12-23-2021

Energy consumption forecasting using machine learning

Mahdi Mohammadigohari
mmm8198@rit.edu

Energy consumption forecasting using machine learning

by

Mahdi Mohammadigohari

A Capstone Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies: Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology
RIT Dubai
December 23, 2021

RIT

Master of Science in Professional Studies:

Data Analytics

Graduate Capstone Approval

Student Name: Mahdi Mohammadigohari

Graduate Capstone Title: **Energy consumption forecasting using machine learning**

Graduate Capstone Committee:

Name: Dr. Sanjay Modak **Date:**

Chair of committee

Name: Dr. Ioannis Karamitsos **Date:**

Member of committee

Acknowledgments

The accomplishment of this dissertation would not have been possible without the great help and encouragement of many people. I would like to take this opportunity to thank those who have supported me during this challenging journey, which has been a unique experience in my life.

First and foremost, I would like to express my special thanks to my supervisor, Dr Sanjay Modak, for his professional guidance, support and encouragement, and for all time and effort that he has contributed to my capstone project. It has been a precious gift to work so closely with him.

I would like to thank my mentor, Dr Ioannis Karamitsos, for his generous support and encouragement during my graduation study. He has always provided me with valuable comments on my research, and future career.

I acknowledge the Data Analytics Faculty and Rochester Institute of Technology for their comprehensive support.

Finally, I would like to express my deepest gratitude and heartfelt thanks to my beloved family: my wife, Masoumeh, and my children, Bonyad and Donya, for their love, encouragement and support over the years. I can’t express just how grateful I am, and I dedicate this thesis to them.

Abstract

Forecasting electricity demand and consumption accurately is critical to the optimal and cost-effective operation system, providing a competitive advantage to companies. In working with seasonal data and external variables, the traditional time-series forecasting methods cannot be applied to electricity consumption data. In energy planning for a generating company, accurate power forecasting for the electrical consumption prediction, as a technique, to understand and predict the market electricity demand is of paramount importance. Their power production can be adjusted accordingly in a deregulated market. As data type is seasonal, Persistence Models (Naïve Models), Seasonal AutoRegressive Integrated Moving Averages with eXogenous regressors (SARIMAX), and Univariate Long-Short Term Memory Neural Network (LSTM) is used to explicitly deal with seasonality as a class of time-series forecasting models.

The main purpose of this project is to perform exploratory data analysis of the Spain power, then use different forecasting models to once-daily predict the next 24 hours of energy demand and daily peak demand. To split the electricity consumption data from 2015 to 2018 into training and test sets, the first three years from 2015 and 2017 were used as the training set, while values from 2018 were used as the test set. The obtained results showed that the machine learning algorithms proposed in the recent literature outperformed the tested algorithms.

Models are evaluated using root mean squared error (RMSE) to be directly comparable to energy readings in the data. RMSE has calculated two ways. First to represent the error of predicting each hour at a time (i.e. one error per-hourly slice). Second to represent the models' overall performance. The results show that electricity demand can be modeled using machine learning algorithms, deploying renewable energy, planning for high/low load days, and reducing wastage from polluting on reserve standby generation, detecting abnormalities in consumption trends, and quantifying energy and cost-saving measures.

Keywords: Short-term electricity demand, Electricity demand forecasting, Exploratory data analysis, Machine learning.

Table of Contents

List of Figures

List of Tables

List of Symbols

$actual_t$	actual value at time t
B^s	backshift operator such that $B^s y_t$=y_{t-s}
c_t	cell activation vector
D	order of the seasonal differencing to make data stationary
d	order of differencing to make data stationary
e_t	error at time t
$forecast_t$	forecasted value at time t
f_t	forget gate
i_t	input gate
n	number of observations in the sample
o_t	order of the seasonal AR term
P	order of the seasonal AR term
p	order of the AR term
$\theta_1, \theta_2, \cdots \theta_q$	coefficients of the MA order
$\theta(B)$	polynomial in B of order q
Q	order of the seasonal MA term
q	order of the MA term
S	number of periods in a season
$\emptyset(B)(1-B)^d$	combined autoregressive operator
$\phi(B)$	polynomial in B of order P
W	weight matrices from the cell to gate vectors
y_d	household electricity load at time instance t on day d
y_t	value of the series at time t
z_t	purely random process, also known as white noise

List of Abbreviations

ACF	AutoCorrelation Function
AIC	Akaike’s Information Criterion
ANN	Artificial Neural Network
AR	Autoregressive
ARIMA	AutoRegressive Integrated Moving Average
ARIMAX	AutoRegressive Integrated Moving Average with eXogenous variables
CLD	Copy-Last-Day
CO2	Carbon Dioxide
DR	Demand Response
EUI	Energy Use Intensity
GDP	Gross Domestic Product
GM	Grey Model
GSI	Global Solar Irradiance
HVAC	Heating, Ventilation, and Air Conditioning
IQR	Interquantile Range
kWh	kilowatt-hour
MA	Moving Average
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MSE	Mean Square Error
NAN	Not a Number
PACF	Partial AutoCorrelation Function
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RT	Regression Tree
SARIMAX	Seasonal AutoRegressive Integrated Moving Averages with eXogenous
SVM	Support Vector Regression
TF	Transfer Function
W	Watts

Chapter 1

1.1 Introduction

During the last years, volatile energy prices, and electricity generation industry deregulation has resulted in introducing electricity load forecasting as a critical issue for the power plants' operation (Jota et al. 2011). Varying time horizons and accurate forecasts are used by Power plants to make sure plant operation while planning for the possible facility expansions to measure future demand optimally and securely (Kyriakides and Polycarpou 2006). As residential and commercial buildings are consuming about 40% of the total energy, according to the US Energy Information Administration, building load forecasting has also become important around the world (US Energy Flow, 2020). As a result, the buildings' energy efficiency improvement is critical to diminish gas emissions. Data power consumption Analyzing has introduced demand response (DR) actions as energy and cost-saving opportunities in the energy sector (forecasting, scheduling, and risk management 2002). Buildings' heating and cooling equipment scheduling as a result of turning off electrical equipment is contained in DR actions as one of the basic tasks. Therefore, at the building level, energy demand forecast has become a heated topic in recent years (Cai et al. 2018). Building energy management to predict future load demand can be a viable task by forecasting demand load. The energy sector uses load shapes to the identification of values timely at each point which represents the load as a time function (Price, 2010). Building redesigning or renovation activities can give more importance to the Load shapes' ability to make informed decisions (Jota et al. 2011). On the other hand, energy consumption prediction can be used in abnormalities detection, intervention or change impact identification in a system, and energy and cost reduction (Moreno et al. 2007). While factors such as weather condition, season, weekday, the behavior of the occupants, and social activities, which are influential in the demand load, accurate short-term demand load forecasting of a building, has been identified as a challenging task (Hor, Watson, and Majithia 2006).

Availability long-term electricity demand projections can also be more profitable to the growth in electricity demand during an economic downturn (Hor, Watson, and Majithia 2006). On the building or the power system level, load forecasting can be generally categorized into three groups (based on time horizon): short term forecasting (ranging from a few hours ahead to a few weeks ahead), medium-term (encompassing a month to one year ahead), and long-term forecasting (ranging from 1 to 20 years ahead) (Kyriakides and Polycarpou, 2006). Long-term forecasts, ranging from 1 to 20 years ahead which are an imperative topic for strategic planning, new generation construction, and transmission capacity (Kandil et al. 2002). Mid-term forecasting encompassing a month to one year ahead which are applied for maintenance and power-sharing agreements scheduling (Friedrich et al. 2014). The last class, short-term load forecasting, ranging from a few hours ahead to a few weeks ahead has great importance in real-time control, plant scheduling, fuel purchasing schemes, short-term maintenance as well as short-term storage usage

(Friedrich et al. 2014). The two major categorized groups of demand forecasting methods are: a) classical techniques, and b) computational methods based on Artificial Intelligence (Kyriakides and Polycarpou, 2006). As a function of historical data, the demand can be modeled by the forecasting techniques of the classical time-series data, working under the assumption that data is linear and stationary (Kyriakides and Polycarpou, 2006). The statistical properties of a stationary time series such as mean, variance, and autocorrelation structures do not change a function of time and are all constant over time. However, the non-stationary and seasonal approach of the electricity demand data makes many classical techniques inadequate. The classical methods can be improved by inputting external variables in the prediction models. The real GDP (gross domestic product) and demography as external variables were combined in a classical technique called ARIMA to forecast Morocco's long-term annual electricity demand (Citroen et al. 2015). Daily electricity demand, in another research, was predicted by entering the temperature as an external variable into the ARIMA forecasting model (Felice et al. 2013), while the data seasonality was not addressed by this technique. However, Cools, Moons, and Wets in 2009 used the Seasonal AutoRegressive Integrated Moving Averages with eXogenous regressors (SARIMAX) method to improve the traditional time-series techniques, dealing with the data seasonality explicitly and accounting for external variables. Therefore, buildings' electricity consumption and demand data can benefit from such an ideal class of models. The day of the week, holidays, and temperature were considered as external variables to be applied by Papaioannou et al. in a SARIMAX prediction model to forecast the electricity demand in Greece in the year 2016. In another study, Taşpinar et al. (2013) showed that the daily residential energy consumption can be highly influenced by ambient temperature and cloud cover in Turkey. This project focuses on building the Persistence, ARIMA, and SARIMAX models (SARIMA with external variables) to forecast the national electricity consumption and peak demand.

1.2 Project goals

The goal of this project is to test whether a general and simple approach based on Machine Learning models, can yield good enough results in a complex forecasting problem, exploring machine learning techniques and developing data-driven models for forecasting energy consumption and performance.

Once a day, electrical grid Transmission Service Operators (TSOs) issue energy demand forecasts to appropriately meet energy demands for the coming 24-hour period. This is a highly relevant problem across implemented every day the world, forecasting the expected maximum energy demand on an hourly basis and consists of 24-hourly slices. Ultrashort term (6 hours or less) forecasts are combined with these forecasts to keep maintain balance in the grid, and to plan supply dispatch for day-ahead bidding processes.

1.3 Aims and Objectives

The fundamental objective of this project is to compare different Machine Learning models on the coming 24-hour forecat mission of electricity load by using past data and evaluate the models' performance. This aim was broken down into as follows:

1. Implement classical statistical forecasting models
2. Implement and gain insight into walk forward validation, forecasting performance, and feature selection.

1.4 Research Methodology

To achieve our desired aim and determine the data type and sources that can be used for this aim, we proposed a supervised learning and defined the purpose of this study clearly in the first step of the project. In our case, we are interested in predicting the hourly energy consumption using the crisp-DM method (Wirth (2000)), which represents an overall process of a data mining project, typically consists of the five iterative phases: business understanding, data understanding, data preparation, model development, and evaluation.

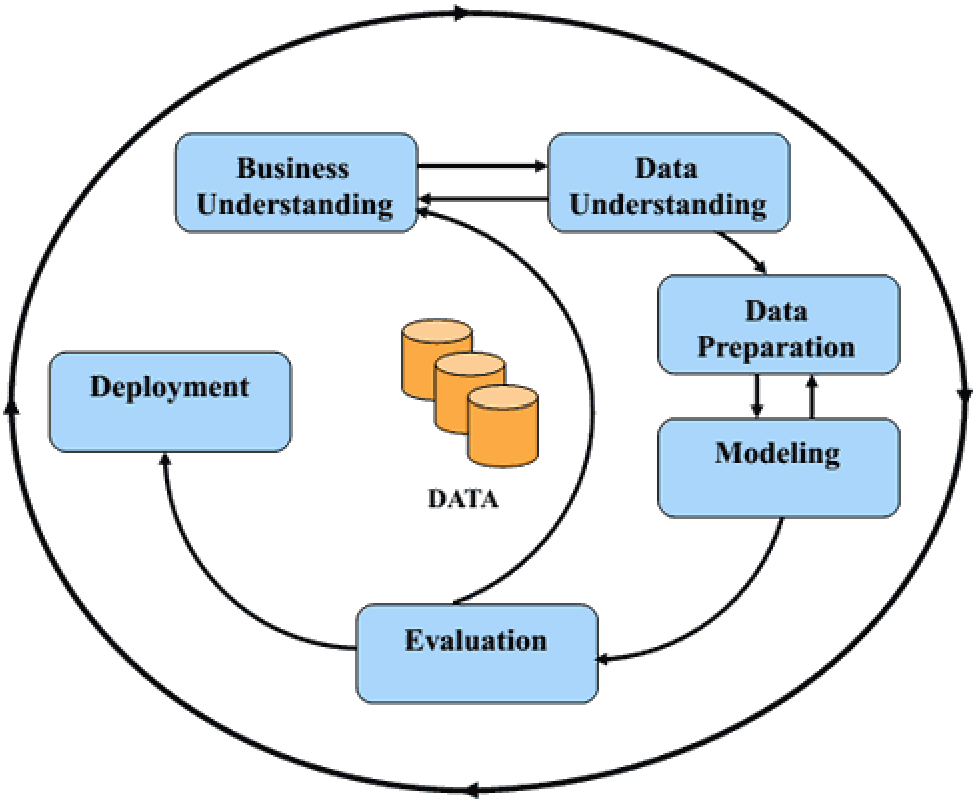


Figure 1- 1 CRISP-DM Methodology (source: Wirth, 2000)

1.4.1 Business Understanding

Business Objective: Predicting the power demand with high accuracy might introduce a great set of values for a country, for a city, or even for households. Stakeholders might adjust their power production accordingly to reduce cost, or they can buy sufficient amounts of energy if they meet their power needs from external sources. In some certain cases, such as in tendering processes in daily energy exchange, the stakeholders may generate additional profit.

1.4.2 Data Understanding

This stage includes an initial collection, description, basic exploration, and verification of the data. In the project, the dataset was used from Kaggle repository.

It contains 4-year electrical consumption, generation, pricing, and weather data for Spain, containing hourly electricity load data and the corresponding TSO load and energy price predictions for future data points which makes it a unique dataset. The data is multivariate time series as it contains multiple features. We have, in this dataset, information about the energy price, national grid total load and the different energy resources' produced amount (in MW).

1.4.3 Data Preparation and Feature Selection

After the collection phase, a four-step data preprocessing and feature selection were conducted to get the available dataset ready towards building the predictive models:

1. Cleaning the data, handling missing values, detecting outliers and treating them properly.
2. Transforming the energy, changing the data type, normalizing the numerical attributes, and creating a window of calendar days consisting of 24-hour segments to predict the next 24 hours in advance.
3. Processing the energy data to generate autoregressive features.

1.4.4 Modelling

This project follows the scheme mentioned below.

1. Identifying the problem of predicting electricity demand and consumption using time-series methods.
2. Using and performance evaluating the SARIMAX, ARIMA, Persistence, and LSTM techniques
3. Demonstrated the concept by applying the forecasting models in this project.
4. Applying the model performance metrics such as mean absolute percentage error (MAPE), and root mean square error (RMSE) to evaluate the models. For the production

phase, understanding the model generalization power to the future unseen data time-series specific cross-validation would also help.

5. Communicating the results through this project.

1.4.4.1 Time Series Forecasting

A collection of observations recorded in a sequence through time called a time series that is categorized based on the measurement frequency into continuous and discrete series in general (Chatfield 2001). By either aggregating the series over a period or sampling, a time series can be turned into a continuous or discrete series (Chatfield 2001). Electricity load demand is transformed to a continuous time series by sampling. The future values of a time series correctly can be achieved through understanding the different types of forecasting methods (Chatfield 2001). The forecasting methods are procedures that can be categorized into three groups:

1. Judgmental forecasts: bases this category on subjective criteria such as judgment or intuition.
2. Univariate forecasts: This category uses only present and past values of the series to predict future values.
3. Multivariate forecasts: uses at least one additional variable (are known as predictors or explanatory variables) to forecast future values.

1.4.4.1.1 Persistence Models (Naïve Models)

To establish reference (baseline) models and compare tests, persistence prediction models are usually applied. It is beneficial in many cases to develop a forecasting model to assess whether it can outperform a baseline model. Persistence models are simple methods of using past data to predict future data points. They are developed to benchmark performance when evaluating more complex techniques, compare the performance of feature engineering, hyperparameter tuning, and model architecture against a set of references. A persistence model can be assumed that the electricity load at the time $t + 1$ is equal to the load at a time t, according to Notton and Voyant (2018). A persistence model with constant and equal load over the next day to the current one, in a day-ahead forecast and with 15 min granularity to define time instances $t + 1$ and t, would most likely fail. Better performance for a persistence model can be reached if it can be taken that the electricity load at time t of day d (briefly (t, d)) would be the same with the corresponding load at the same time t on the previous day $d - 1$ or the previous same day $d - 7$. Another variation of such a model would also consider more than one previous day.

Let $y_d(t)$ be a household electricity load at time instance t on day d. Then, for a 1-day ahead persistence model, we have

$$\tilde{y}_d^{PM} = y_{d-1}(t).$$

Also, a N-day ahead persistence model can be defined as follows

$$\tilde{y}_d^{PM} = \frac{1}{N}\sum_{i=d-N}^{d-1} y_{i-1}\,(t).$$

It takes an average of the N previous days' load, in other words, at the same time.

By only considering the N previous same days, an N-day persistence model can be further improved the above model as electricity load is highly correlated with the presence of the residents in a household. Given that we are interested in a day-ahead forecast for time t, while d is corresponded to a Monday, the average load at the same time on the most recent N previous Mondays is needed to be created. As a copy-last-days persistence model (CLD), we will refer to this model according to which the forecasts are computed as follows:

$$\tilde{y}_d^{PM} = \frac{1}{N}\sum_{i=d-7N}^{d-7} y_{i-1}\,(t).$$

1.4.4.1.2 ARIMA Models

ARIMA (autoregressive integrated moving average) model is one of the most widely used time series models its statistical properties' use, well-known Box & Jenkins methodology, and GM (Grey model) in the modeling process (Zhang 2003). In forecasting energy consumption in economies multiple regression models and artificial neural network models are other models that can be used. A purely autoregressive (AR) and moving average (MA) process are combined to form the ARMA process, dealing with using only the past values of the time series and the current and past values of a random process to predict future values. To stationary data, however, the data is non-stationary in most real-world cases, we can apply this model. A value with its difference from previous values is replaced to deal with non-stationary data in an employed technique known as differencing.

Nonseasonal ARIMA models are written as ARIMA (p, d, q) where (Chatfield 2001):

- p is the order of the AR term.
- d is the order of differencing needed to make the data stationary.
- q is the order of the MA term.
- Here, the "order" is the number of previous values in the time series that are used in determining each term.

An ARIMA model's building blocks and its mathematical representation are explained below.

Let $AR(p)$ be an autoregressive process of order p, it can mathematically be represented as a weighted linear sum of the past p values plus white noise (Chatfield 2001).

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots \emptyset_p y_{t-p} + z_t.$$

Where $\emptyset_1, \emptyset_2, \cdots \emptyset_p$, denote the AR order's coefficients while Z_t denotes the error term with 0 mean and variance.

The $AR(p)$can be written as (Chatfield 2001):

$$By_t = y_{t-1},$$
$$\emptyset(B)y_t = z_t,$$
$$\emptyset(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 \cdots - \emptyset_p B^p,$$

applying the back-shift operator while $\emptyset(B)$ is a polynomial in B of order p.

An $MA(q)$ is mathematically represented as a weighted linear sum of the last q white noise error

$$y_t = \theta_1 z_{t-1} + \theta_2 z_{t-2} + \cdots \theta_q z_{t-q} + z_t.$$

Where the coefficients of MA order and the white noise terms with 0 mean and constant variance can be denoted by $\theta_1, \theta_2, \cdots \theta_q$, and $Z_t$respectively.

Using the back-shift operator B, the $MA(q)$ can be represented as (Chatfield 2001)

$$\emptyset(B) = 1 + \theta_1 B + \cdots \theta_1 B^q.$$

Where $\theta(B)$ is a polynomial in B of order q.

A mixed autoregressive moving average model of p autoregressive terms and q moving average terms are combined to build an ARIMA (p, d, q) model which are differenced d times (Chatfield 2001)

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots \emptyset_p y_{t-p} + z_t + \theta_1 z_{t-1} + \theta_2 z_{t-2} + \cdots \theta_q z_{t-q}.$$

Applying the back-shift operator B, the ARIMA (p, d, q) can be mathematically as (Chatfield 2001)

$$\emptyset(B)(1 - B)^d y_t = \emptyset(B) z_t,$$

where $\phi(B)(1 - B)^{d}$ is the combined autoregressive operator.

1.4.4.1.3 SARIMAX Models

To deal with non-stationary data, Autoregressive Integrated Moving Average (ARIMA) models are used, working with stationary and linear data. The Seasonal ARIMA (SARIMA), a generalized form of the ARIMA, is used to deal explicitly with seasonality in data by using seasonal AR, MA, and differencing terms in the model. External variables can also be input to the seasonal ARIMA which enables the user to input the external variables' effects to the model. The weather is considered an exogenous variable which is defined as variables that may influence a model but are not influenced by it.

A SARIMAX model is written as SARIMAX (p, d, q) (P, D, Q)s where:

- p is the order of the AR term.
- d is the order of differencing needed to make the data stationary.
- q is the order of the MA term.

- P is the order of the seasonal AR term.
- D is the order of the seasonal differencing needed to make data stationary.
- Q is the order of the seasonal MA term.
- S is the number of periods in a season.

