by a comprehensive dataset comprising:

Electricity Demand Data includes detailed, half-hourly electricity demand records for NSW from January 2010 to March 2021

Forecast Demand Data includes half-hourly increments for NSW.

Air Temperature Data includes air temperature in NSW. The time interval for each observation is not consistent in this dataset.

Solar Exposure and Rainfall Data were obtained from the *Bureau of Meteorology*[5]

Public Holidays Data includes information on public holidays in Australia[6]. (*Plazzer, M., n.d.)*

Gross Domestic Productivity includes the GDP for Australia[7]

4.3 Data Loading

The forecast data was provided in two large parts. To facilitate easier management and integration with R, we divided this data into five smaller, zipped files. This division was necessary to comply with GitHub's file size constraints and to streamline the process of loading the data into R for subsequent analysis.

4.4 Data Cleaning

Identify and eliminate null values and handling missing data, were two key areas of data cleaning activities the team performed.

Identification and Removal of Null Values: An anomaly was detected in the Solar Exposure Data, where a singular instance of a missing value (NA) was identified. Given the potential impact of missing data on the accuracy of our analysis, this NA value was promptly removed to maintain data integrity.

Handling Missing Data in Rainfall Amounts: The Rainfall Amount Data presented a more significant challenge, with 93 instances of missing values (NAs) identified. In this scenario, simply removing these values was not viable due to the volume of missing data. Instead, a strategic approach was taken by replacing these missing values with the median of the dataset. The choice of the median, specifically a

[5] Bureau of Meteorology (BOM), n.d. Climate Data Online. Available at: http://www.bom.gov.au/climate/data/index.shtml
[6] Plazzer, M., n.d. Australian Public Holiday Data. Available at: https://www.michaelplazzer.com/datasets/australian-public-holiday-data/
[7] World Bank, n.d. GDP (current US$) - Australia. Available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=AU

value of zero, was driven by its robustness to outliers, ensuring that our replacement strategy did not introduce bias or distort the underlying distribution of the data.

4.5 Data Transformation

Inconsistency amongst various datasets including missing values, additional column data, and non-matching time scales wee few updates the team performed for data transformation.

Temperature data: Since the frequency of temperature data is non-constant, our data cleaning phase involved aligning it with the electricity demand dataset, which is recorded at consistent 30-minute intervals. To synchronize these datasets, we adjusted temperature timestamps to align with the nearest demand interval, particularly for times within the 45 to 14 minutes and 15 to 44 minutes ranges. This method allowed us to aggregate temperature readings and calculate mean temperatures for each interval, enabling a coherent analysis of how temperature variations impact electricity demand and enhancing the dataset's robustness. During this process, we identified 559 instances where temperature data was missing for existing electricity demand entries. Instead of excluding these data points, we implemented a method to estimate missing temperature values based on available data. This estimation was conducted by interpolating between known data points, effectively filling in gaps and preserving the continuity and completeness of our dataset for robust statistical analysis.

Forecast Data (Benchmark): We employed a refined data cleaning approach to ensure accuracy in our analysis. Each record in the dataset includes a `LASTCHANGED` timestamp indicating the most recent update. To eliminate duplicates and maintain data integrity, only the entries with the latest `LASTCHANGED` timestamp for each `DATETIME` were retained. This method ensures that our analysis is based on the most current and relevant data available. Similar to the temperature data, the Forecast Demand dataset also had instances of missing values. These gaps were addressed by interpolating between known data points, which provided a continuous and complete dataset for analysis.
Adjusting Data Scales: Our analysis includes data like GDP, rainfall, and solar exposure, which come in different scales, such as yearly and daily measurements. We carefully adjusted these to ensure they could be accurately compared and analysed together, maintaining data consistency.

Restructuring GDP Data: The original GDP data was spread across multiple columns, each representing a different year. We reorganized this data into a more manageable format by merging these columns into a single 'Year' column. This step made it easier to integrate the GDP data with other time-sensitive information in our analysis.

4.6 Exploratory Analysis

4.6.1 Descriptive Statistics

The electricity demand data comprises 196,513 half-hourly records from NSW, showing a minimum demand of 5,075 MW and a maximum of 14,580 MW. The first quartile is at 7,150 MW, the median at 8,053 MW, and the third quartile at 8,959 MW, with an overall average demand of approximately 8,113 MW.

```
   DATETIME              TOTALDEMAND        REGIONID
 Length:196513        Min.   : 5075    Length:196513
 Class :character     1st Qu.: 7150    Class :character
 Mode  :character     Median : 8053    Mode  :character
                      Mean   : 8113
                      3rd Qu.: 8959
                      Max.   :14580
```

The temperature dataset for New South Wales consists of 220,326 records, with temperatures ranging from a minimum of -1.3°C to a maximum of 44.7°C. Key statistics include a first quartile of 13.4°C, a median of 17.7°C, and a mean slightly lower at 17.42°C, indicating a symmetric distribution with a slight skew towards cooler temperatures. The third quartile is at 21.3°C, reflecting the typical upper range of temperatures observed.

```
   LOCATION              DATETIME              TEMPERATURE
 Length:220326         Length:220326         Min.   :-1.30
 Class :character      Class :character      1st Qu.:13.40
 Mode  :character      Mode  :character      Median :17.70
                                             Mean   :17.42
                                             3rd Qu.:21.30
                                             Max.   :44.70
```

4.7 Univariate Analysis – Demand

The time series plot shows the fluctuation of electricity demand in New South Wales over 11 years. A clear seasonal pattern is observed, with peaks during winter and summer months, reflecting higher heating and cooling needs.

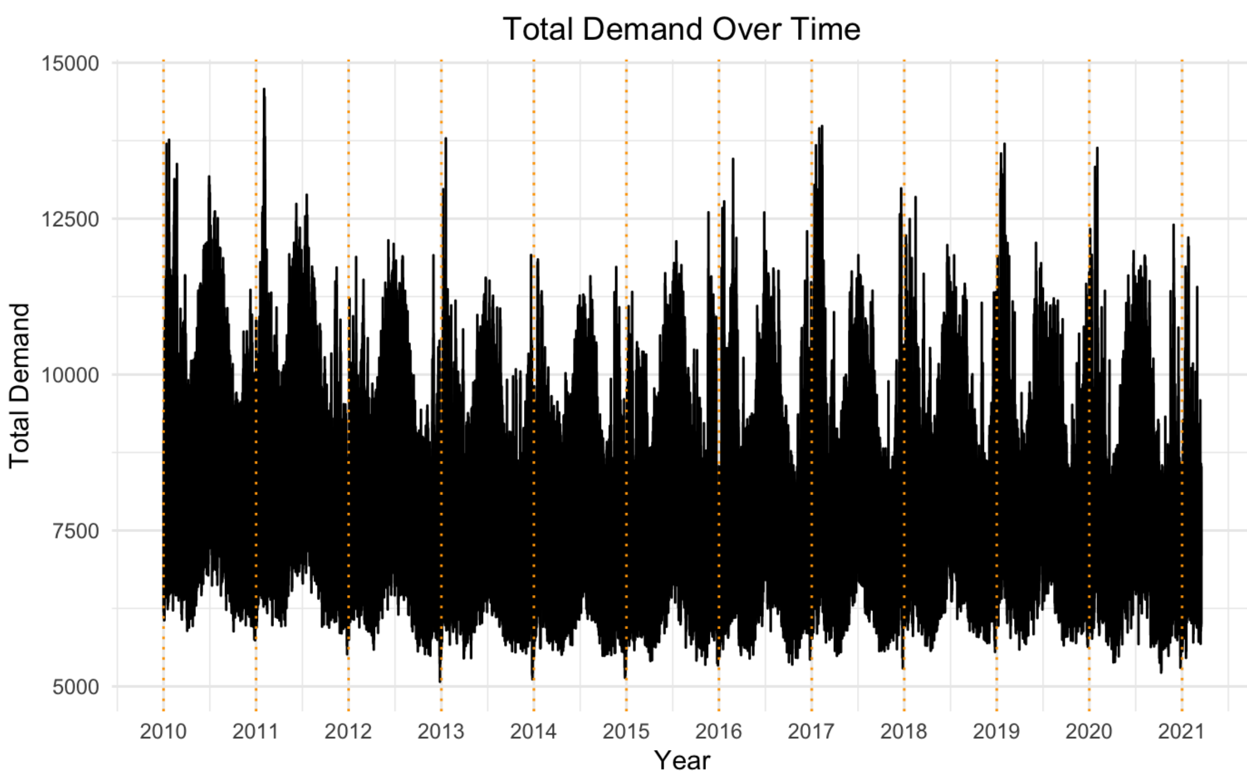


Figure 2: Fluctuation of electricity in NSW

Further illustrations of plots provide details of demands across defined time.

4.7.1 Weekly-of-the-month demand pattern

This set of line graphs illustrates the average electricity demand throughout a typical week in each month of the year, with lines representing different weeks within each month. Upon examination of the graphs, several insights can be drawn:

Daily Patterns: There is a consistent daily pattern across all months, with demand increasing during the day and peaking in the early evening before dropping off at night. This likely reflects increased household and business activity during daylight hours and reduced activity overnight.

Seasonal Variations: There are some differences in the daily patterns across the months, which seem to be influenced by seasonal changes. We will dive deeper into this in Figure 3.

Weekly Consistency: Within each month, the weekly patterns remain relatively consistent, although there are slight variations in demand from week to week. This consistency indicates a stable electricity usage pattern within a month across different weeks.

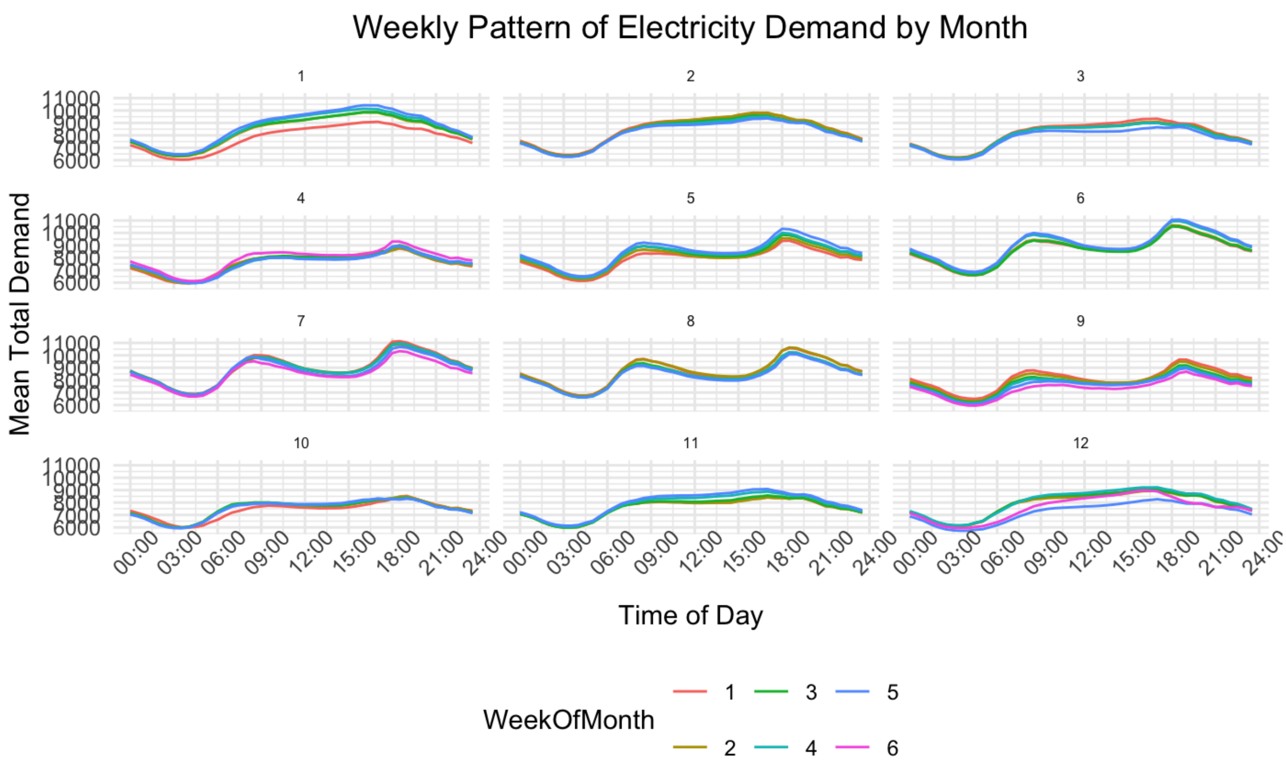


Figure 3: Weekly demand patterns of historic data.

