LEARNING PEAKS FOR COMMERCIAL & INDUSTRIAL ELECTRIC LOADS

by

B Hari Kiran Reddy

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

School of Electrical and Computer Engineering

West Lafayette, Indiana

December 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Junjie Qin, Chair

School of Electrical and Computer Engineering

Dr. Steven D. Pekarek

School of Electrical and Computer Engineering

Dr. Shreyas Sundaram

School of Electrical and Computer Engineering

Approved by:

Dr. Dimitrios Peroulis

ACKNOWLEDGMENTS

I am incredibly grateful to my advisor, Dr. Junjie Qin, for his invaluable suggestions, continuous support, and patience during my master's study. His immense knowledge and great experience have encouraged me all the time in completing this study. Moreover, whenever I was stuck, he offered me the liberty to discuss my thoughts openly and guided me to the best possible direction. I would also like to thank Dr. Steven D. Pekarek and Dr. Shreyas Sundaram for being a part of my master's advisory committee.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

As on 2017, US Energy Information Administration (US EIA) claims that 50 % of the total US energy consumption are contributed by Commercial and Industrial (C&I) end-users. Most of the energy consumption by these users are in the form of the electric power. Electric utilities, who usually supply the electric power, tend to care about the power consumption profiles of these users mainly because of the scale of consumption and their significant contribution towards the system peak. Predicting and managing the peaks of C&I users is crucial both for the users themselves and for utility companies.

In this research, we aim to understand and predict the daily peaks of individual C&I users. To empirically understand the statistical characteristics of the peaks, we perform an extensive exploratory data analysis using a real power consumption time series dataset. To accurately predict the peaks, we investigate *indirect* and *direct* learning approaches. In the indirect approach, daily peaks are identified after forecasting the entire time series for the day whereas in the direct approach, the daily peaks are directly predicted based on the historical data available for different users during different days of the week. The machine learning models used in this research are based on Simple Linear Regression (SLR), Multiple Linear Regression (MLR), and Artificial Neural Networks (ANN).

1. INTRODUCTION

'Electric Power' is one of the top leagues' inventions of all time. The ability to physically transmit it to larger distances makes it exceptional compared to other forms of energy. The three stages of delivering the electric power are: Generation, Transmission, and Distribution. Conventionally, it has been generated at massive power plants and moves through a complex system called 'grid' of the electric substations, transformers, and power lines supplying consumers the required power at any given time. The challenging aspect about the electric grid lies in maintaining its stability as most of the associated parameters are changing almost instantaneously. The resilience of the grid is confronted mostly during the times of peak power (or energy) consumption where the scope for any mismanagement is very little. Furthermore, with the large scale deployment of Distributed Energy Resources (DERs) such as roof-top solar, wind, batteries, EV chargers in the recent times and their integration with the existing electric grid, the potential of situations like grid failure increases by numerous folds, especially during the peak times.

1.1 Motivation

This research focuses on learning the peak power consumption values and their corresponding times which are essential to manage for both electric utility companies and end-users (mainly the commercial and industrial users). This section discusses the importance of the peak values from both sides of the coin.

1.1.1 Electric Utility Companies's Point of View

The primary objective of the electric utility companies is to maintain a proper equilibrium between the generation and the consumption of electric power at all times. Behind-the-meter distributed energy resources (DERs), transportation electrification, and the increasingly frequent extreme climate events present novel challenges to these companies as these possess the capability of producing high fluctuations in terms of the power demand in the power system. The smart technologies such as energy management systems (EMSs) and smart meters associated with DERs and transportation electrification infrastructure enable them to respond to time-varying prices which are usually set by these utility companies. Smart households or industrial consumers tend to store energy in large storage units during the low price rate duration and later consume the same during high price rates. This leads to a situation where the peaks are created artificially at irregular times. Furthermore, DERs adoption and the growing penetration of fast EV charging increasingly strains local distribution network, makes it important to not only predict the aggregate peaks but also peaks of smaller communities (e.g., behind the same local transformer). Therefore, studying the peaks of the influential consumers becomes more important than just focusing on the aggregate peaks of all connected consumers together which are more predictable and continuous. Moreover, having an good estimation of these peak power consumption values supports in planning the distribution investments efficiently with keeping an eye on the total system capacity along with network constraints.

1.1.2 Commercial & Industrial Users's Point of View

Today, most of the required infrastructure in the power distribution industry is owned by electric utility companies. These companies are responsible for the lines, poles, meters, power outages, repairs, and other issues with how the electricity is delivered to all types of consumers: residential, commercial, industrial, and agricultural. These companies read consumer energy meters and bills on a monthly basis. Different utility companies charge at different rates based on their current liabilities, fixed and variable costs, insurance, and rent expenses. These vary based on the region, season, type of consumer, and, more importantly, on the availability of electric power in the electricity markets. Normally, these companies charge a residential or an agricultural user based on the total energy consumption at a volumetric charge (usually defined in $ per kilowatt-hour) monthly. A typical electricity bill of a residential user is shown in the figure 1.1 for reference. However, the same companies charge a commercial or an industrial user with an additional charge called 'Demand Charge' (usually defined in $ per kilowatts). These demand charges accounts for the maximum power consumption value by the individual on a certain month. These charges contribute significantly towards the total monthly payment for any commercial or industrial user which they try to reduce to the maximum extent possible. A typical electricity bill for the commercial user is shown in figure 1.2.

Bill Detail Sample

Electric Service **Billing Date:** Jan 4, 2016 **Days of Service:** 28 **Read Dates:** Nov 20, 2015 - Dec 18, 2015

Meter	Current Read	Last Read	Usage
Electric kWh	37570	37570	0
Electric kWh	44084	43577	507

1 Cost of Basic Service		$13.50
2 Residential Charge	507 KWH @ $0.103	52.22
3 Power Cost Adjustment		6.45
Coon Rapids Excise Tax		2.89
MN State Tax		5.16
Transit Tax		0.19
Total Cost for Electric Service (Actual Charges)		**$80.41**

Figure 1.1. Electricity bill of a residential user

Bill Detail Sample

1 **Electric Service** **Billing Date:** Jun 9, 2015 2 **Days of Service:** 32 **Read Dates:** Apr 24, 2015 - May 26, 2015

Meter	Current Read	Last Read	Mult	Usage
Electric kWh	4630	4465	120	19800
Demand kW	0.67	0.52	120	80.40

3 Cost of Basic Service		$30.00
4 Energy Charge	19800.00KWH @ 0.05940	1,176.12
5 Power Cost Adjustment		202.87
6 Demand Charge	80.40KW @ 7.40000	594.96
MN State Tax		137.77
Transit Tax		5.01
Total Cost for Electric Service (Actual Charges)		**$2,146.73**

Figure 1.2. Electricity bill of a commercial user

In the figure shown above, clearly demand charge for this C&I user contributes around 30% to the total cost which is fairly significant. Just as a side note, a electric utility company not only charges the consumer for their power consumption but also inculcates other service charges like maintenance, connections, and adjustments in conjunction with the local government regulations. The collection of electricity rates and additional service charges is termed as a tariff.

1.2 Literature Review

Historically, forecasting of the load has been broadly classified into four categories based on the span of the forecasting: (a) very-short term (up to next 15-30 minutes), (b) short term (up to next 24 hours), (c) medium term (up to one year ahead), and (d) long term (up to 5 to 10 years ahead). Load forecasting is typically based on a mathematical model which involves different predictors and their combinations. To build this model, different methods have been explored in the past. Regression methods, time series regression methods, support vector regression, fuzzy logic and neural network are just to name a few. This section discusses about these methods briefly to provide a high level understanding on each one of them.

Regression method is one of the widely used statistical techniques. These methods are used to model the power consumption profile depending on the features like weather conditions, day of week and customer classes [1][2][3]. This method is built on the assumption that the power consumption can be regressed based on the linear relationship of the features using the least squares method which minimizes the sum of the residuals to find the best fit for the dataset.

Another form of the regression methods usually used in the areas of predicting the weather data, heat rate monitoring, stock prices, and electric load forecasting is time series regression methods. These are used in a specific way of analyzing a sequence of data points to capture auto-correlation, trend, or seasonal variation rather than just recording the data points intermittently or randomly. Some of the time series models include Auto Regressive Integrated Moving Average (ARIMA), exponential smoothing, state space models and spectral models. [4] discusses the principles and the practices that is typically followed for time series analysis across domains. An example on forecasting the short term load based on the time series analysis is discussed in [5].

One of the most common machine learning methods used in the recent times is Support Vector Regression (SVR) techniques. Their capability of not explicitly forming the feature vectors rather forming them implicitly using kernels saves a lot of computational time and provides the scope to include large number of relevant features that are useful to explain the output variable. The application of Support Vector Regression method in the power system domain for short term load forecasting is discussed in [6] whereas in [7], the same method is applied for forecasting the electricity peaks.

Fuzzy logic is a technique of developing the qualitative inputs and output by assigning numerical values to them. The main two types of this technique are many-valued logic or Boolean logic. An input or output under many-valued logic takes a real number between 0 and 1 whereas under Boolean logic, it takes on a truth value of "0" or "1". For example, a temperature being 'hot', 'pleasant', or 'cold' can be assigned 0, 0.5, and 1 under many-valued logic. After the logical processing of fuzzy inputs, qualitative outputs can be obtained by 'defuzzification'. [8] discusses the short term load forecasting factoring weather parameters using fuzzy logic.

Neural Networks like Artificial Neural Networks (ANN) are inspired by the biological neural networks, have an extensive application because of their capability of evolution, self-organizing, and adaption. The network comprises of the inputs, hidden layers, and output/s. In each hidden layer, there are many neurons. These neurons are connected with the neurons of the adjacent hidden layers and are associated with a weight and a threshold. [9] discusses the short term forecasting using artificial neural network. [10] discusses the application Recurrent Neural Network, which captures the sequential information present in the input data, for short term load forecasting.

There are other methods which combines the usage of different methods into one. For example, [11] discusses the approach where ANN is used as the first step to predict the load and then SVM forecasting model is created on the basis of data points whose load type is the same as the predict point. Another example combining two different methods is discussed in [12]. In this paper, ANN structure with one input provided by a fuzzy weather controller. The use of fuzzy logic will enhance the performance of the system as well as make it more transparent and adaptable.

In some cases, optimization algorithms can be applied on methods like SVM and ANN to further improve the forecasting on the parameters obtained after training the model. In [13], particle swarm optimization is chosen as the optimization tool that is applied on the weight matrix of ANN to improve results whereas genetic algorithm is used to optimize the parameters for SVM model in [14].

1.3 Thesis Outline

The remainder of this thesis constitutes of three chapters namely Exploratory Data Analysis, Machine Learning for Time Series Data, and Machine Learning for Peaks Data.

In chapter 2, basic insights about the dataset used for this research are discussed. Precisely, the variation of power consumption profiles for different users has been studied comprehensively for the entire time series. Also, elementary analysis on the daily peak values and their corresponding times for different users has been discussed. By the end of this chapter, the necessity of applying advanced machine learning algorithms on these power consumption profiles will be realised.

Further in chapter 3, machine learning techniques, some of which are discussed in section 1.2, are performed on different users to forecast their power consumption profiles in a robust manner for the entire 24-hour period. Specifically, first order auto regressive models and artificial neural networks have been used as the mathematical models for forecasting the load and more importantly, for forecasting the daily peak values and their corresponding times.

Finally in chapter 4, machine learning models like simple linear regression, multiple linear regression and artificial neural network have been applied on the peaks data to forecast the daily peaks directly without predicting the entire time series for different users.

2. EXPLORATORY DATA ANALYSIS

In this chapter, primitive observations of the dataset used for this research have been discussed using basic statistical concepts such as mean, median, and mode. These observations provide a rough idea about the users's power consumption profiles along with the distribution of their peak values over different periods. To start with, this chapter first describes the dataset which is followed by the exploratory analysis of the power consumption values for the entire time series and then for the daily peak values and their corresponding times.

2.1 Data Description

A dataset containing Irish smart meter data of 6000 electricity users from August 2009 to December 2010 is used for our study [15]. The metering interval (Δ) for this dataset is 30 mins which makes the number of metering intervals equal to 48 for 24 hours. Precisely, 12:00 am to 12:29 am corresponds to the 1st metering interval, 12:30 am to 12:59 am to the 2nd and so on. All 6000 users have been ranked according to their peak energy consumption for the entire 17 months and a subset of 75 users of the highest peaks is obtained specifically for this study which are more likely to be Commercial and Industrial (C&I) users.

2.2 Exploratory Data Analysis for Power Consumption Time Series

As the power consumption values of all the users were recorded in a continuous manner, the dataset is classified based on different metering intervals, days, weeks, and months. This section broadly shows the variation of these power consumption values across different (a) Time-of-Day, and (b) Day-of-Week for these 17 months. Also, a clustering analysis has been performed where the users are grouped together on different days of the week depending on their power consumption profiles. This analysis helps in understanding the correlation among the users on different days.

2.2.1 Time-of-Day Variations

A scatter plot and the box plot shown in figure 2.1 and 2.2 represents the power consumption values of a C&I user (user '1') against time for all 518 days present between August 2009 to December 2010. It is evident from the interquartile (IQR) range (25^{th} - 75^{th} percentile) of each box corresponding to the metering interval that the power consumption values for this user are low and possess small variances during metering intervals 1-11 and 43-48, i.e., from 12:00 am to 05:29 am and from 09:00 pm to 11:59 pm. Moreover, very few outliers (marked as '+') are present during these metering intervals. However, the scenario completely changes for the metering intervals 12-42, i.e., from 05:30 am to 08:59 pm. The power consumption values during these metering intervals are higher, possess large variances and have a significant number of outliers. It is trivial to understand the fact that narrow variations are preferred to predict the power consumption value at certain metering interval.