A SARIMAX (p, d, q) (P, D, Q)s is mathematically represented as (Chatfield 2001):

$$y_t = \beta_0 + \beta_1 X_{1.t} + \beta_2 X_{2.t} + \cdots + \beta_k X_{k.t} + \frac{(1- \mathit{\Theta}_1 B^s - \mathit{\Theta}_2 B^{2s} \cdots - \mathit{\Theta}_Q B^{Qs})(1- \theta_1 B - \theta_2 B^2 \cdots - \theta_q B^q)}{(1- \mathit{\Phi}_1 B^s - \mathit{\Phi}_2 B^{2s} \cdots - \mathit{\Phi}_P B^{Ps})(1- \emptyset_1 B - \emptyset_2 B^2 \cdots - \emptyset_p B^p)} z_t,$$

where:

- y_t denotes the value of the series at time t.
- $X_{1.t}, X_{2.t}, \dots, X_{k.t}$ denote observations of the exogenous variables.
- $\beta_0, \beta_1, \dots, \beta_k$ denote the parameters of the regression part.
- $\emptyset_1, \emptyset_2, \dots, \emptyset_p$ denote the nonseasonal autoregressive terms' weights.
- $\mathit{\Phi}_1, \mathit{\Phi}_2, \dots, \mathit{\Phi}_P$ denote the seasonal autoregressive terms' weights.
- $\theta_1, \theta_2, \dots, \theta_q$ denote the nonseasonal moving average terms' weights.
- $\mathit{\Theta}_1, \mathit{\Theta}_2, \dots, \mathit{\Theta}_Q$ denote the seasonal moving average terms' weights.
- B^s denotes the backshift operator such that $B^s y_t = y_{t-s}$.
- z_t denotes the white noise terms.

1.4.4.1.4 LSTM Models

When it comes to modeling long-range dependencies, an LSTM neural network can be more appropriate as a specific type of RNN, introduced in 1997, by Hochreiter and Schmidhuber. Instead of hidden units, memory blocks are contained in the architecture of LSTM. Nonlinear sigmoidal gates are applied multiplicatively to modulate memory cells that are contained in a memory block.

The same gates are shared by memory cells to diminish the parameters. Whether the model keeps the values at the gates (if the gates evaluate to 1) or discards them (if the gates evaluate to 0) are determined by these gates, leading to being exploited long-range temporal contexts by the network (Hochreiter and Schmidhuber 1997).

To compute a mapping sequence to the output $y = (y_1, y_2, \cdots, y_T)$, let $x = (x_1, x_2, \cdots, x_T)$, be the input sequence. The unit activations can be determined by the following equations:

$$i_t = \sigma\,(W_{xi} x_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i)$$

$$f_t = \sigma\,(W_{xf} x_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f)$$

$$c_t = f_t c_{t-1} + i_t \tanh\,(W_{xc} x_t + W_{hc} h_{t-1} + b_c)$$

$$f_t = \sigma\,(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o)$$

$$i_t = o_t \tanh\,(c_t)$$

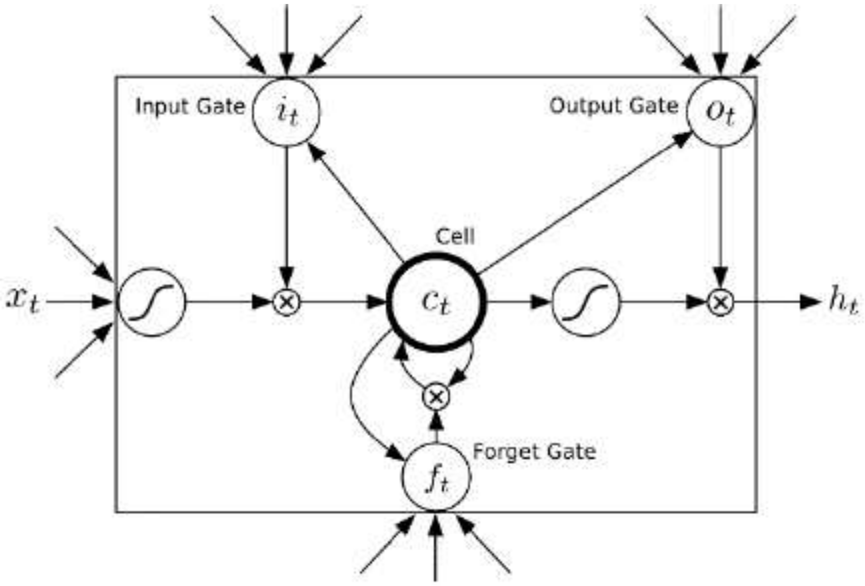


Figure 1- 2 LSTM cell

Where:

- σ denotes the logistic sigmoid function.
- $i, f, o, c,$ denote the input gate, forget gate, output gate, and cell activation vector, respectively.
- W terms denote the weight matrices from the cell to gate vectors (e.g., W_{si}).
- tanh denotes the output activation function.

Our LSTM network minimizes the usual root mean squared error (Hochreiter and Schmidhuber 1997), compiling an iterative gradient descent algorithm.

1.4.4.2 Univariate Models

A univariate model describes a single variable based on its relationship with its past values and white noise (Chatfield 2001). The following sections discuss some of the widely adopted models.

1.4.4.3 Multivariate Models

To explain the interrelationships between the time series, multivariate models can be used to multivariate datasets (Chatfield 2001). For instance, in economics, an increase in prices leads to an increase in wages, which will lead to an increase in prices again (Chatfield 2001). This phenomenon that the outputs affect the inputs is presented in closed-loop systems (Chatfield 2001).

1.4.5 Forecasting Accuracy Measures: Evaluation Method

In the final step after training the model with the train set of data, the model was needed to be evaluated with the test set to calculate the accuracy of the model. For the production phase, time-series specific cross-validation would also help for understanding the generalization power of the model to the future unseen data. In addition, there may be many domain-specific features or some fundamental features that highly affecting the model performance, a few examples of these might be given as hourly weather condition, Vacation and Special days, features regarding energy consuming factories and sun set and rise data.

In-sample errors and out-of-sample errors are used to measure the accuracy of prediction models. The in-sample error, which refers to the training phase, is a measure of how well the model fits the data. However, the out-of-sample error, which refers to the test phase, is the preferred option to evaluate and compare the forecasting techniques' strength, predicting future values (Chatfield 2001).

MAPE is the most frequently used measure to evaluate the model accuracy (Hahn et al.). While MSE and RMSE measures are dependent on the data scale and unit, MAPE is scale-independent, making them better choices for model comparison from different scales (Armstrong and Collopy 1992). By taking the square of the errors, MSE, which calculates the mean of the errors squared disregards the direction of errors. It is calculated as (Chatfield 2001):

$$MSE = \frac{1}{n}\sum_{1}^{n} e_t^2.$$

Where, n is the number of observations in the sample, and e_t is the error at the time t, such that

$$e_t = actual_t - forecast_t.$$

RMSE has the same unit of measurement as the data, taking the root of the MSE. It is calculated as (Chatfield 2001):

$$RMSE = \sqrt{\frac{1}{n}\sum_{1}^{n} e_t^2}.$$

MAPE, which reports the average of the absolute errors as a percentage of the actual values, is calculated as (Chatfield 2001):

$$MAPE = \frac{1}{n}\sum_{1}^{n}\left|\frac{e_t}{actual_t}\right| \times 100.$$

In this research, models are evaluated using root mean squared error (RMSE) to be directly comparable to energy readings in the data. RMSE has calculated in two ways. First to represent the error of predicting each hour at a time (i.e. one error per-hourly slice). Second to represent the model overall performance (one value).

In addition, forecasts are produced with a walk forward method. Walk forward makes predictions by moving stepwise through the samples making a forecast at each step. After a forecast is made, the test value is added to the end of the training set and reused, which is shown this process in the figure 1-3.

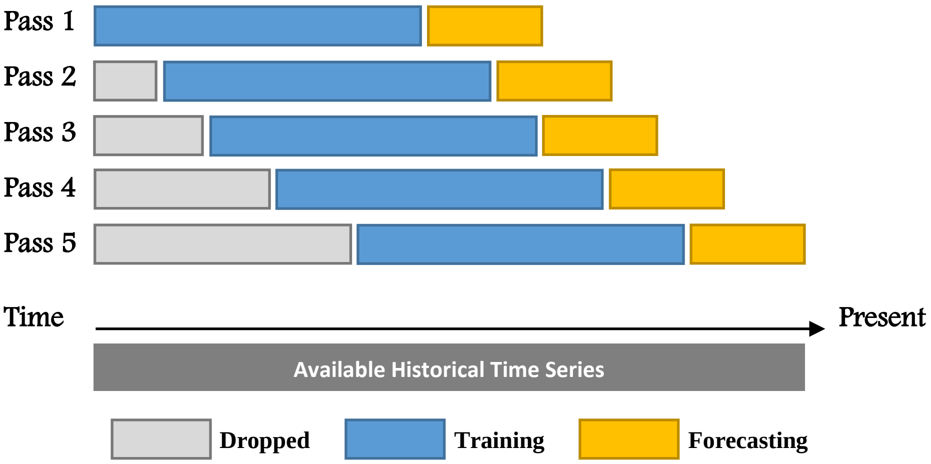


Figure 1- 3 Walk forward method

1.5 Limitations of the Study

Some of the limitations of this research and the modeling approach are:

1. Several relevant variables, which can be considered by models to predict the next 24-hours demand, as the influential factors and might be of interests are:
    - Weather variables in major cities as the major consumers of energy. The correlation between weather features and electricity demand can be analyzed and applied as an influential input in the model.
    - Electricity demand in Portugal and France, as the shared energy regions with Spain as the shared energy regions. The power transfers between these regions and their correlation might influence the model performance.
    - Encoded correlated weekday categories, which implicitly is considered by the model by applying the 7^{th}, 14^{th}, 21^{th}, etc, lag features, and holidays feature.
2. Multiple seasonality in the data, for instance, both weekly and yearly trends cannot be allowed by SARIMAX models.

Chapter 2 – Literature Review

2.1 Introduction

In recent years, artificial intelligence (AI) in general and machine learning (ML) techniques in specific terms as well as a growing trove of publicly available energy consumption data have been proposed for accurate power forecasting and optimal decision making in energy planning. It enables generating companies to manage energy demand effectively to cost.

Energy consumption is on the increase and has a significant impact on the environment. Current predictions show that the growing trend of CO2 emissions which is often held responsible for most of the Earth's progressive warming will continue (Leduc et al. 2016). Therefore, international political, economic, and environmental research has focused on energy consumption reduction and energy efficiency improvement to cope with the problem of global warming and over-exploitation of natural resources.

2.2 Energy Optimization Methods

This section presents a review of the existing literature about energy optimization methods and energy consumption forecasts in terms of their unpredictability. There are numerous studies have been done in optimizing energy performance so far of the new or existing building and two categorized methods in the development of data-driven prediction techniques. These data-driven prediction models can be learnt from simulation-collected data using building performance simulation tool such as TRNSYS and ESP-r to collect energy-related data to train data-driven models and they can be trained through real data using smart buildings equipment to collect data to train models (Crawley et al. 2008; Zhao and Magoulès 2012; Foucquier et al. 2013). While most common optimization methods are simulation-based, they have their own restrictions and strengths. The literature is analyzed towards time series analysis as the most popular approach for forecasting demands using machine learning techniques. Among all data-driven techniques, ANN can be seen as the most widely used.

Rodrigues et al. (2014) used Artificial Neural Networks (ANN) to predict daily and hourly Short-Term Load and household electricity consumption, taking into account apartment location, occupants' numbers, electric appliance consumption and, hourly meter system. In this study, a feed-forward ANN and the Levenberg-Marquardt algorithm showed a good performance.

Friedrich and Afshari (2015) used a time series forecasting approach in developing a Transfer Function (TF) model for forecasting the city's electricity load of Abu Dhabi using hourly measured weather variables including global solar irradiance (GHI), wind speed, temperature, and specific humidity as inputs. The proposed model with various combinations of exogenous inputs is compared with Autoregressive Integrated Moving Average (ARIMA) model and to an ANN model all demonstrating better accuracy for all tested forecasting horizons.

Zhang et al. (2016) proposed a novel hybrid model including eps-SVR and nu-SVR models to develop a performance prediction model for forecasting the electricity load of buildings. They employed a differential evolution (DE) algorithm to optimize the performance of the Support Vector Regression (SVR) model, finding the best model parameters and corresponding weights for both eps-SVR and nu-SVR models to forecast both half-hourly as well as daily electricity consumption for an institutional building in Singapore. The results of their proposed model showed a lower mean absolute percentage error (MAPE) for both the daily and half-hourly energy consumption data.

Ahmad et al. (2017) compared the ML model's accuracy in predicting the hourly HVAC energy consumption of a hotel in Madrid, Spain, utilizing two machine learning-based methods, namely artificial neural networks and random forests (RF). It was found that ANN is capable of performing marginally better than RF with a root-mean-square error (RMSE) of 4.97 and 6.10 respectively. However, it was seen that both of the models can be feasible and effective in building energy prediction.

Lusis et al. (2017) showed the forecasting granularity and one day-ahead load forecasting accuracy for residential customers can be affected by the calendar and training set scale respectively. Statistical analysis has been shown that the regression trees approach substantially outperforms ANN and support vector regression techniques despite the similarity of average root mean square error (RMSE) for all techniques.

Deng et al. (2018) applied and compared six data mining techniques including Support Vector Machine and Random Forest for estimating Energy Use Intensity (EUI) for commercial office buildings in the US and the plug loads, and lighting loads of HVAC, based on the 2012 CBECS microdata. The machine learning algorithms SVM and RF provided more predictive model accuracy based on a large number of outliers in the CBECS dataset.

To overcome the ML techniques' challenge of getting stuck in the local optimum which negatively affects the optimization model's performance, Divina et al. (2018) proposed to apply an ensemble learning approach to forecast the short-term electrical consumption. In this study, Divina et al. considered an ensemble learning scheme to achieve very accurate predictions.

Amasyali and El-Gohary (2019) developed hybrid machine learning prediction models, training from both collected real data from an office building (e.g. building energy consumption, outdoor weather conditions, and occupant behavior) and simulation-generated data. The hourly prediction regression model outputs of the outdoor weather-related factor and occupant behavior-related factor were fed to an ensemble model to forecast cooling load consumption.

Divina et al. (2019) compared and analyzed the statistical and Machine Learning based strategies' performance in predicting energy consumption in non-residential smart buildings using the electricity energy consumption data collected from thirteen smart buildings located on a university campus in Spain. The authors showed the highly accurate predictions can be reached in favor of strategies based on Machine Learning approaches and the historical window's optimal size optimization.

Wang et al. (2018) developed an RF model for predicting hourly building energy. The hourly electricity consumption of two educational buildings in North Central Florida was predicted based on an adopted homogeneous ensemble approach. To train RF models, different input variables were compared to search the feature space that has a critical impact on the prediction model's performance. RF prediction model was show better performance in comparison with regression tree (RT) and SVR models. Based on yearly and monthly data to train the RF model, energy usage prediction could be enhanced considering the changes in semesters' energy behavior.

Chapter 3- Project Description

3.1 Introduction

In recent decades, many countries around the world have suggested several methods to the improvements in the buildings' energy efficiency and power demand prediction which have been, for instance in Europe, the largest area in consuming energy. Li et al. (2010) have shown that the energy performance improvement of buildings is the key instrument to save 60 billion Euros annually. On the other hand, the growing demand for generating energy and constructing new buildings caused by the rapid growth in the world population has been considered a leading contributor to greenhouse emissions. Therefore, energy efficiency in the building sector and power demand forecasting in energy demand management with high accuracy have gained considerable attention to minimize the amount of harmful gas and fossil fuel consumption (Li and Wen, 2014).

Hence, the prediction and optimization of energy consumption have always been a long-lasting concern of many researchers due to the overwhelming growth in the number of reliable datasets leveraging machine learning (ML) models (Seyedzadeh et al. 2018).

Load forecasting techniques for projecting future electricity demands as a heated topic of research become a fundamental subject for operations and power systems planning. The operating cost of power generating companies can be negatively affected by the lack of accuracy of load forecasts (Haida and Muto, 1994). There have been several distinguished classes in this context based on the lead-time of the forecast.

3.2 Energy Demand and Consumption

While power is the rate at which work is done, energy is the capacity to do work. While the speed at which a person walked would be analogous to power, the distance traveled, if a person goes from point A to point B, would be analogous to energy. Energy is recorded using watthour or kilowatt-hour (kWh), while, power in the context of a building is usually measured in watts (W) or kilowatts. Energy demand is generally categorized into a) electricity demand and b) heating/cooling demand.

The building electricity demand is the amount of electricity consumed to operate the electrical equipment in a building. The electricity demand can be affected by the ventilation system, the electrical equipment's efficiency, and the behavior of the occupant. To reach a defined level of comfort in buildings, the units of thermal energy, as heating and cooling loads, needed to be fed to or removed from space by the heating, ventilation, and air conditioning system (HVAC) system (Burdick 2012). The heating and cooling loads of a building can be influenced by some factors such as location, orientation, time of the year, and the building's indoor design conditions (Burdick

2012). In the building management system, both inside load data of a building are commonly recorded in kilowatts or megawatts.

3.3 Demand Load Forecast

The availability of sufficiently reliable electricity short-term and long-term demand predictions is required for planning electricity generation and transmission systems Properly. The entire real GDP (gross domestic product) can be directly linked to these projections for estimating its growth in general. From the societal progress aspect, electricity of paramount importance as a basic human need. In recent decades, it has been introduced as a tradeable commodity to the market while many countries' power industry has been deregulated. The whole stakeholders, in Spain, were exposed to uncertainty in a high amount, by the "Electric Power Act 54/1997 law", due to the countless factors which directly has been linked to electricity price and the unachievable task of electricity storage in large quantities (Ortiz et al. 2016). Therefore, the reliability of forecasting techniques at all scales (hourly, daily, long-term, etc.) in this new market of generation, demand, and especially prices, to be able to participate in them more efficiently, has become a basic need. The changes analysis of income elasticity in Spain, from a short-term perspective, can be allowed by a simple framework which is presented in this section to describe the main features of Spain's electricity consumption and to apply the index decomposition methodology. Additionally, to predict the electricity load for the next day in the most precise way, we will develop a reliable forecasting tool.

Chapter 4- Data Analysis

4.1 Project description

This chapter compared the forecasting capabilities of classical statistical models versus modern neural network implementations on a realistic task of short-term energy demand forecasting. The main question the project asks is:

What forecasting model and supervised learning problem formulation gives the lowest MAE given constrained computation power?

Timeseries forecasting models implemented in addressing this question are:

1. Naïve
2. ARIMA- Autoregressive Integrated Moving Average
3. SARIMAX-Seasonal Autoregressive Integrated Moving Average with eXogenous Parameters
4. Long-Short Term Memory Neural Network

Features used to generate forecasts include autocorrelated hourly energy consumption. A detailed description of each feature is the energy demand lags ranging between 7 days (168 hours) and 1 month.

The output of each model was always the peak expected demand per hour for the next 24-hour period. This forecast was generated from 00:00 each day, throughout the testing period (see cross validation).

4.1.1 Supervised Learning Problem Framing

To predict easily the next hour in advance, the data was isolated from the energy dataset in the format:

Table 4- 1 Isolated features

time	day_forecast	actual_load
2016-01-01 00:00:00	23273.0	22431.0
2016-01-01 01:00:00	22495.0	21632.0
2016-01-01 02:00:00	21272.0	20357.0

A window of calendar days consisting of 24-hour segments was created to predict the next 24 hours in advance. Each hour from $day - 1$, is used to forecast each hour of the current day.

The “actual_load” feature is isolated to reach the following format:

Table 4- 2 Hour-by-hour transform

Date	$\boldsymbol{h00}$	$\boldsymbol{h01}$	$\cdots$	$\boldsymbol{h23}$
2016-01-01	22431.0	21632.0	$\cdots$	24000.0
2016-01-01	22113.0	20515.0	$\cdots$	26029.0

Hence, according to the problem definition, the aim is to forecast loads of the next day at any given hour using the hourly loads of the previous day, reducing a multiple input, multiple outputs problem into 24-univariate naive forecasts.

4.1.2 Walk Forward Validation

Using the above structure, we can establish a walk forward method of predicting the next value. The table 4-3 shows how for each hour of the day, there is a separate model to predict the next day's predicted maximum load for the given hour. In this case ARIMA is a distinct statistical model for hours h0, h1, ... h23.

Table 4- 3 ARIMA model prediction for each hour of the day

Date	$h00$	$h01$	$\cdots$	$h23$
2016-01-01	22431.0	21632.0	$\cdots$	24000.0
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$ARIMA - h0$	$ARIMA - h1$	$\cdots$	$ARIMA - h2$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
2016-01-01	22113.0	20515.0	$\cdots$	26029.0

4.1.3 Feature Engineering (Windowing)

The data is windowed per day to prepare the data for the walk forward validation model structure, shifting the hourly data at the time t to obtain the time $t - 1$. The data from time $t - 1$ is used to fed the model.

In a similar way to define more features, we shifted the data by x steps, and removed the data’ last x steps. A similar transform result can be seen in the below table.

Table 4- 4 Feature creation by shifting the data

Date	$\boldsymbol{h00}$	$\boldsymbol{h01}$	$\cdots$	$\boldsymbol{h23}$
$\boldsymbol{t}$	22431.0	21632.0	$\cdots$	24000.0

Days shifted by x steps

Table 4- 5 Shifted days

Day 5	4	3	2	1
Day 6	5	4	3	2
Day 7	6	5	4	3

Days shifted and truncated

Table 4- 6 Shifted and truncated days

Date	t	$t-1$	$t-2$	$t-3$
Day 4	3	2	1	0
Day 5	4	3	2	1
Day 6	5	4	3	2
Day 7	6	5	4	3

Where each set of t represents a vector of length (number of days, 24 hours).

Two variants of this lagging were implemented. The SARIMA and Persistence models used direct lags following the linear sequence of data through time (i.e. to lag one complete day was 24 lags). The neural network used a different structure where one lag represented the same hourly segment, one day prior.

The previous time steps are constructed by the SARIMA and Persistence models, shifting the sequence of energy one hour at a time and output a sequence of 24 values corresponding with the demand of the next day. The table 4-7 describes the learning problem's structure.