4.7.2 Seasonal demand

This figure provides a focused examination of the daily electricity demand patterns across different seasons, which were hinted at in our earlier monthly analysis. The box plots for each season reveal the following patterns:

Autumn (Red) and Spring (Green): These seasons exhibit relatively consistent electricity demand with slight increases in the evening hours, likely reflective of the typical residential consumption patterns during these milder months.

Summer (Cyan): The demand profile during Summer shows a peak in midday to afternoon demand, establishing the anticipated increase in electricity use for cooling as the day reaches its highest temperatures.

Winter (Purple): Winter shows distinct peaks in demand in the early morning and evening, likely due to heating requirements during the colder parts of the day, which aligns with our initial observations of potential seasonal impacts on electricity use.
The differentiation in daily patterns by season underscores the influence of weather-related factors on electricity consumption.

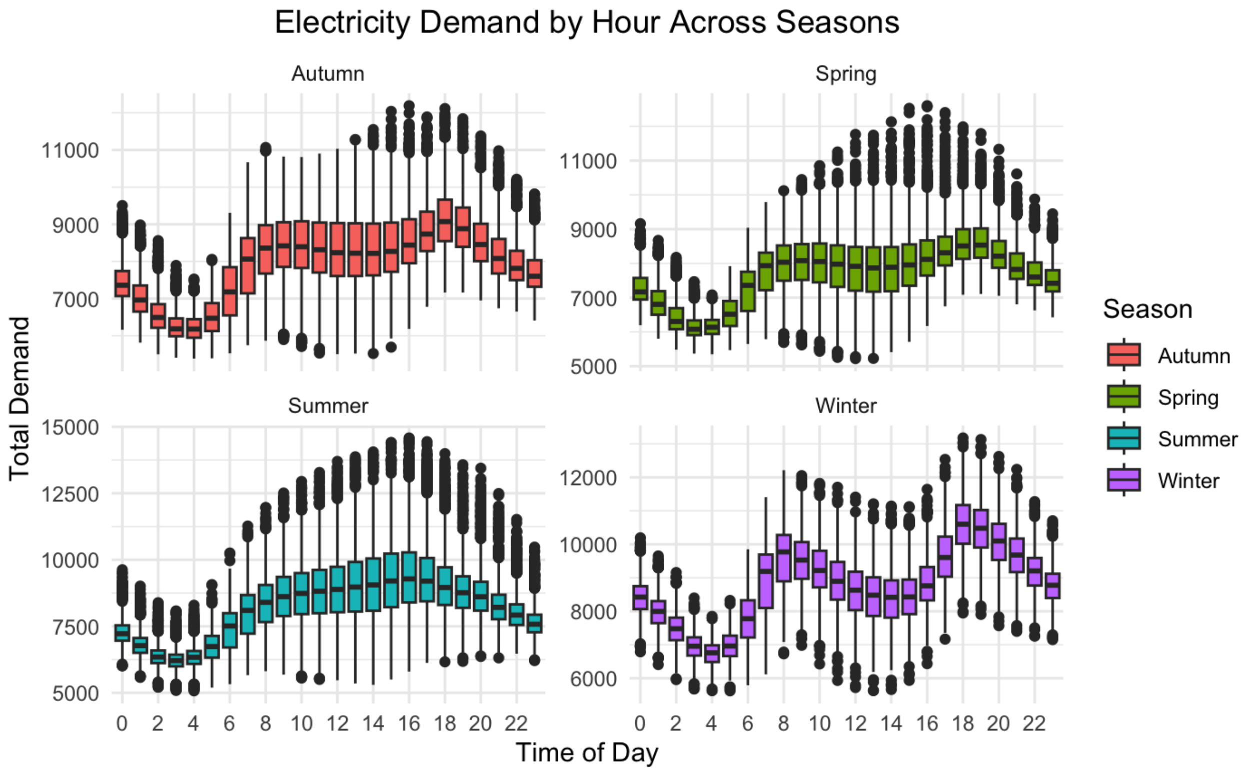


Figure 4: Seasonal demand of historic data

4.7.3 Average demand during a 24-hour period

This line graph compares the average electricity demand at each hour throughout the day for different days of the week. The following patterns can be observed:

Weekday Demand (Monday to Friday): There is a pronounced double-peaked pattern on weekdays, with electricity demand rising sharply in the morning, tapering slightly midday, and then peaking again in the early evening. This likely reflects routine domestic and commercial activities, such as people getting ready for work and school in the morning and then returning home and using various appliances in the evening.
Weekend Demand (Saturday and Sunday): The weekend pattern, while still showing two peaks similar to the weekdays, has less overall demand.

Overall Trends: Across all days, there is a significant drop in demand late at night, when most residential and commercial activities cease, until early morning when the demand begins to rise with the start of a new day.

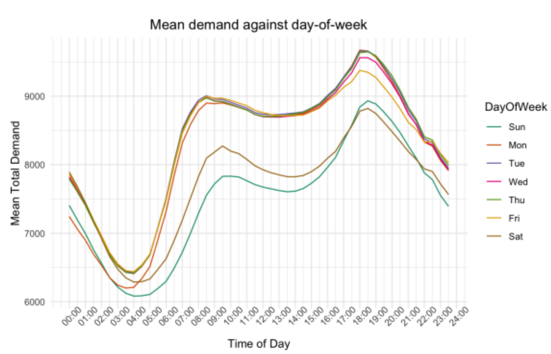


Figure 5: Mean demand plotted against day-of-week

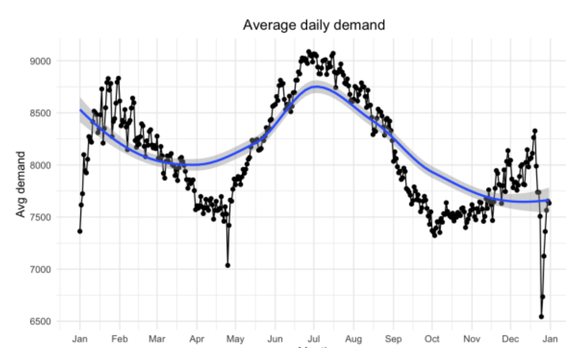


Figure 6: Average daily demand

4.8 Univariate Analysis - Temperature

The temperature line graph indicates strong seasonality correlation across the years of data provided. The pattern and peak of usage have been consistent across multiple years.

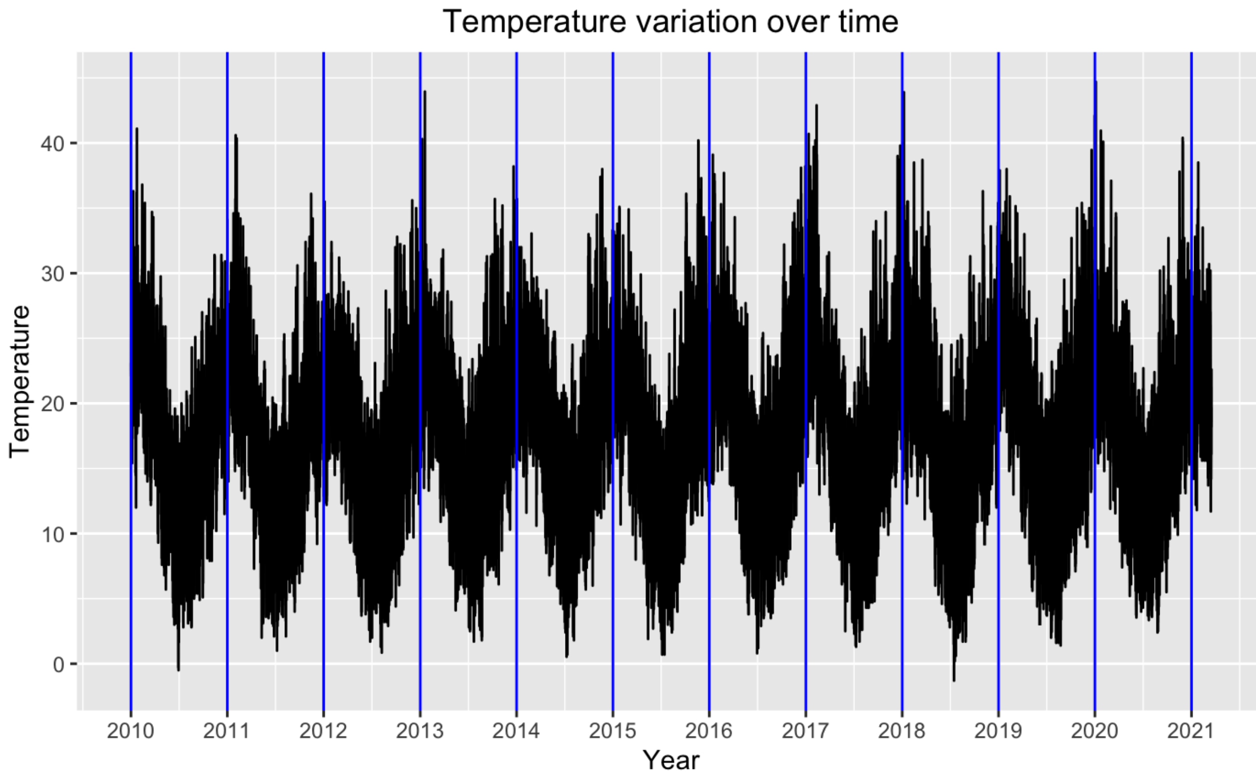


Figure 7: Temperature variation over time

4.9 Bivariate Analysis - Demand vs Temperature

To identify strong correlation for temperature, identify the p-value with the demand. Pearson's test has been considered.

```
#p-value for Demand vs Temp
> cor.test(df_nsw$TEMPERATURE, df_nsw$TOTALDEMAND, use = "pearson")
```

```
Pearson's product-moment correlation

data:  df_nsw$TEMPERATURE and df_nsw$TOTALDEMAND
t = 66.701, df = 195952, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1446675 0.1533262
sample estimates:
   cor
   0.1489997
```

Null hypothesis is rejected based on small p-value for temperature. Temperature has correlation to demand requirement.

The scatter plot shows a distinct U-shaped relationship between temperature and electricity demand. The demand is clear indication that higher (beyond 30 degrees) and lower (under 10 degrees) temperatures significantly increases the electricity demand. Lowest demand is observed during 20 degrees Celsius. The temperature variability clearly shows the reliability of air conditioning during the extreme hot and extreme cold temperatures. The scatter plot below illustrates a discernible U-shaped correlation between temperature and electricity demand. Notably, both exceptionally high (above 30 degrees Celsius) and exceptionally low (below 10 degrees Celsius) temperatures coincide with a significant increase in electricity demand. Conversely, the lowest demand is observed at approximately 20 degrees Celsius. This observation underscores the impact of temperature extremes on electricity consumption, particularly evident in the reliability of air conditioning during periods of extreme heat and cold. Moreover, it highlights that moderate temperatures correspond to a more moderate demand level.