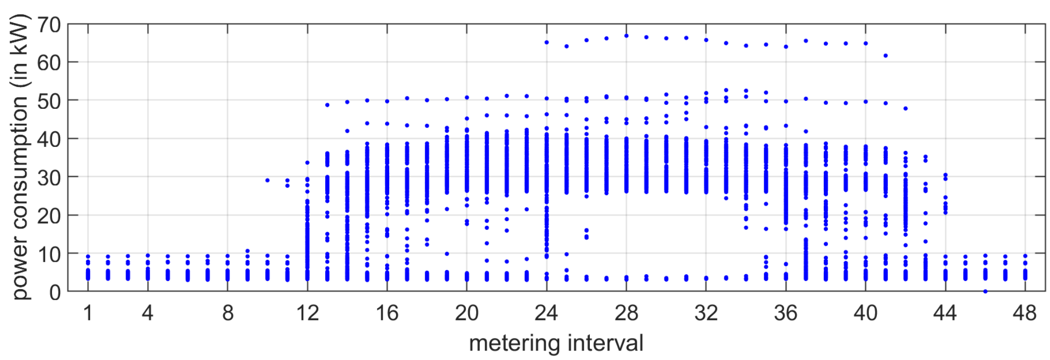


Figure 2.1. Scatter plot of user 1's power consumption profile for all 518 days

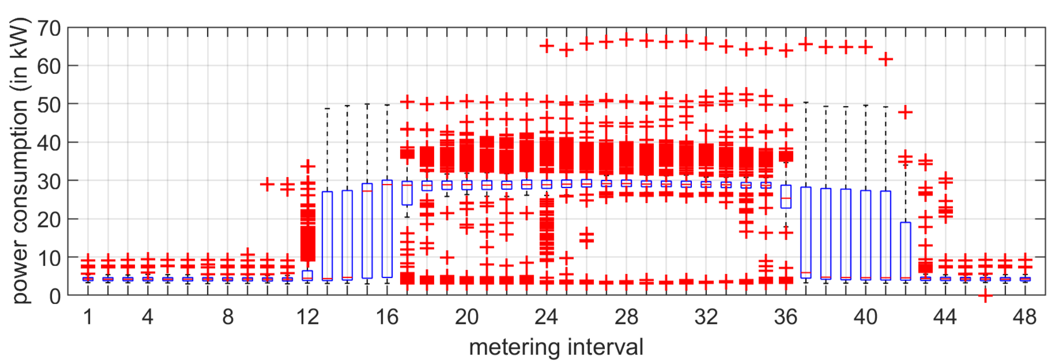
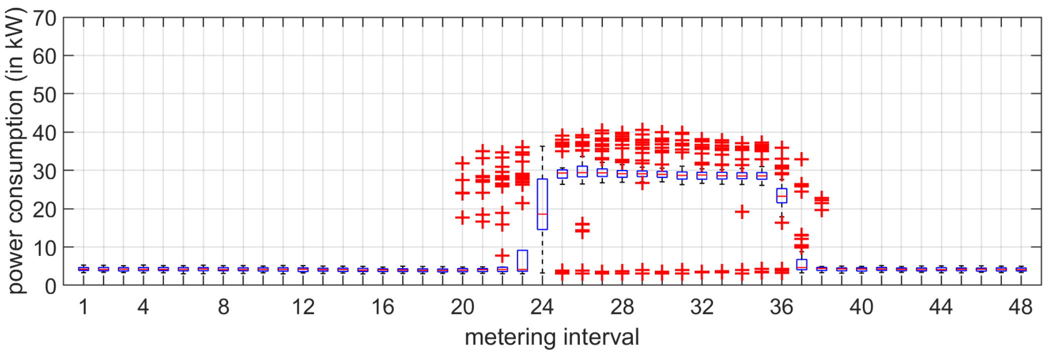


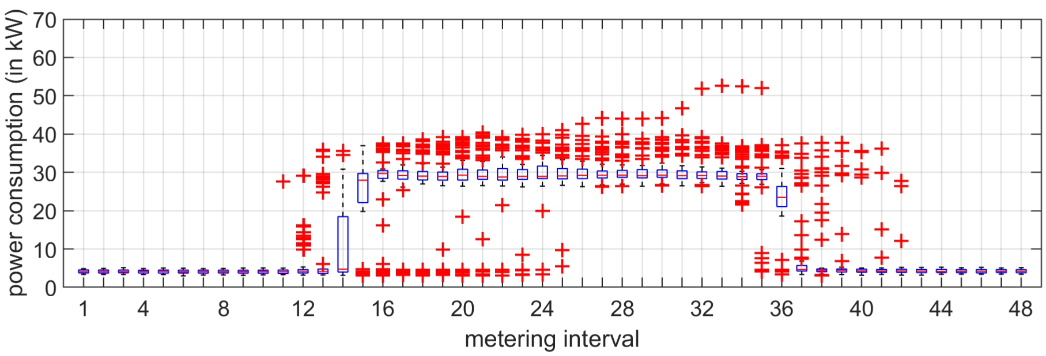
Figure 2.2. Box plot of user 1's power consumption profile for all 518 days

2.2.2 Day-of-Week Variations

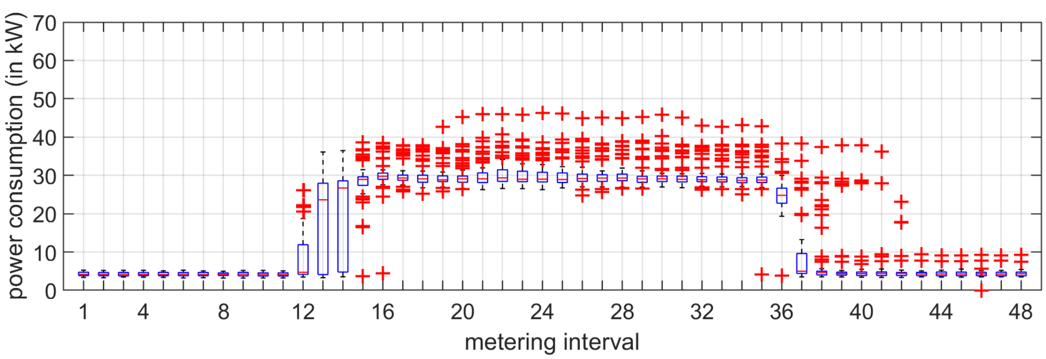
To overcome large variances across different metering intervals, the dataset has been further divided based on the days of the week. Precisely, day 1 corresponds to 'Monday', day 2 corresponds to 'Tuesday', and so on. The total number of weeks present during August 2009 to December 2010 is 75. Figure 2.3 shown below displays the box plots of the same user (user '1') for all seven days of the week.



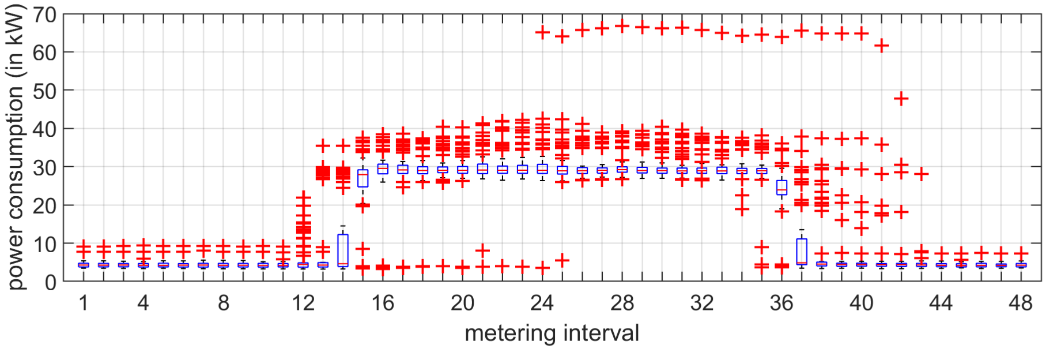
(a) Mondays



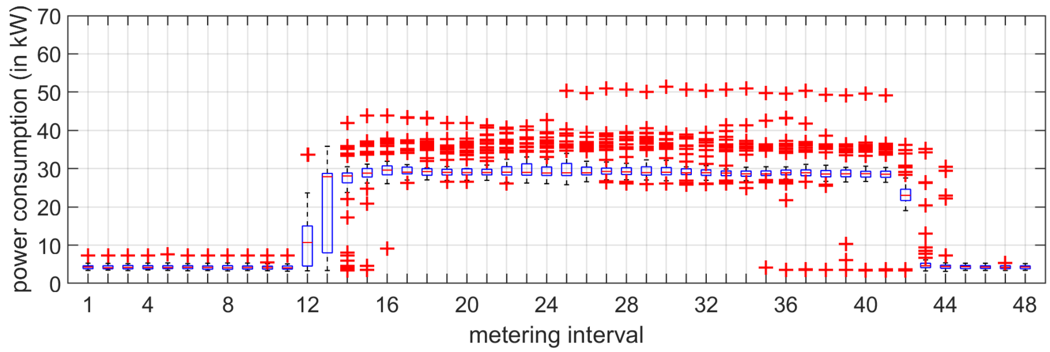
(b) Tuesdays



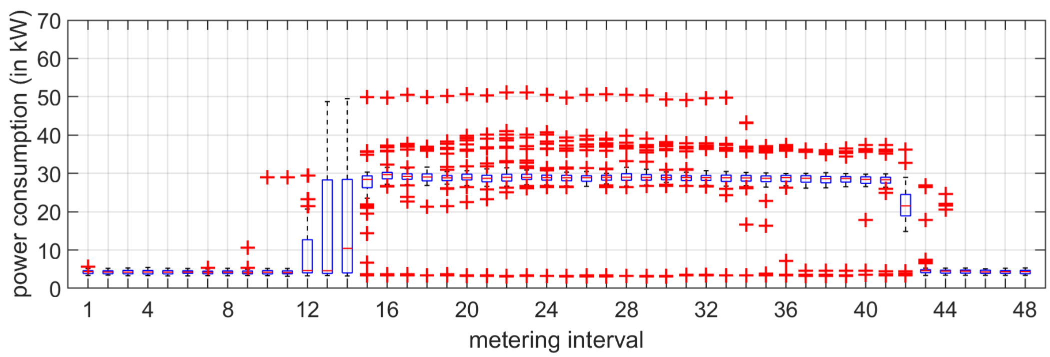
(c) Wednesdays



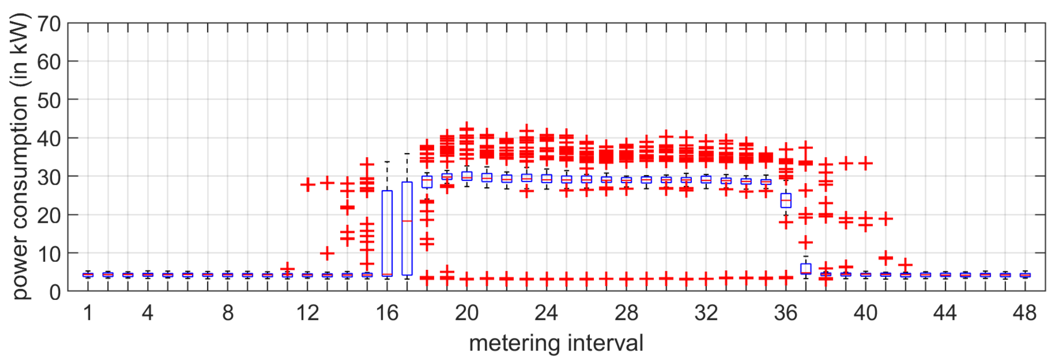
(d) Thursdays



(e) Fridays



(f) Saturdays



(g) Sundays

Figure 2.3. Power consumption profile of user '1'

From figure 2.3, it can be clearly observed that the number of metering intervals with the broader variation has been decreased as the dataset is further divided into different days of the week. The reason is very straightforward. The power consumption by a specific user during any day of the week does not necessarily have to be the same as on the other day of the week. For example, this user on Mondays consumes very little power when compared to Tuesdays at the 18^{th} metering interval.

Secondly, an important observation for this user is the trend of the power consumption values when plotted against time for all 518 days is similar to that of the power consumption values when plotted against time for different days of the week which is not necessarily true for all other users. For example, as shown in figures 2.4 and 2.5, user '2' has an entirely different trend of the power consumption values were plotted against time for all 518 days to that of the power consumption values when plotted only for Mondays.