Table 4- 7 The data structure of the supervised problem

Input Lagged Features						**Outputs**
30 days prior	$\cdots$	7 days prior	2 days prior	Previous day	$\gg MODEL \gg$	Forecast
$\boldsymbol{h0 \cdots h24}$	$\cdots$	$h0 \cdots h24$	$h0 \cdots h24$	$h0 \cdots h24$	$\gg MODEL \gg$	$h0 \cdots h24$

While feature vectors are created and stacked by lagging the original sequence at different intervals and aligning similar lags respectively, the data needed for a forecast of a single day, constituted by one 2D matrix of lagged features. In the figure 4-1, the output as seen is a 24-hourly predictions' row vector.

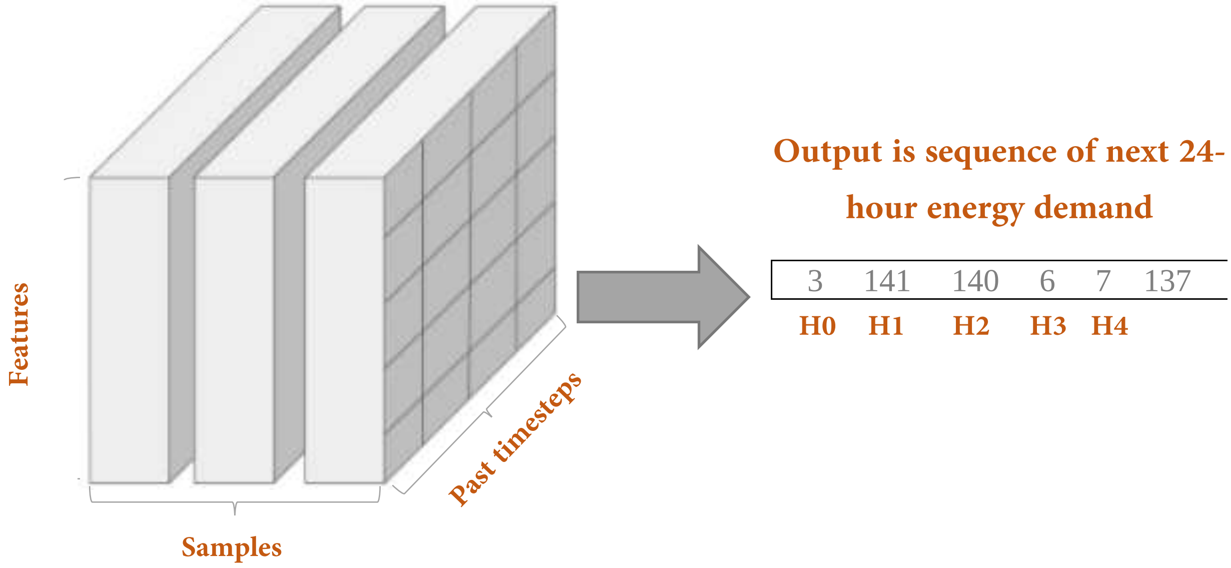


Figure 4- 1 Flatten the features and lags into a single vector

Based on the assumption that each hour of the day had a stronger autocorrelation with the same hour a day prior than the hour prior, the shifted features was correlated. However, the feature input of the LSTM problem framing, which was different from the SARIMAX, was the same. The structure of the supervised problem is seen in the table 4-8.

Table 4- 8 The data structure of the LSTM problem

I/O	date	h00	h01	···	h23
lag features	2015-12-01	21331.0	20622.0	···	25101.0
	···	···	···	···	···
lag features	2016-01-01	22431.0	21632.0	···	24000.0
		⋮	⋮	⋮	⋮
	LSTM-h0	LSTM-h1	···	LSTM-h23	2527
		⋮	⋮	⋮	⋮
Forecast	2016-01-02	22113.0	20515.0	···	26029.0

4.1.4 Summary of Used Functions

- transform_to_windows: converts the data from row data into windowed rows where each row is a day with 24 columns representing each hour of the day.
- plot_hour: helper function to view series data
- shift_by_days: helper function to make_shifted_features. calls pd.shift on the input dataframe to shift the data x number of rows.
- make_shifted_features:
  - calls shift_by_days for a list of shift values.
  - shortenes the resulting dataframe
- trim_length: helper function to make_shifted features. Shortens the length of the final dataframe of fatures so there are no NaNs.
- rename_cols: Helper function used in make_shifted_features. Labels the columns of the shifted dataframes with an appropriate label indicating the shift value.

4.2 Energy Dataset

The data can be found in the mentioned link and is published via ENTSOE and REE. This data can play an important role in predicting supply and demand balance. In acknowledgement of the progressions in the existing studies and the great importance of household energy consumption forecast, we present a case study for modelling electrical consumption predication based on analytical data. This comes with changes in electricity consumption patterns that are also affected by energy efficiency improvements and changes in household consumption behavior.

The data is extracted from Kaggle, containing information about Spain’s hourly electricity production and weather from 2015 to 2019.

Weather data was purchased from “OpenWeatherApi”. Data from the five largest cities in Spain was purchased for the previous 8 years. Data includes hourly measurements of temperature (min, max), humidity, precipitation $(1h, 3h)$, snow $(1h, 3h)$, and general description of weather status (text format).

4.2.1 Energy Dataset Explanation

The statistics and descriptive information of energy dataset are provided in table 4-9.

Table 4- 9 Statistics and descriptive information of energy dataset

	Variable	Count	Mean	STD	Min	25%	50%	75%	Max	Dtype
1	time	35064	----	----	----	----	----	----	----	object
2	generation biomass	35045	383.513540	85.353943	0	333	367	433	592	float64
3	generation fossil brown coal/lignite	35046	448.059208	354.568590	0	0	509	757	999	float64
4	generation fossil coal-derived gas	35046	0	0	0	0	0	0	0	float64
5	generation fossil gas	35046	5622.737488	2201.830478	0	4126	4969	6429	20034	float64
6	generation fossil hard coal	35046	4256.065742	1961.601013	0	2527	4474	5838	8359	float64
7	generation fossil oil	35045	298.319789	52.520673	0	263	300	330	449	float64
8	generation fossil oil shale	35046	0	0	0	0	0	0	0	float64
9	generation fossil peat	35046	0	0	0	0	0	0	0	float64
10	generation geothermal	35046	0	0	0	0	0	0	0	float64
11	generation hydro pumped storage aggregated	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	float64
12	generation hydro pumped storage consumption	35045	475.577343	792.406614	0	0	68	616	4523	float64
13	generation hydro run-of-river and poundage	35045	972.116108	400.777536	0	637	906	1250	2000	float64
14	generation hydro water reservoir	35046	2605.114735	1835.199745	0	1077.25	2164	3757	9728	float64
15	generation marine	35045	0	0	0	0	0	0	0	float64
16	generation nuclear	35047	6263.907039	839.667958	0	5760	6566	7025	7117	float64
17	generation other	35046	60.228585	20.238381	0	53	57	80	106	float64
18	generation other renewable	35046	85.639702	14.077554	0	70	88	97	119	float64
19	generation solar	35046	1432.665925	1680.119887	0	71	616	2578	5792	float64
20	generation waste	35045	269.452133	50.195536	0	240	279	310	357	float64
21	generation wind offshore	35046	0	0	0	0	0	0	0	float64
22	generation wind onshore	35046	5464.479769	3213.691587	0	2933	4849	7398	17436	float64
23	forecast solar day ahead	35064	1436.066735	1677.703355	0	69	576	2636	5836	float64
24	forecast wind offshore eday ahead	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	float64
25	forecast wind onshore day ahead	35064	5471.216689	3176.312853	237	2979	4855	7353	17430	float64
26	total load forecast	35064	28712.1299	4594.10085	18105	24793.75	28906	32263.25	41390	float64
27	total load actual	35028	28696.939905	4574.987950	18041	24807.75	28901	3219241015	41015	float64
28	price day ahead	35046	49.874341	14.618900	2.06	41.49	50.52	60.53	101.99	float64
29	price actual	35064	57.884023	14.204083	9.33	49.3475	58.02	68.01	116.8	float64

Overall, there are a total number of 35,064 observations and 29 variables in this dataset, including 292 missing values. The minimum load volume is 18041 MWh and the maximum load volume is 41015 MWh along with the average volume of 28696.939905 MWh while the minimum, maximum, and average price accounts for 9.33, 116.8, and 57.884023 respectively. There is no duplicate row in the dataset but four columns that are constituted by zeroes values.

4.2.2 Data Preprocessing: Energy Dataset

The dataset contains hourly electricity load data and the respective TSO load and energy price forecasts for future data points. We focus on predicting electrical consumption better than the already present forecast in the data. The metrics we are using for comparison are Mean Absolute Percentage Error or MAPE. To achieve this aim, “total load actual” and “total load forecast” features were extracted from the dataset that needs preprocessing as models’ input variables.

This section describes the process used to construct and clean the dataset. Processes are completed by applying the “format_data” function to rename the columns, shorten the text identifier for times, and convert to a Datetime index; and the “interpolate_nans” function to fill the missing values using a linear interpolation method.

In dealing with NAN values, it is important not to change the structure of the data. This can occur through dropping values changes the number of observations in a day. A number of daily observations per day needs to line up with the days before and after or filling missing values with a single value (i.e. series mean value) is not representative of the temporal nature of the data. As there were only a total number of 36 NAN values in the dataset of the length 35064, we used a linear interpolation function without changing the structure of the distribution.

To check for duplicated timestamps, since we were working with sequence data, there needed to be, the correct number of values per 24-hour period. If not, the data could at some point offset and become a source of error for the model.

In this case, each day is 24 hours and contains 24 readings. Therefore, we can calculate how many data points we should have in a given period. Namely, for the 5 years cleaned in this example we have the years 2015, 2017, 2018, 2019 as non-leap years, while 2016 is, so we have to account for it. Hence,

$$(365 \times 2 + 366) \times 24 = 26304 \text{ hours.}$$

Therefore, we have 3 duplicated entries, dropping the extra values and take the first occurrence of the data point by default.

4.3 Analysis of Energy Load Data

In this section, we analyze energy load data as a whole by plotting yearly line, grouped by month line, and grouped by days of the week-mean. Then, repeat this section with each of the hourly slices and investigate the stationarity of hourly segments.

The objective was to understand the structure of the energy demand data as a whole. Questions that help guide us are:

- What is the level, and noise variance, does the data contain any seasonal or trending parts?
- Is the data stationary? If not, can we make it stationary?
- What relationship is there between months and days of the year?
- What differences are there between the actual load and the predicted loads from Spain's TSO?

We see in table 4-10 that the current short-term forecasts capture the same distribution as the actual. We see this in both their respective levels and interquartile ranges (28723 MWh and 28810 MWh for the level of forecasts and demand respectively).

In supplying energy demand, we were also interested in the baseload. The baseload is defined as the minimum demand over some time (usually weekly). In this case, our global baseload is the minimum over the 4 years of data (baseload is 18000 MWh).

Table 4- 10 Statistics and descriptive information of the isolated features

	day_forecast	actual_load
count	35064	35064
mean	28698.281385	28712.129962
std	4575.828854	4594.100854
min	18041	18105
25%	24807	24793.75
50%	28902	28906
75%	32194.25	32263.25
max	41015	41390

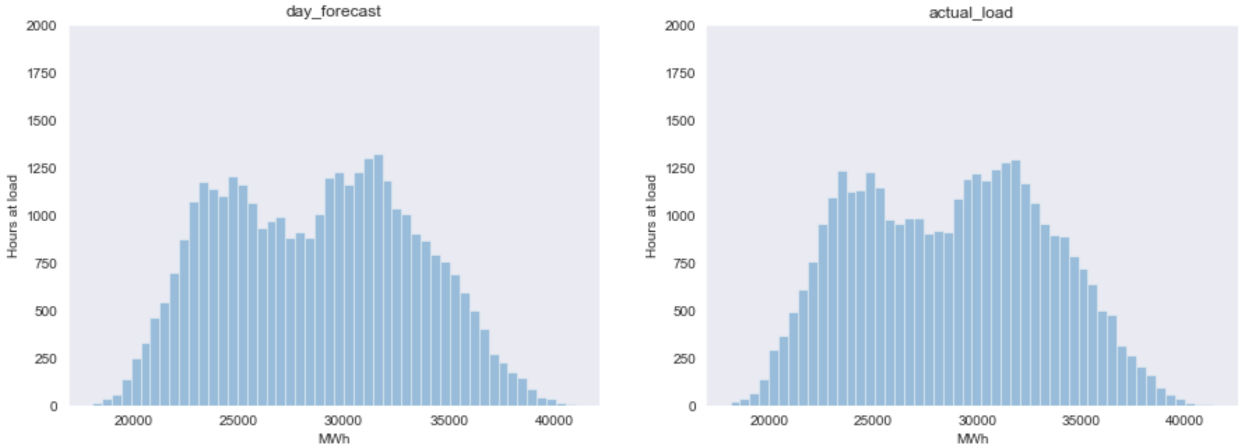


Figure 4- 2 Feature distributions

We see both actual and forecast load have bimodal distributions.

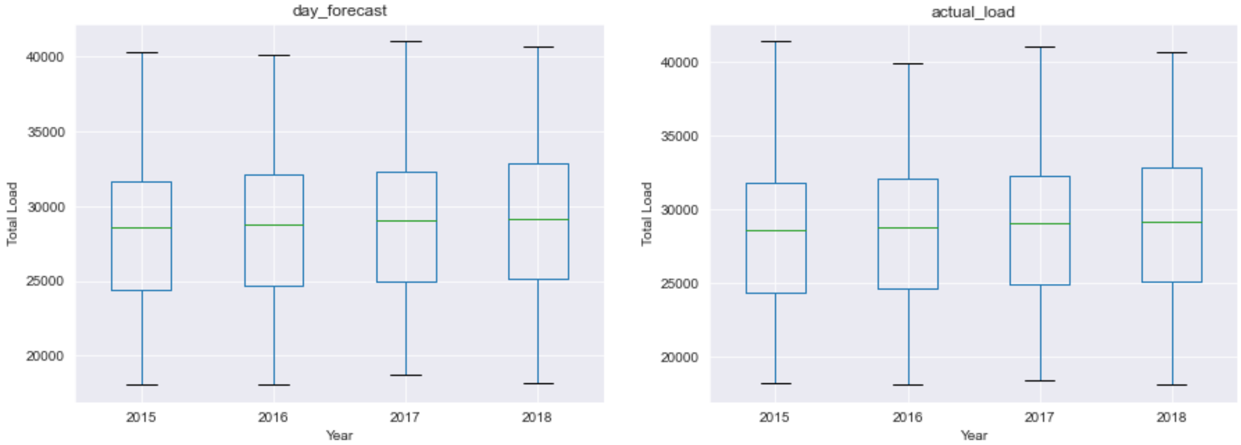


Figure 4- 3 Yearly variability in the distribution

According to the figure 4-3, yearly variability in the distributions is consistent in terms of the IQRs and median values. There appears a small trend of rising median actual load from 2015 to 2018. There are no outliers as would be expected (an outlier in energy distribution would be a shortage or excess of power, both resulting in grid imbalance and a high risk of downtime).

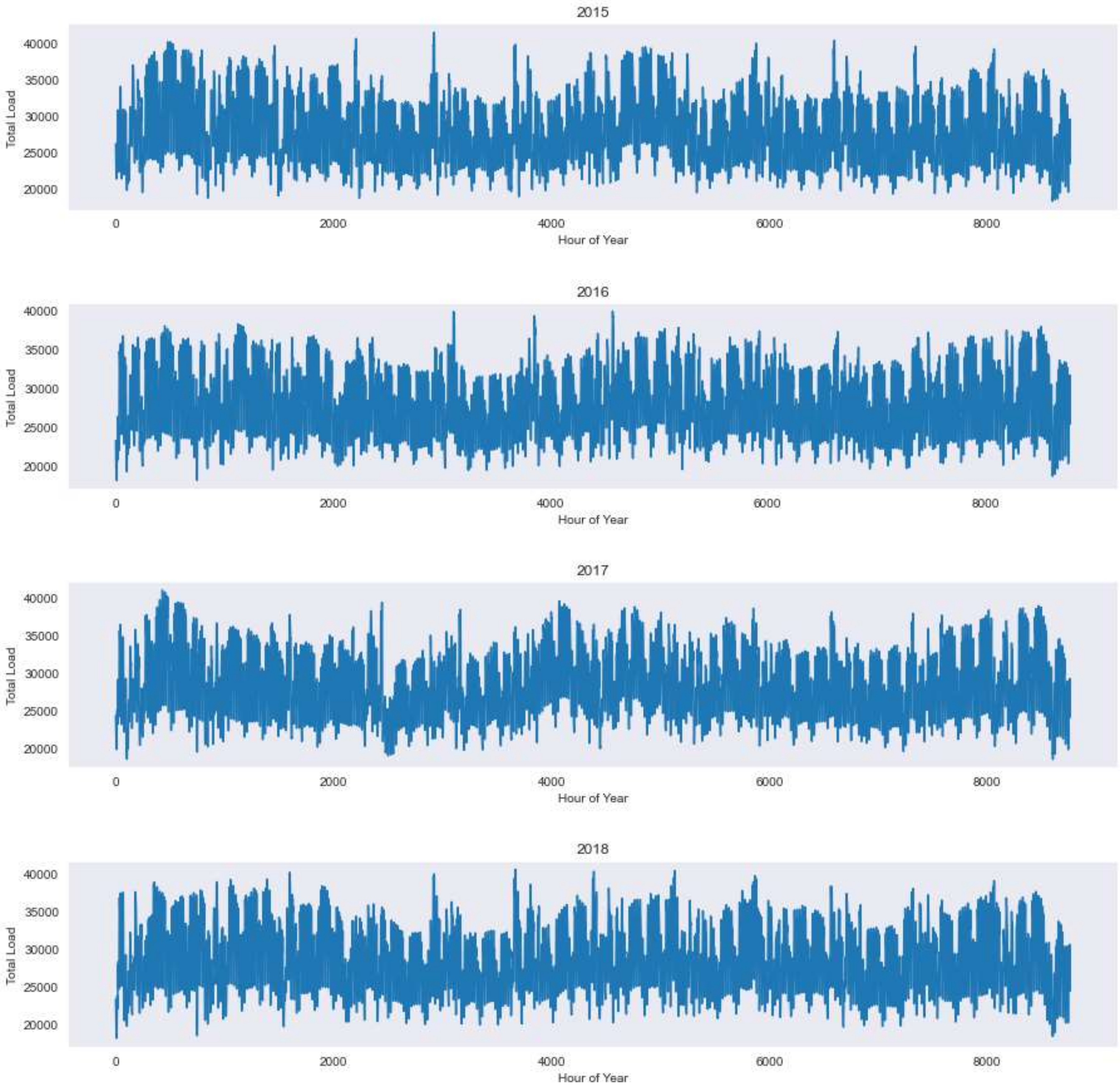


Figure 4- 4 Global load and seasonality

Looking at the global load any seasonality is not obvious. In the summers of 2015 and 2017, there appears a spike in overall hourly electricity consumption.

4.3.1 Demand Variability

We want to look at monthly demand, inter-week demand, and daily profile variability.

4.3.1.1 Monthly Demand Variability

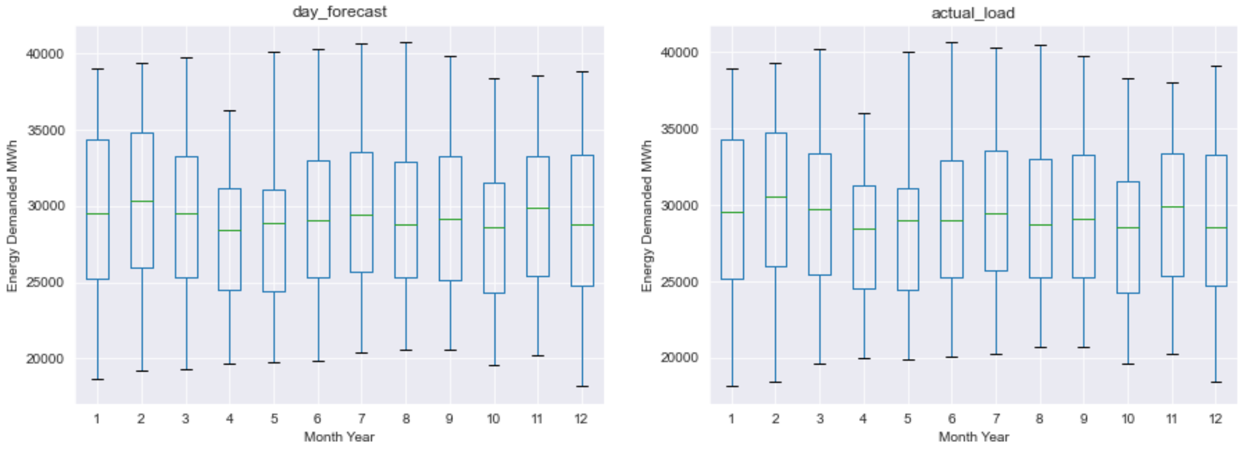


Figure 4- 5 Monthly demand variability

According to the figure 4-5, actual loads show clear seasonal trends throughout the year. The median actual load is highest in months 1,2,3 and 6,7,8, and 11. This corresponds to the winter and summer months and is worth looking closer at temperature data to see if there is a correlation.