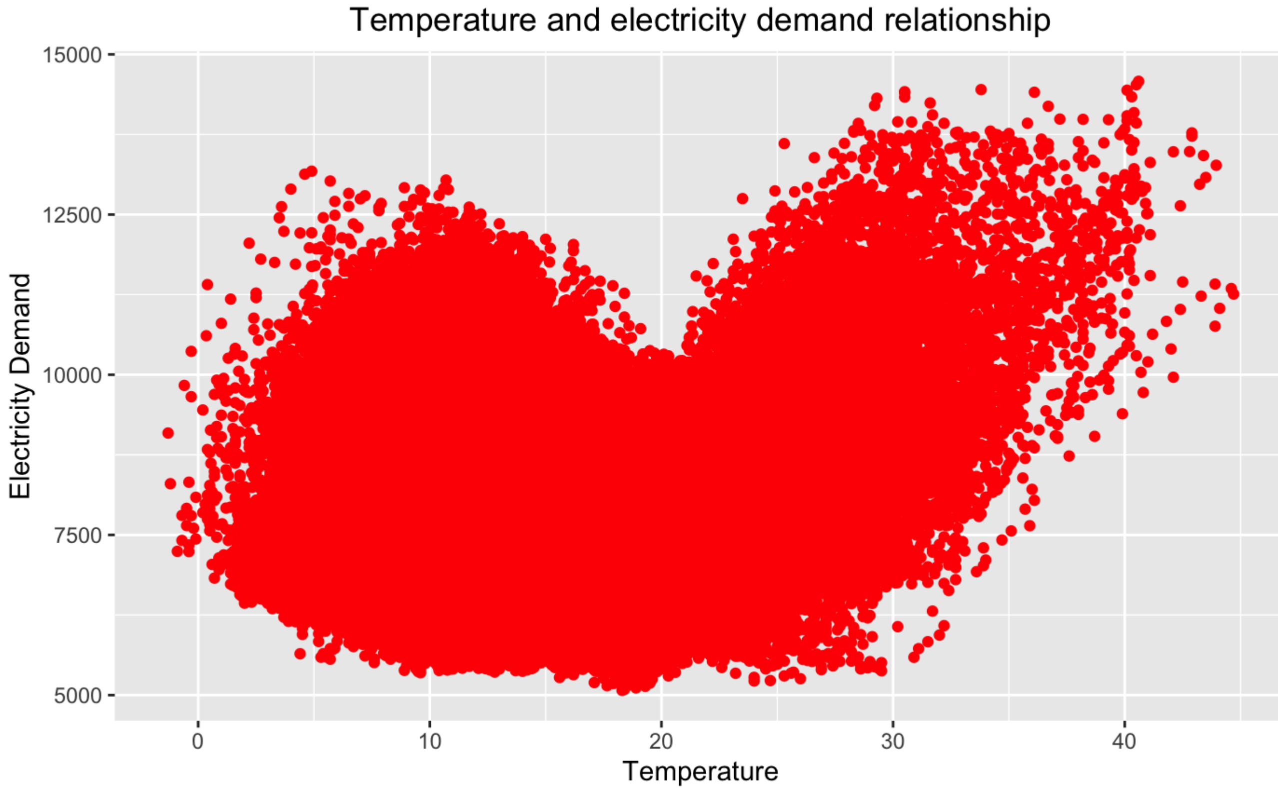


Figure 8: Temperature versus electricity demand relationship.

4.10 Bivariate Analysis – Holiday on Demand

The box plot compares electricity demand on holidays versus non-holidays. It appears that the median electricity demand on holidays is lower than on non-holidays. There are several extreme values that deviate from the typical ranges on both types of days. As part of our analysis, holidays did not have impact on demand.

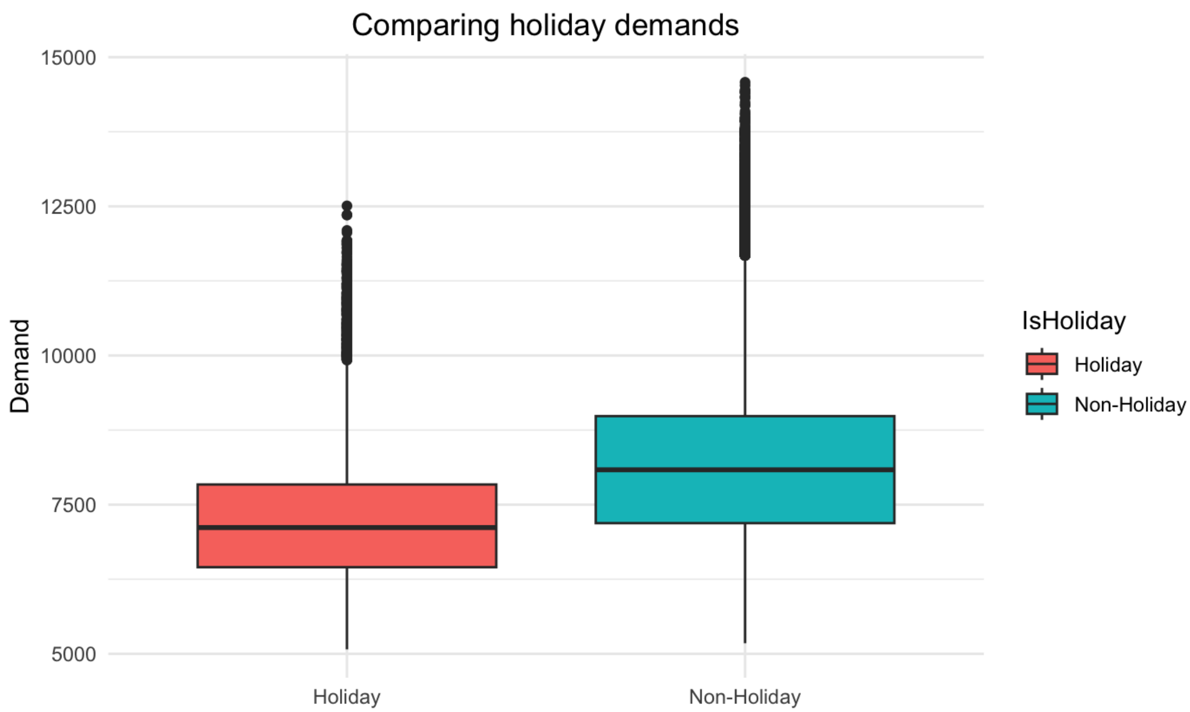


Figure 9: Demand during holidays.

4.11 Actual Demand vs Forecast Demand

To further evaluate the pattern, random days were selected in a single year (2017) The plot compares Forecast Demand and Total Demand over six randomly selected days in 2017. The key findings are:

- **Midnight Demand Forecast:** During the midnight hours, the forecast consistently predicts lower demand compared to the actual demand. This underestimation suggests that the activity during these hours is not fully captured by the forecasting model, indicating potential for improvement by incorporating factors that may lead to increased demand during these times.
- **Evening Demand Forecast:** In the evening hours, particularly from 4 PM to 8 PM, the forecast often overestimates the actual demand. This could indicate that the model is perhaps weighing certain indicators too heavily, which are not as influential as predicted in driving demand during these hours.

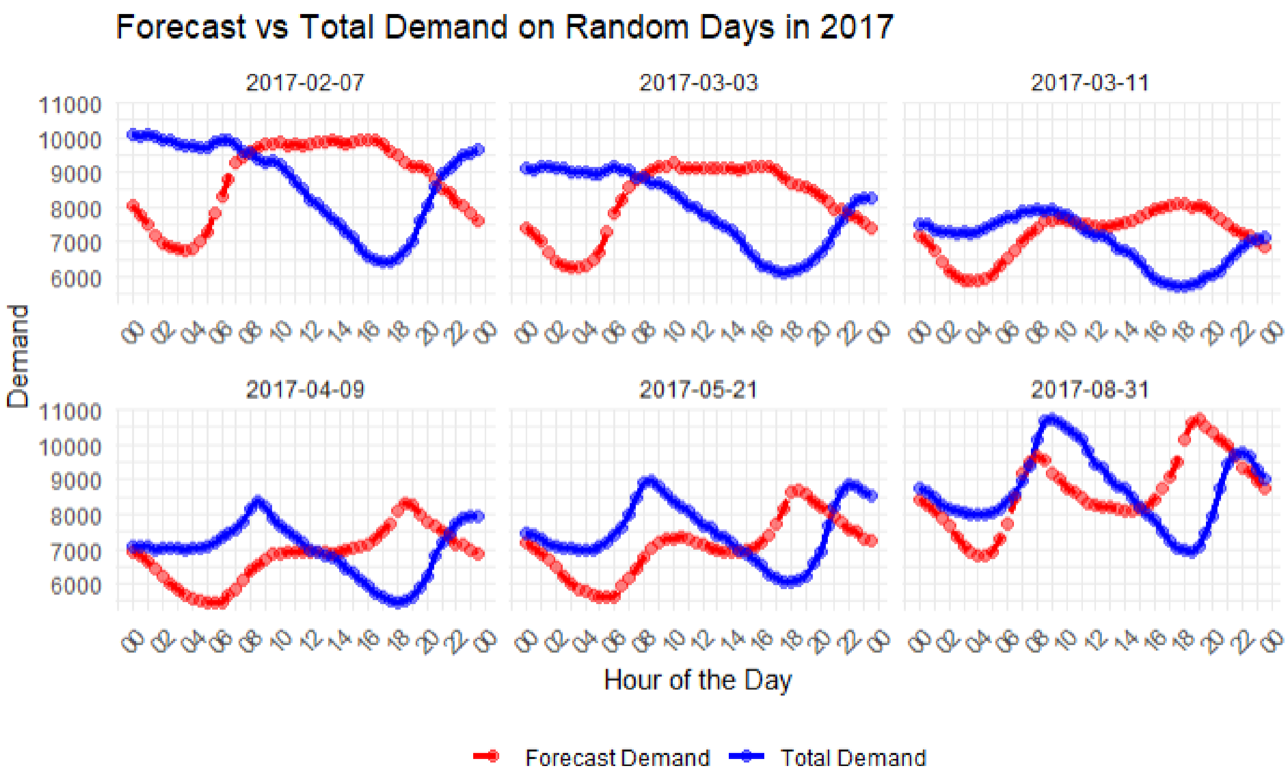


Figure 10: Forecast vs Total Demand (2017)

Interestingly, shifting the benchmark 10 hours backwards aligns it more closely with actual demand. This adjustment significantly improves the benchmark's predictive performance and, as a result, will be adopted for future benchmark comparison.

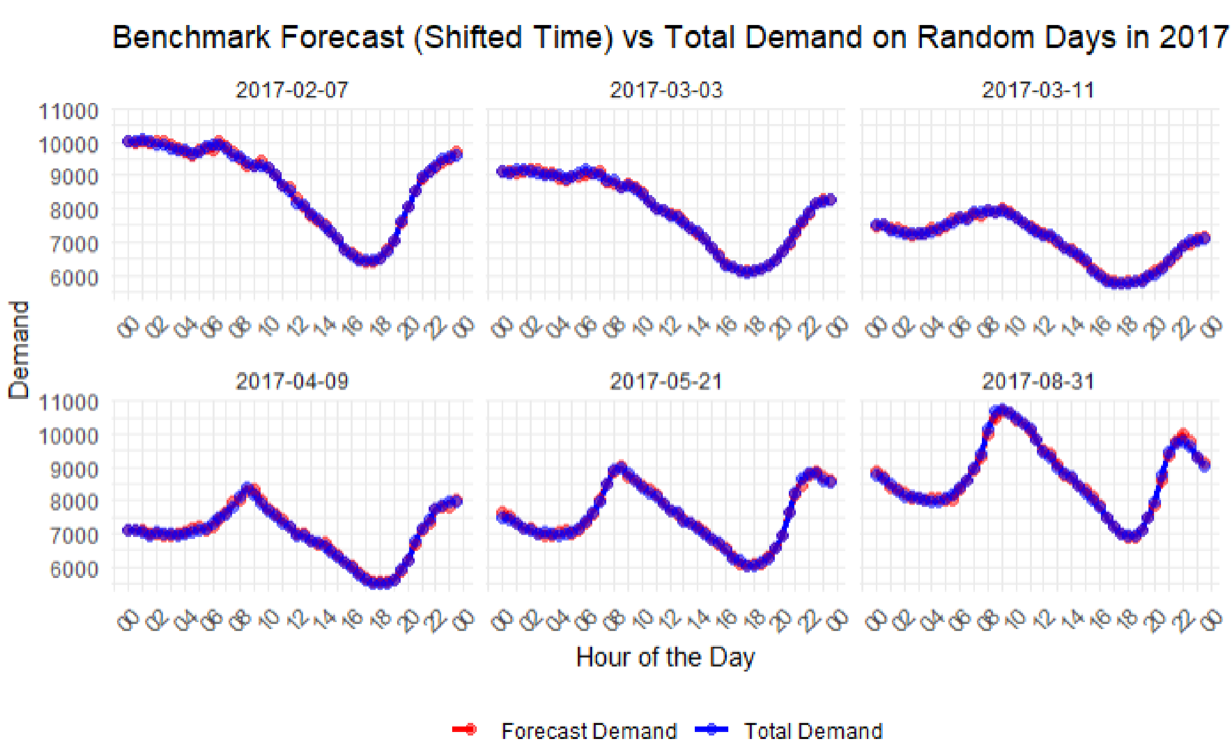


Figure 11: Time Shifted Forecast vs Demand (2017)

4.12 Daily Demand vs Solar Exposure

These three scatter plots display the relationship between solar exposure and electricity demand (maximum, mean, and minimum daily demand) in New South Wales. In each plot, there is a trend line suggesting that as solar exposure increases, electricity demand first decreases and then increases slightly. This pattern could imply that more sunlight reduces the need for artificial lighting and heating, up to a point, after which demand rises potentially due to cooling needs. Each plot exhibits a wide dispersion of data points, indicating variability in demand at similar levels of solar exposure.