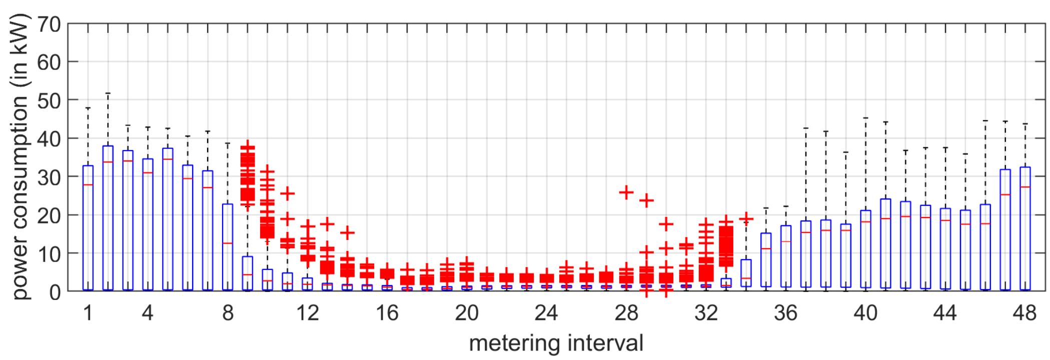


Figure 2.4. Power consumption profile of user '2' for all days

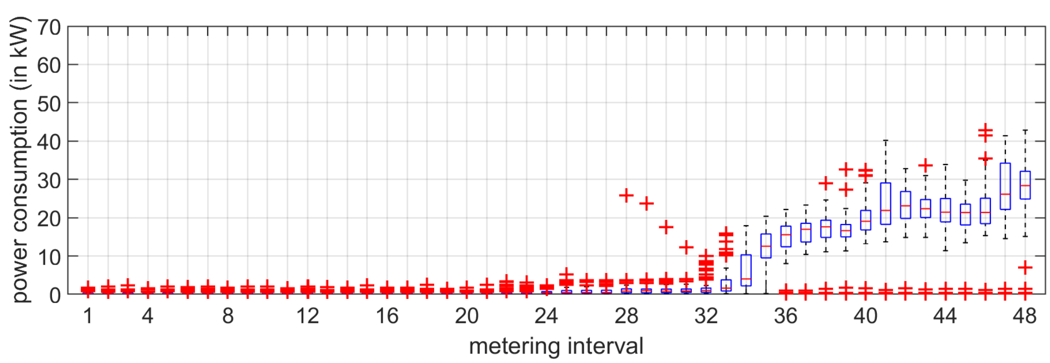
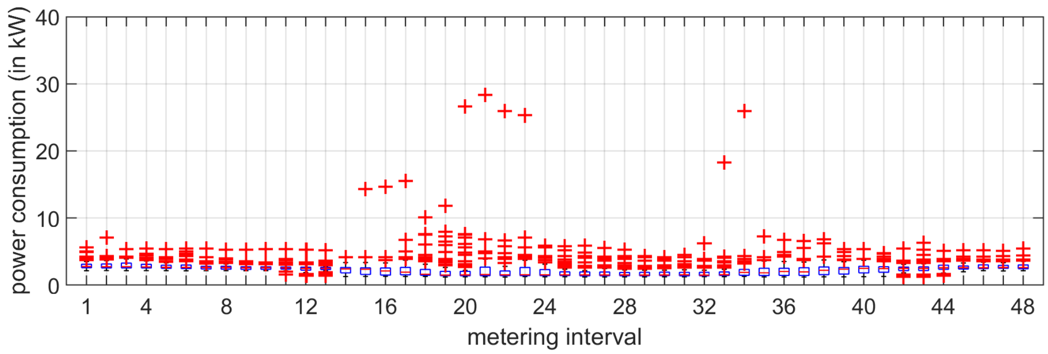
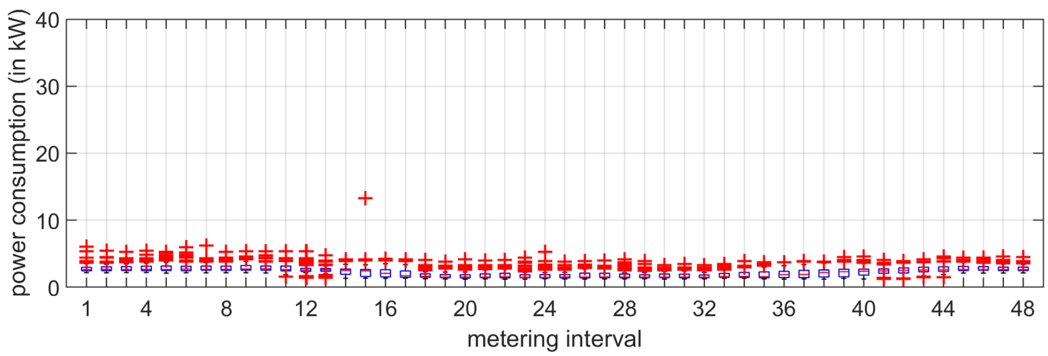


Figure 2.5. Power consumption profile of user '2' on Mondays

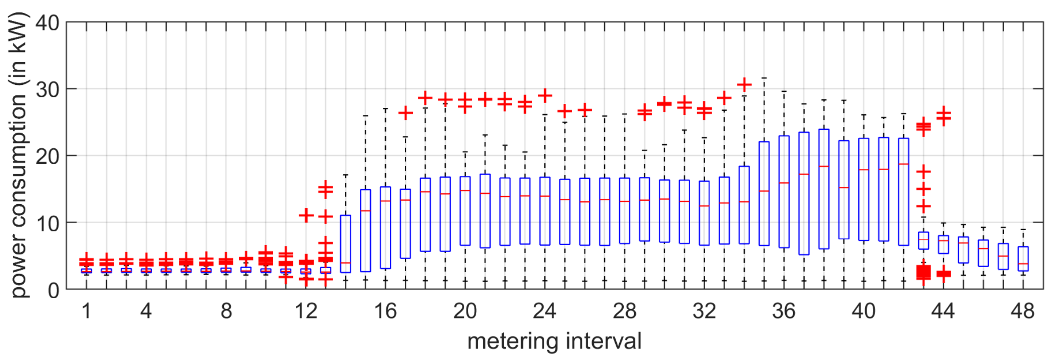
Finally, dividing the data into different days of the week also helps in identifying the pattern of power consumption values among different days of the week. For example, user '25' consumes small amount of power on Sundays and Mondays, whereas a similar power consumption pattern can be observed on all other days of the week.



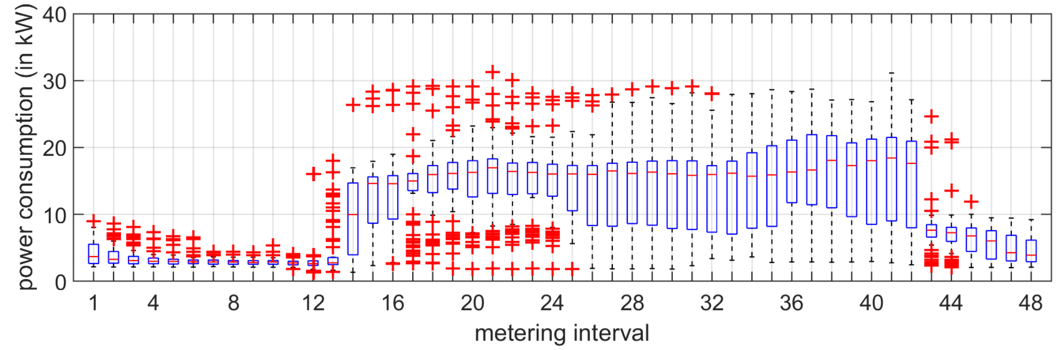
(a) Sundays



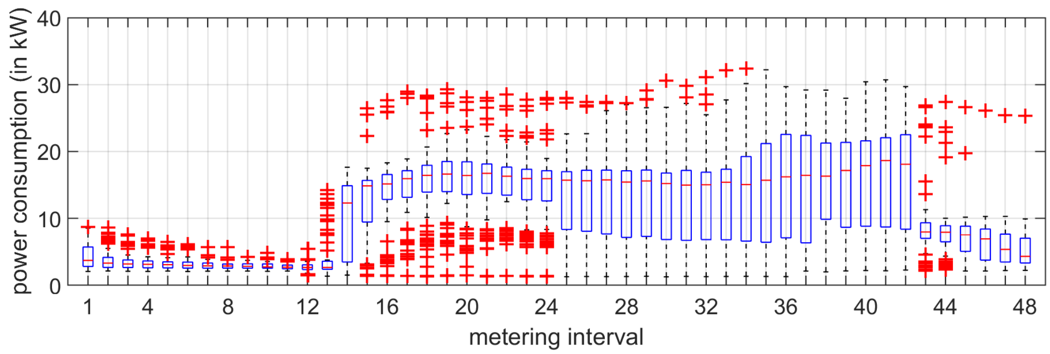
(b) Mondays



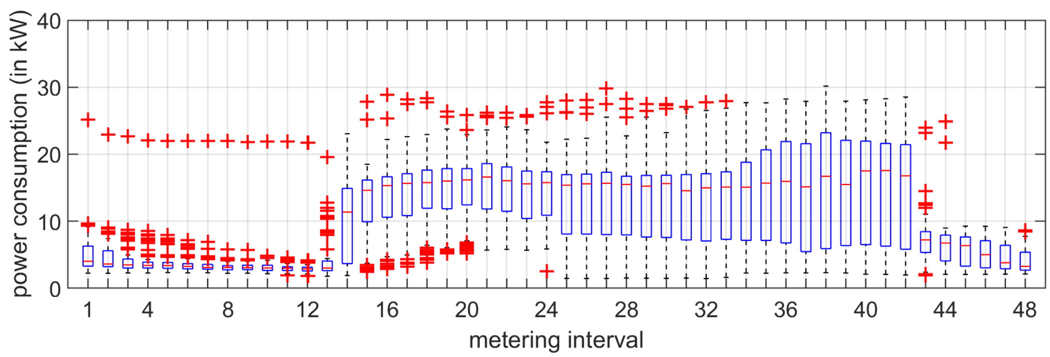
(c) Tuesdays



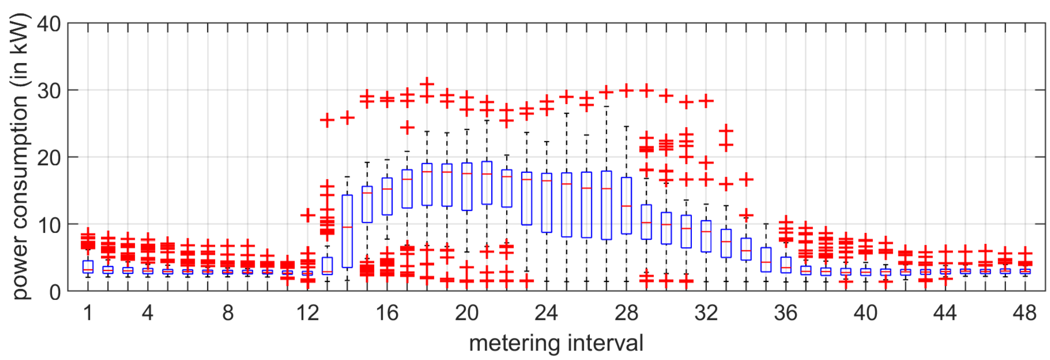
(d) Wednesdays



(e) Thursdays



(f) Fridays



(g) Saturdays

Figure 2.6. Power consumption profile of user '25'

2.2.3 Clustering Analysis

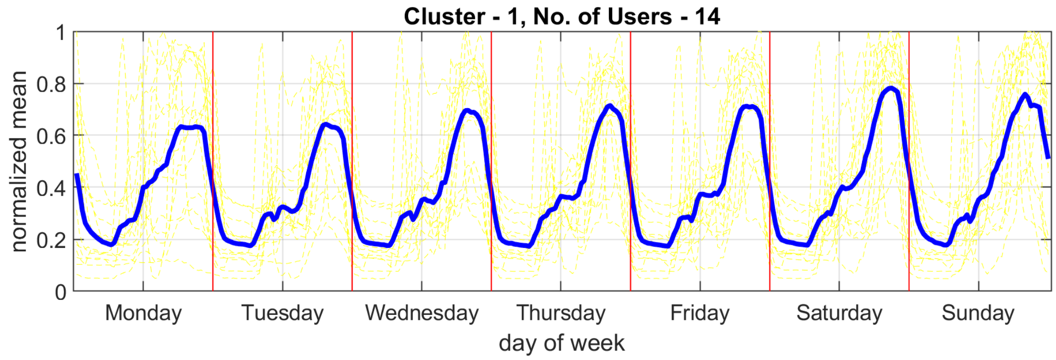
Most of the times, there exists a strong correlation among some users as they share several factors together. For example, all commercial users in the fireworks business will experience the same market demand for similar products during the same period of the year. So, it is usually recommended to form clusters where the users with similar power consumption profiles can be grouped together to enhance the load predictability. Clustering helps in determining the internal structure and identifying the hidden patterns of the dataset. This also helps in saving the computational time during the analysis of determining the peak values and peak times by skipping the users with similar power consumption profiles. This section discusses the analysis of cluster formation on the dataset used. To perform the same, '`kmeans`' algorithm, as shown in figure 2.7, is used on the normalised mean values of the power consumption values of each metering interval for all days of the week for all users.

Algorithm 1 k-means algorithm

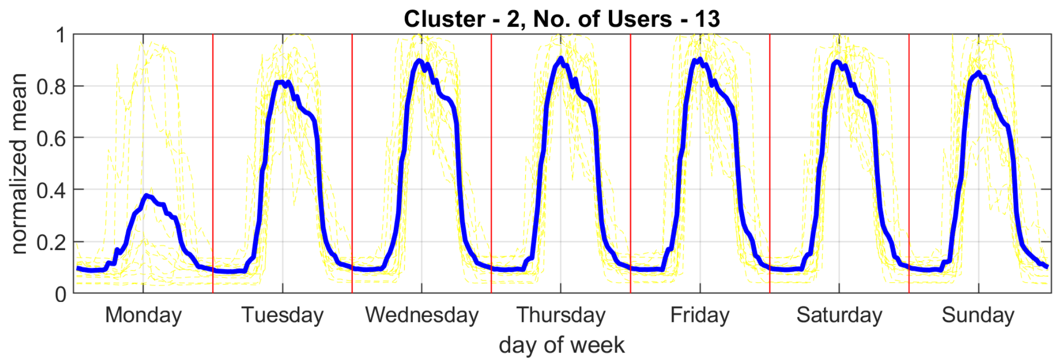
Specify the number k of clusters to assign.
Randomly initialize k centroids.
repeat
  expectation: Assign each point to its closest centroid.
  maximization: Compute the new centroid (mean) of each cluster.
until The centroid positions do not change.

Figure 2.7. K-means algorithm

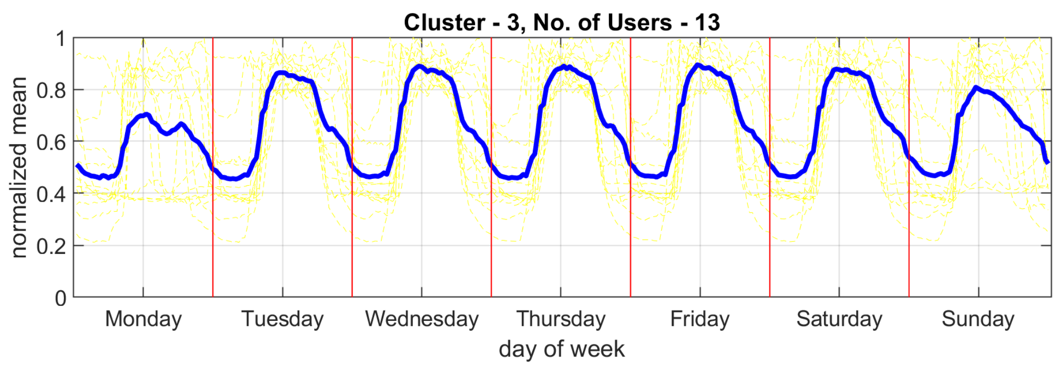
For this study, all 75 users are categorized into 5 clusters as shown in figure 2.8. The number of users in cluster 1, cluster 2, cluster 3, cluster 4, and cluster 5 are 14, 13, 13, 11, and 24 respectively. Users belonging to different clusters show peculiar characteristics on different days of the week. Users in cluster 5 consume a small amount of power on Sundays and Mondays whereas the power consumption pattern on other days of the week is the same throughout the day whereas the users in cluster 1 and cluster 4 display identical pattern on all days of the week. Additionally, users belonging to cluster 2 and 3 exhibit same pattern on all days of the week except Mondays. From the box plots of user '1' and user '25' shown in figures 2.3 and 2.6, it can be easily identified that user '1' belongs to cluster 2 whereas user '25' is a member of cluster 5.