4.3.1.2 Inter-week Demand Variability

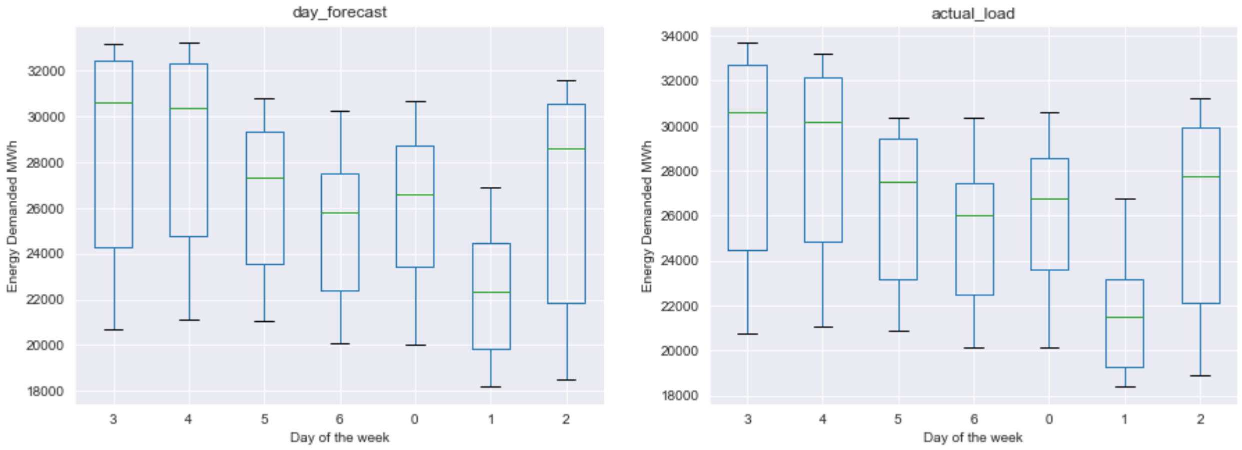


Figure 4- 6 Inter-week demand variability

Figure 4-6 shows that, days of the week show that weekends (days 5 and 6) have lower overall consumption. This is expected because in general businesses are not operating. Also, notice the median is in the upper range of the IQR. This is an indication that most of the power demand is occurring in the upper band. This corresponds to the shape of the daily profile which will be seen in the next section. This observation supports the use of day vectors as features to identify the likely load demanded.

4.3.1.3 Daily Demand Variability

In this section, we transformed the data into the second form.

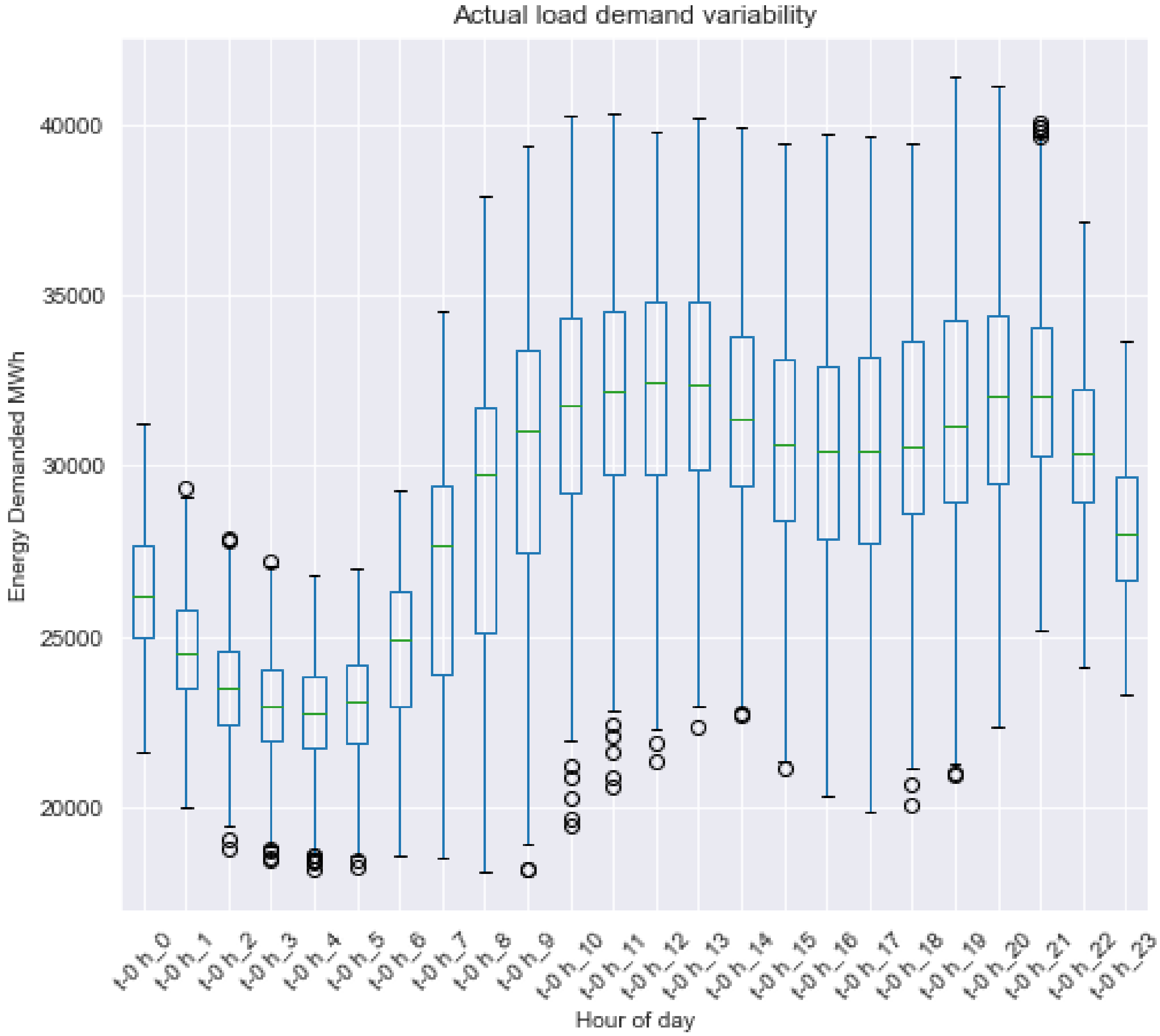


Figure 4- 7 Daily demand variability

From the daily demand variability plot, we can observe how the load remains low over the night and then starts increasing as people wake up, and then continues increasing during the office hours and peaks in the evening when everyone returns home and turns on the electrical appliances.

4.3.2 Load Profile and Shape

4.3.2.1 Mean Yearly Profile

The mean load profile is our target for forecasting. As explained in the problem definition the objective is to model the short-term load, 24 hours in advance.

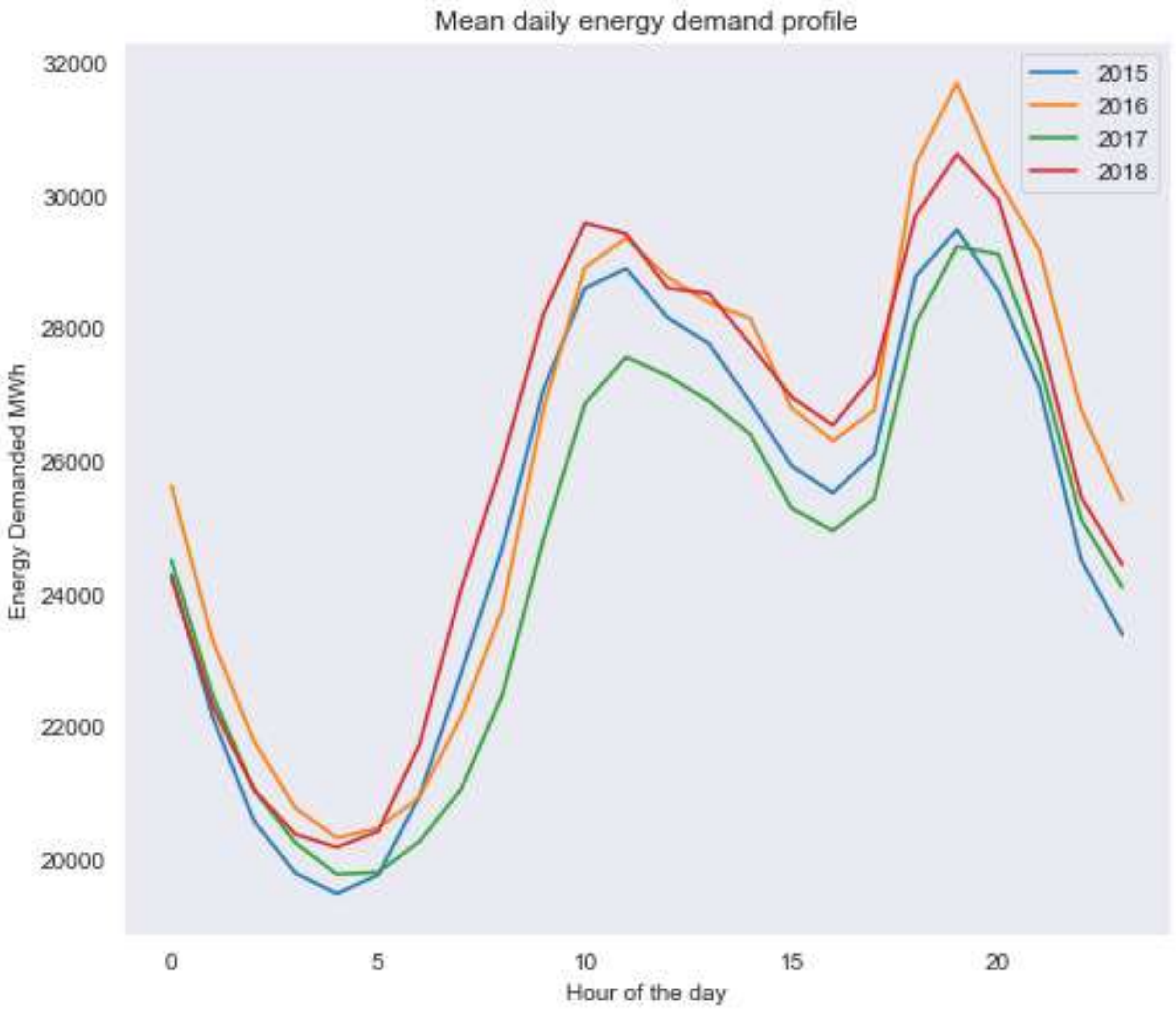


Figure 4- 8 Mean yearly profile

4.3.2.2 Monthly Mean Load Profile

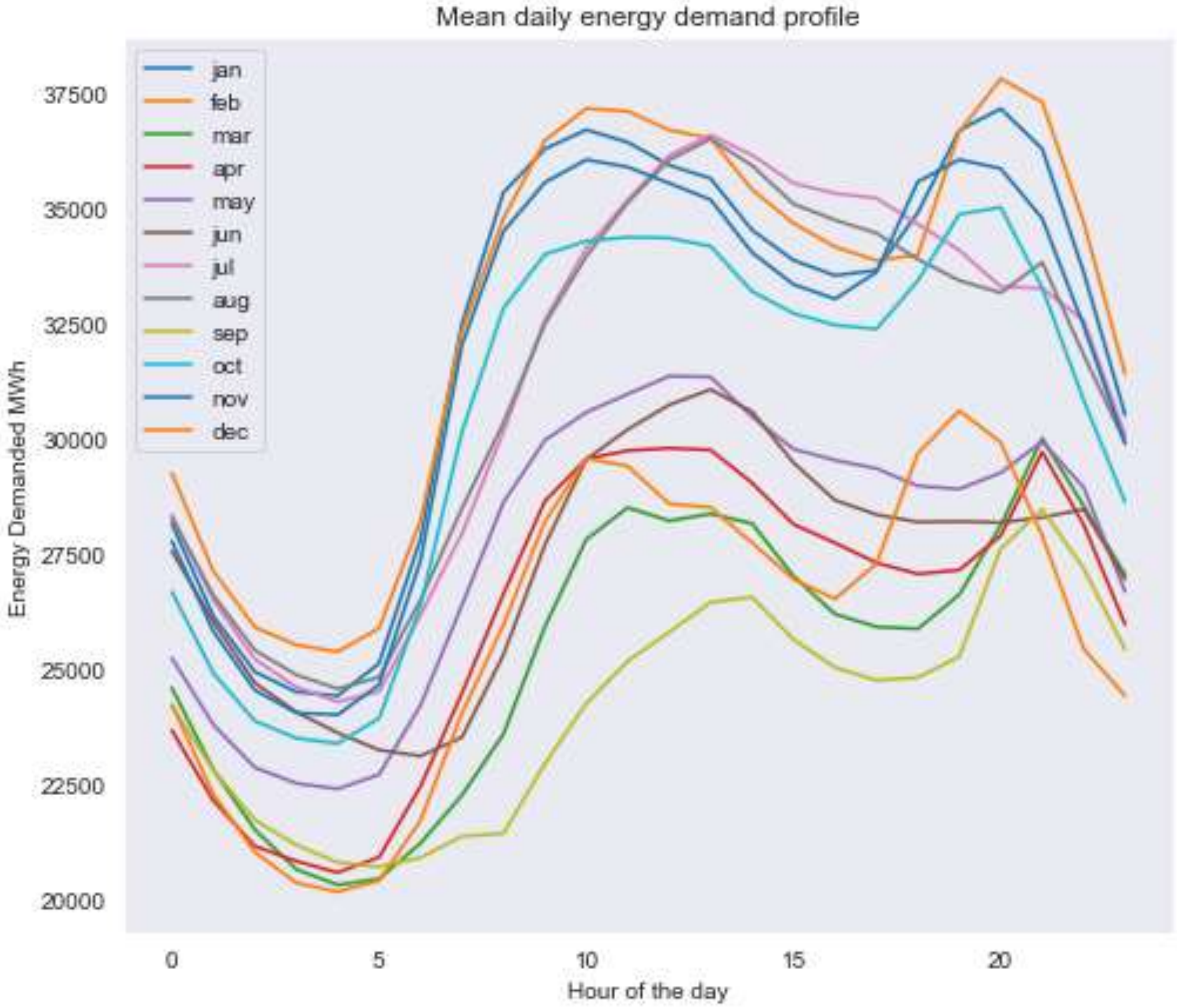


Figure 4- 9 Monthly mean load profile

According to the figure 4-9, mean energy load profile by month show a clear difference in seasonal profile. The months of Jan, Feb, Jul, Aug, Oct, Nov are on average seeing higher baseloads, and higher sustained load during peak hours. The months of Mar, Apr, May, Jun, and Sep have comparably lower.

4.3.2.3 Inter-week Average Load

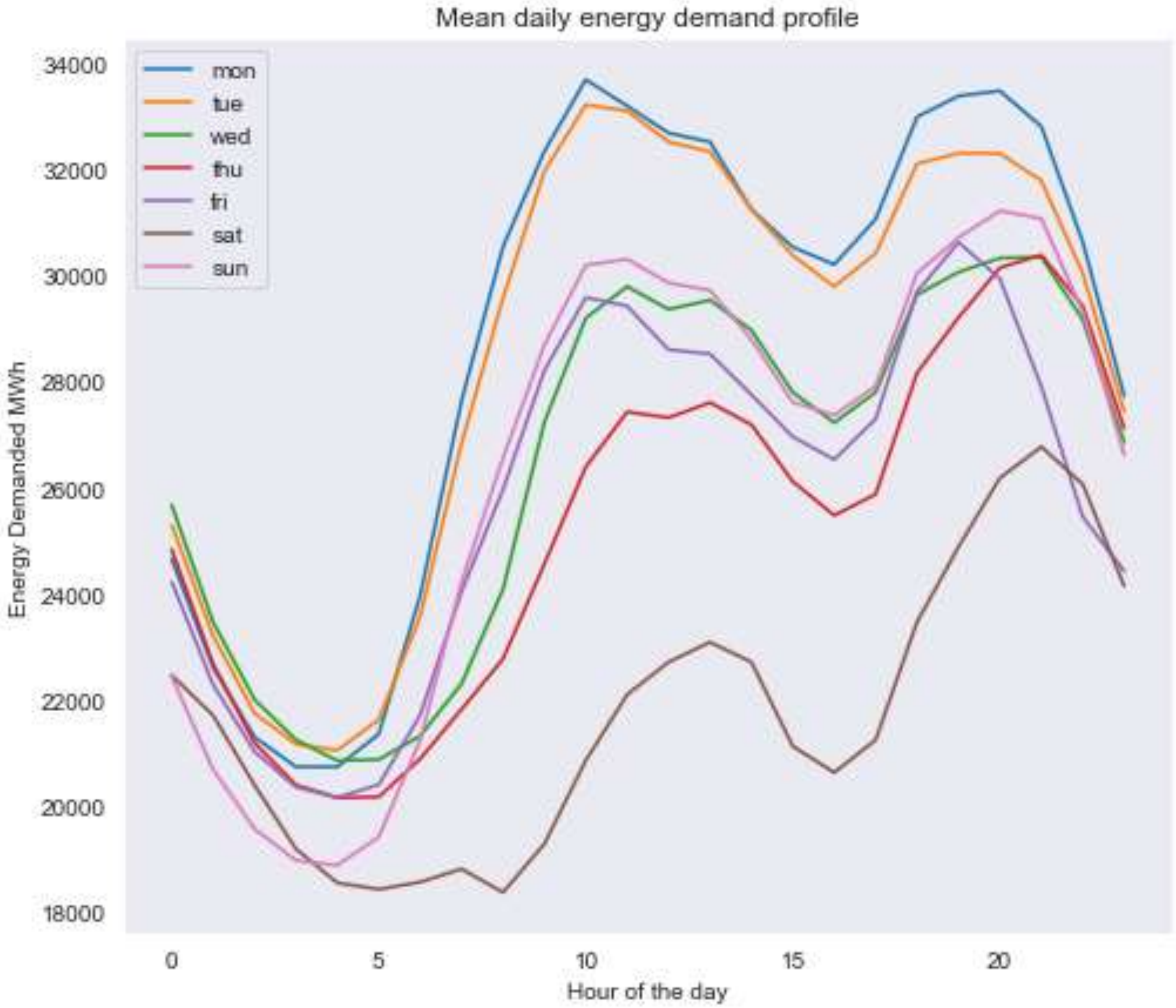


Figure 4- 10 Inter-week average load

4.3.3 Augmented Dickey-Fuller Test for Stationary

As noted above there appeared to be some seasonal anomalies in 2015 and 2017. The ad fuller test is a hypothesis test for time-series stationarity. In this case, we will test on the daily mean energy demanded.

Null Hypothesis: The dataset is non-stationary and therefore differencing must be carried out.

If the p-value is < 0.05 (two-tailed test), we reject the null and assume that the time series is stationary. When studying the individual hourly time series this will be repeated. The results are shown in table 4-11.

Table 4- 11 AD Fuller test results

Test Statistic	**-5.779773e+00**
P-value	5.158246e - 07
#Lags	9.400000e+01
Observation	1.366000e + 03

Therefore, we reject the null based on the p-value < 0.05. This means that as a whole the time series is stationary and does not need to be differenced. However, this might not be the case when analyzing the individual hourly slices.

4.3.4 Hourly Energy Loads

The problem description describes how we are looking at forecasting 24 hours in advance. To do this we think about the problem as 24 individual forecasts. That is why we use hour 0 of yesterday to forecast hour 0 of today.

The analysis above indicated that the general data was bimodal, stationary, showed differences in mean consumption by day of the week, and month of the year.

The following analysis applies the same concepts to the hourly slices. The goal is to understand:

- Which hourly slices are stationary and should be differenced?
- To what extent do daily profiles change through the year?
- What are reasonable lag order and degrees of differencing for each hourly slice (used in ARIMAX model)

4.3.4.1 Mean Hourly Segments

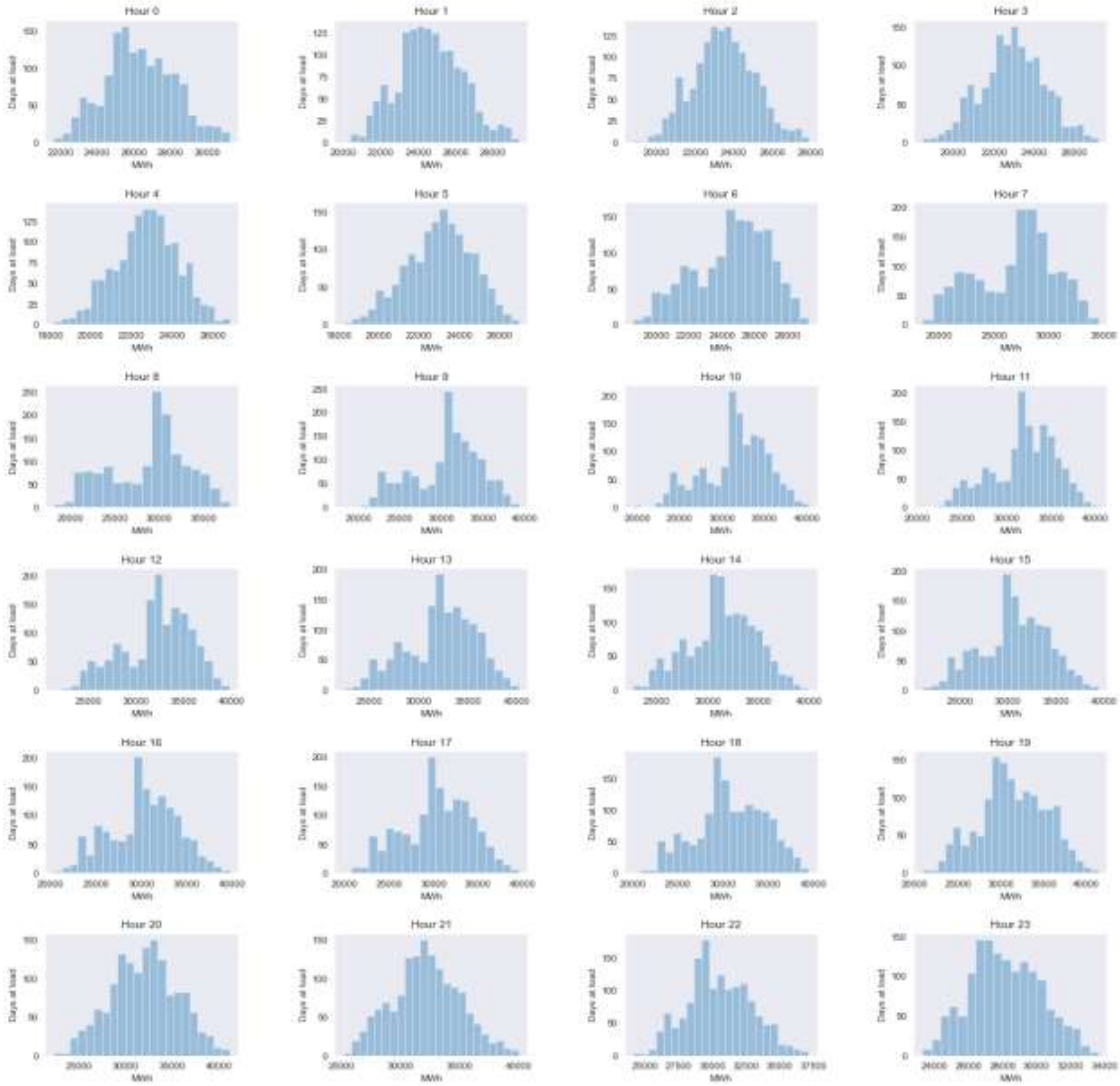


Figure 4- 11 Mean hourly distribution

Figure 4-11 shows that:

- Hours 1-5: Closest to a normal distribution.
- Hours 6-15: Left skewed non-normal.
- Hours 16-19: Possibly slight right skew and bimodal.
- Hours 20-21: Again, close to normal distribution.
- Hours 22-23: Starting to look like a right skew.