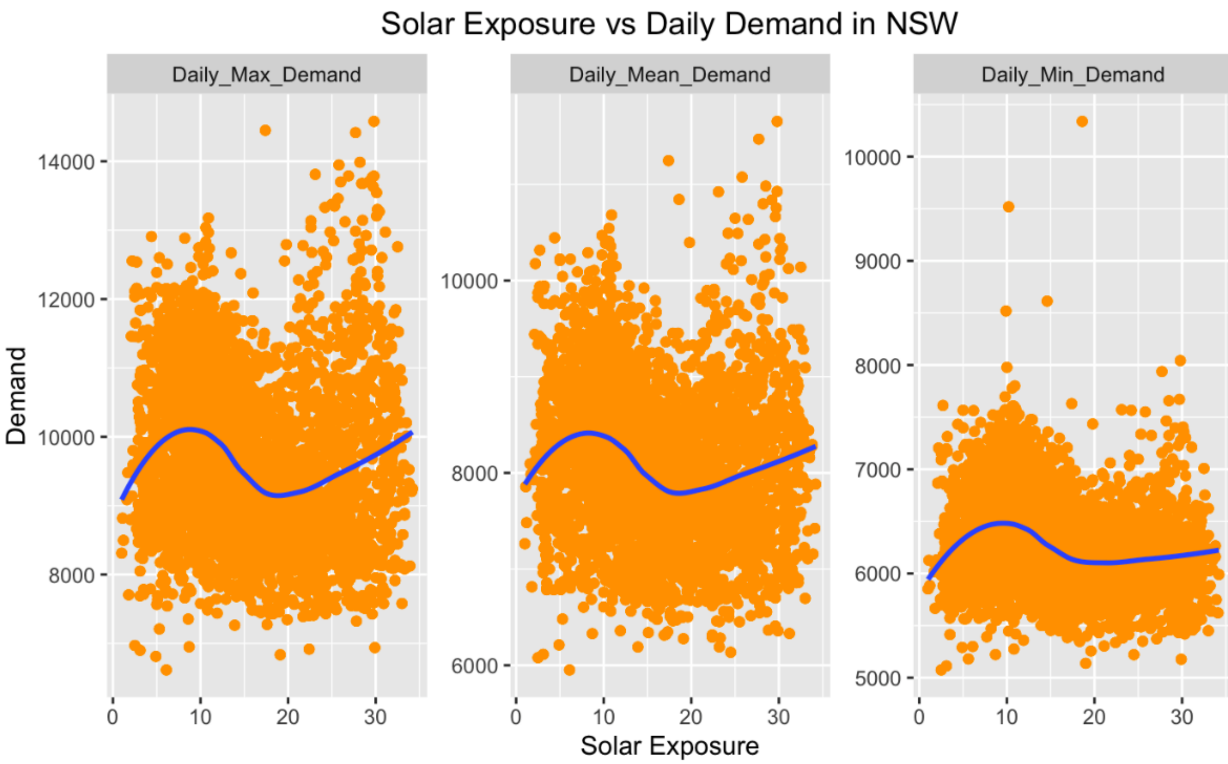


Figure 12: Solar exposure vs daily demand.

4.13 Daily Temperature vs Solar Exposure

The scatter plots correlating solar exposure with temperature in New South Wales for maximum, mean, and minimum daily temperatures show a clear pattern that supports our previous statement. As solar exposure increases, temperature also rises, which is consistent across all three measures of temperature. The trend lines indicate a non-linear relationship with a marked increase in temperature at higher levels of solar exposure, which is expected as more solar radiation generally results in warmer conditions.

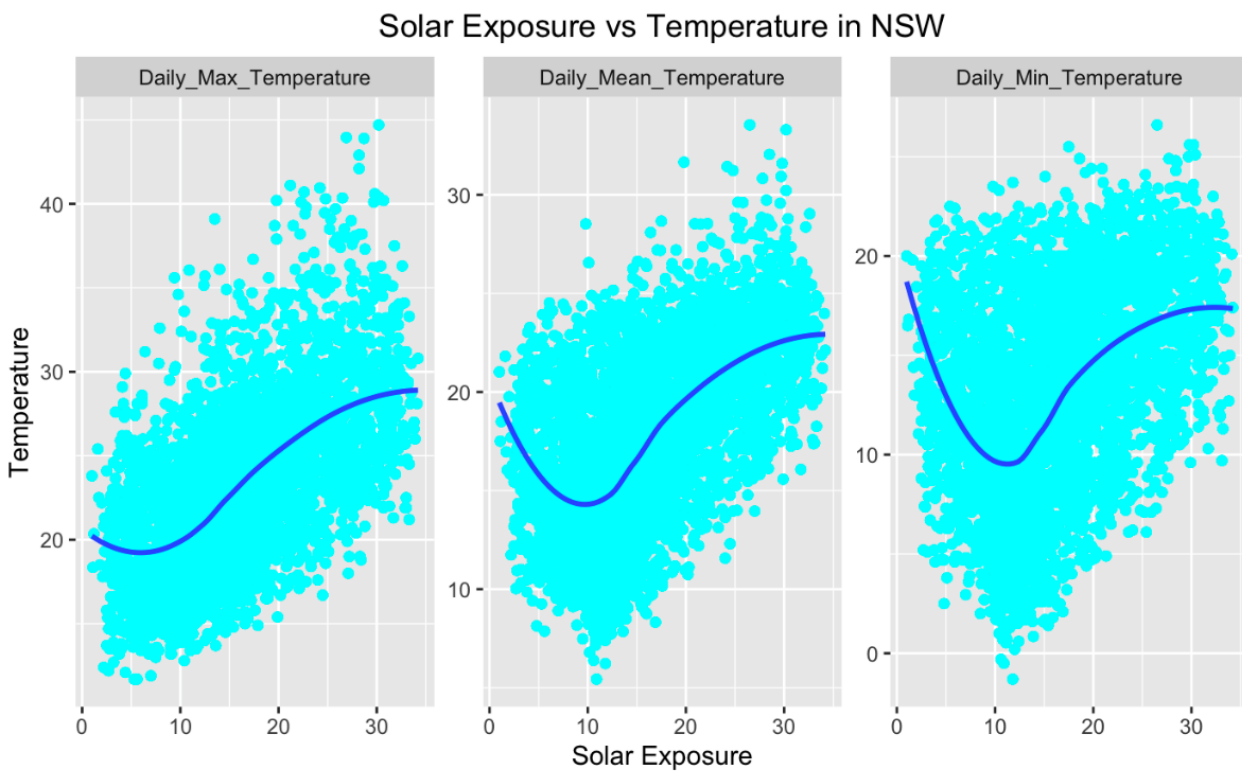


Figure 13: Solar Exposure vs Temperature.

4.14 Correlation of Different Time Frequencies and Demand

The heatmap analysis focusing on "TOTALDEMAND" reveals several correlations with other variables:

- A moderate positive correlation (0.48) with "TimeOfDay" indicates that demand peaks at certain hours, supporting our understanding of daily consumption patterns.
- A moderate negative correlation (-0.21) with "Year" suggests a potential decrease in demand over the years covered in the dataset, possibly due to improvements in energy efficiency or shifts in consumption habits.
- The correlation with "Month" at -0.12, although weak, hints at slight seasonal variations in demand, consistent with previous observations.
- Weak positive correlations with "DayOfWeek" (0.04) and "IsHoliday" (0.12) suggest that the day of the week and holidays have minimal impact on total demand in this dataset.

Overall, the data highlights hourly patterns as the most significant factors influencing "TOTALDEMAND." However, correlations with "Year" and "Month" also indicate potential long-term and seasonal trends warranting further investigation.

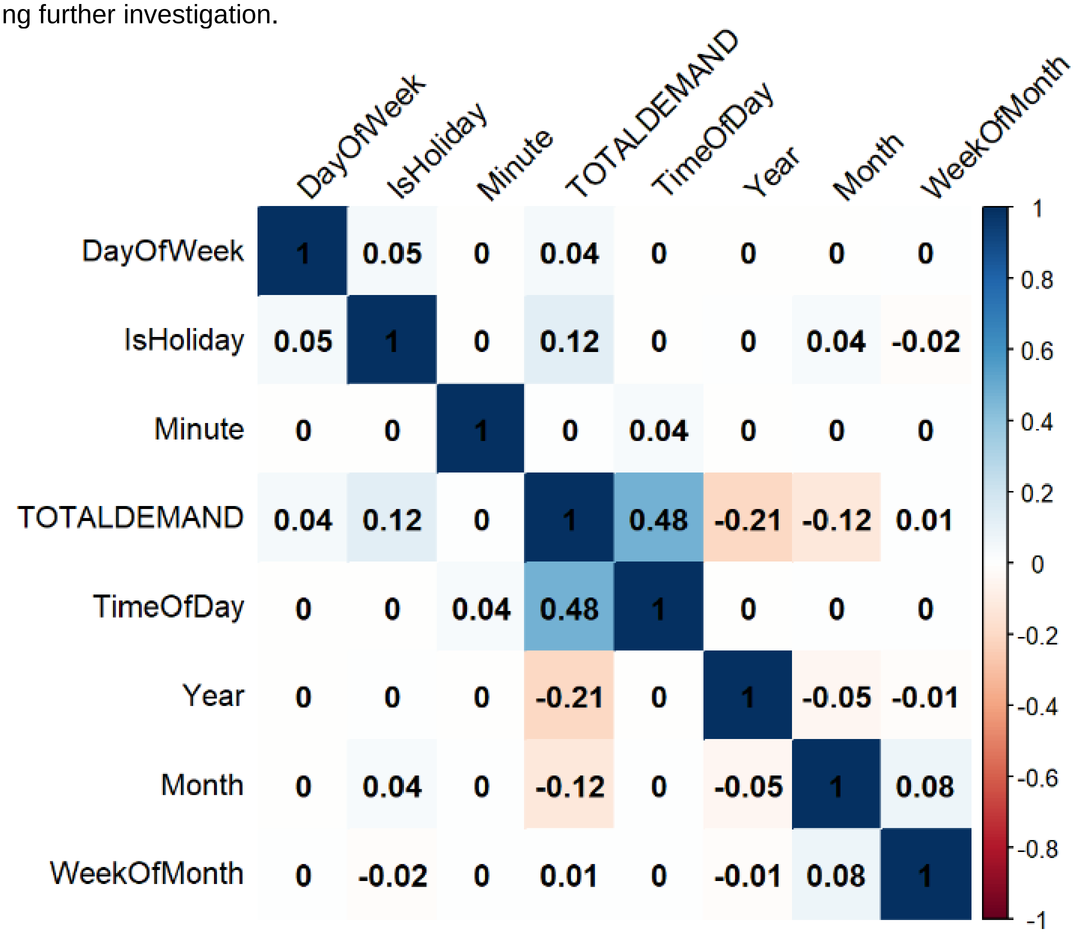


Figure 14: Correlation – Time vs Demand

4.14.1 Correlation of Demand and Daily Frequencies Variables

Solar Exposure: There's a strong positive correlation between solar exposure and temperature variables. It's quite expected since higher solar exposure typically leads to higher temperatures. Solar exposure shows a slight negative correlation with energy demand indicators, suggesting that higher solar exposure could potentially lead to lower energy consumption, perhaps due to reduced heating needs or greater solar energy generation. The negative correlation between solar exposure and rainfall suggests that increased solar exposure is typically associated with drier conditions.