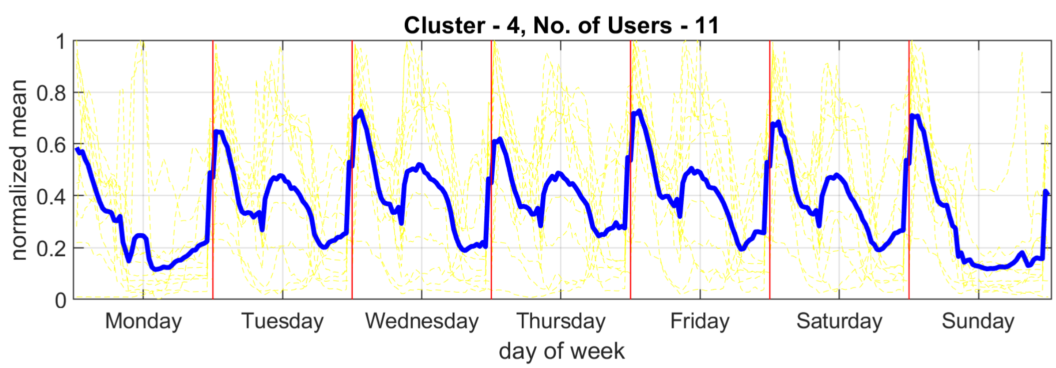
(a) Cluster 1



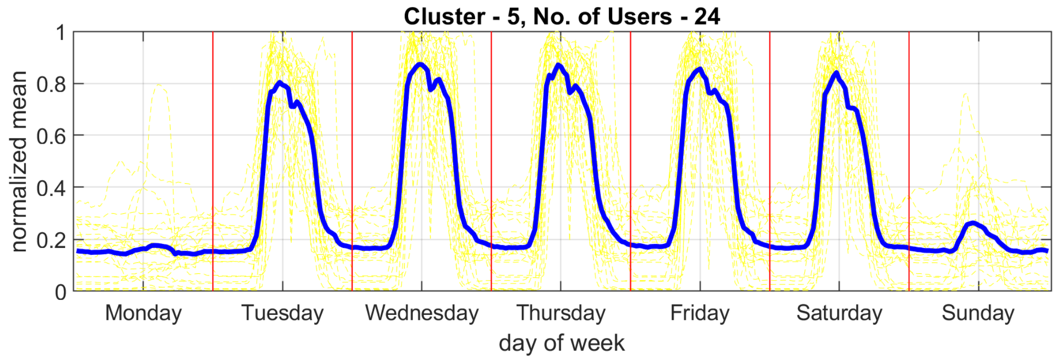
(b) Cluster 2



(c) Cluster 3



(d) Cluster 4

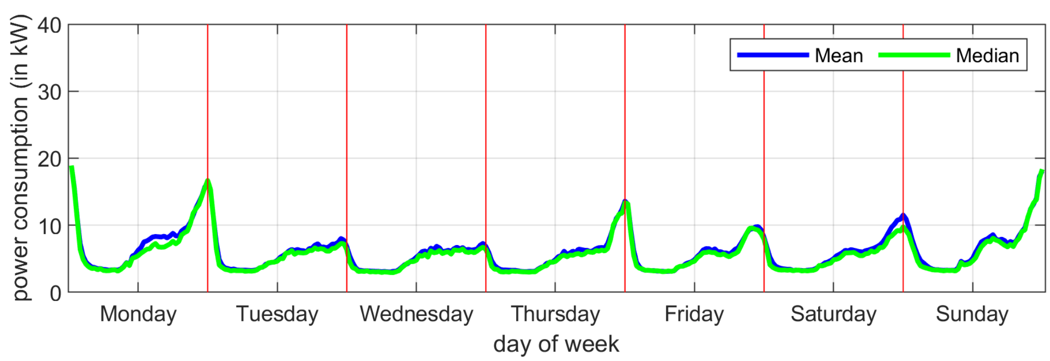


(e) Cluster 5

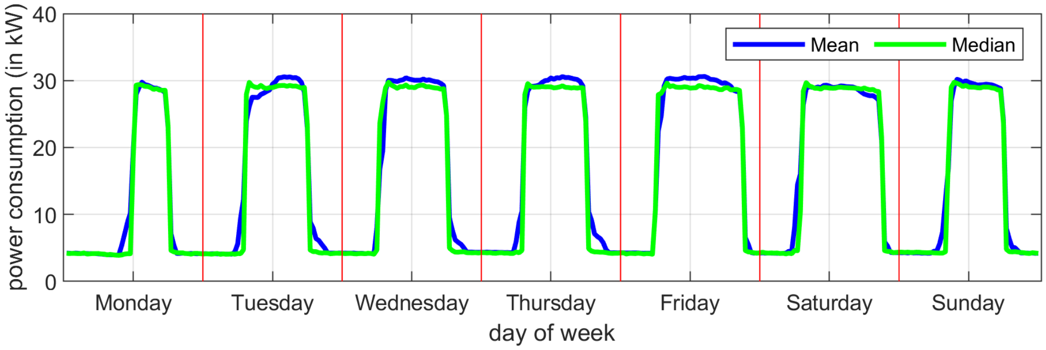
Figure 2.8. Clusters based on mean values of every metering interval for all days of the week

This analysis not only suggests the pattern for the entire week but also provides with a good estimation within the day of the week. For example, most of the users in cluster 2, cluster 3, and cluster consumes the maximum power during the middle of the day while users belonging to cluster 1 and cluster 4 consumes the maximum close to midnight.

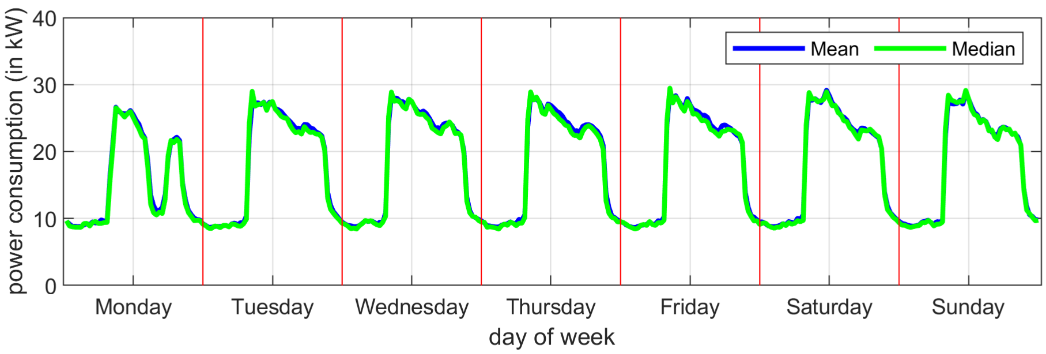
Same analysis has also been carried over the normalised median values, instead of normalised mean values, of every metering interval which resulted in almost the same categorization of users and profile patterns when divided into 5 clusters. This is because the mean and the median values values of every metering interval during all days of the week for most of the users are very close. The same has been shown in figure 2.9 for typical users belonging to each cluster.



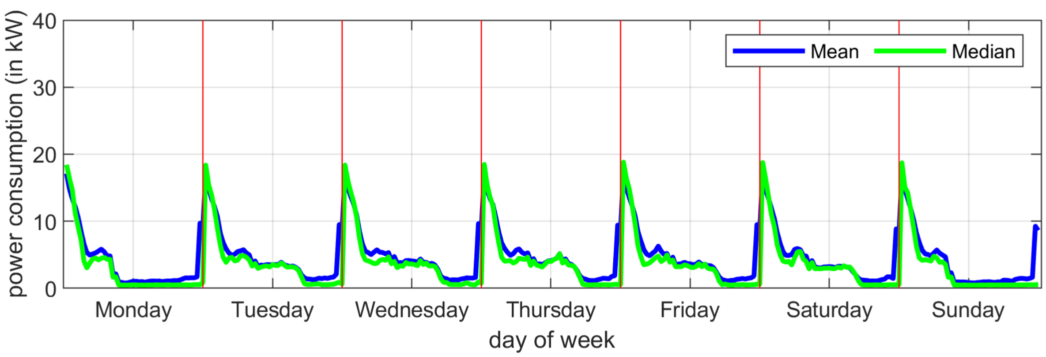
(a) User 19 in Cluster 1



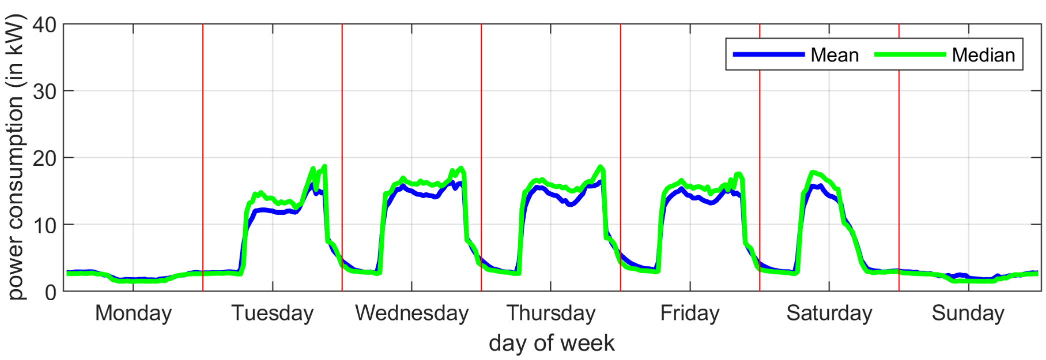
(b) User 1 in Cluster 2



(c) User 8 in Cluster 3



(d) User 23 in Cluster 4

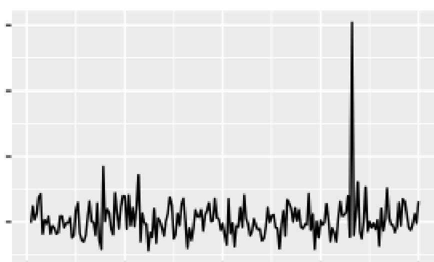


(e) User 25 in Cluster 5

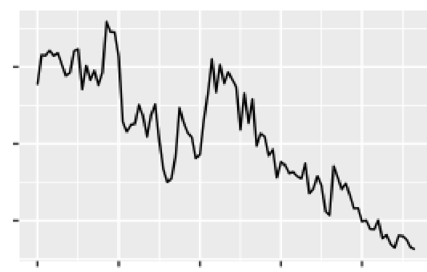
Figure 2.9. Mean and Median values of typical users from different clusters

2.2.4 Stationary Tests

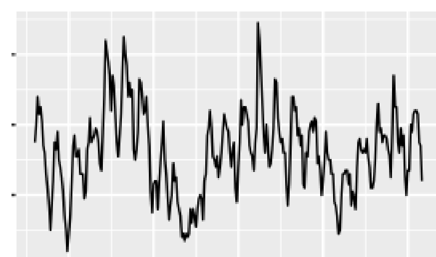
Stationary is an important aspect of time series analysis. By definition, a stationary time series is one whose properties do not depend on the time at which the series is observed. Time series with trends, or with seasonality, are not stationary. Figure 2.10 shows different types of time series as examples. The time series shown is (a) is stationary whereas (b) and (c) are not stationary as they possess a downward trend and seasonality respectively.



(a) Stationary time series



(b) Time series with trend



(c) Time series with seasonality

Figure 2.10. Different types of time series

One way to determine more objectively whether the time series is non stationary is through statistical hypothesis tests known as unit root tests. A unit root is a stochastic trend in a time series. A unit root test tests whether a time series variable is non-stationary and possesses a unit root. The null hypothesis is generally defined as the presence of a unit root and the alternative hypothesis is either stationarity, trend stationarity or explosive root depending on the test used. Unit root tests include (but limited to) Augmented Dickey–Fuller (ADF) Test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test. These tests are discussed to check the stationarity on the used dataset for all users on different days of the week in the following subsections.

Augmented Dickey–Fuller Test

An augmented Dickey–Fuller test tests that a unit root is present in a time series sample. The model used for this test with $y(t)$ as the time series sample is mentioned in eqn. 2.1.

$$y_t = c + \delta t + \phi y_{t-1} + \epsilon_t \tag{2.1}$$

where c is the drift component, δt is the trend component, ϕy_{t-1} is the first order auto regression component, ϵ_t represents the mean zero innovation process.

The null hypothesis and the alternative hypothesis for this test are given as the following

Null Hypothesis, H_o: $\phi = 1$, meaning there is a possibility of a unit root,

Alternative Hypothesis, H_a: $\phi < 1$, meaning no possibility of a unit root.

It should be noted that the model with $\delta = 0$ has no trend component and with $c = 0$ has no drift component. Figure 2.11 shows the number of stationary or trend-stationary days for different users on different days of the week where the null hypothesis is rejected at a significance level of 5%.

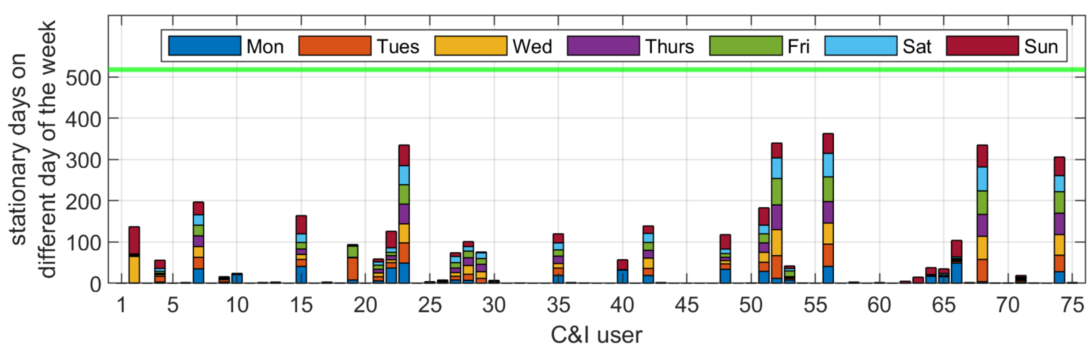


Figure 2.11. Augmented Dickey–Fuller Test

From the figure above, it is clear that there are quite a few users whose power consumption profiles are either stationary or trend-stationary but for most of the users, power consumption profiles are not stationary on different days of the week.