Average variability for the 4 years for each hour, 24 rows and 1 column, showing year progress of consumption

4.3.4.2 Inter-week Variability of Each Hourly Slice

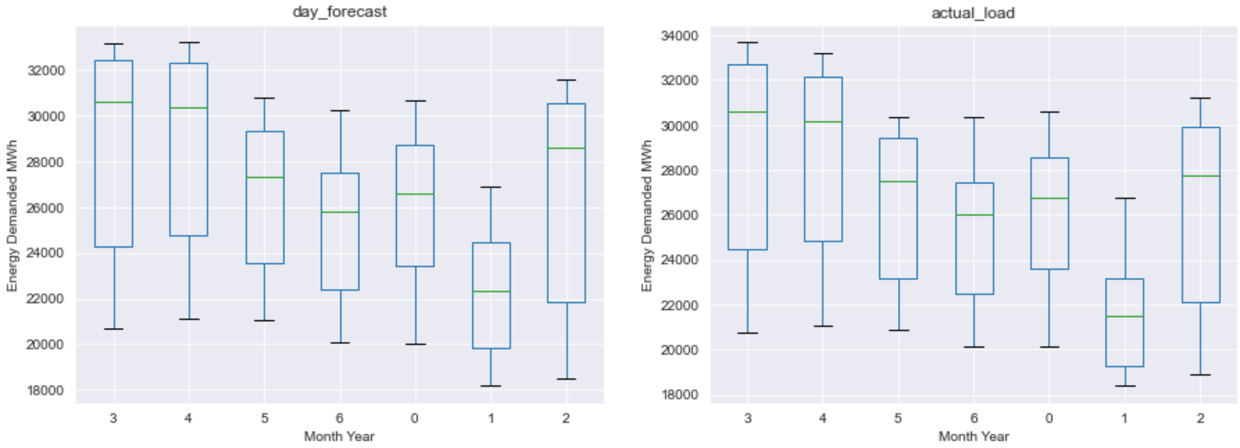


Figure 4- 12 Inter-week variability of distribution of each hourly slice

4.4 Machine Learning Models

4.4.1 Persistence Models

Persistence models (naive models) are simple techniques to forecast future data points by using past data. They are developed to benchmark performance when evaluating more complex methods, compare the performance of feature engineering, hyperparameter tuning, and model architecture against a set of references.

This section implements the walk forward test pipeline that all multi-step models will use. The general process is the following:

1. Import data and transformed it into windows (24 hours in day's window).
2. Split data into train and test.
3. Walk forward prediction generations.
4. Model evaluation.
5. Plot errors.

Persistence Models Evaluated

1. Previous day hour-by-hour.
2. Last 3-day average.
3. Year ago day hour-by-hour.

We have data available from 2015-01-01 to 2019-08-25. For simplicity, while training and evaluating models we fixed the sample from 2015-01-01 to 2018-12-31, a period of exactly 4 years.

The first three years (2015-2017) were used as the standard training set, leaving the final year (2018) as the testing set.

Finally, the forecast horizon is set to 24 hours in advance. Therefore, the problem is defined as predicting the next day's 24-hour slices of expected energy demand.

4.4.1.1 Persistence Models 1: Previous day hour-by-hour

Table 4-12 shows a method that, the energy loads from the previous day were used by the previous day hour-by-hour model to predict the next day on an hour-by-hour basis.

Table 4- 12 Hour-by hour forecasting technique of the model 1

Hour	Current day	-->	Forecast
h0	450	-->	450
h1	389	-->	389
…	…	-->	...
h23	345	-->	345

We defined the previous day persistence model 1, set the “train_test_split” function to split the first 3 years as the train dataset. Figure 4-13 shows the previous day persistence model’s outcome.

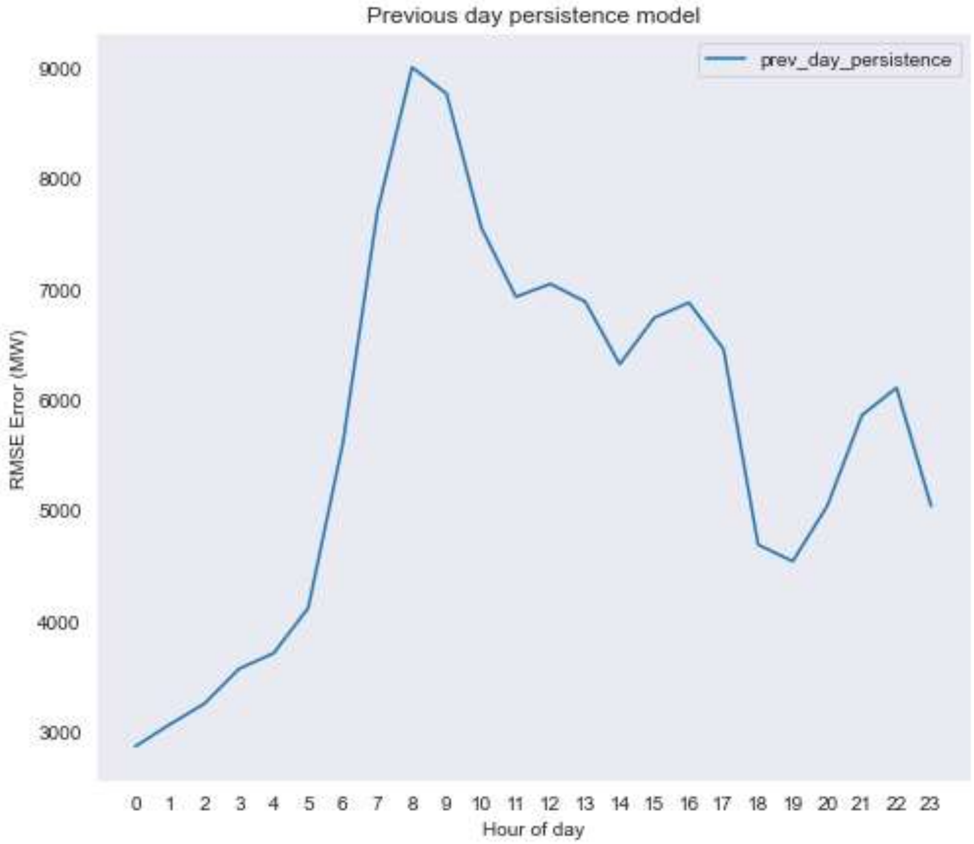


Figure 4- 13 Previous-day persistence model 1's outcome

Table 4-13 provides the RMSE score of our prediction for the first 3 hours with this method.

Table 4- 13 The results of the model 1 for the first 3 hours

	Prev_day_persistence
H0	2858.969991
H1	3058.548695
H2	3246.803937

Models are evaluated using root mean squared error (RMSE) to be directly comparable to energy readings in the data. RMSE has calculated two ways. First to represent the error of predicting each hour at a time (i.e. one error per-hourly slice). Second to represent the model's overall performance (one value).

Forecasts are produced with a walk forward method. Walk forward makes predictions by moving stepwise through the samples making a forecast at each step. After a forecast is made, the test value is added to the end of the training set and reused.

4.4.1.2 Persistence Models 2: Moving average last 3 days

We defined the previous day persistence model 2 (moving average), split the data as model 1. Figure 4-14 shows the model 2's outcome.

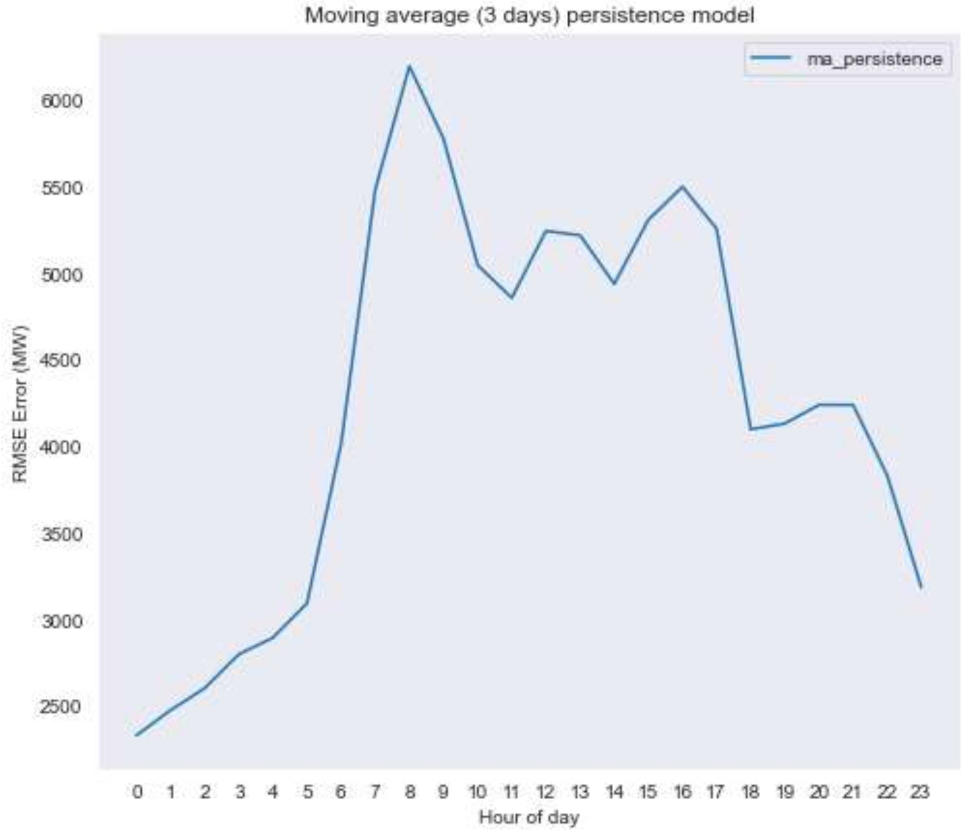


Figure 4- 14 Previous-day persistence model 2's outcome

4.4.1.3 Persistence Models 3: Same day previous year hour-by-hour

The same day previous year uses the energy loads from the previous day to forecast the next day on an hour-by-hour basis. Figure 4-15 illustrates the results of the persistence model 3.

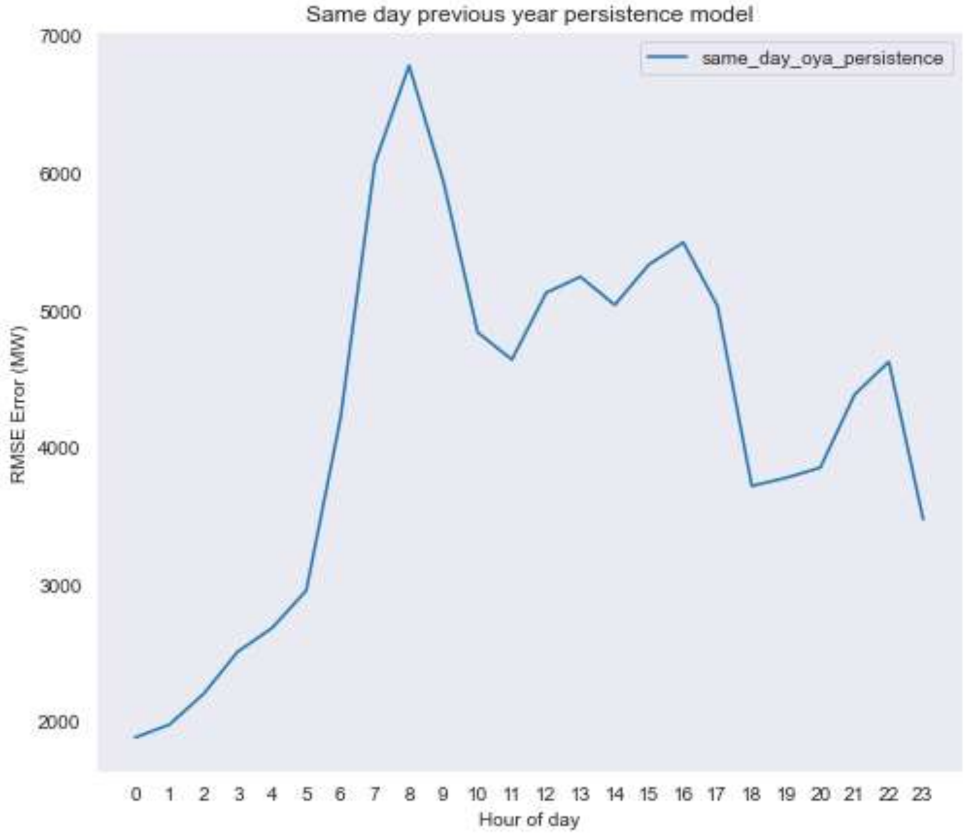


Figure 4- 15 Persistence model 3's outcome

4.4.1.4 Compare the Persistence Models

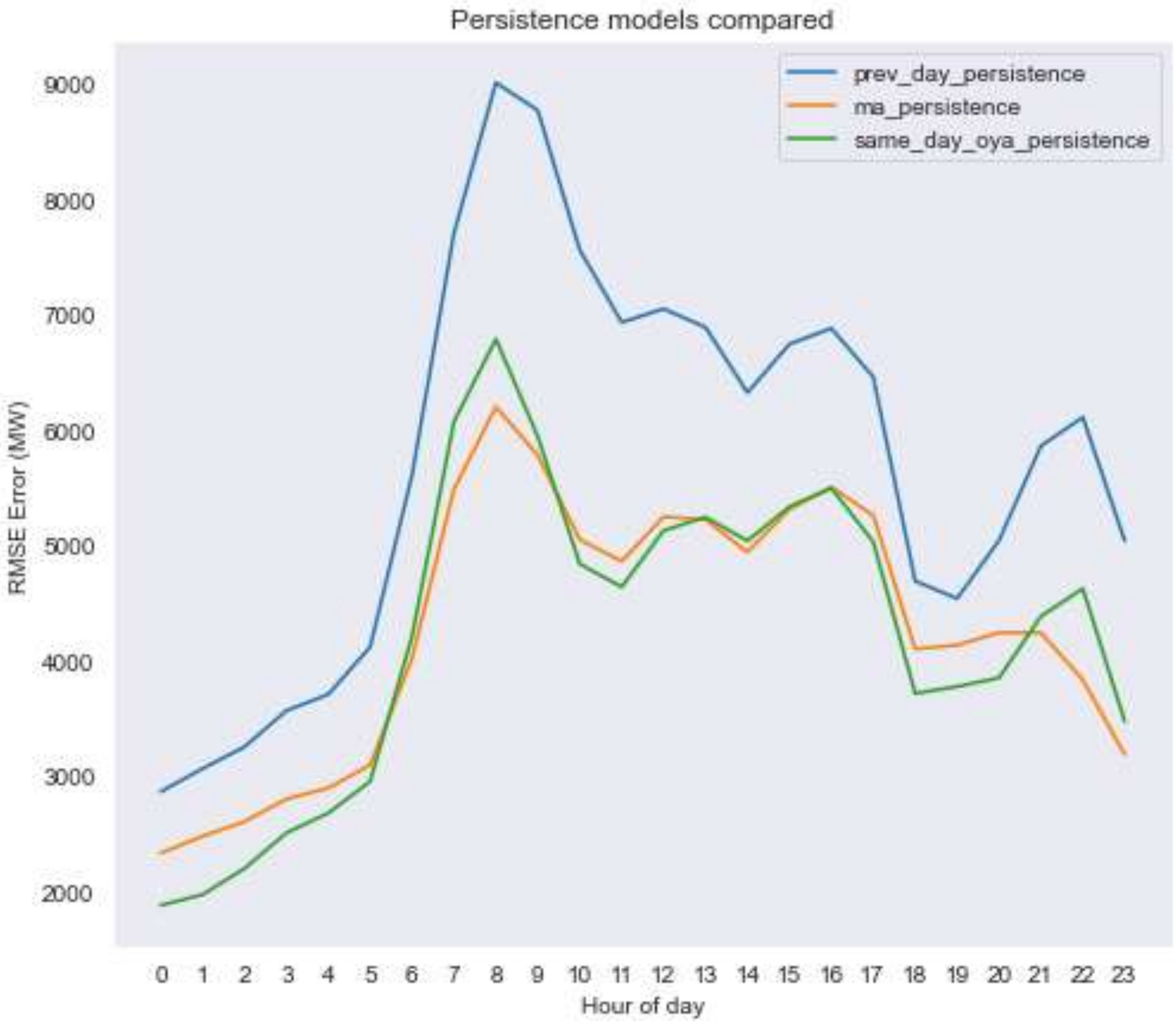


Figure 4- 16 Persistence models' comparison

According to the figure 4-16, we see that the previous-year model outperforms the other models to predict the next day's power demand.

4.4.2 ARIMA Model

ARIMA stands for Autoregressive Integrated Moving Average. Compared with the above model it uses a linear combination of past time steps and moving averages to predict t.

ARIMA takes only a stationary time series. As explored in the data analysis section the load data can be made stationary analysis results.

We applied the ARIMA model from “statsmodels.api” which takes the following arguments:

- p represent the lag order.
- d represent the degree of differencing.
- q represent The order of moving average.

As described in the stationary test in the data analysis section, set of daily mean data was stationary. Here we test if the hourly data is stationary using the “adfuller” test over 1 week of lags (24×7).

Table 4- 14 “adfuller” test results

Test Statistic	-7.804115e + 00
P-value	7.359355e - 12
#Lags	1.680000e+02
Observation	2.613500e + 04

Table 4-14 illustrates the p-value of the test is significantly smaller than the threshold of 0.05 and therefore we reject the null and assume a stationary dataset. Therefore, the default model parameter for (the differencing value) is 0. We investigated a differencing parameter of 24 and 168, corresponding with the previous day and the previous week.

4.5.2.1 Autocorrelation and partial autocorrelation

Description of the plots:

ACF - Describes the direct and indirect relationships between lagging (shifted) autoregressive features. In the relationships between $t, t-1, t-2, t-3$, etc. taking into account the interrelationships between features, in this case, $t-1$, $t-2, t-3$, etc.

PACF - Describes only the direct relationships between lagging (shifted) and autoregressive features.

$p\ (AR)$: Determining the autoregressive hyperparameter value p, is best described as the number of lags beyond which there is no significant relationship. This is seen in the ACF as the point at which plot values lie outside the significance band (light blue horizontal band)

$q\ (MR)$: Determining the moving average hyperparameter value q, is described as the direct relationship between the lag feature and the feature.

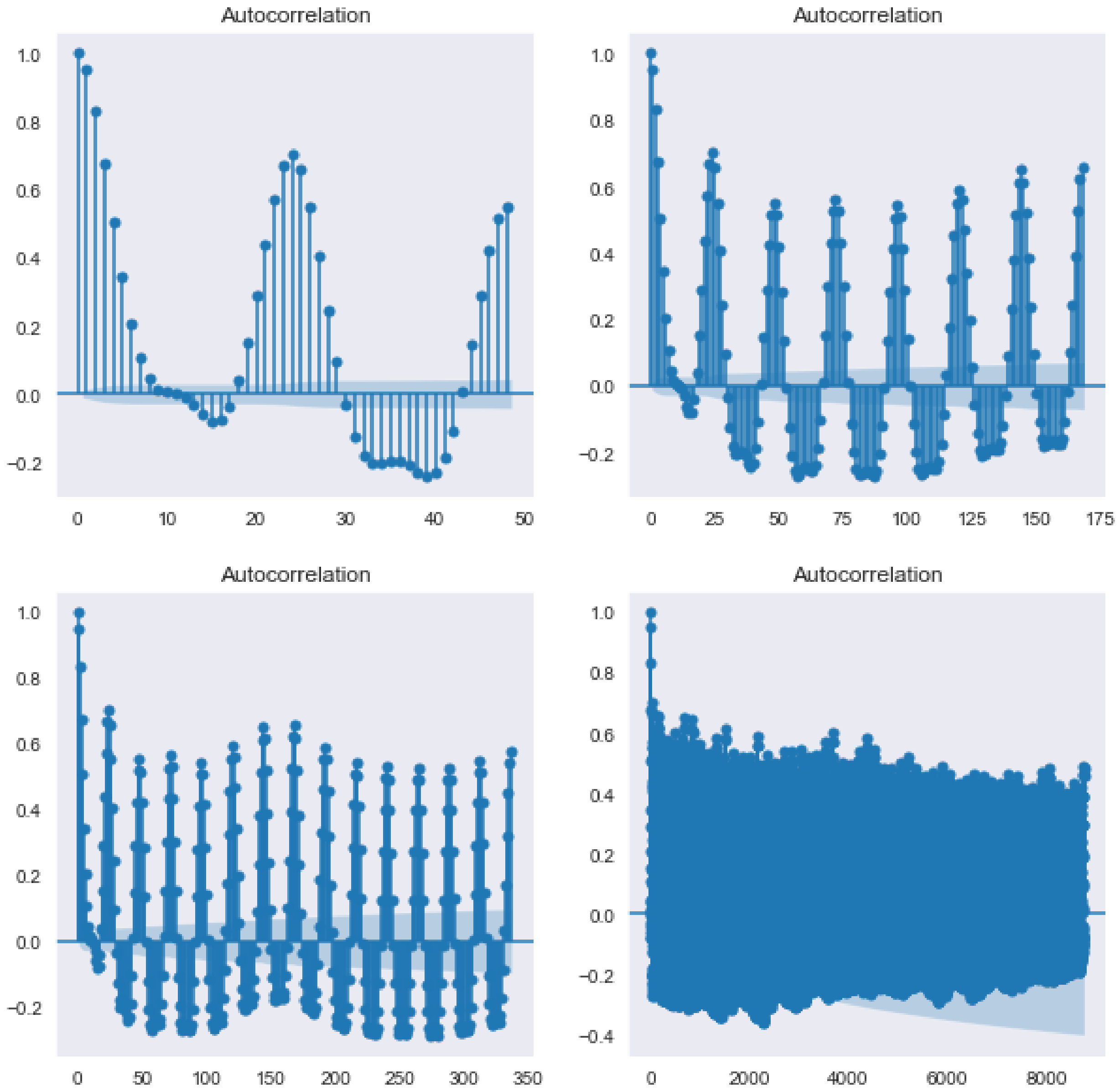


Figure 4- 17 ACF test results

According to the above figure, the autocorrelation plot shows significant positive correlations in the first 9 lags. However, in this problem we are attempting to forecast the next 24 hours and using a lag less than 24 does not capture the full context of a next 24-hour demand forecast.