Rainfall's Impact: Rainfall shows very weak correlations with most variables, suggesting it has a minimal linear relationship with both temperature and energy demand in this dataset.

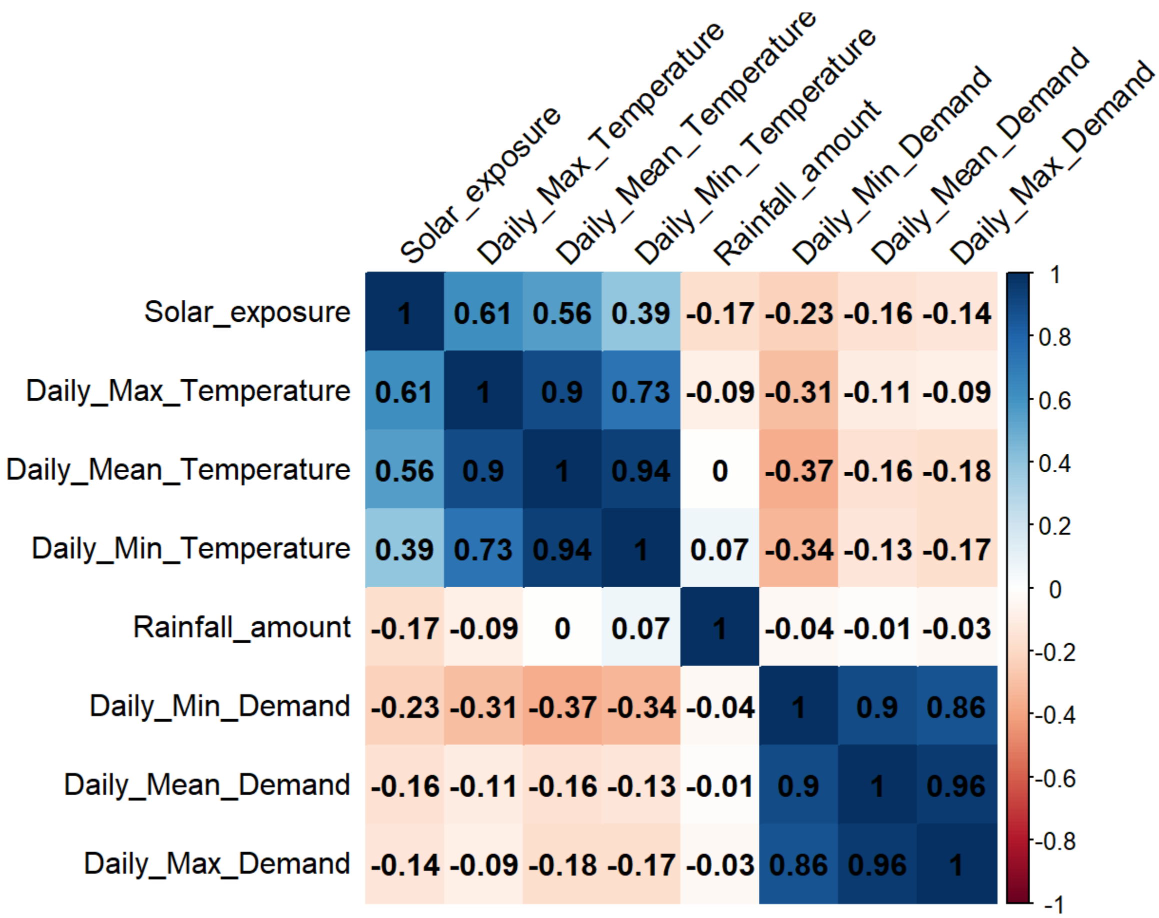


Figure 15: Correlation - Demand vs *v*ariables

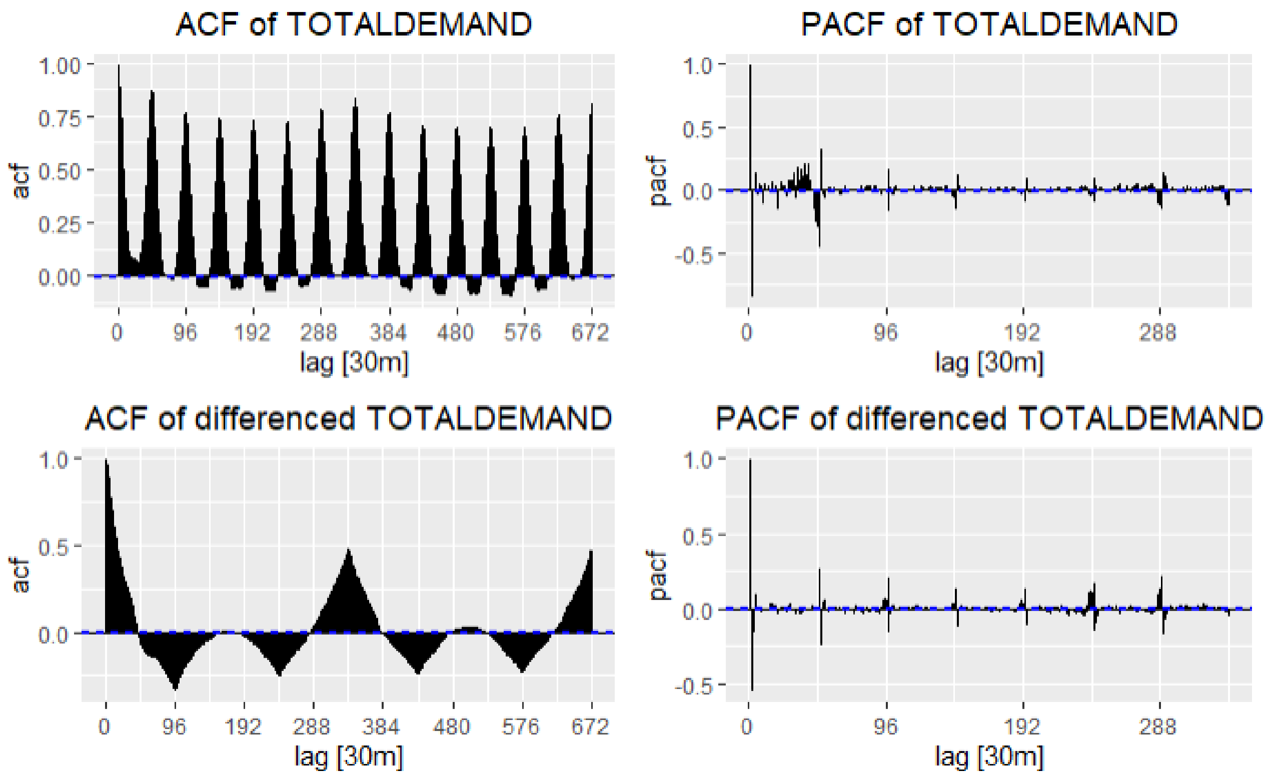
4.15 Autocorrelations

ACF: Shows strong and consistent pattern of autocorrelation that persists across multiple lags (48, 96,144...), with a periodicity that suggests a strong daily seasonal pattern in the data.

PACF: The PACF plot exhibits significant autocorrelation at the initial lag, suggesting a substantial 30-minute influence on electricity demand. Additionally, within the range of 0 to 48 lags, there are several significant spikes, which indicate that, aside from the strong immediate past influence, there are other influential lags within the first 24-hour period. These significant lags between 0 to 48 underscore the presence of shorter-term cyclic patterns within the daily cycle

ACF seasonally differenced: Comparing the ACF of the non-differenced TOTALDEMAND data to the ACF of the differenced data, there is a notable reduction in daily seasonality after differencing, as the high-frequency fluctuations observed initially have been mitigated. However, while the daily seasonality is no longer evident, the ACF of the differenced data reveals the persistence of other seasonal patterns, most notably on a weekly basis, as indicated by the spike at lag 336. Additionally, the identified lags at 96 and 240 in the ACF of the differenced TOTALDEMAND data suggest the presence of additional seasonality beyond the weekly pattern seen at lag 336. Specifically, the lag at 96 indicates a 2-day seasonal pattern, while the lag at 240 points to a 5-day seasonality.

PACF seasonally differenced: The PACF of the seasonally differenced TOTALDEMAND data presents a pronounced autocorrelation at the initial lag, followed by a notable decay. This decay is then interrupted by spikes at every 48th lag (e.g., 48, 96, 144, etc.), and at one lag prior to these points (e.g., 47, 95, 143, etc.). These spikes indicate a significant autocorrelation at a daily cycle and the immediate time step before that cycle. It suggests that even after removing the broader seasonal trend, there is a remaining pattern where the demand is not only influenced by the demand of the previous day at the same time but also slightly by the demand 30 minutes prior to that.



4.15.1 Stationarity

The KPSS unit root test was applied to the electricity demand data to assess its stationarity, revealing that the data is non-stationary, both in terms of seasonal and non-seasonal components. This non-stationarity indicates the presence of unit roots, suggesting that the data is influenced by trends and seasonal patterns. However, after applying both seasonal and non-seasonal differencing to the dataset, a subsequent KPSS test confirmed that the data achieved stationarity. This transformation effectively removed the underlying trends and seasonal effects, stabilizing the mean and variance over time and making the data suitable for further time series forecasting and analysis.

Non-seasonal

```
> dfts_demand_nsw %>% features(TOTALDEMAND, unitroot_kpss)
# A tibble: 1 x 2
  kpss_stat kpss_pvalue
      <dbl>       <dbl>
1      66.3        0.01
>
> dfts_demand_nsw %>%
+     mutate(diff_TOTALDEMAND = difference(TOTALDEMAND)) %>%
+     features(diff_TOTALDEMAND, unitroot_kpss)
# A tibble: 1 x 2
  kpss_stat kpss_pvalue
      <dbl>       <dbl>
1 0.0000830         0.1
```

Seasonal

```
> dfts_demand_nsw %>%
+     features(TOTALDEMAND, unitroot_kpss)
# A tibble: 1 x 2
  kpss_stat kpss_pvalue
      <dbl>       <dbl>
1      66.3        0.01
>
>
> # 0.1 Stationary after first order differencing
> dfts_demand_nsw %>%
+     mutate(diff_TOTALDEMAND = difference(TOTALDEMAND, 48)) %>%
+     features(diff_TOTALDEMAND, unitroot_kpss)
# A tibble: 1 x 2
  kpss_stat kpss_pvalue
      <dbl>       <dbl>
1   0.00116         0.1
```

5 Modelling

This section of the report details the modelling process undertaken to forecast total demand over a 24-hour period using ARIMA-based models. Given the cyclic nature of demand influenced by time (daily and weekly patterns) and external factors such as temperature, several models were developed and assessed to identify the most effective approach for accurate forecasting.

5.1 Models Considered

A series of ARIMA models were configured to explore various influences on demand patterns:

- ARIMA with Fourier Terms (arima_fr): Incorporates seasonal decomposition using Fourier terms to model both daily and weekly demand fluctuations.
- Following Seasonal ARIMA (SARIMA) variants were included to choose the appropriate model;

Variants	Description
Basic SARIMA (sarima_212210)	utilises seasonal and non-seasonal differencing to stabilise the mean of the series
SARIMA with Fourier Terms (sarima_212210_fr)	enhances the basic SARIMA model by adding Fourier terms to capture weekly seasonal patterns
SARIMA with Temperature Adjustments (sarima_auto_temp)	introduces temperature and its squared term as explanatory variables, reflecting the impact of environmental factors
SARIMA with Temperature and Fourier Terms (sarima_212210_temp_fr)	combines both Fourier terms and temperature variables for a comprehensive model

Table 8: SARIMA variants considered for modelling.