Kwiatkowski–Phillips–Schmidt–Shin Test

This test tests that an observable time series is stationary around a deterministic trend (i.e. trend-stationary) against the alternative of a unit root using the following model

$$y_t = c_t + \delta t + u_{1t}, \tag{2.2}$$

$$c_t = c_{t-1} + u_{2t}, \tag{2.3}$$

where c_t is the random walk term, δt is the trend component, u_{1t} models the stationary process, and u_{2t} represents the independent and identically distributed process with mean 0 and variance σ^2.

The null hypothesis and the alternative hypothesis for this test are given as the following

$$\text{Null Hypothesis, } H_o\text{: } \sigma^2 = 0,$$

$$\text{Alternative Hypothesis, } H_a\text{: } \sigma^2 > 0.$$

It should be noted that the null hypothesis $\sigma^2 = 0$ implies that the random walk term (c_t) is constant and acts as the model intercept whereas the alternative hypothesis $\sigma^2 > 0$ introduces the unit root in the random walk. Figure 2.11 shows the number of trend-stationary days for different users on different days of the week where the null hypothesis cannot be rejected at a significance level of 5%.

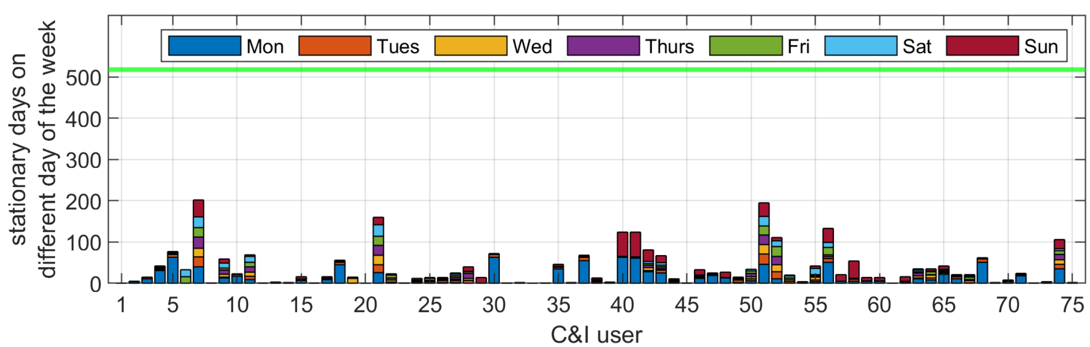


Figure 2.12. Kwiatkowski–Phillips–Schmidt–Shin Test

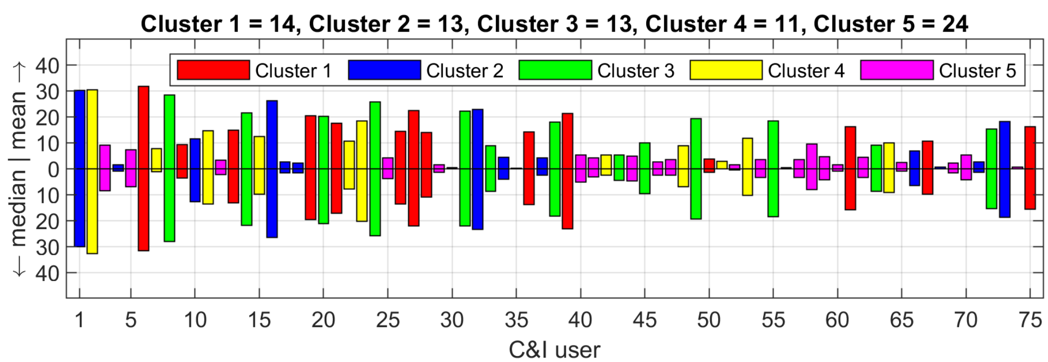
From the above figure, it can be said that most of the users's power consumption profiles are not stationary around a deterministic trend on different days of the week.

2.3 Exploratory Data Analysis for Peak Values and Peak Times

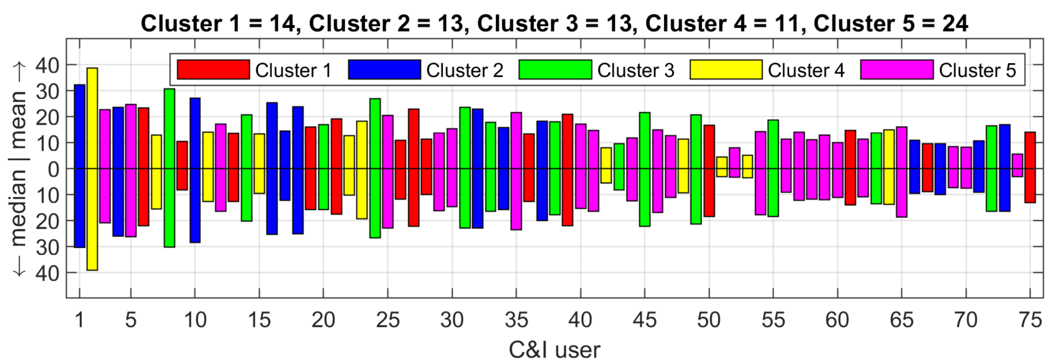
In this section, distribution of the daily peak values and their corresponding times for all users belonging to different clusters is discussed.

2.3.1 Empirical Distribution of Peak Values

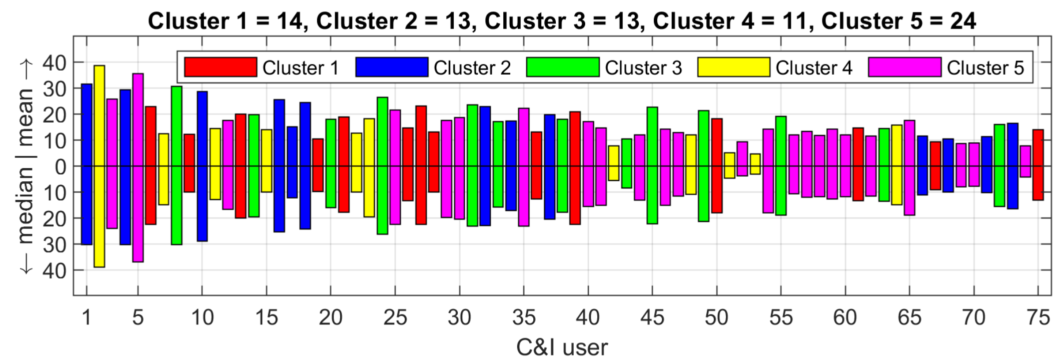
Before analysing the distribution of peak values, it is important to understand the scale of the peak power consumption values of the users representing the commercial and industrial businesses specific this region. Figure 2.13 represents the mean and the median peak power consumption values of all users on different days of the week.



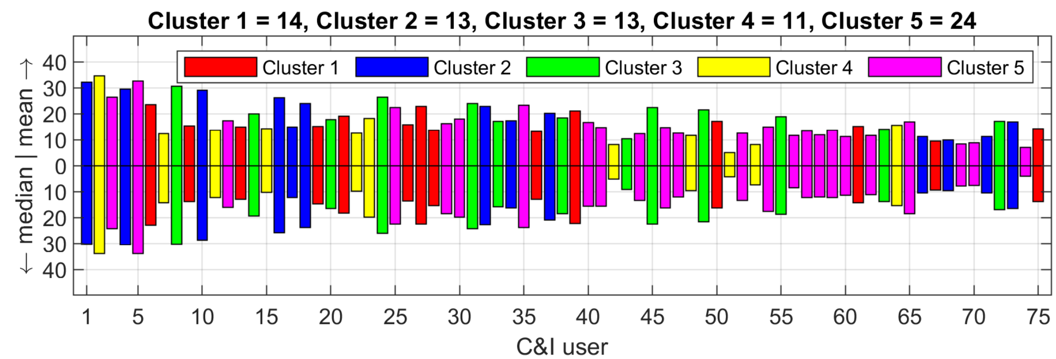
(a) Mondays



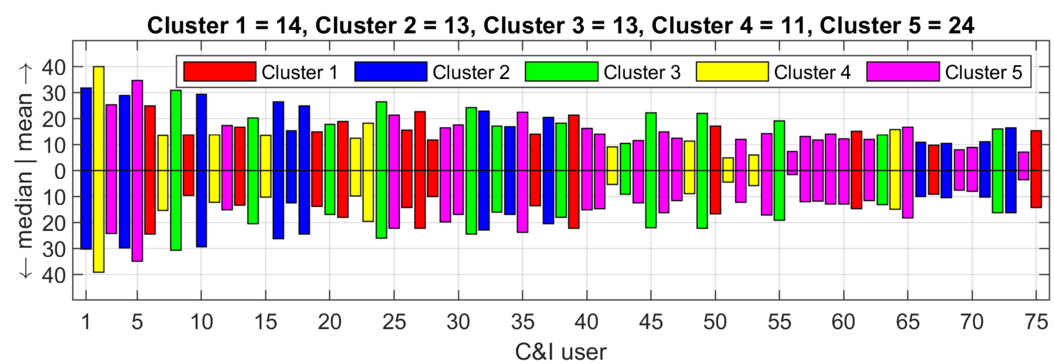
(b) Tuesdays



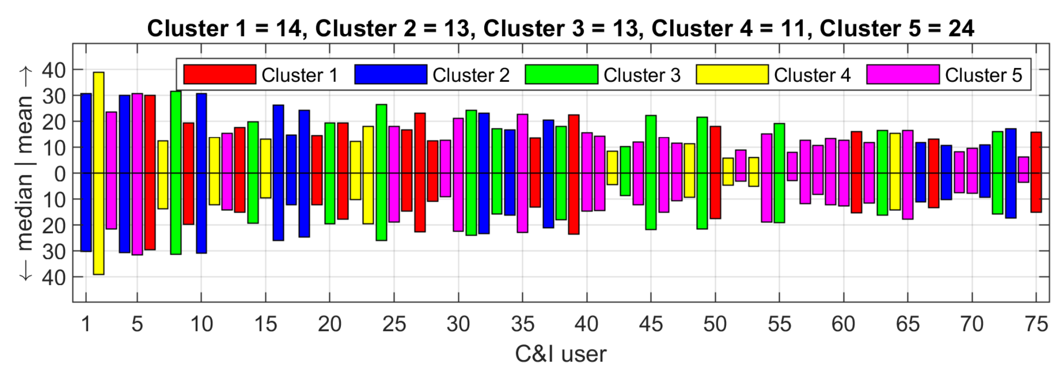
(c) Wednesdays



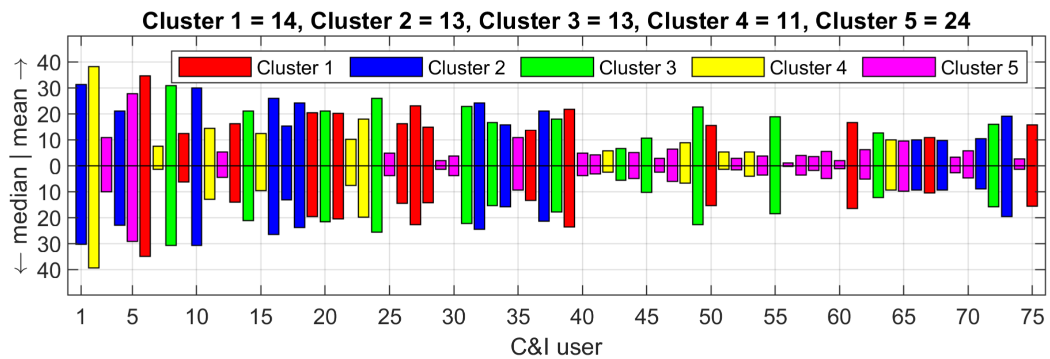
(d) Thursdays



(e) Fridays



(f) Saturdays



(g) Sundays

Figure 2.13. Mean and median peak power consumption values

From the figures above, it can be depicted that both mean and median peak values for most of the users coincide and are in the range of 10kW to 40kW. It should also be noted that these values for some users may vary on different days of the week depending on the cluster they belong to.

Historical data can provide a good understanding on the distribution of the peak values for some users. For example, the peak power consumption values for user '16' (belonging to cluster 2) is close to 30 kW during all days of the week which can be inferred from the box plot shown below in figure 2.14.