The cyclic pattern of the autoregressive features is apparent in the plots. In the ARIMA model the assumption is that beyond the chosen lag point (p), there is no correlation. Observing the bottom right plot, we see this point occurs around approximately 4000 lags (approximately $1/2$ year).

For the scope of this project calculating with 4000 lags is not feasible. We will focus on lag points 24 (previous day), 48 (previous two days), 168 (previous week).

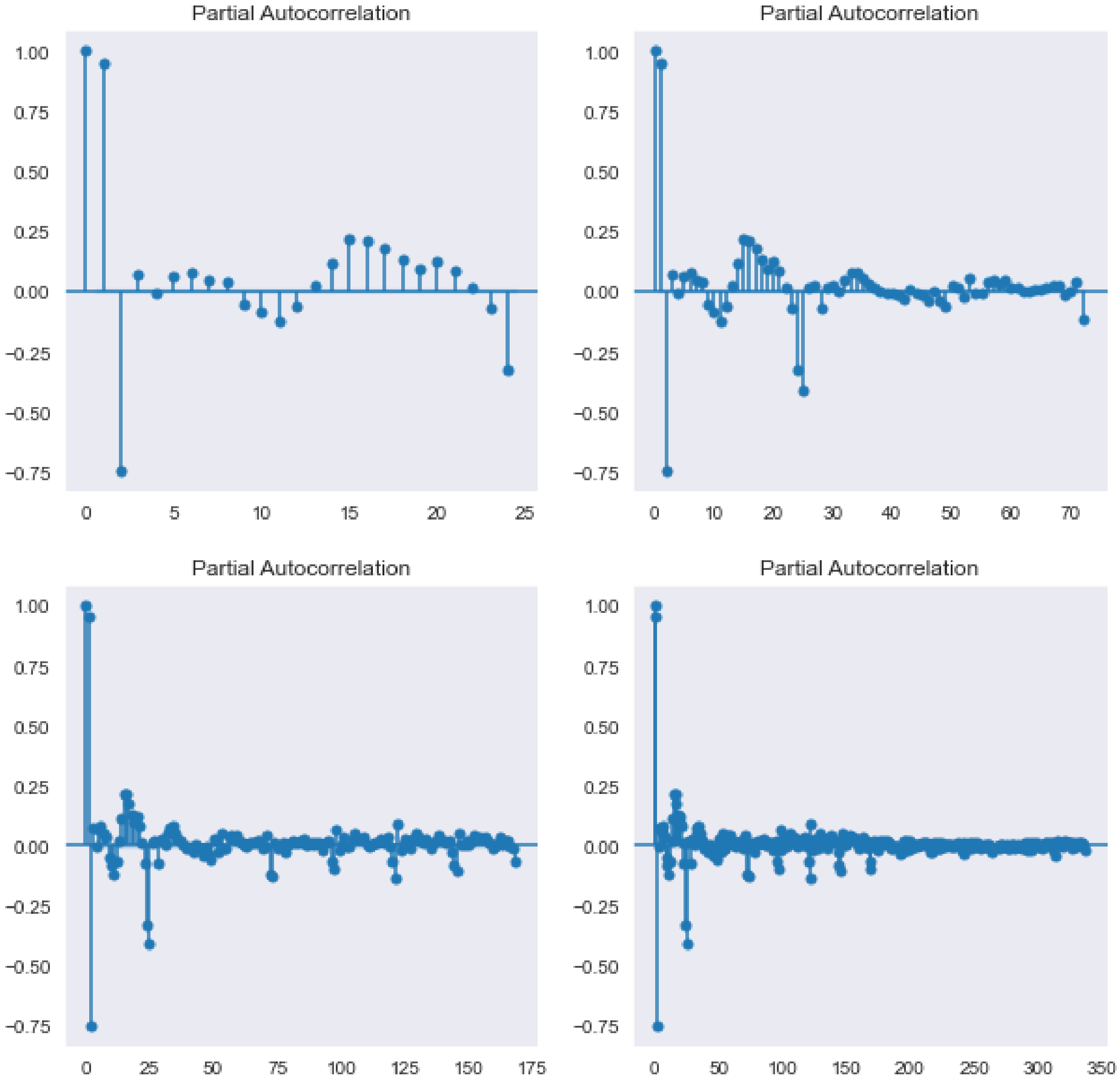


Figure 4- 18 PACF test results

The partial autocorrelation plot shows, by looking at the figure 4-18, that beyond 24 lags there is no significant partial autocorrelation. Considering this we will investigate lags 2, 3, 12, and 24 for the moving average values.

4.4.2.1 ARIMA Model: Baselines

To determine the baseline parameters for the ARIMA model, we set the lag value (p) to 24 for autoregression as using a lag less than 24 does not capture the full context of a short-term day ahead forecast based on the ACF test results in figure 4-17, a differencing order (d) of 0 as we assumed the hourly data is stationary using the "adfuller" test, and a moving average model (q) of 0 to avoid the potential for incorrectly specifying the MA order.

This implies 24 autoregressive features are computed for each day in the training set. Because this is computationally intensive, we run this for $1/365$ test cycle of the walk-forward validation and log the training time.

We implemented a condensed version of the walk-forward validation set. The specific details of the datasets are described below:

- Train: 2017-01-01 to 2017-12-31.
- Test: 2018-01-01 to 2018-03-31.
- Model: ARIMA $(24,0,0)$, prediction of the first $1/90$ days of the test set.

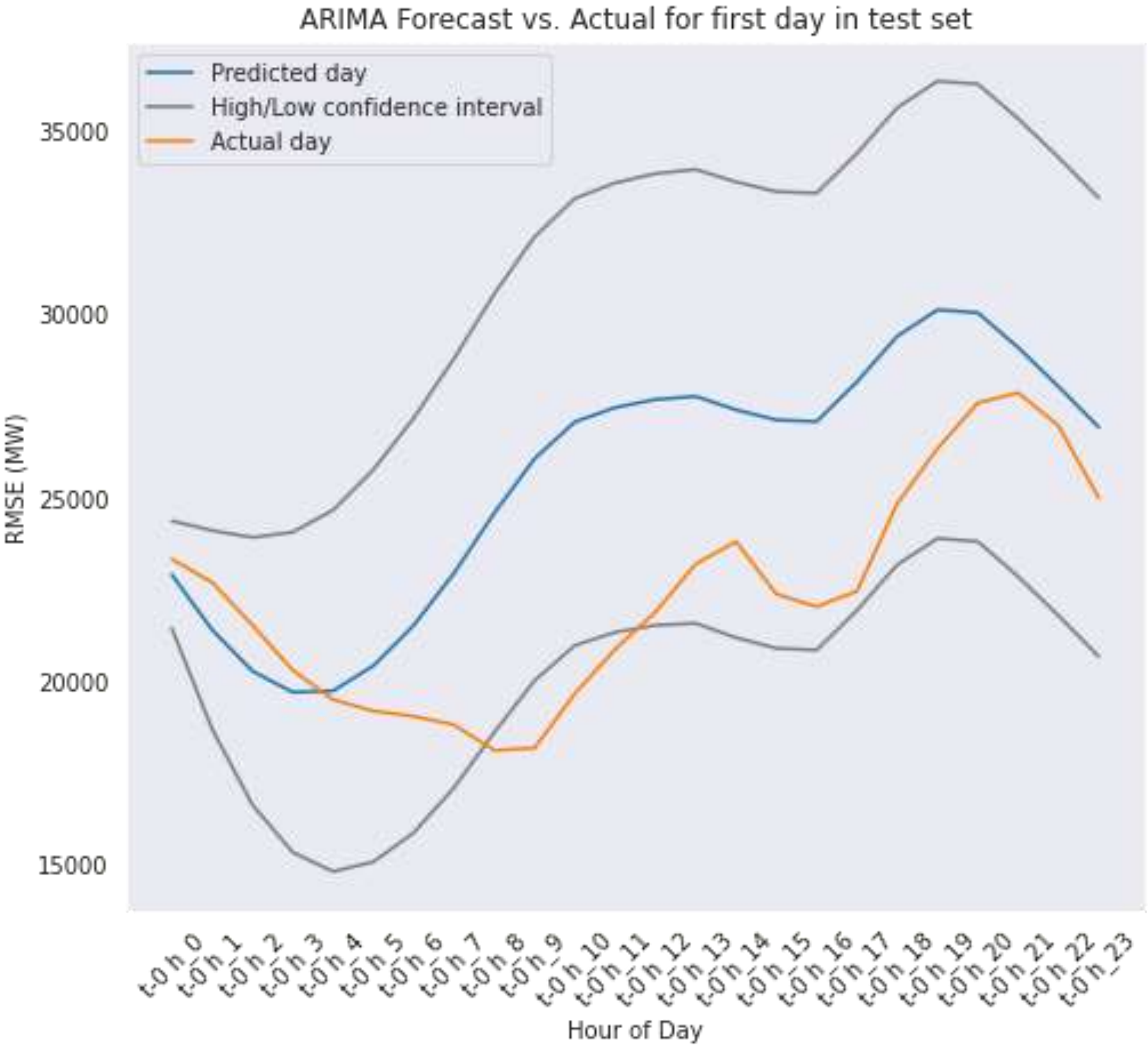


Figure 4- 19 The results of the ARIMA model CHANGE Legend on x axis and y axis

Figure 4-19 shows the first test day forecast vs actual values after training on the whole 3 years of training data.

4.4.3 SARIMAX Model

The SARIMAX model (Chatfield 2001) is more complex and uses mode features resulting in a larger state space to calculate. According to the documentation, the ARIMA model is maintained at a minimum while the SARIMA model has newer implementations. It is not clear that the SAIRMAX is a faster algorithm. In this section, we run a test against an ARIMA and compare it.

As mentioned, the SARIMAX model is possibly a faster implementation. Functionally, the model also offers an additional layer of hyperparameters, $P/D/Q/m$, about seasonality.

- P: Seasonal autoregressive order.
- D: Seasonal difference order.
- Q: Seasonal moving average order.
- m: The number of time steps for a single seasonal period.

The additional features allow us to reframe the forecasting problem to each m periods is a season. Within the season we can set P, D, Q respectively as functions of the season.

The seasonal parameters were chosen based on knowledge of the problem. Baseline seasonal hyperparameter values described below:

- m: 24 to represent the cyclic pattern of energy demand every 24 hours.
- P: 1 to take the autoregressive features from the previous season (i.e. previous day).
- D: 1 to consider the differencing between consecutive seasons (i.e. days).
- Q: 0 to consider that consecutive seasonal forecasts are independent.

The datasets specific details are described below:

- Train: 2017-01-01 to 2017-12-31.
- Test: 2018-01-01 to 2018-03-31.
- Model: SARIMAX $(1,1,0,24)$ predicting the first $1/90$ days of the test-set.

The model results are shown in figure 4-20.

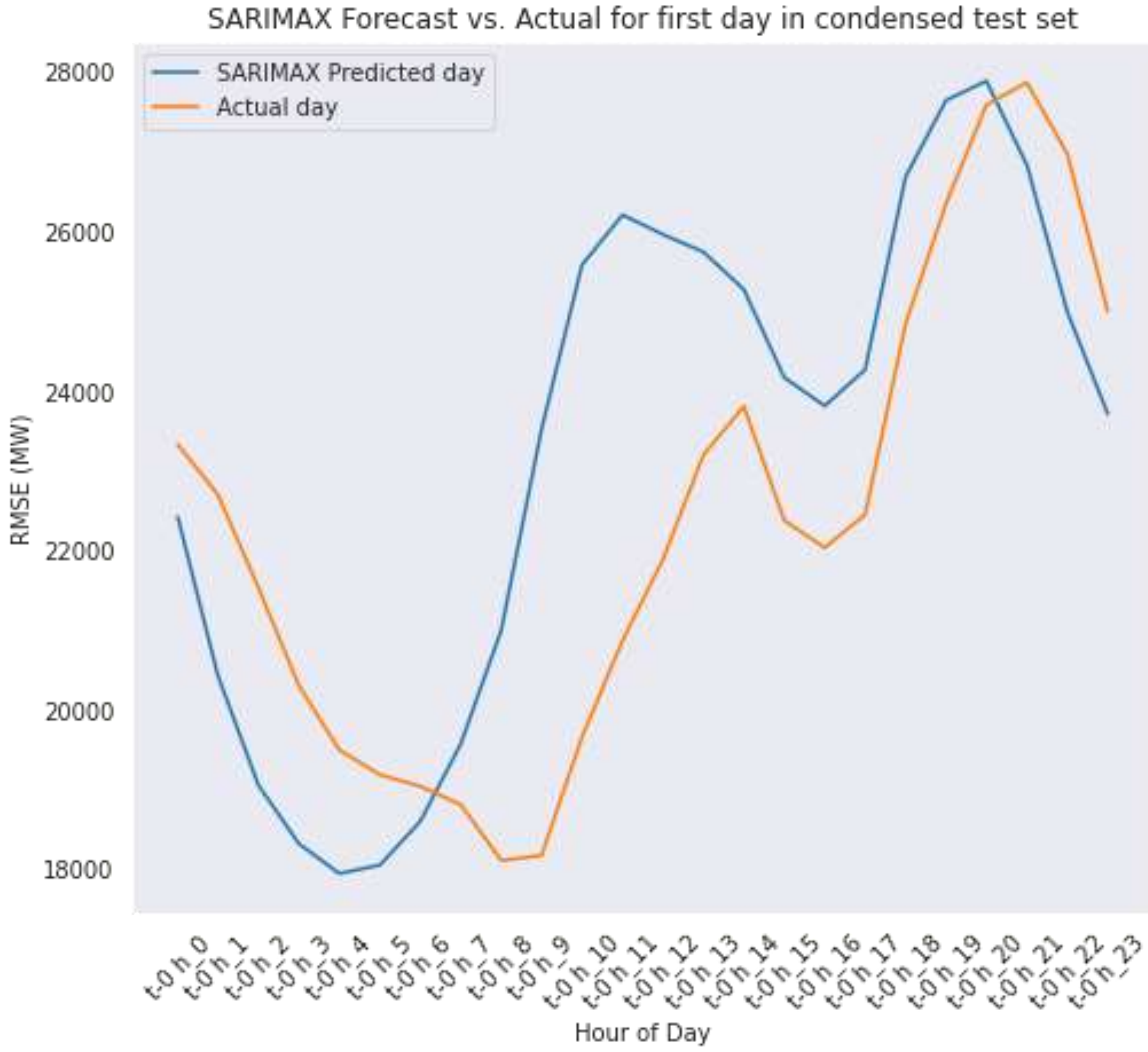


Figure 4- 20 The results of the SARIMAX model CHANGE Legend on x axis and y axis

RSME score is 2642.88 with SARIMAX model. Applying trend and seasonality function gives better results in our dataset.

4.4.4 LSTM Model: Univariate Time Series Forecasting with Keras

In this section, we perform a LSTM model (Hochreiter and Schmidhuber 1997) as simple univariate technique with "Keras", a deep learning application programming interfaces written in Python, running on top of the machine learning platform "TensorFlow". Each hour of the day is structured a univariate sequence as the model's input and output.

The demand of the next 24 hour is forecast by the predictive model. The model makes 24 forecasts corresponding to each hour of the day. The benefits of the method are as follows:

- we take advantage of stronger direct (partial) autocorrelations between $h0, h1, \ldots h23$ of today, the day prior, and so on (compared with the autocorrelation between, $h0, h1, h2, h3$, etc)
- compared with the autocorrelation between days' hour, stronger (partial) autocorrelations between hours of today, yesterday, and so on is taken into account.
- we can train the model on smaller datasets and capture the effects of seasonality.

The data structure for the univariate case is described by the following diagram. In this case, we are predicting hour-by-hour using previous data from the same hour. In this way, each hour becomes a dataset on its own. We can combine these multiple sets into one single block of data with the shape:

- Input (samples, lags, hour slices).
- Output (samples, hour slices).

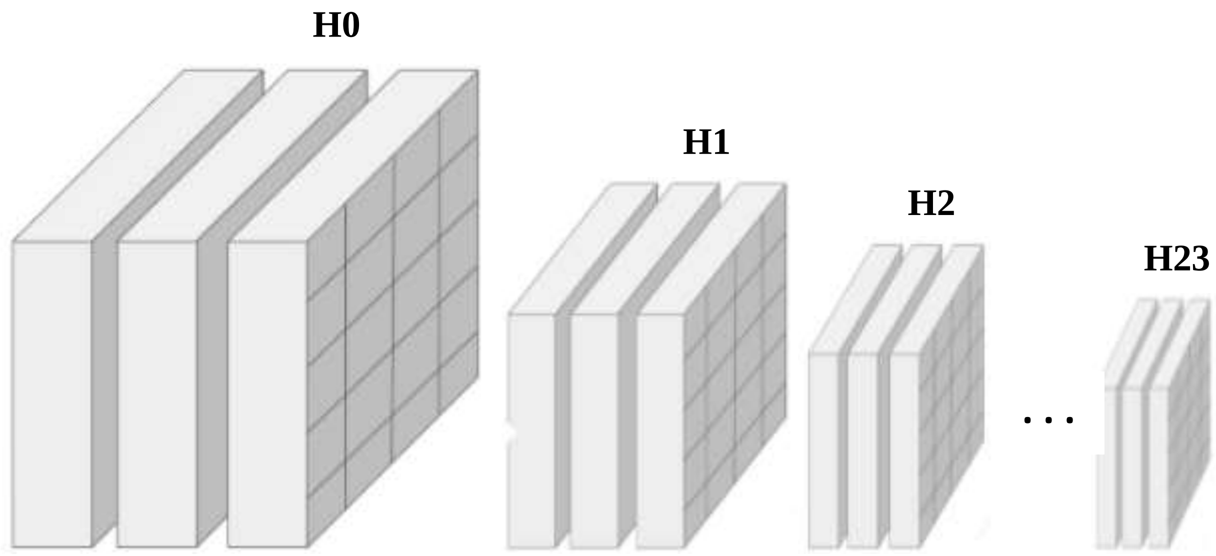


Figure 4- 21 The data structure of the univariate LSTM problem

4.4.4.1 Autocorrelation and Partial- Autocorrelation analysis

We can see electricity demand as 24 hours in each day of the year or as a given hour of everyday of the year for each hour in the day.