The models were evaluated based on their statistical fit and forecast accuracy, with the following key outcomes:

5.1.1 Statistical Metrics

The SARIMA model with Temperature and Fourier Terms showed the lowest AIC and BIC values (615804 and 615937 respectively), indicating the best fit among the models tested.

```
  .model                 sigma2  log_lik     AIC    AICc     BIC
  <chr>                   <dbl>    <dbl>   <dbl>   <dbl>   <dbl>
1 sarima_212210_temp_fr   5354. -307887. 615804. 615804. 615937.
2 sarima_212210_fr        5357. -307900. 615825. 615825. 615941.
3 sarima_auto_temp        5375. -307992. 616002. 616002. 616082.
4 sarima_212210           5377. -308004. 616023. 616023. 616085.
5 arima_fr               10744. -326939. 653917. 653917. 654086.
```

Figure 16: Statistical metric comparison of models.

5.1.2 Forecast accuracy comparison for the models

Parameters	Description
RMSE	SARIMA with Temperature and Fourier Terms achieves 234, slightly higher than the SARIMA with Fourier Terms alone, which recorded the best RMSE of 230
MAE and MAPE	It recorded MAE of 207 and MAPE of 2.29%, demonstrating robust performance across various metrics. Despite the slightly better RMSE for the SARIMA with Fourier Terms model, the combined model's MAE and MAPE showcase its overall effectiveness in a balanced account of error metrics

Table 9: Accuracy comparison of various models

	.model	.type	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE	ACF1
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	sarima_212210_fr	Test	-169.	230.	203.	-1.87	2.25	0.629	0.544	0.817
2	sarima_212210_temp_fr	Test	-178.	234.	207.	-1.97	2.29	0.642	0.554	0.814
3	sarima_212210	Test	-249.	292.	251.	-2.70	2.72	0.776	0.690	0.849
4	sarima_auto_temp	Test	-251.	292.	252.	-2.72	2.73	0.781	0.691	0.847
5	arima_fr	Test	-94.2	607.	543.	-1.68	6.00	1.68	1.43	0.931

Figure 17: Accuracy for models considered.

5.2 Selection of the Final Model

The **SARIMA with Temperature and Fourier Terms** was selected as the final model. This decision was based on its comprehensive ability to incorporate significant seasonal and external influences effectively, thereby providing the most accurate and reliable forecasts.
The final model configuration includes the following factors;

Factors	Details
Seasonal Adjustments	pdq (2,1,2) for non-seasonal components and PDQ (2,1,0) for seasonal components with a period interval of 48, fine-tuned to best capture inherent demand cycles
External Factors	Temperature and its squared term are used to account for nonlinear impacts of environmental conditions on demand
Seasonal Decomposition	Fourier terms are specifically used to model weekly demand patterns, enhancing the model's ability to predict peak and off-peak variations accurately

Table 10: SARIMA with temperature and Fourier defining factors.

Mathematical equation for the SARIMA with temperature and Fourier

$$(1-\varphi_1 B-\varphi_2 B^2)(1-\Phi_1 B^{48}-\Phi_2 B^{96})(1-B)(1-B^{48})Demand_t = \alpha+\beta_1 * \text{Temperature}_t+\beta_2 * \text{Temperature}_t^2 + \beta_3 * \Sigma_{k=1}^{3}\left(a_k * \cos\left(\frac{2\pi kt}{336}\right)+b_k * \sin\left(\frac{2\pi kt}{336}\right)\right)+(1+\theta_1 B+\theta_2 B^2)\varepsilon_t$$

Where;

- t: Time index.
- α: Constant term.
- ε_t: Error term at time t.
- β: Coefficient for exogenous variable
- B: Backshift operator, where $B^k X_t = X_{t-k}$.
- φ_i: Coefficients of the autoregressive (AR) terms for the i-th lag.
- θ_i: Coefficients of the moving average (MA) terms for the i-th lag.
- Φ_i: Coefficients of the seasonal autoregressive terms for the i-th seasonal lag.
- Periodicities:
  - 48: Represents the periodicity for daily cycles
  - 336: Represents the periodicity for weekly cycles

Table 11: Mathematical model

5.3 Model Performance and Benchmarking

Rolling forecast was employed to assess the accuracy of the final model. This validation technique involved progressively re-estimating the model parameters as newer data is made available, mimicking practical forecasting scenarios. The model's performance was evaluated across five consecutive test days, each containing 48 observations.

5.3.1 Model Performance

The alignment between forecasted and actual demand values over multiple cycles suggests the model is effective at tracking the underlying seasonal trends and fluctuations in demand. There are also some noticeable discrepancies at certain points where the model fails to match the peaks and troughs precisely. In particular, the model occasionally underestimates the highest demand peaks and overestimates the lowest troughs. These variations could indicate potential areas for model refinement, such as adjusting for volatility or considering additional variables that may affect demand.

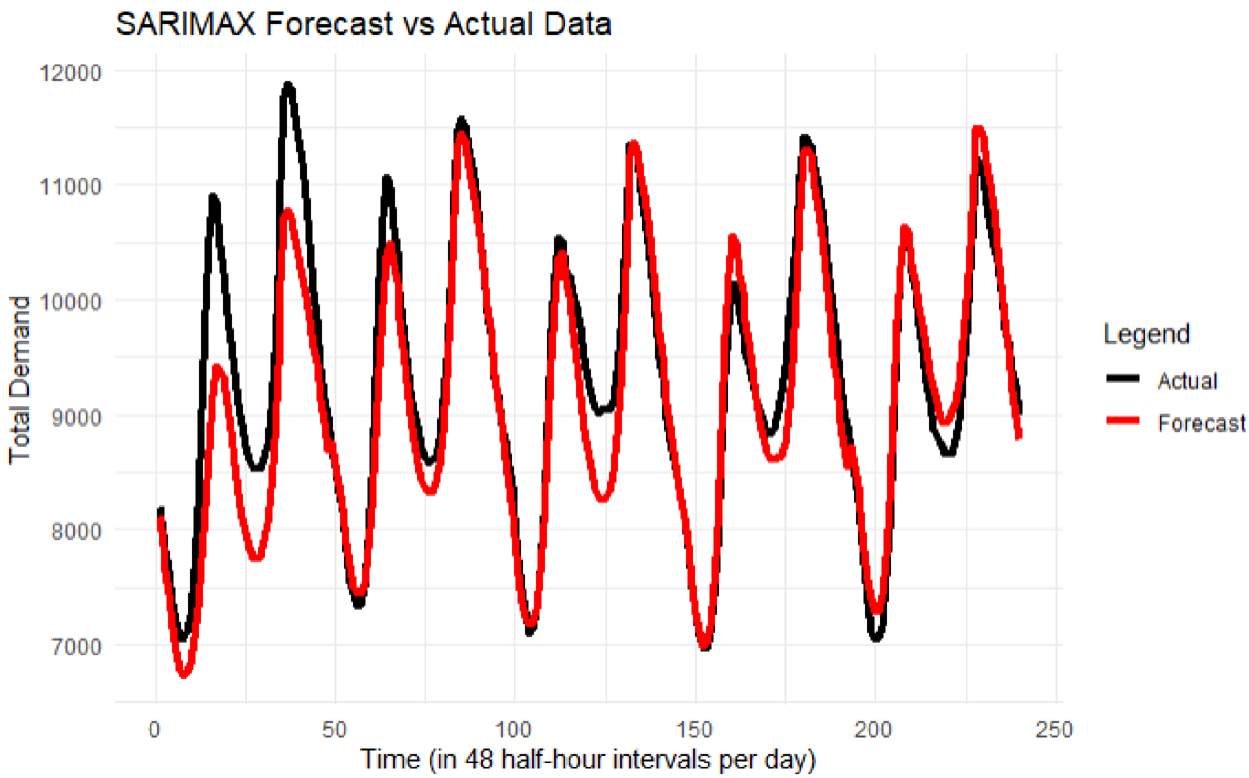


Figure 18: Model performance for 24 hours (48 intervals)

Comparatively, when measured against the benchmark provided, the model demonstrates a larger mean deviation in predictions, suggesting that while it effectively tracks overall demand trends, pinpointing the exact demand at given points could be further improved. The elevated ME, RMSE, and MAE metrics underscore this aspect. Additionally, the increased MPE and MAPE values reveal that our forecasts tend to vary more from the actuals in proportional terms than those of the benchmark. Collectively, these insights affirm the model's proficiency in capturing demand rhythms but also highlight the need for refinement to bolster its precision and accuracy at capturing peak and trough demands.

	ME	RMSE	MAE	MPE	MAPE
Our Forecast	207.84938	466.80030	335.87966	2.1178338	3.5245671
Benchmark	-18.28552	96.56456	75.64261	-0.2047934	0.8000381

Figure 19: Benchmark forecast accuracy.

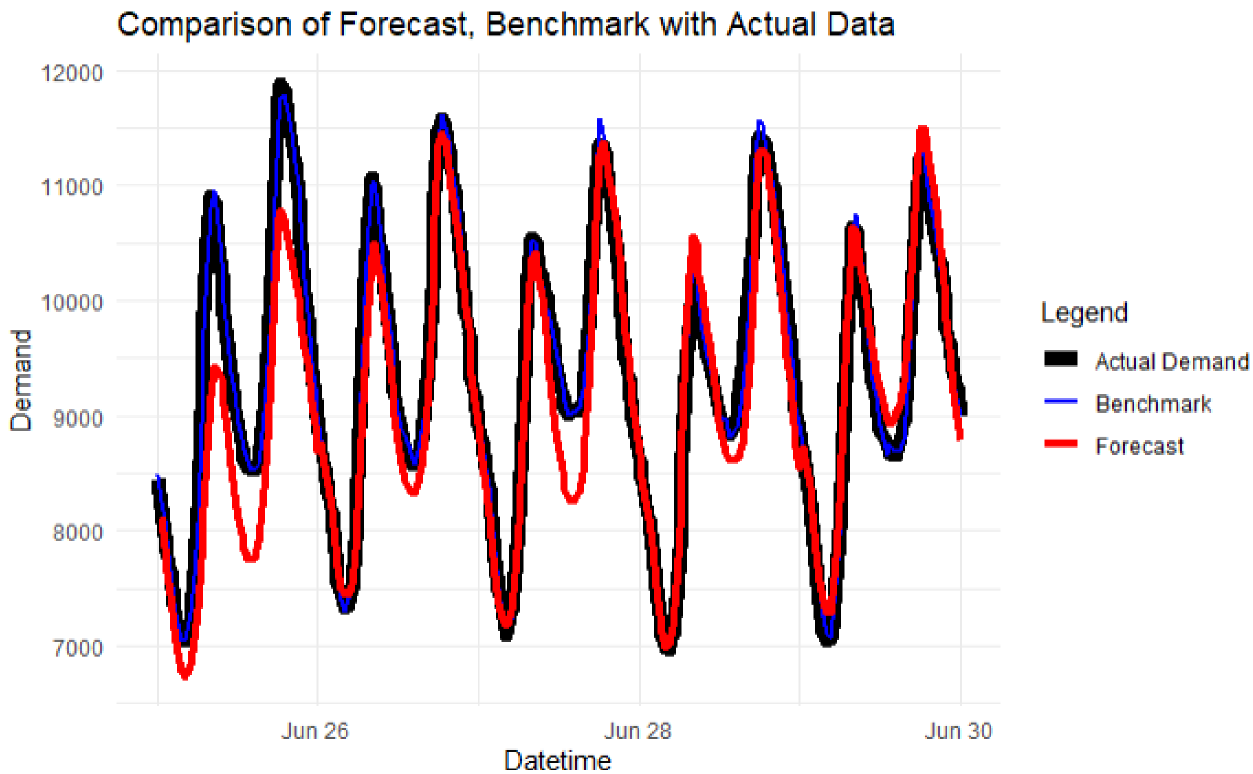


Figure 20: Model Forecast and benchmark with actual data.

While the current model demonstrates a satisfactory fit to the benchmark, opportunities for improvement lie in refining both the model structure and its parameters. Specifically, enhancing the model's ability to accurately capture the peaks and valleys of actual demand patterns through further research and experimentation is crucial. By iteratively refining forecasting methodologies and fine-tuning model parameters, we can expect to achieve higher levels of accuracy and precision in predicting future demand trends. These enhancements will bolster predictive capabilities and furnish decision-makers with more reliable insights, empowering them to make informed and strategic decisions with greater confidence.