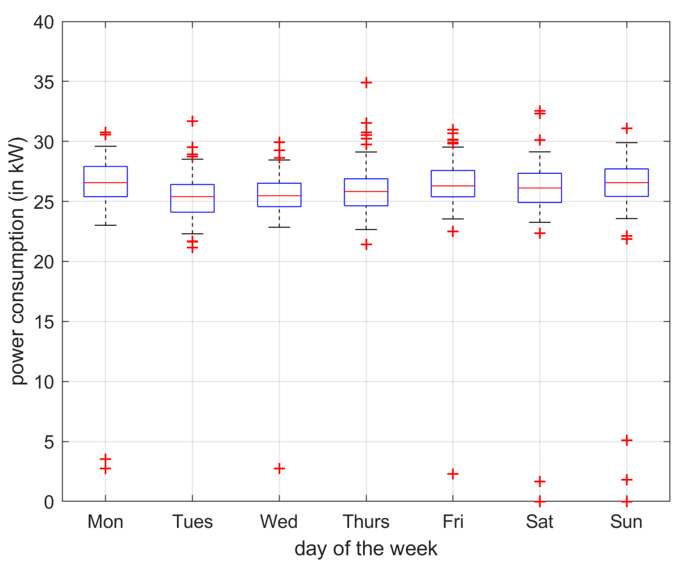
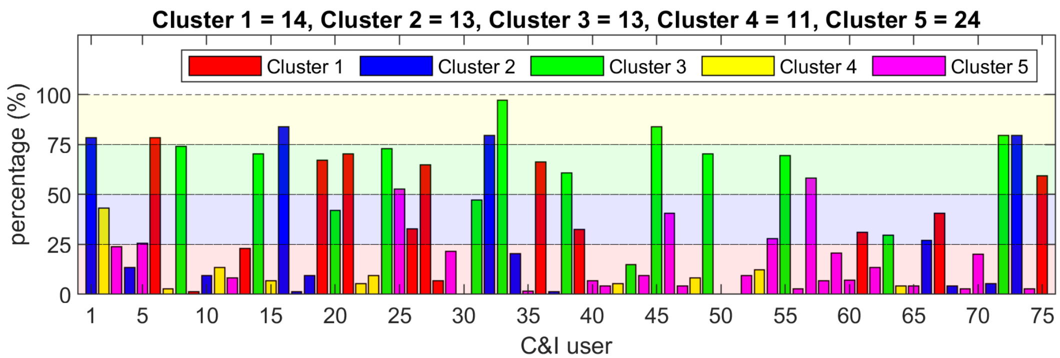
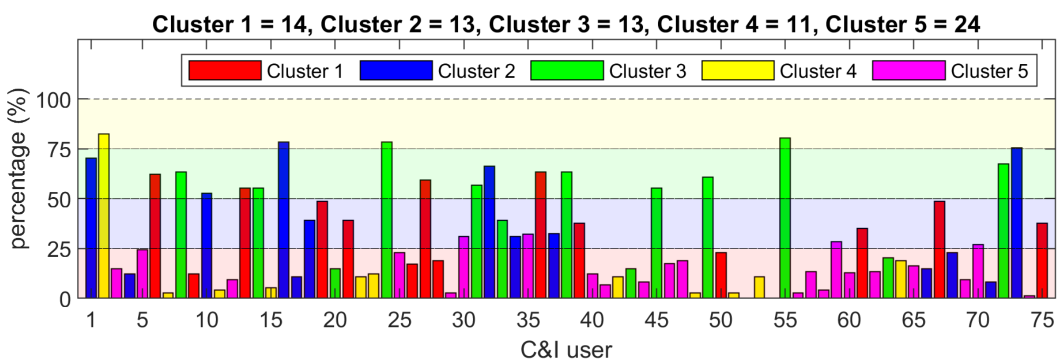


Figure 2.14. Peak power consumption values of user '16' on different days of the week

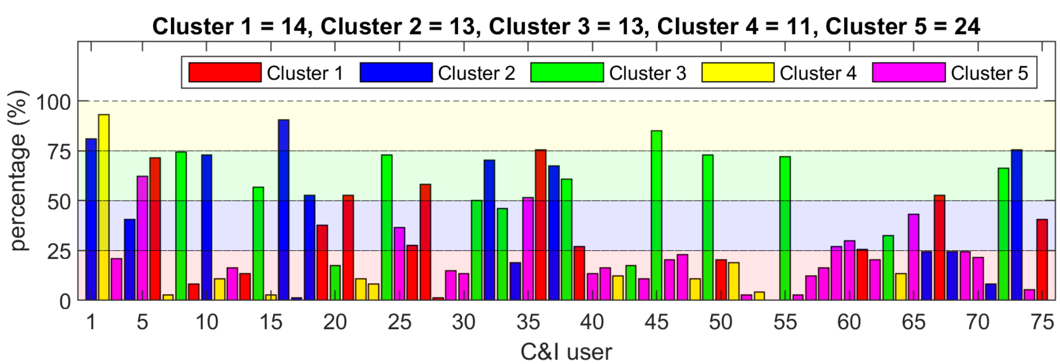
However, not every user exhibits the same characteristics as user '16'. To understand how concentrated the peak values are for different users on all days of the week, percentage of the peak values in the interval of +/- 10% of the mean peak value of the user on the particular day of the week is calculated and plotted in figure 2.15.



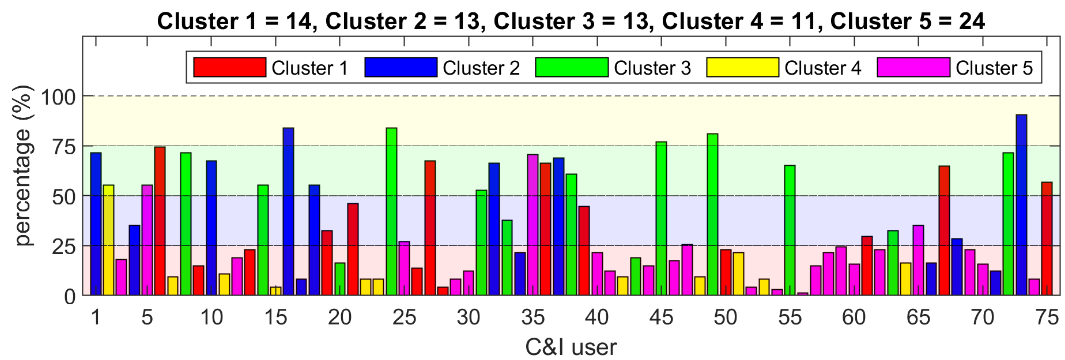
(a) Mondays



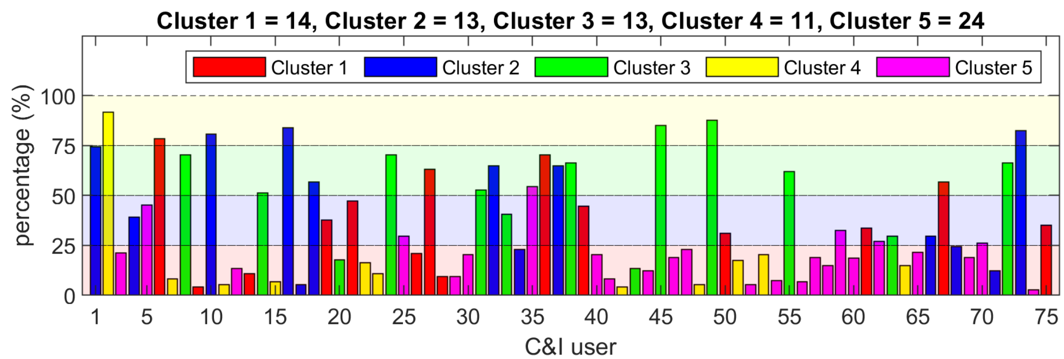
(b) Tuesdays



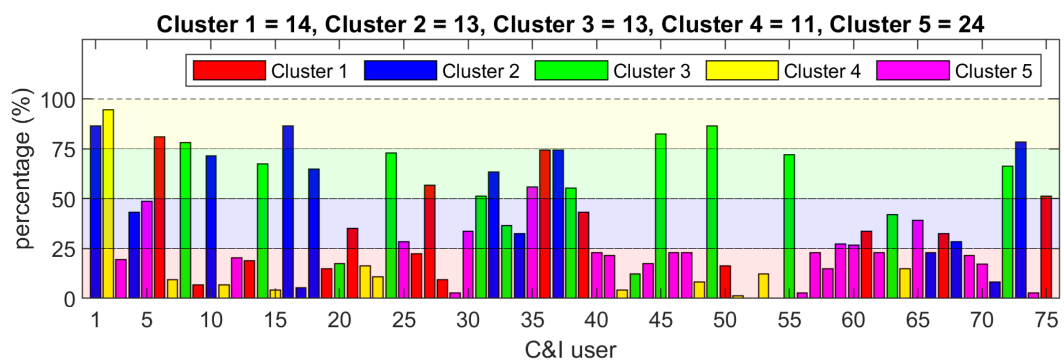
(c) Wednesdays



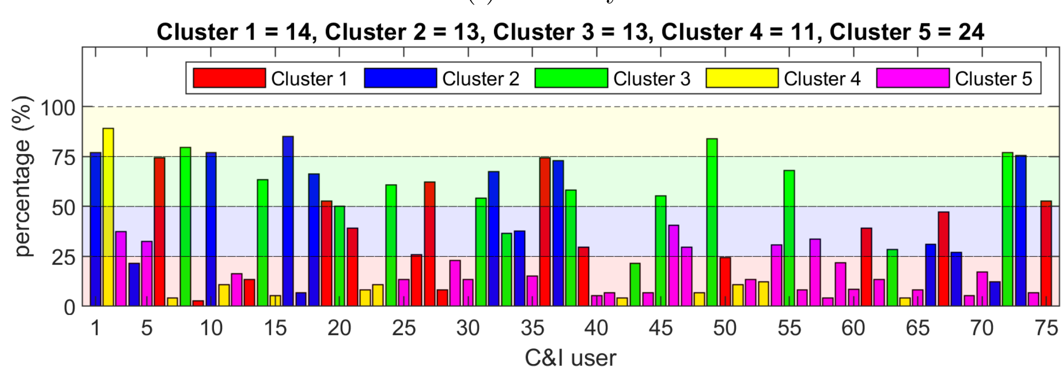
(d) Thursdays



(e) Fridays



(f) Saturdays



(g) Sundays

Figure 2.15. % of peak values in the interval of +/- 10% around user's mean peak value

It can be observed that the peak values for most of the users on different days of the week do not lie in the vicinity of the mean peak value. Very few users on different days of the week cross the 75% benchmark among which the users from cluster 2 and cluster 3 dominate clearly. It should also be noticed that the percentages of the peak values lying near the mean peak value for the (different) users are not the same for different days of the week. This means the peak values are distributed differently on different days of the week around its mean.

Another natural way to analyse the distribution of the peak values for different users on different days of the week is by testing whether these values come from a normally distributed population. This can be examined using Lilliefors test. Figure 2.16 indicates the distribution of number of users from each cluster whose peak values follow normal distribution on different days of the week at a confidence level of 95%. Clearly, there are not many users whose peak values are normally distributed on different days of the week.

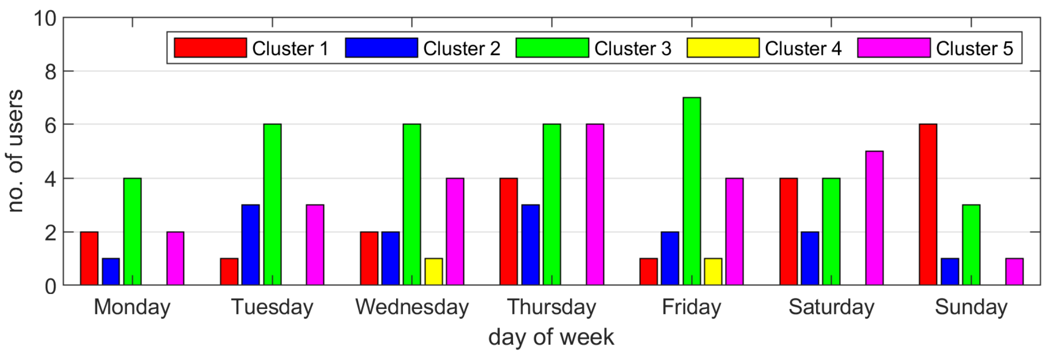
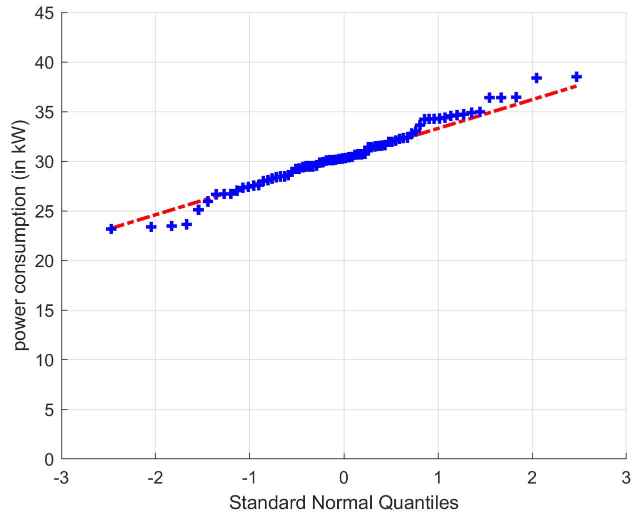
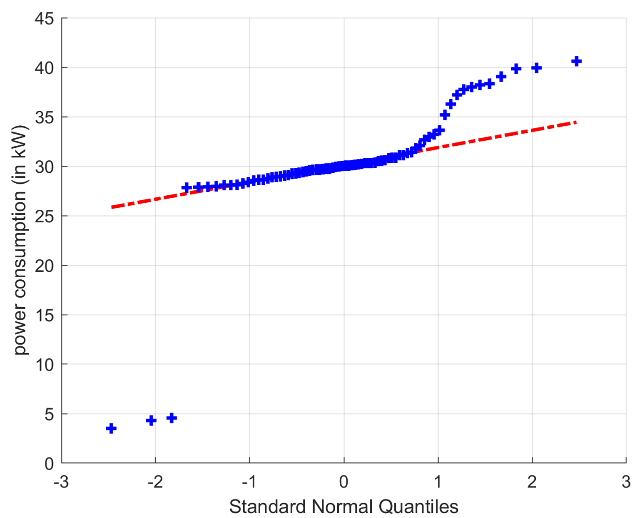


Figure 2.16. No. of users whose peak values follow normal distribution

To have a direct comparison of the peak values against the standard normal quantities graphically, QQ plots are used. Figure 2.17 (a) shows a user whose peak values are aligned with the standard normal quantities on a particular day of the week whereas figure 2.17 (b) implies that this user's peaks values follow a heavy tail distribution on a different day of the week, which is not trivial to identify.