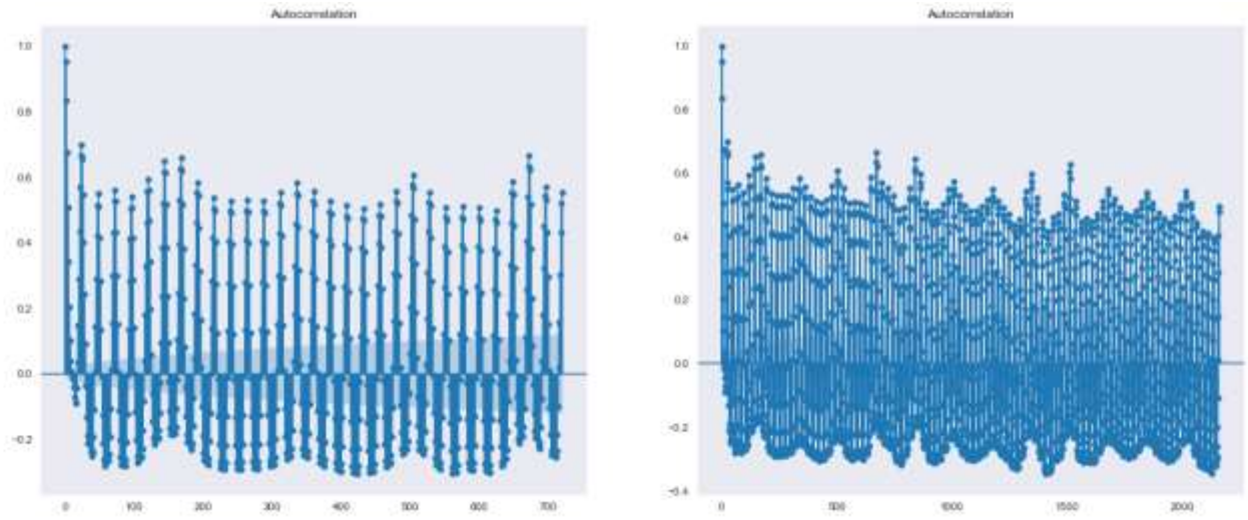
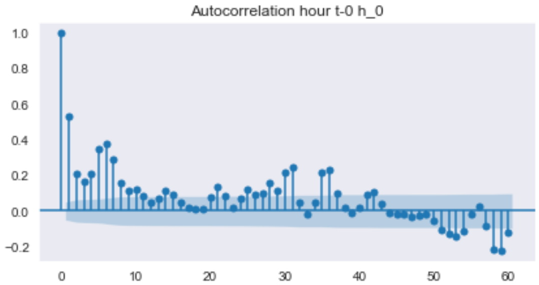
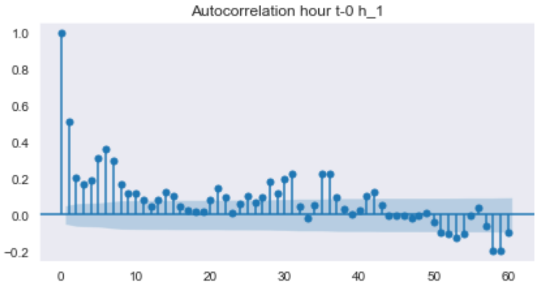
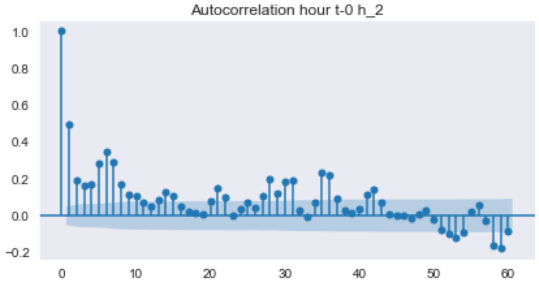


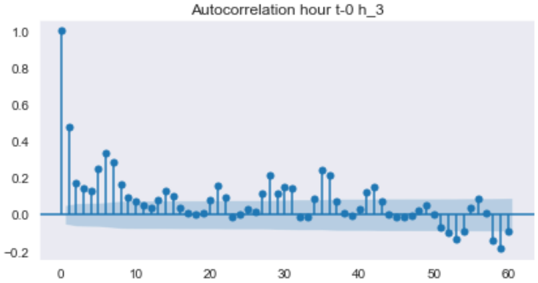
Figure 4- 22 Autocorrelation plot

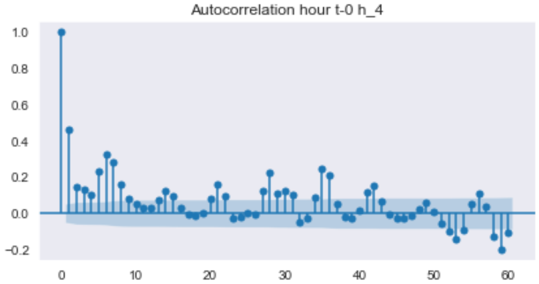
In the above plot, we can see that the consecutive hours have strong influence on each other by using the data in this view.

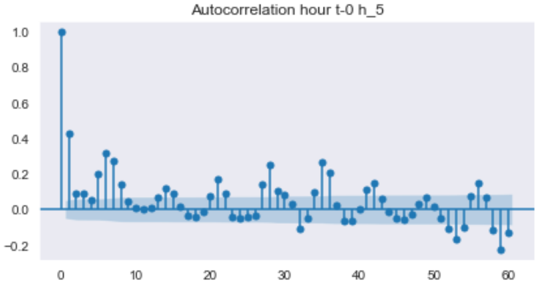

Autocorrelation hour t-0 h_0
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60

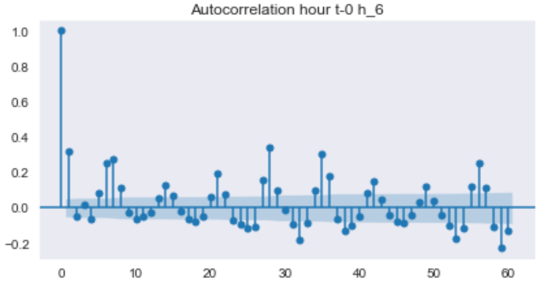

Autocorrelation hour t-0 h_1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60

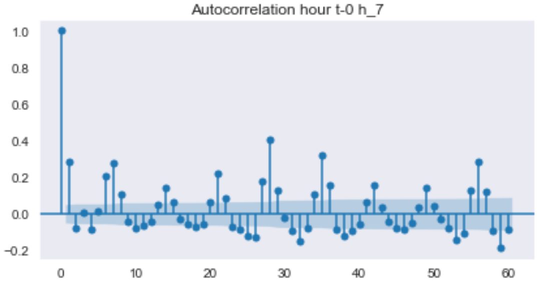

Autocorrelation hour t-0 h_2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60

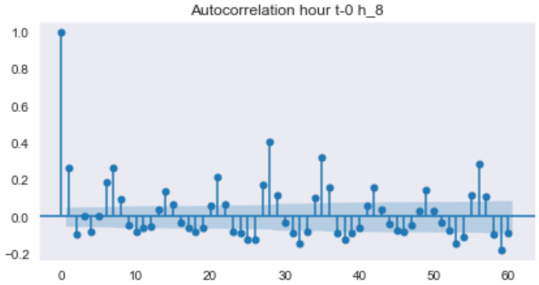

Autocorrelation hour t-0 h_3
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60

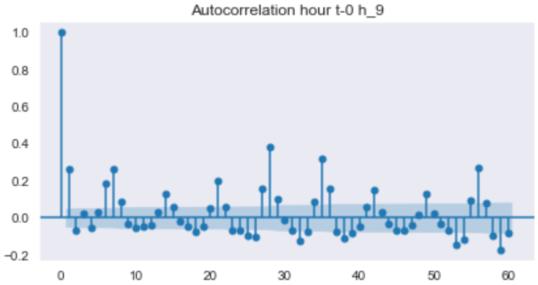

Autocorrelation hour t-0 h_4
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60

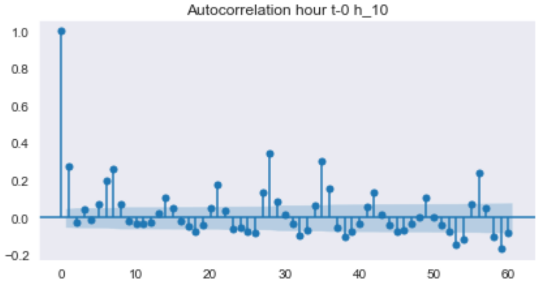

Autocorrelation hour t-0 h_5
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60

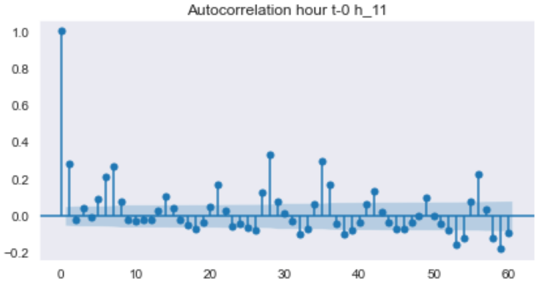

Autocorrelation hour t-0 h_6
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60

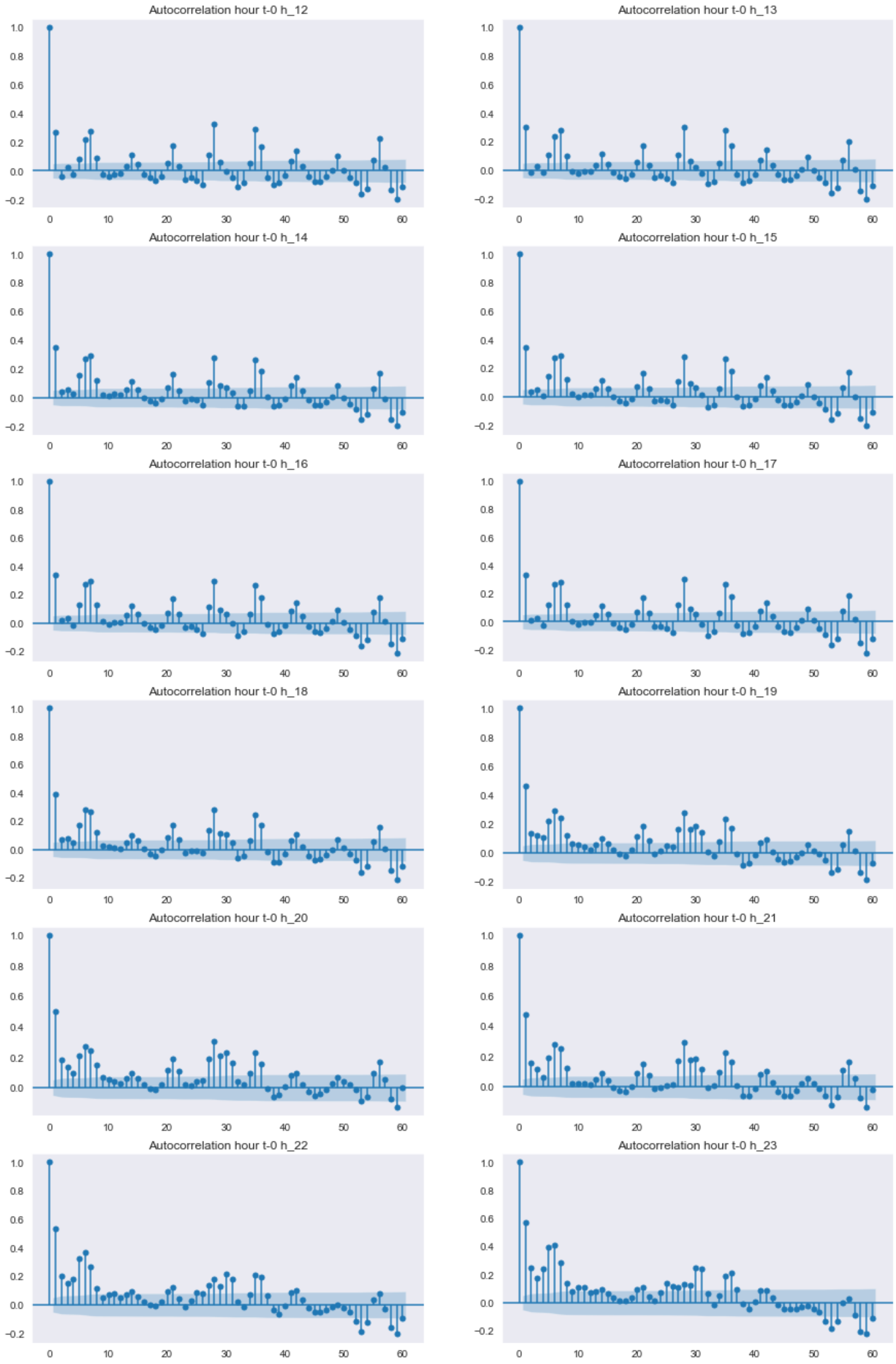

Autocorrelation hour t-0 h_7
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60


Autocorrelation hour t-0 h_8
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60


Autocorrelation hour t-0 h_9
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60


Autocorrelation hour t-0 h_10
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60


Autocorrelation hour t-0 h_11
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60



Autocorrelation hour t-0 h_12

Autocorrelation hour t-0 h_13

Autocorrelation hour t-0 h_14

Autocorrelation hour t-0 h_15

Autocorrelation hour t-0 h_16

Autocorrelation hour t-0 h_17

Autocorrelation hour t-0 h_18

Autocorrelation hour t-0 h_19

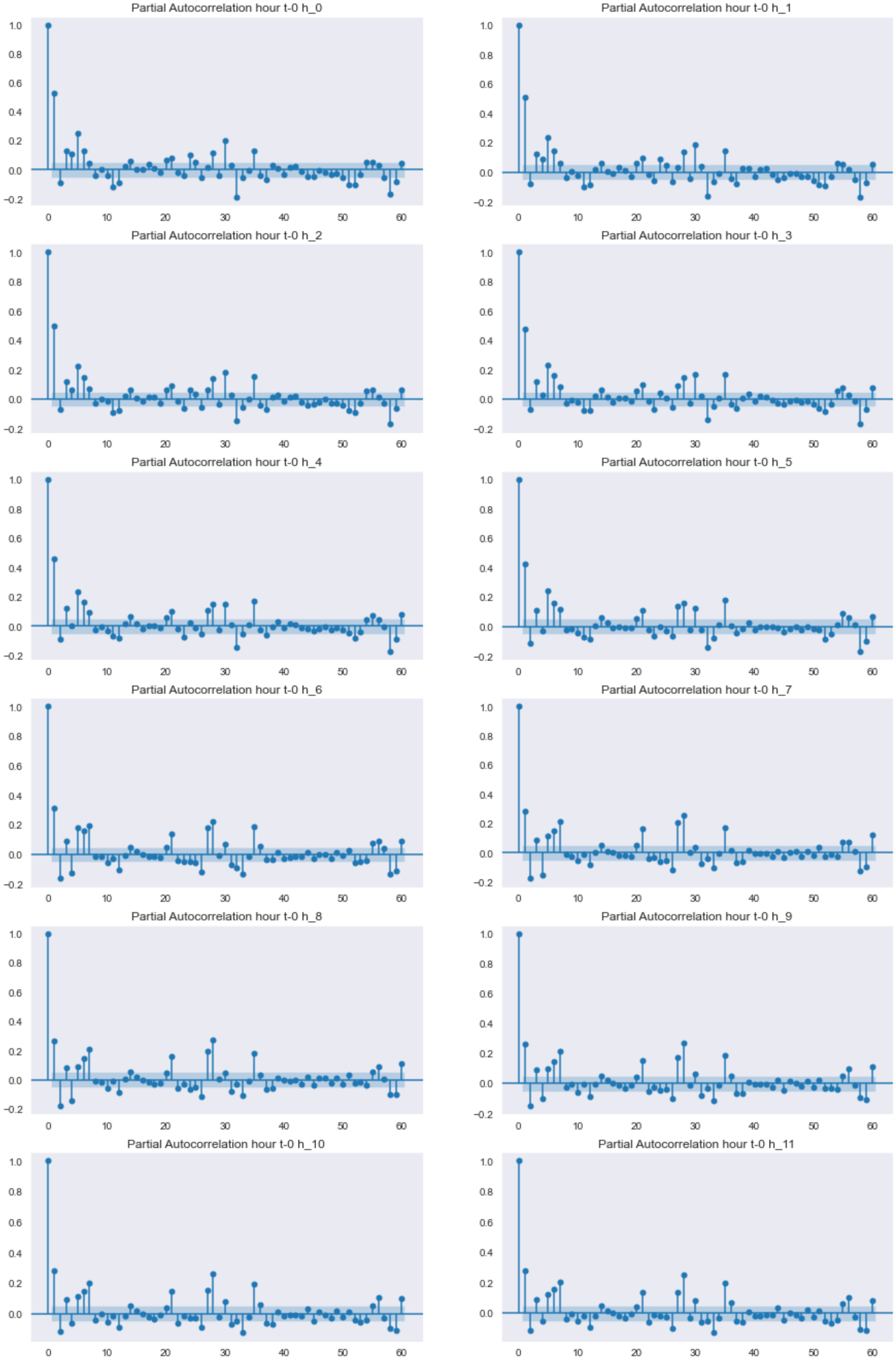
Autocorrelation hour t-0 h_20

Autocorrelation hour t-0 h_21

Autocorrelation hour t-0 h_22

Autocorrelation hour t-0 h_23

Figure 4- 23 Hourly autocorrelation plots


Partial Autocorrelation hour t-0 h_0
Partial Autocorrelation hour t-0 h_1
Partial Autocorrelation hour t-0 h_2
Partial Autocorrelation hour t-0 h_3
Partial Autocorrelation hour t-0 h_4
Partial Autocorrelation hour t-0 h_5
Partial Autocorrelation hour t-0 h_6
Partial Autocorrelation hour t-0 h_7
Partial Autocorrelation hour t-0 h_8
Partial Autocorrelation hour t-0 h_9
Partial Autocorrelation hour t-0 h_10
Partial Autocorrelation hour t-0 h_11
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
50
60

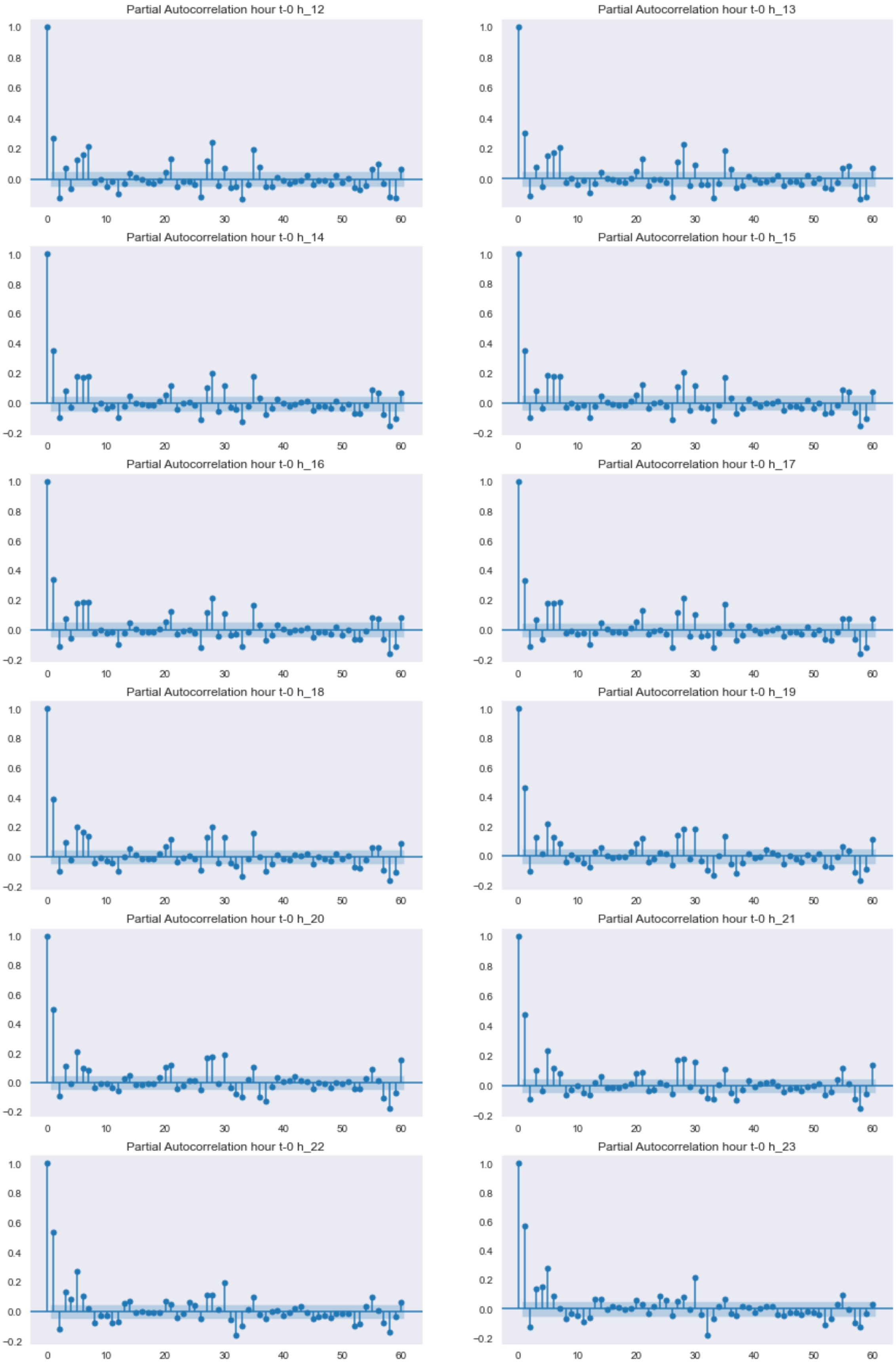


Figure 4- 24 Hourly (partial) autocorrelation plots

By looking (partial) autocorrelation plots, we can see that there is a strong correlation's indication to a moving average process, comparing with the consecutive hours plot; as well

- every 7 day, there is a cyclic autocorrelation for hours $2 - 21.$ Hence, a good feature would be multiples of 7 days up to $30 - 60$ days.
- the structure of the remain hours shows that the last $21 - 30$ days instead would be a better choice.

4.4.4.2 Normalization and Sample Creation

To build the LSTM model, we normalized the values and created paired windows of X as rows of the past data and Y as the electricity load of the target day.

To define a cross-validation testbench, the model is trained on a small amount data between 2015 to 2018.

4.4.4.3 Performance Evaluation

The MAE is calculated for each hour forecast to evaluate the model performance. We can calculate total model MAPE to compare model runs. For each hour, we take the mean of the sum of all errors. The results are shown in figure 4-25.