6 Discussion

The project's development of a sophisticated energy demand forecasting model in New South Wales (NSW) aimed not only at enhancing technical capabilities but also at advancing socio-economic and environmental considerations. By integrating these models into broader operational and strategic frameworks, the potential extends beyond mere technical refinement to providing robust insights that can transform energy management practices. Below we synthesise the modelling results with the broader objectives of cost management, resource allocation, sustainability, and consumer empowerment, while emphasising complementary insights that enhance the value of the client's existing forecasting tools.

6.1 Complementary insights and strategic enhancements

Economic and consumer impact (addressing rising energy costs): While the client's existing model outperforms in accuracy, our Fourier-enhanced SARIMAX model provides complementary insights, especially in detecting complex seasonal variations. This model can support dynamic pricing and demand response strategies by providing alternative forecasting perspectives that may better capture atypical demand patterns or sudden shifts due to external factors. These insights are crucial during volatile energy market conditions, helping to mitigate high energy expenditures and improve economic efficiency.

Promoting socia-economic equity: Our model provides detailed insights into the correlations between temperature fluctuations and energy demand, highlighting opportunities for targeted policy interventions. These insights are particularly valuable for ensuring energy affordability and preventing disparities during peak demand periods, thereby supporting social equity. Predictability of the forecast will ensure appropriate resource management for service providers assuring optimised costs to end users.

Sustainability through optimised renewal integration: The explicit integration of temperature data within our forecasting model provides a strategic lever to align energy demand with renewable energy production. This capability not only optimises the use of renewable resources but also minimises environmental impacts by reducing reliance on carbon-intensive energy sources during peak times. Enhancing this model feature could substantially advance sustainability efforts, making it a critical component of environmental management strategies. Australia's over reliance on fossil fuel such as coal can be minimised by , having a highly predictive forecast to cater peak demand using renewable energy resources. For example, a 40 degree Celsius day, could use solar energy for peak load whilst wind energy can sustain for a cold night. Our analysis suggests approximately 60% of the demand are in these two categories, having alternate source of energy generation during this time, will sustain Australia's climate policies.

Operational efficiency and infrastructure planning: Our model offers alternative predictive outputs that validate or refine the client's infrastructure planning and resource allocation strategies. By providing a different analytical perspective, our model acts as a complementary tool that enhances decision-making processes, leading to more informed, resilient, and efficient energy management practices.

Consumer empowerment and energy conservation: Although our model shares the half-hourly data frequency with the client's, it brings unique insights into the interaction between temperature variations and energy consumption. These insights can drive the development of targeted energy conservation strategies and dynamic pricing models that empower consumers to make informed decisions aligned with environmental cues, fostering a culture of conservation and responsible energy usage.

6.2 Strategic enhancements and further integration proposals

Building upon the project's findings, and to ensure that our contributions provide actionable value not covered by existing enhancements in the conclusion, we propose several strategic initiatives:

- **Ensemble Forecasting Integration:** Leveraging both the client's and our forecasting models in an ensemble method can enhance prediction accuracy and provide a more comprehensive view of potential demand fluctuations. This integrated approach reduces prediction errors and broadens the analytical capabilities available for decision making.

- **Advanced Scenario Planning Capabilities**: Enhancing our model's scenario planning capabilities can provide the client with deeper insights under various hypothetical conditions. This enhancement is essential for strategic planning and can help in assessing the impacts of extreme weather events, policy changes, or significant technological innovations.
- **Specialised Demand Response Strategies:** Utilizing the detailed insights from our temperature-demand analysis to develop specialised demand response strategies can optimize energy consumption more effectively during critical periods. These strategies can be tailored to activate based on specific environmental triggers identified by our model.

7 Conclusion and Further Recommendations

Energy forecasting is a critical component in energy management and policy formulation, requiring accurate and reliable models to anticipate demand fluctuations. Our analysis identified a strong correlation between temperature and energy demand, highlighting the significant impact of temperature variations on consumption patterns. Given the non-stationary nature of the data and its complex seasonality, we decided to employ the SARIMA model enhanced with Fourier series to address these seasonal intricacies. Additionally, we incorporated temperature as an exogenous variable to capture the non-linear relationship between temperature and demand more effectively. This approach allowed us to refine our predictive model to better forecast energy demand over a short-term, specifically across a 24-hour interval.

Some advantages of using the model were -

- **Improved Seasonal Pattern** - effectively capturing seasonal variations in energy consumption data, allowing for more precise forecasting compared to traditional SARIMA models.
- **Enhanced Forecast Accuracy** - achieving better accuracy in predicting energy demand, particularly in scenarios with pronounced seasonal fluctuations.
- **Robustness to Seasonal Changes** - the model exhibited robust performance in scenarios where seasonal patterns evolve over time, making it suitable for forecasting in dynamic environments.

Within our constrained timeframe and resources, Team Echo developed a hybrid forecast model that demonstrates high accuracy compared to the benchmark. While we acknowledge the success of our current model, we also recognise the potential for further enhancement in forecast accuracy. We offer the following recommendations to advance our forecasting capabilities:

Incorporating Exogenous Variables: while the Fourier SARIMA model excels in capturing seasonal patterns, integrating exogenous variables such as weather data, economic indicators, and demographic factors could further enhance its predictive capabilities. This approach would enable the model to account for external factors influencing energy consumption.

Exploring Hybrid Models: investigating hybrid models that combine Fourier SARIMA with other advanced forecasting techniques, such as machine learning algorithms or neural networks, may lead to even greater accuracy and robustness. Hybrid approaches leverage the strengths of different methods to overcome limitations and improve overall performance.

Fine-Tuning Model Parameters: conducting thorough parameter optimisation and sensitivity analysis can help fine-tune the Fourier SARIMA model for specific forecasting tasks and datasets. Adjusting model parameters based on historical performance and data characteristics can optimise forecasting accuracy and reliability.

Continuous Model Evaluation: regularly evaluating the performance of the Fourier SARIMA model against new data and benchmarking it against alternative approaches ensures ongoing refinement and improvement. Continuous model evaluation allows for the identification of potential weaknesses or areas for enhancement, guiding iterative model development.

Combining the quadratic modelling with Fourier series: within the SARIMA framework, can assist in leveraging the strengths of both approaches to capture complex seasonal patterns in energy demand, leading to more accurate and robust frameworks. The model will become complex and hence due consideration should be provided to data characteristics, computational complexity, and parameter estimation/tuning.

In conclusion, the Fourier SARIMAX model presents a promising approach to energy consumption forecasting, offering improved accuracy and robustness in capturing seasonal patterns. While the model shows considerable potential, further research and experimentation are needed to fully exploit its capabilities and address remaining challenges. By incorporating the above recommendations, we can continue to advance energy forecasting methodologies and contribute to more effective energy management strategies.

7.1 Exogenous variables analysed during EDA

In addition to temperature, Team Echo conducted analysis two additional variables: rainfall and solar exposure (especially in min and max temperature mean). Facing time constraints and a reduced team size, the team was unable to complete the full integration of these variables into the final model. Team strongly advocates integration of solar exposure, which correlates directly with temperature, and the analysis of rainfall, which impacts weather patterns in New South Wales (NSW). These variables are crucial for understanding renewable energy effectiveness and weather patterns in NSW. While unable to complete the integration, Team Echo's approach sets a solid foundation for future work, emphasising the importance of these variables and their potential impact on renewable energy usage and weather forecasting.

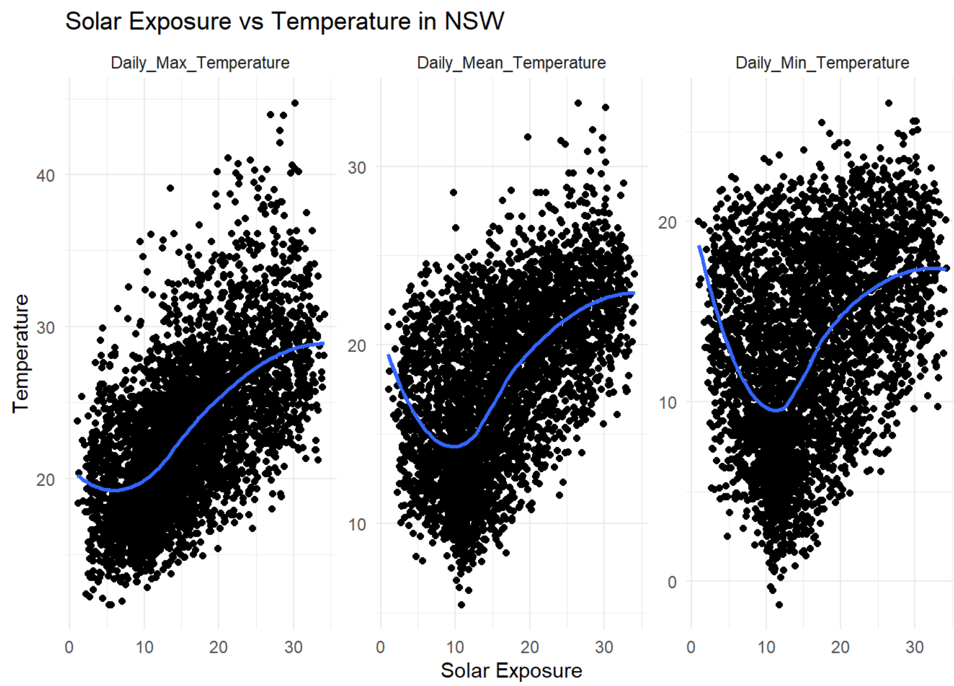


Figure 21: Mean solar exposure

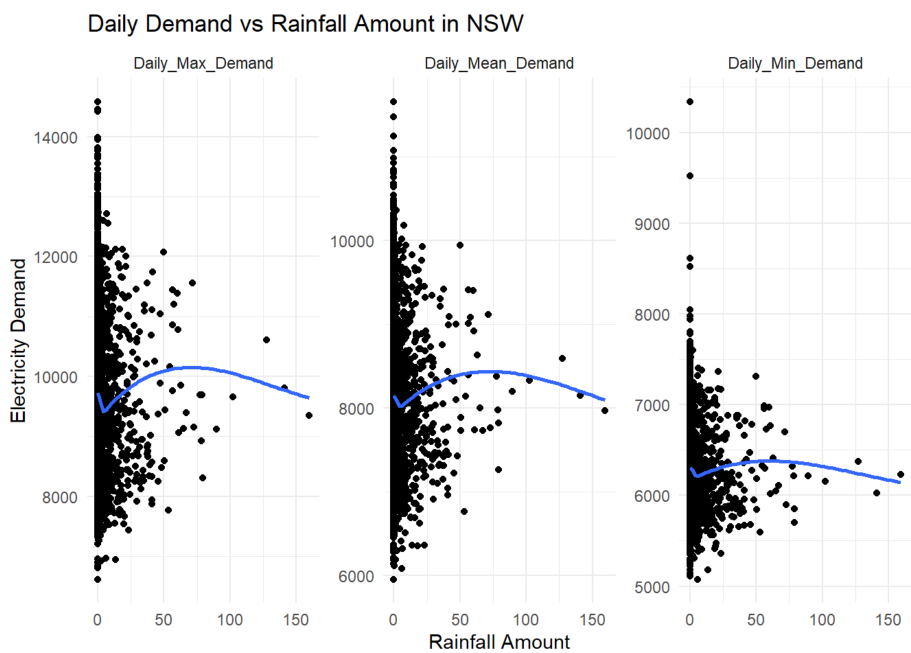


Figure 22: Rainfall impact

8 Bibliography

A. Azadeh, M. Z. (2015). A neuro-fuzzy algorithm for improved gas consumption forecasting with economic, environmental and IT/IS indicators. *Journal of Petroleum Science and Engineering*, 716-739.

Abdel-Aal, R., & Al-Garni, A. (1997, November). *Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis*. Retrieved from Science Direct: https://www.sciencedirect.com/science/article/abs/pii/S0360544297000327

Ahmad, T., Chen, H., Yabin, G., & Jiangyu, W. (2018, April). *A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review*. Retrieved from Science Direct.