(a) Peaks values follow normal distribution

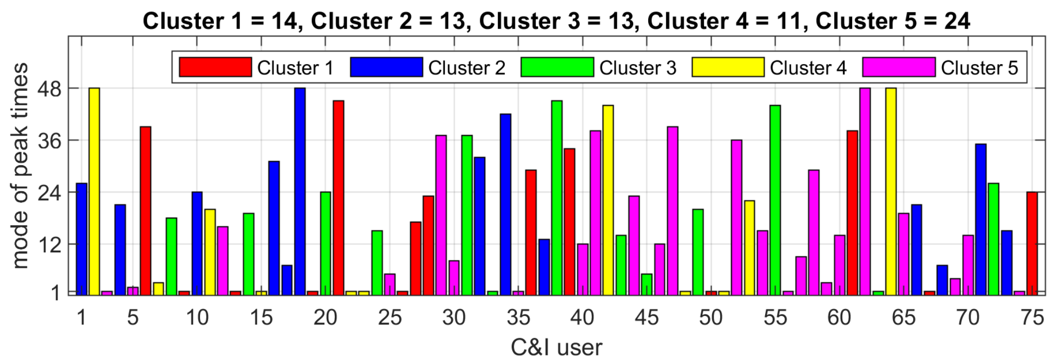


(b) Peak values do not follow normal distribution

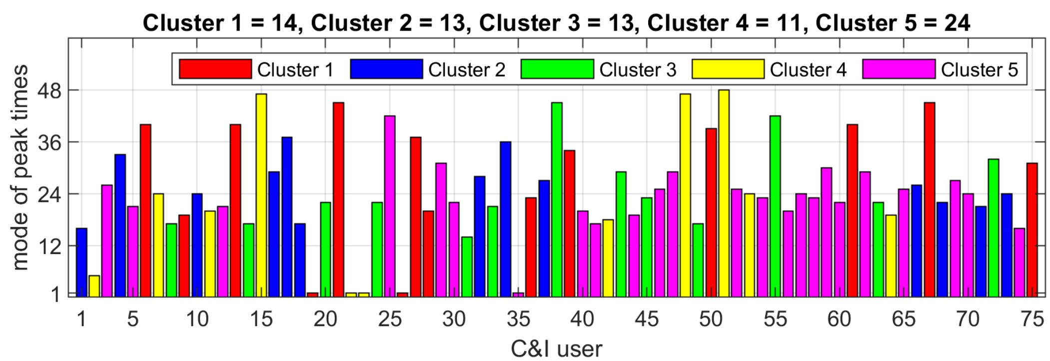
Figure 2.17. Comparison against normal distribution

2.3.2 Empirical Distribution of Peak Times

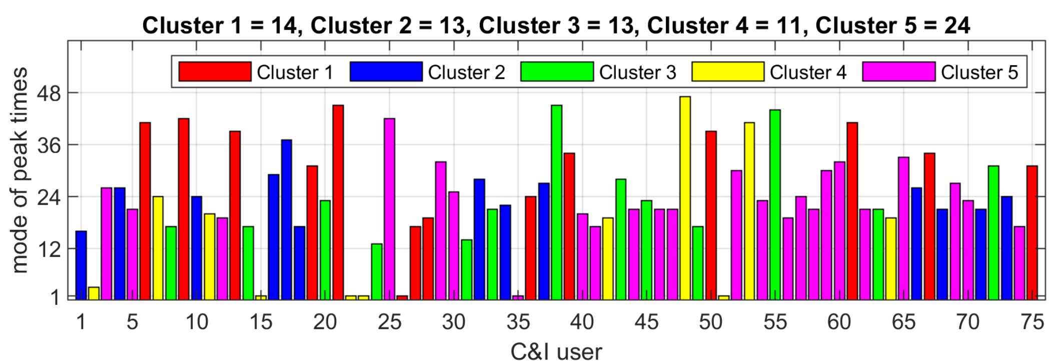
In the previous subsection, distribution of peak values was discussed. But it is also important to know when do these peaks occur. Like the mean and the median values used for developing an understanding about the peak values, mode is used for analysing the peak times. Figure 2.18 shows the times of all users where the peak happens with the maximum frequency on different days of the week.



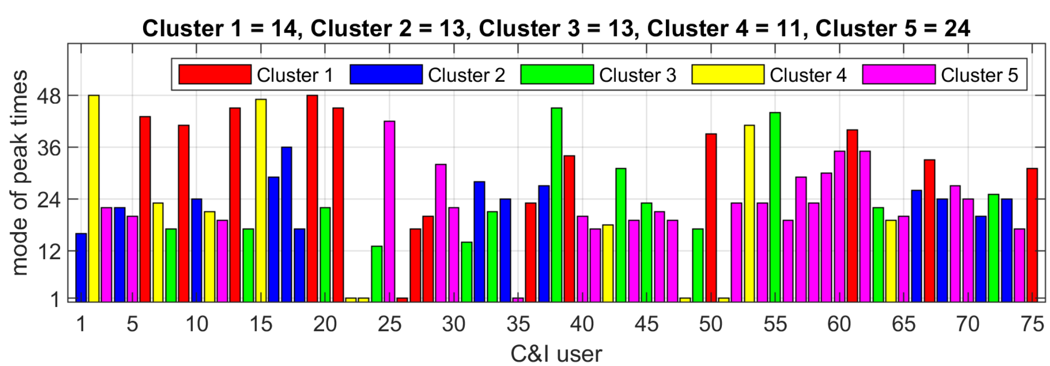
(a) Mondays



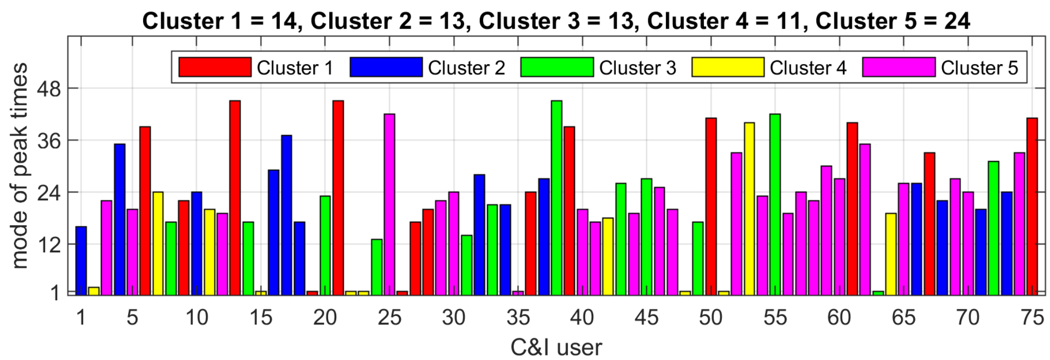
(b) Tuesdays



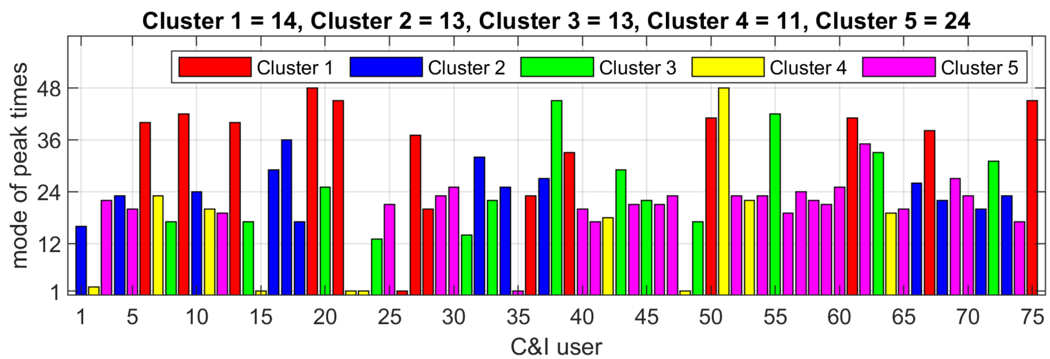
(c) Wednesdays



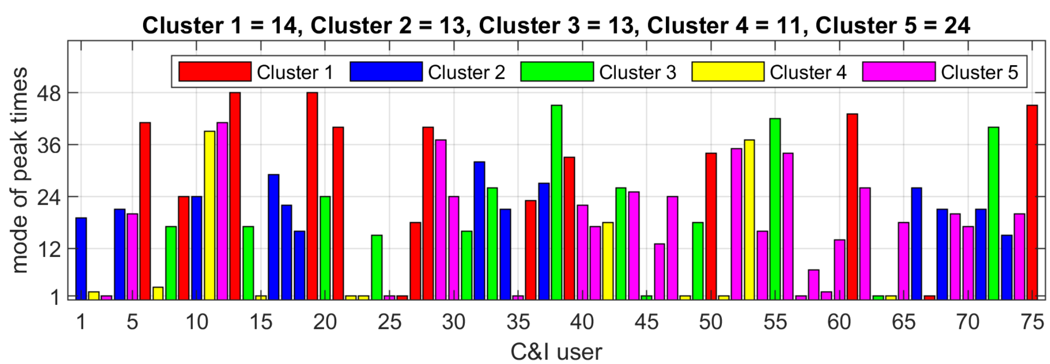
(d) Thursdays



(e) Fridays



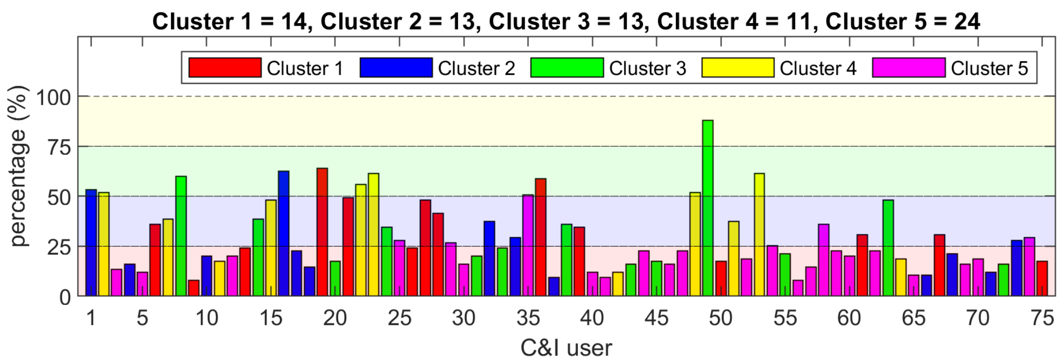
(f) Saturdays



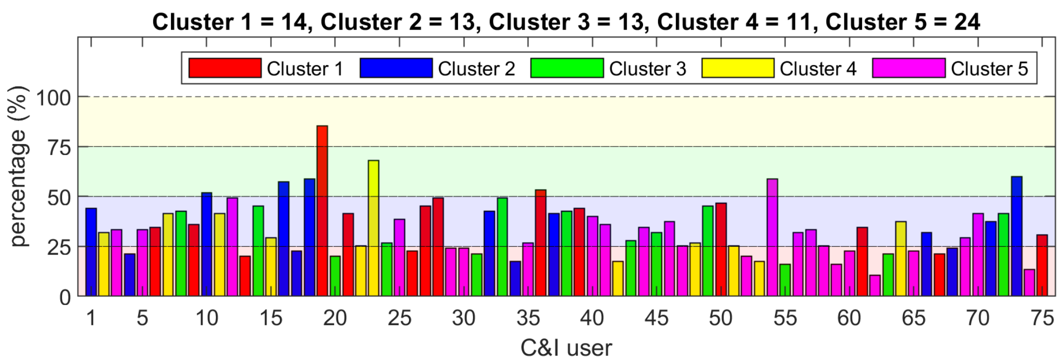
(g) Sundays

Figure 2.18. Mode of peak times

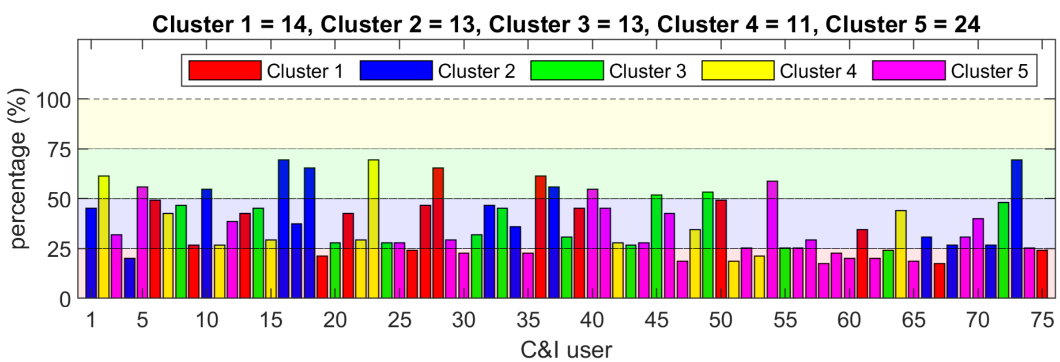
It can be observed that different users consume the maximum power during different metering intervals. Clustering helps in identifying these peak metering intervals for the majority of the users belonging to a specific cluster on different days of the week, but not any every user within the cluster follow the same general trend of the cluster. This explains how scattered these peak times are! To understand the gravity of the same numerically, percentage of peak times in the interval of +/- 1 of the mode peak time is plotted in figure 2.19 for all users on different days of the week, where most of the users fail to cross the 50% benchmark.