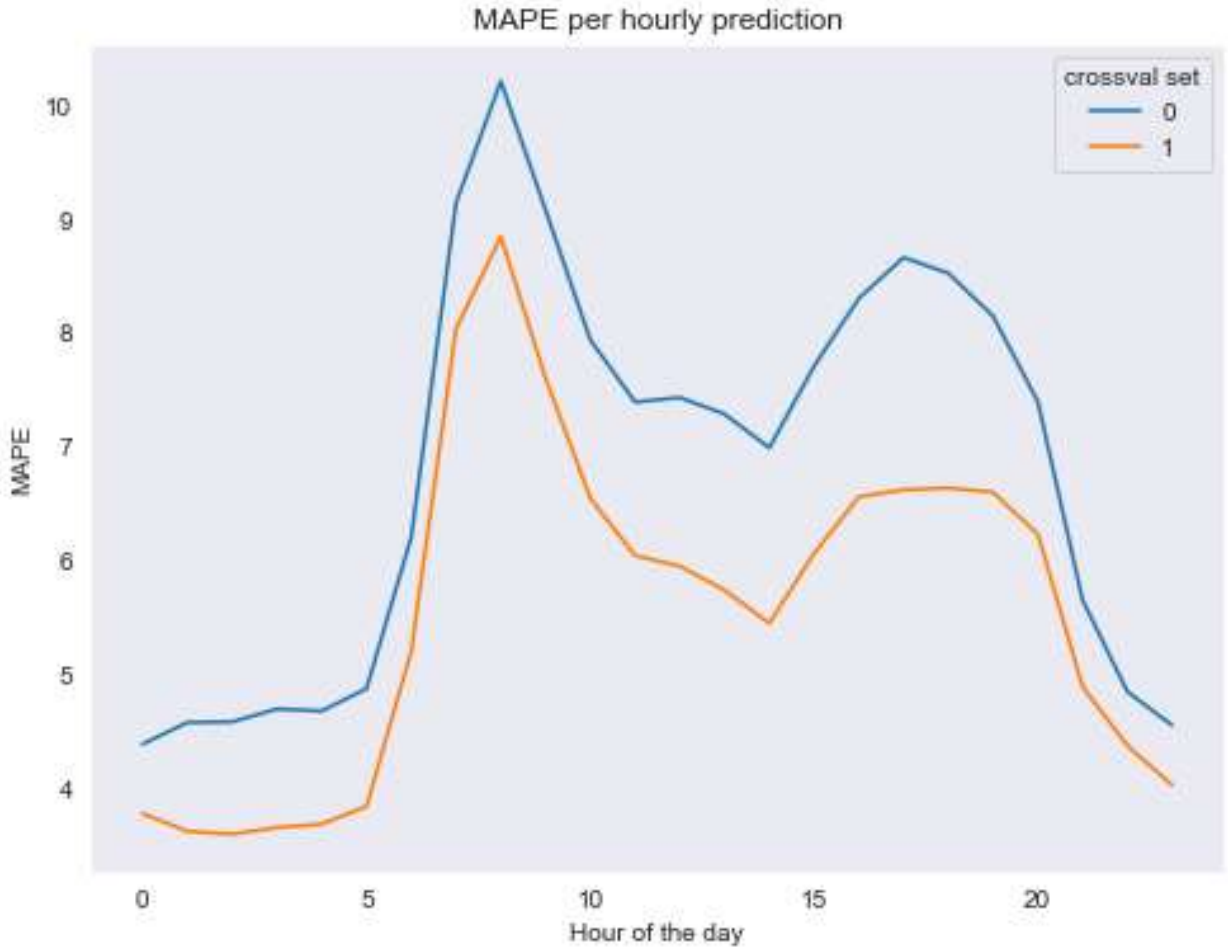


Figure 4- 25 MAPE results for per hourly forecasting by LSTM model

4.4.4.4 Predicting on Testing Dataset

For having an evaluation on the model performance with unknown data, we create a typical train and test sets by the preprocessing pipeline, the specific details of the train datasets were described below:

- Train: 2015-01-01 to 2017-12-31.
- Test: 2018-01-01 to 2018-03-31.
- Training set dimensions: X $(1036{,}26{,}24)$, and Y $(1036{,}24)$.

The dimension of the test set with training data needed for prediction was $(425,24)$, while the dimension of the testing set was X $(365,26,24)$, and Y $(365,24)$ respectively. We used the entire training set to learn the model. The model outcomes are shown in figure 4-26.

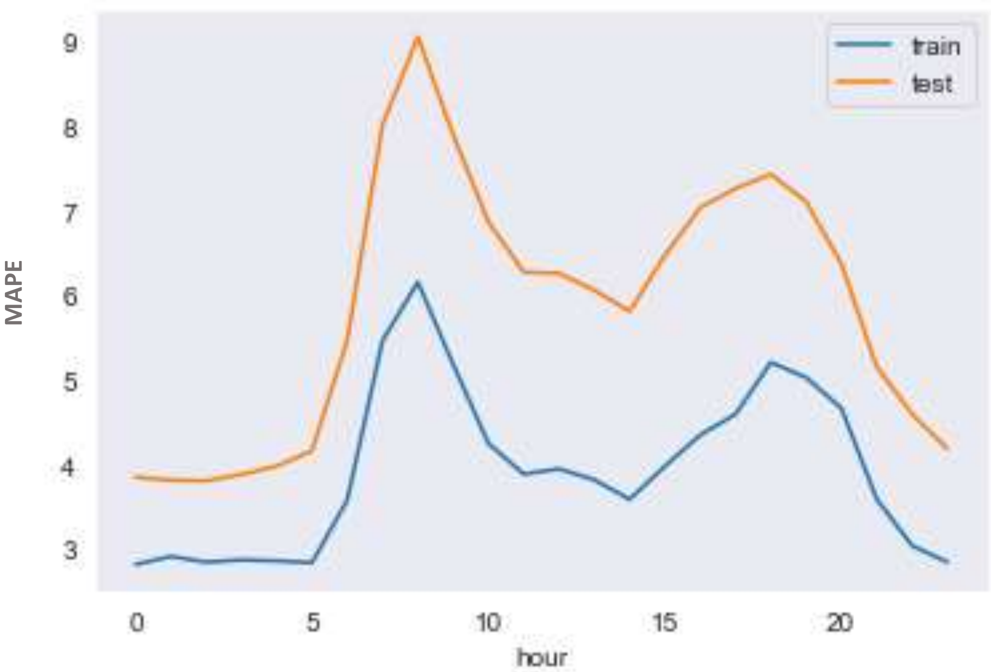


Figure 4- 26 Prediction outcomes of the LSTM model on an unseen test set Missing y axis legend

4.4.4.5 Plot Holdout Test Prediction

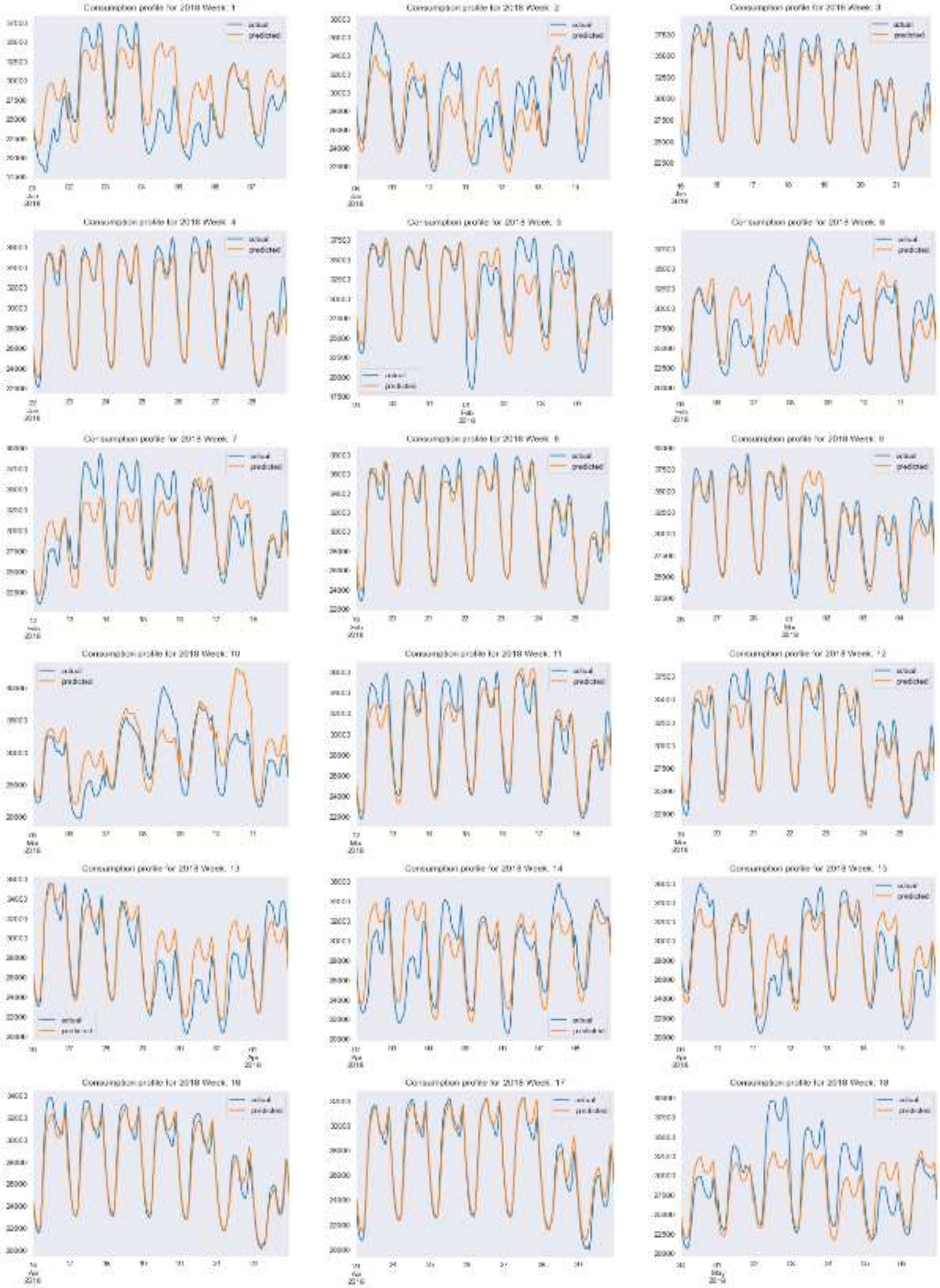


Figure 4- 27 Holdout test prediction

According to the above weekly plots, the model performance is very well on hours 12 to 23 and in some weeks. However, on select hours early in the day and the other weeks, we can observe a weak performance for the model. This means that the model did not perform well outside scope of univariate lag features.

Chapter 5- Conclusion

5.1 Conclusion

In this project, the hourly electricity load consumption is used to forecast future load electricity demands. As such, traditional techniques may not be able to forecast future values accurately. The hourly electricity load values between 01/01/2015 to 31/12/2018, are reported in Spain's energy dataset. In chapter 1, we summarized the importance of demand forecasting and related literature.

To explore the dataset's characteristics, we started with exploratory data analysis, providing descriptive information in the second section of chapter 4. In the data cleaning process, we replaced null values with mean values, extracted redundant attributes, and aggregated hourly load values daily level to see the trend and seasonality functions more clearly.

In section 4 of this chapter, four machine learning approaches are applied to the dataset with Python programming language.

To predict future data points by using past data, the Persistence models (naive models) which can be considered as simple techniques, were developed. The first three years (2015-2017) were used as the standard training set, leaving the final year (2018) as the testing set. The three persistence models, namely the previous-day hour-by-hour persistence, moving average (3 days) persistence, and same-day previous day persistence were used to resolve the problem of predicting the next day's 24-hour slices of expected energy demand.

Table 5- 1 The outcomes of the persistence model 1, 2, and 3 for the first 3 hours

	Prev_day_persistence	ma_persistence	Same_day_oya_persistence
H0	2858.969991	2326.951955	1874.826006
H1	3058.548695	2472.876323	1968.313603
H2	3246.803937	2599.912892	2193.094612

In the summary table 5-2, our persistence models' results are given for the first three hours. We can observe that, the previous-year model did a better job on average to predict the next day's power demand. This is the potential evidence of yearly seasonal patterns in the data. Both models have difficulty predicting the morning hours between 6 am and 10 am. The forecast gets progressively better through the day.

The second and third methods are called ARIMA which is a traditional time-series modeling technique, and SARIMAX which can be considered as a class of time series models that automatically deals with seasonality in data. We implemented a condensed version of the walk-forward validation set.

The performance of each model is presented in Table 5-2.

Table 5- 2 The results of the ARIMA & SARIMAX models

Model	Test Set	RMSE (Single Step)	Computation Time (min)
ARIMA	Condensed	4215.73	20:40
SARIMAX	Condensed	2642.88	01:21

While SARIMAX might not have been faster, the forecast was substantially better both in terms of the RMSE on the first walk forward validation step and the look of the forecast in the plot.

In the last method, we developed a simple univariate LSTM model with "Keras". In this case, to predict hour-by-hour using previous data from the same hour, each hour becomes a dataset on its own. We combined these multiple sets into one single block of data. The model performance is significantly well on some hours and some weeks.

This project shows us, electricity demand can be modeled using machine learning algorithms, and the models can be used to predict future electricity demand. Models that take into account trend and seasonality functions for electricity demand forecast would give better accuracy scores in future studies.

5.2 Applications

The models developed in this work have several impactful applications. They can be used by energy sector managers for the following applications:

5.2.1 Detect Abnormalities in Consumption Trends

The models can be used to calculate what the consumption values and patterns should be in the coming day. As the day starts, data can be compared to the model, and any significant deviations can be flagged as abnormal consumption.

5.2.2 Quantify Energy and Cost-Saving Measures

The models can be used to create forecasts of energy consumption under the current conditions. Once the energy measures take effect, the values under new conditions can be compared to that of the model, and the differences can be easily calculated.

5.2.3 Make Informed Decisions

The models can be used to forecast future values. If the consumption levels rise to a critical level in the forecast, then decisions can be made to recommission or renovate relevant parts of the power system. Improved forecasting is beneficial to the deployment of renewable energy, planning for high/low load days, and reducing wastage from polluting on reserve standby generation (typically inefficient gas or coal-fired powerplants).

5.3 Recommendations

Deciding the characteristics of a short-term energy consumption forecasting model in a real-world scenario will be heavily influenced by the requirements of the utility, where factors such as the risk appetite, supply requirements and finances will play an important role. It is without doubt a topic that will grow more important as the smart grid develops, where all parties involved should reap the environmental and economic benefits of progressing short-term electrical consumption forecasting.

Appendix

Python Coding:

https://colab.research.google.com/drive/1VYq_gL5MlwbSF4taIFD46YZKdwjY9TeF#scrollTo=dVgD_qRdFSf-

Bibliography

Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*, *147*, 77–89. https://doi.org/10.1016/j.enbuild.2017.04.038

Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, *8*(1), 69–80. https://doi.org/10.1016/0169-2070(92)90008-W

Amasyali, K., & El-Gohary, N. (2019). Predicting Energy Consumption of Office Buildings: A Hybrid Machine Learning-Based Approach. In *Advances in Informatics and Computing in Civil and Construction Engineering* (pp. 695–700). Springer International Publishing. https://doi.org/10.1007/978-3-030-00220-6_83

Burdick, A. (2012). Strategy guideline: Accurate heating and cooling load calculations. In *Guidelines for Improved Duct Design and HVAC Systems in the Home* (pp. 103–138). Nova Science Publishers, Inc.

Cai, M., Pipattanasomporn, M., & Rahman, S. (2019). Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. *Applied Energy*, *236*, 1078–1088. https://doi.org/10.1016/j.apenergy.2018.12.042

Chatfield, C., *Time-series Forecasting*. Chapman & Hall/CRC, 2001

Cools, M., Moons, E., & Wets, G. (2009). Investigating the Variability in Daily Traffic Counts through use of ARIMAX and SARIMAX Models: Assessing the Effect of Holidays on Two Site Locations. *Transportation Research Record*, *2136*(1), 57–66. Retrieved from https://doi.org/10.3141/2136-07

Citroen, N., Ouassaid, M., & Maaroufi, M. (2015). Long term electricity demand forecasting using autoregressive integrated moving average model: Case study of Morocco. In *Proceedings of 2015 International Conference on Electrical and Information Technologies, ICEIT 2015* (pp. 59–64). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/EITech.2015.7162950

Crawley, D. B., Hand, J. W., Kummert, M., & Griffith, B. T. (2008). Contrasting the capabilities of building energy performance simulation programs. *Building and Environment*, *43*(4), 661–673. https://doi.org/10.1016/j.buildenv.2006.10.027

Deng, H., Fannon, D., & Eckelman, M. J. (2018). Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata. *Energy and Buildings*, *163*, 34–43. https://doi.org/10.1016/j.enbuild.2017.12.031

Divina, F., Gilson, A., Goméz-Vela, F., Torres, M. G., & Torres, J. F. (2018). Stacking ensemble learning for short-term electricity consumption forecasting. *Energies*, *11*(4). https://doi.org/10.3390/en11040949

Divina, F., Torres, M. G., Vela, F. A. G., & Noguera, J. L. V. (2019). A comparative study of time series forecasting methods for short term electric energy consumption prediction in smart buildings. *Energies*, *12*(10). https://doi.org/10.3390/en12101934

De Felice, M., Alessandri, A., & Ruti, P. M. (2013). Electricity demand forecasting over Italy: Potential benefits using numerical weather prediction models. *Electric Power Systems Research*, *104*, 71–79. https://doi.org/10.1016/j.epsr.2013.06.004

Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*. https://doi.org/10.1016/j.rser.2013.03.004

Friedrich, L., & Afshari, A. (2015). Short-term Forecasting of the Abu Dhabi Electricity Load Using Multiple Weather Variables. In *Energy Procedia* (Vol. 75, pp. 3014–3026). Elsevier Ltd. https://doi.org/10.1016/j.egypro.2015.07.616

Friedrich, L., Armstrong, P., & Afshari, A. (2014). Mid-term forecasting of urban electricity load to isolate air-conditioning impact. *Energy and Buildings*, *80*, 72–80. https://doi.org/10.1016/j.enbuild.2014.05.011

Haida, T., & Muto, S. (1994). Regression based peak load forecasting using a transformation technique. *IEEE Transactions on Power Systems*, *9*(4), 1788–1794. https://doi.org/10.1109/59.331433

Hahn, H., Meyer-Nieberg, S., & Pickl, S. (2009). Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, *199*(3), 902–907. https://doi.org/10.1016/j.ejor.2009.01.062

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hor, C. L., Watson, S. J., & Majithia, S. (2006). Daily load forecasting and maximum demand estimation using ARIMA and GARCH. In *2006 9th International Conference on Probabilistic Methods Applied to Power Systems, PMAPS*. https://doi.org/10.1109/PMAPS.2006.360237

Jota, P. R. S., Silva, V. R. B., & Jota, F. G. (2011). Building load management using cluster and statistical analyses. *International Journal of Electrical Power and Energy Systems*, *33*(8), 1498–1505. https://doi.org/10.1016/j.ijepes.2011.06.034

Kandil, M. S., El-Debeiky, S. M., & Hasanien, N. E. (2002). Long-term load forecasting for fast developing utility using a knowledge-based expert system. *IEEE Transactions on Power Systems*, *17*(2), 491–496. https://doi.org/10.1109/TPWRS.2002.1007923

Kyriakides, E., & Polycarpou, M. (2006). Short term electric load forecasting: A tutorial. *Studies in Computational Intelligence*, *35*, 391–418. https://doi.org/10.1007/978-3-540-36122-0_16

Leduc, M., Damon Matthews, H., & De Elía, R. (2016). Regional estimates of the transient climate response to cumulative CO 2 emissions. *Nature Climate Change*, *6*(5), 474–478. https://doi.org/10.1038/nclimate2913

Li, X., Bowers, C. P., & Schnier, T. (2010). Classification of energy consumption in buildings with outlier detection. *IEEE Transactions on Industrial Electronics*, *57*(11), 3639–3644. https://doi.org/10.1109/TIE.2009.2027926

Li, X., & Wen, J. (2014). Review of building energy modeling for control and operation. *Renewable and Sustainable Energy Reviews*. Elsevier Ltd. https://doi.org/10.1016/j.rser.2014.05.056

Lusis, P., Khalilpour, K. R., Andrew, L., & Liebman, A. (2017). Short-term residential load forecasting: Impact of calendar effects and forecast granularity. *Applied Energy*, *205*, 654–669. https://doi.org/10.1016/j.apenergy.2017.07.114

Market operations in electric power systems: forecasting, scheduling, and risk management. (2002). *Choice Reviews Online*, *40*(03), 40-1574-40–1574. https://doi.org/10.5860/choice.40-1574

Notton, G. and Voyant, C. (2018). Forecasting of Intermit-tent Solar Energy Resource. In Advances in Renewable Energies and Power Technologies, 77-114.

Ortiz, M., Ukar, O., Azevedo, F., & Múgica, A. (2016). Price forecasting and validation in the Spanish electricity market using forecasts as input data. *International Journal of Electrical Power and Energy Systems*, *77*, 123–127. https://doi.org/10.1016/j.ijepes.2015.11.004

Papaioannou, G., Dikaiakos, C., Dramountanis, A., & Papaioannou, P. (2016). Analysis and Modeling for Short- to Medium-Term Load Forecasting Using a Hybrid Manifold Learning Principal Component Model and Comparison with Classical Statistical Models (SARIMAX, Exponential Smoothing) and Artificial Intelligence Models (ANN, SVM): The Case of Greek Electricity Market. *Energies*, *9*(8), 635. https://doi.org/10.3390/en9080635

Price, P. (2010). Methods for Analyzing Electric Load Shape and its Variability. *California Energy Commission*, (May), 1–63.

Rodrigues, F., Cardeira, C., & Calado, J. M. F. (2014). The daily and hourly energy consumption and load forecasting using artificial neural network method: A case study using a set of 93 households in Portugal. In *Energy Procedia* (Vol. 62, pp. 220–229). Elsevier Ltd.https://doi.org/10.1016/j.egypro.2014.12.383

Seyedzadeh, S., Rahimian, F. P., Glesk, I., & Roper, M. (2018, December 1). Machine learning for estimation of building energy consumption and performance: a review. *Visualization in Engineering*. Springer. https://doi.org/10.1186/s40327-018-0064-7

Taşpinar, F., Çelebi, N., & Tutkun, N. (2013). Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods. *Energy and Buildings*, *56*, 23–31. https://doi.org/10.1016/j.enbuild.2012.10.023

US Energy Information Administration, “US Energy Flow.” [Online]. Available: https://www.eia.gov/totalenergy/data/monthly/pdf/flow/total_energy.pdf. [Accessed: 21-May-2020].

Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., & Ahrentzen, S. (2018). Random Forest based hourly building energy prediction. *Energy and Buildings*, *171*, 11–25. https://doi.org/10.1016/j.enbuild.2018.04.008

Zhang, F., Deb, C., Lee, S. E., Yang, J., & Shah, K. W. (2016). Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique. *Energy and Buildings*, *126*, 94–103. https://doi.org/10.1016/j.enbuild.2016.05.028

Zhao, H. X., & Magoulès, F. (2012, August). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*. https://doi.org/10.1016/j.rser.2012.02.049

Zhang, P. G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0