Australian Renewable Energy Agency. (2014, May). *Australian electricity market analysis report to 2020 and 2030.* Retrieved from www.arena.com.au: https://arena.gov.au/assets/2017/02/CSIRO-Electricity-market-analysis-for-IGEG.pdf

Azadeh, A., Ghaderi, S. F., & Sohrabkhani, S. (2008). A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran. *Energy Policy*, 2637-2644.

Bianco, V., Manca, O., Nardini, S., & Mineaa, A. A. (2010). Analysis and forecasting of nonresidential electricity consumption in Romania,. *Applied Energy*, 3584-3590.

Bourdeau, M., Zhai, X., Nefzaoui, E., Guo, X., & Chatellier, P. (2019, July). *Modeling and forecasting building energy consumption: A review of data-driven techniques*. Retrieved from Science Direct.

Boyle, E. (2022, July 31). *PAW Climate 2022 - Myst AI: How to build accurate electricity demand forecasts*. Retrieved from YouTube: https://www.youtube.com/watch?v=Kw5Qhqs9VcE

Cao, G., & Wu, L. (2016). Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting,. *Energy*, 734-745. doi:https://doi.org/10.1016/j.energy.2016.09.065

Chae, Y. T., Horesh, R., Hwang, Y., & Lee, Y. M. (2016). Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy and Buildings*, 184-194.

Commonwealth Scientific and Industrial Research Organisation. (2014, May). *Australian electricity market analysis report to 2020 and 2030.* Retrieved from https://arena.gov.au/assets/2017/02/CSIRO-Electricity-market-analysis-for-IGEG.pdf

Connor, W., Chan, C. W., & Laforge, P. (2012, October 17). *Towards Developing a Decision Support System for Electricity Load Forecast*. doi:10.5772/51306

Crompton, P., & Wu, Y. (2005). Energy consumption in China: past trends and future directions. *Energy Economics*, 195-208. doi:https://doi.org/10.1016/j.eneco.2004.10.006

Dahya, T. (2020). *Predicting US Electricity Usage*. Retrieved from Medium: https://medium.com/@tdahya2/predicting-us-electricity-usage-a-comparison-of-two-approaches-34303b81a4aa

Daut, M. A., Hassan, M. Y., Abdullah, H., Rahman, H., Abdullah, M., & Hussin, F. (2017, April). *Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review*. Retrieved from Science Direct: https://www.sciencedirect.com/science/article/abs/pii/S1364032116310619

Debnath, K. B., & Mourshed, M. (2018, May 01). *Forecasting methods in energy planning models*. doi:https://doi.org/10.1016/j.rser.2018.02.002

Ding, Y., Zhang, Q., & Yuan, T. (2017). Research on short-term and ultra-short-term cooling load prediction models for office buildings,. *Energy and Buildings,*, 254-267. doi:https://doi.org/10.1016/j.enbuild.2017.08.077

Dobson, J. (2023). *On reading and interpreting black box deep neural networks*. Retrieved from SpringerLink: https://link.springer.com/article/10.1007/s42803-023-00075-w#citeas

Dudek, G. (2022). *A Comprehensive Study of Random Forest for Short-Term Load Forecasting*. Retrieved from MDPI: https://www.mdpi.com/1996-1073/15/20/7547

Energy Council - Australia. (n.d.). *Extreme weather and electricity supply.* Retrieved from www.energycouncil.com.au: https://www.energycouncil.com.au/media/hizc3e5y/aec051-extreme-weather-factsheet.pdf

Ervural, B. C., Beyca, O. F., & Zaim, S. (2016). Model Estimation of ARMA Using Genetic Algorithms: A Case Study of Forecasting Natural Gas Consumption. *Procedia - Social and Behavioral Sciences*, 537-545. doi:https://doi.org/10.1016/j.sbspro.2016.11.066

Fasiolo, M. (2021, June 15). *Forecasting electricity demand*. Retrieved from University of Bristol: https://www.youtube.com/watch?v=4Z0vpt7u7MM

Forouzanfar, M., Doustmohammadi, A., Menhaj, M. B., & Hasanzadeh, S. (2010). Modeling and estimation of the natural gas consumption for residential and commercial sectors in Iran,. *Applied Energy*, 268-274.

Graves, A. (2012). Long short-term memory. In *Supervised sequence labelling with recurrent neural networks* (pp. 37-45). Springer. doi:https://doi.org/10.1007/978-3-642-24797-2_4

Guefano, S., Jean, G. T., Tchitile, E. W., & Monkam, L. (2021, January 1). *Forecast of electricity consumption in the Cameroonian residential sector by Grey and vector autoregressive models.* Retrieved from https://pdf.sciencedirectassets.com/306208/1-s2.0-S2215016120X00021/1-s2.0-S2215016121000893/main.pdf?X-Amz-Security-

Token=IQoJb3JpZ2luX2VjEM%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJHMEUCIQC4XIyPKmxEG831NJ76VLTAT3XkmKBNv5BLjeXqvFAJ5gIgf61%2F05pl

H. Fan and I.F. MacGill and A.B. Sproul. (2015). *Statistical analysis of driving factors of residential energy demand in the greater Sydney region, Australia.* Retrieved from https://www.sciencedirect.com/science/article/pii/S0378778815301419

He, W. (2017). Load Forecasting via Deep Neural Networks. *Procedia Computer Science*, 308-314. doi:https://doi.org/10.1016/j.procs.2017.11.374

He, Y., & Lin, B. (2018). Forecasting China's total energy demand and its structure using ADL-MIDAS model. *Energy*, 420-429. doi:https://doi.org/10.1016/j.energy.2018.03.067

Hussain, A., Rahman, M., & Memon, J. (2016, March). *Forecasting electricity consumption in Pakistan: the way forward*. Retrieved from Science Direct: https://www.sciencedirect.com/science/article/abs/pii/S0301421515302068

Kermanshahi, B. (1998). Recurrent neural network for forecasting next 10 years loads of nine Japanese utilities. *Neurocomputing*, 125-133. doi:https://doi.org/10.1016/S0925-2312(98)00073-3

Kumar, U., & Jain, V. (2010). Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India,. *Energy*, 1709-1716.

Limanond, T., Jomnonkwao, S., & Srikaew, A. (2011). Projection of future transport energy demand of Thailand,. *Energy Policy*, 2754-2763.

Liu, X., Moreno, B., & García Ana Salomé. (2016). A grey neural network and input-output combined forecasting model. Primary energy consumption forecasts in Spanish economic sectors, . *Energy*, 1042-1054.

Meer van der, D., Shepero, M., Svensson, A., Widen, J., & Munkhammar, J. (2018, March 1). *Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Processes*. Retrieved from www.ScienceDirect.com: https://www.sciencedirect.com/science/article/abs/pii/S0306261917318275

Miljanovic, M. (2012). Comparative analysis of Recurrent and Finite Impulse Response Neural Networks. *Inian Journal of Computer Science and Engineering (IJCSE)*. Retrieved from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=eef71b7017d34cebb83d5367bb6993f7b01093c0

NSW State of Environment (EPA). (2020). *Energy Consumption (demand for renewal energy in NSW)*. Retrieved from www.soe.epa.nsw.gov.au: https://www.soe.epa.nsw.gov.au/all-themes/human-settlement/energy-consumption

Pao, H. T. (2009). Forecasting energy consumption in Taiwan using hybrid nonlinear models,. *Energy*, 1438-1446.

Pappas, S., Ekonomou, L., Karampelas, P., Karamousantas, D., Katsikas, S., Chatzarakis, G., & Skafidas, P. (210). Electricity demand load forecasting of the Hellenic power system using an ARMA model. *Electric Power Systems Research*, 256-264. doi:https://doi.org/10.1016/j.epsr.2009.09.006

Qing, X., & Niu, Y. (2018). Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy*, 461-468. doi:https://doi.org/10.1016/j.energy.2018.01.177

Radharani Panigrahi, N. R. (2022). *Regression model-based hourly aggregated electricity demand prediction.* Retrieved from Science Direct: https://www.sciencedirect.com/science/article/pii/S2352484722019382

Regulator, Australian Energy. (2023, November 30). *Annual retail markets report 2022-23.* Retrieved from www.aer.gov.au: https://www.aer.gov.au/system/files/2023-11/Annual%20Retail%20Market%20Report%202022-23%20-%2030%20November%202023_1.pdf

Sachin, M. M. (2020). *Analysis of energy consumption using RNN-LSTM and ARIMA Model*. Retrieved from IOPScience: https://iopscience.iop.org/article/10.1088/1742-6596/1716/1/012048/meta

Shaikh, F., Ji, Q., Shaikh, P. H., Mirjat, N. H., & Uqaili, M. A. (2017). Forecasting China’s natural gas demand based on optimised nonlinear grey models,. *Energy*, 941-951. doi:https://doi.org/10.1016/j.energy.2017.09.037.

Shayeghi, H., Ghasemi, A., Moradzadeh, M., & Nooshyar, M. (2015). Simultaneous day-ahead forecasting of electricity price and load in smart grids,. *Energy Conversion and Management*, 371-384. doi:https://doi.org/10.1016/j.enconman.2015.02.023

Shine, P., Murphy, M. U., & Scully, T. (2018). Machine-learning algorithms for predicting on-farm direct water and electricity consumption on pasture based dairy farms,. *Computers and Electronics in Agriculture*, 74-87. doi:https://doi.org/10.1016/j.compag.2018.03.023

Sim, S., Tay, K., Huong, A., & Tiong, W. (2019, July 26). *Forecasting Electricity Consumption Using SARIMA Method in IBM SPSS Software.* Retrieved from https://www.hrpub.org/download/20191230/UJEEEB14-14991055.pdf

Sujjaviriyasup, T. (2017). A new class of MODWT-SVM-DE hybrid model emphasizing on simplification structure in data pre-processing: A case study of annual electricity consumptions,. *Applied Soft Computing*, 150-163. doi:https://doi.org/10.1016/j.asoc.2017.01.022

Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks,. *Energy, 85*, 208-220.

Thatcher, M. J. (2007). Modelling changes to electricity demand load duration curves as a consequence of predicted climate change for Australia. *Energy*, 1647-1659. Retrieved from Science Direct: https://www.sciencedirect.com/science/article/pii/S0360544206003495

Wang, J., Du, P., Lu, H., & Yang, W. N. (2018). An improved grey model optimized by multi-objective ant lion optimization algorithm for annual electricity consumption forecasting,. *Applied Soft Computing*, 321-337.

Wei, N., Li, C., Duan, J., Liu, J., & Zeng, F. (2019, 12). *Daily Natural Gas Load Forecasting Based on a Hybrid Deep Learning Model*. doi:10.3390/12020218

Wei, N., Li, C., Duan, J., Liu, J., & Zeng, F. (2019). Daily Natural Gas Load Forecasting Based on a Hybrid Deep Learning Model. *Energies*. Retrieved from https://www.mdpi.com/1996-1073/12/2/218

Weron, R. (2023, November 30). *Value in Energy Data #23: Rafał Weron - Recent Advances in Electricity Price Forecasting*. Retrieved from Energy Systems Catapult: https://www.youtube.com/watch?v=fvz2JR5BDko

Wiki Article. (2024, January). *https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average*. Retrieved from Autoregressive integrated moving average.

Xu, G., & Wang, W. (2010). Forecasting China's natural gas consumption based on a combination model. *Journal of Natural Gas Chemistry*, 493-496.

Yang, Y., Chen, Y., Wang, Y., Li, C., & Lian, L. (2016). Modelling a combined method based on ANFIS and neural network improved by DE algorithm: A case study for short-term electricity demand forecasting,. *Applied Soft Computing*, 663-675.

Yaslan, Y., & Bican, B. (2017). Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting. *Measurement*, 52-61. doi:https://doi.org/10.1016/j.measurement.2017.02.007

Zhu, S., Wang, J., Zhao, W., & Wang, J. (2011, November). *A seasonal hybrid procedure for electricity demand forecasting in China*. Retrieved from Science Direct: https://www.sciencedirect.com/science/article/abs/pii/S0306261911002960

Ziel, F., & Steinert, R. (2016). Electricity price forecasting using sale and purchase curves: The X-Model. *Energy Economics*, 435-454.

***** **End of Document** *****