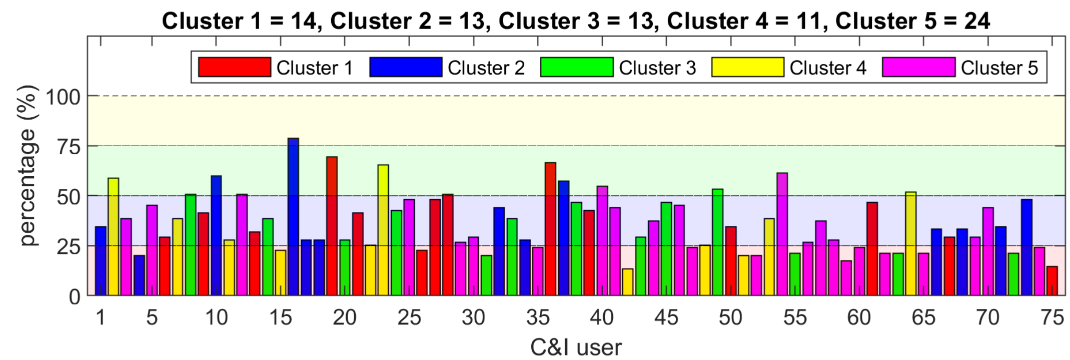
(a) Mondays



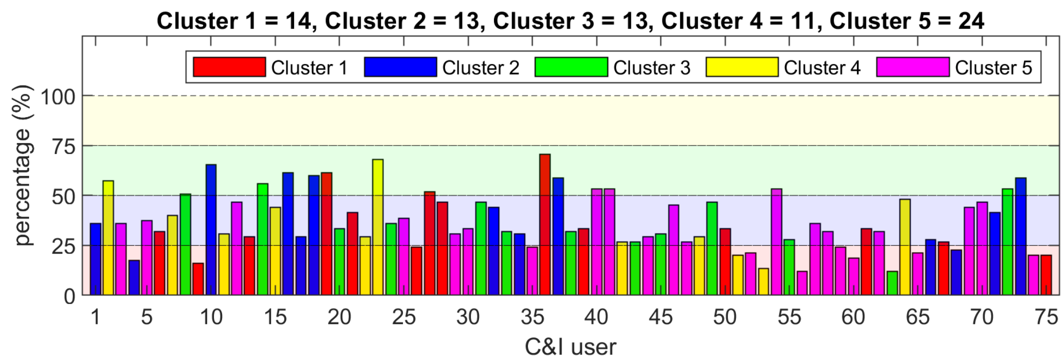
(b) Tuesdays



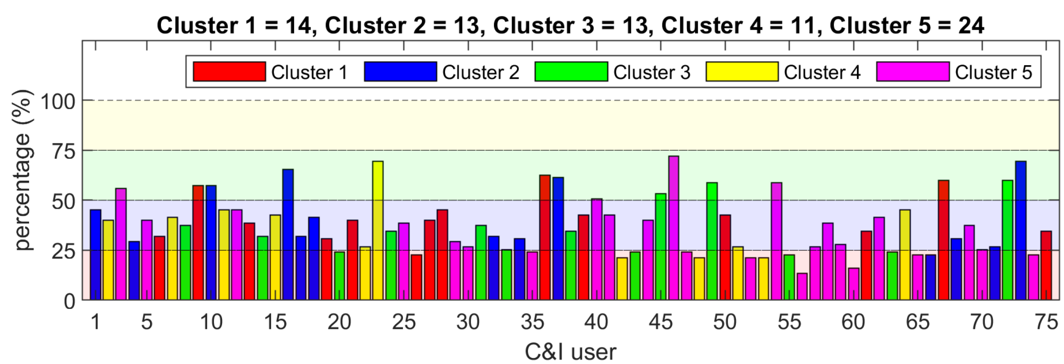
(c) Wednesdays



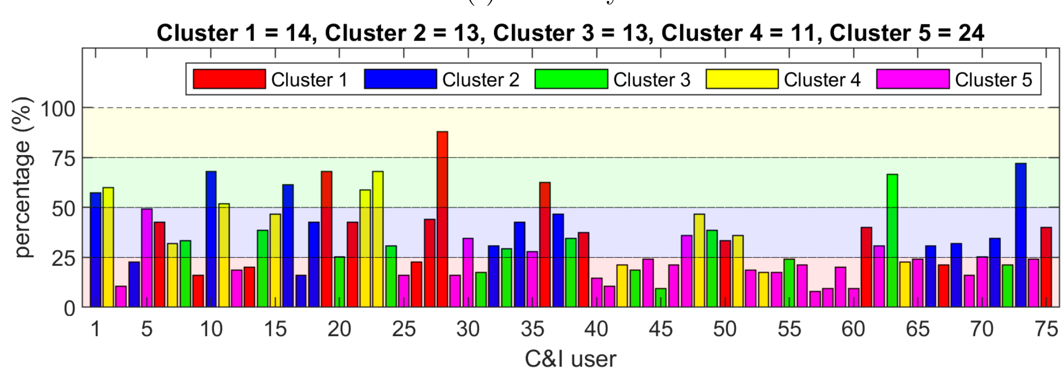
(d) Thursdays



(e) Fridays



(f) Saturdays



(g) Sundays

Figure 2.19. % of peak times in the interval of +/- 1 around user's mode peak time

2.4 Summary

The highlights of this chapter are the following:

1. Dividing the dataset of the power consumption values into different days of the week helps in reducing large variances and in understanding the correlation among different days of the week.

2. Clustering analysis assists in identifying strong correlation among users based on the pattern recognition of the power consumption profiles.

3. Almost all power consumption profiles pertaining to different users on different days of the week are not stationary.

4. Daily peak power consumption values and their corresponding metering intervals for most of the users are widely distributed during different days of the week. Elementary statistical methods such as mean, median and mode are not sufficient enough to predict the same. Advanced machine learning methods are required.

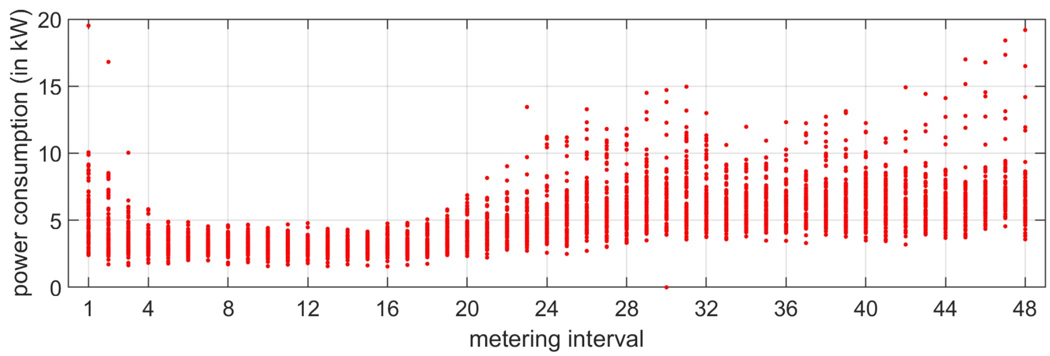
3. MACHINE LEARNING FOR TIME SERIES DATA

The primary goal of this study is to forecast the daily peak power consumption values and their corresponding metering intervals of every user on different days of the week. This is achieved by taking two different approaches. The first approach is by forecasting the entire time series for 24 hours, starting from 12:00 am to 11:59 pm, for each user every day and then computing the peak power consumption values using the predicted time series which is referred as the *indirect approach* whereas the second approach is by directly forecasting the peak power consumption values based on the historical data which is referred as the *direct approach*. For both of these approaches, different machine learning techniques have been explored and are discussed in chapter 3 (current chapter) and in chapter 4 (the following chapter) respectively.

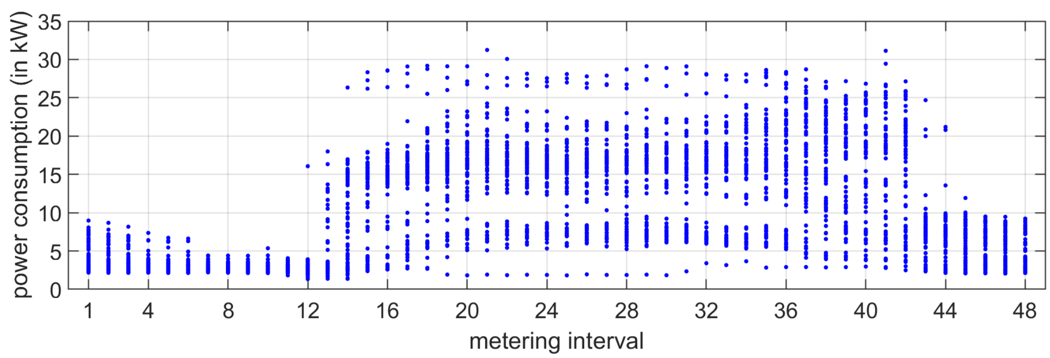
3.1 Actual data pertaining to case study

Before moving to the specifics of the machine learning algorithms, this section discuss the information related to the case study used for this and for the following chapter. The size of the dataset being large, some typical users on typical day of the week is chosen as a part of case study. Specifically, the data pertaining to user '19' and user '25' on Wednesdays is chosen for the case study. The reason for choosing these users is a part of cluster analysis discussed in section 2.2.3. The idea is to choose the users from cluster 1 and cluster 5 as these two clusters accommodate the 50% of total users (38 out of 75) when compared with the other clusters. Therefore, user '19' and user '25' are chosen from cluster 1 and cluster 5 respectively. Secondly, the reason for choosing 'Wednesdays' as the typical day of the week can also be explained using cluster analysis. Users belonging to cluster 1 have the same power consumption profile on all days of the week, so choosing any other day from the week other than Wednesdays will not result in a significant difference in the analysis whereas the users belonging to cluster 5 barely consume any power on Mondays and Sundays which also means that the peak power consumption is very low during these days of the week. The main objective of this study being learning the peak power consumption values of C&I users which are significant for both electric utilities and to the users themselves for the applications like demand charge reduction, near-zero peak on a weekend day does not matter much. Therefore, choosing a typical working day makes more sense than merely choosing a non-working day which justifies the selection of 'Wednesdays' for the case study.

To have visual understanding of the actual data, figure 3.1 represents the power consumption profiles of user '19' and user '25' on Wednesdays for all 75 weeks present during the period from August 2009 to December 2010.



(a) User '19'



(b) User '25'

Figure 3.1. Power consumption profiles of user '19' and user '25' on Wednesdays

As discussed in section 2.2.4, figure 3.2 displays the number of stationary days according to the ADF and KPSS unit root tests for user '19' and user '25' on Wednesdays. ADF tests suggests that there are no stationary or trend stationary days for both users while KPSS test suggests that there are quite a few days which are stationary around a deterministic trend. This supports the argument of choosing these users as the typical users as most of the power consumption profiles for different users on different days of the week are not stationary.

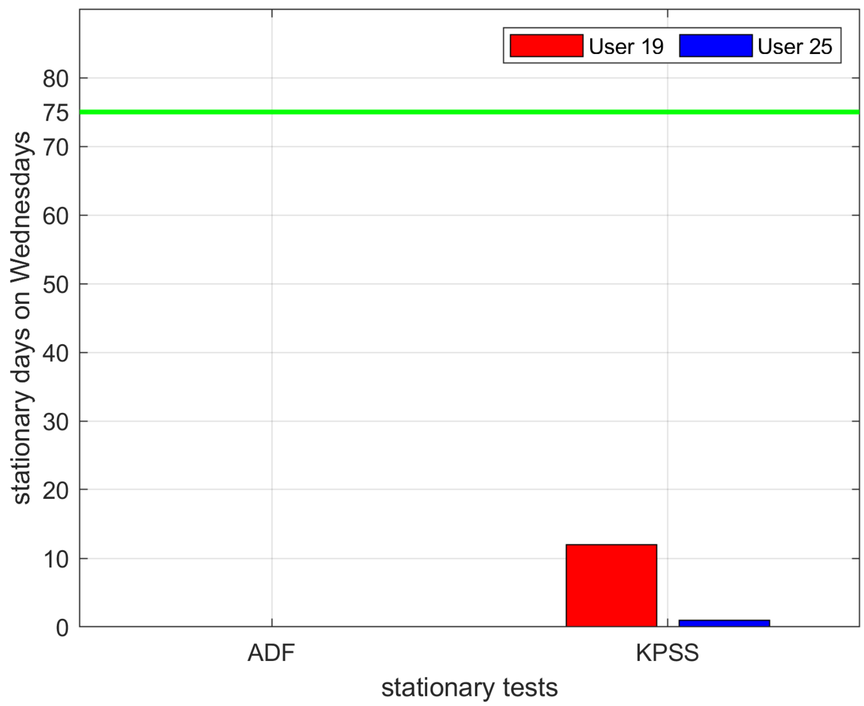
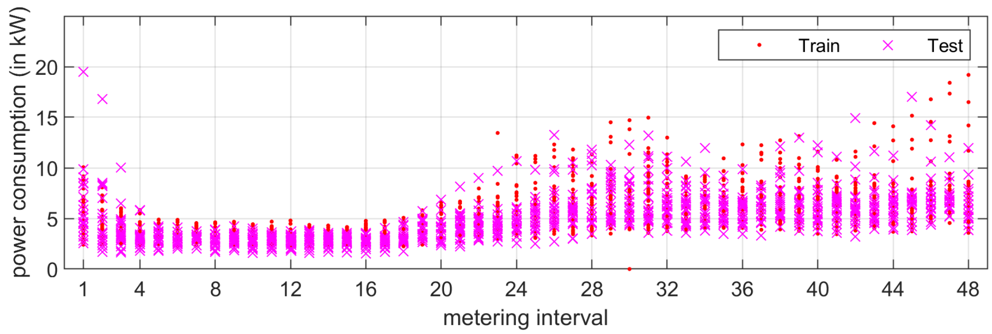


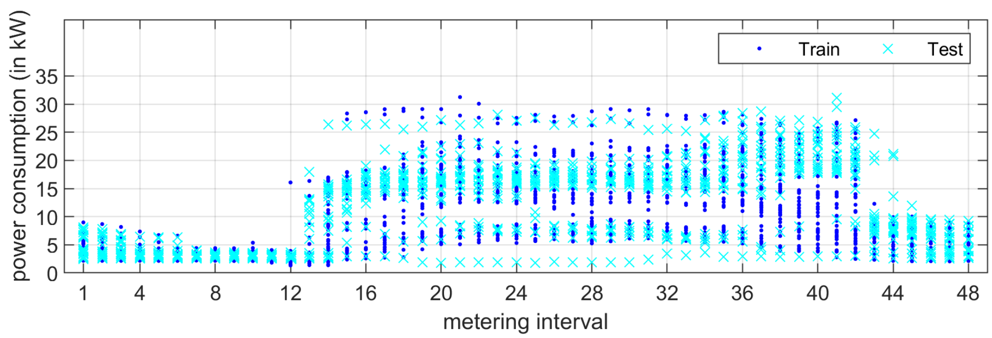
Figure 3.2. Stationary days of user '19' and user '25' on Wednesdays

3.2 Machine Learning Models

Often times in machine learning techniques, the entire dataset is categorized into two main groups, namely *training* and *test* in order to train and test the performance of the model. In general, 70-80% of the dataset is used as the training set, whereas 20-30% is used as the test set. In this study, the total number of weeks for each user and each day of the week being equal to 75, the first 55 weeks are considered as the training weeks, whereas the last 20 weeks (approximately 27% of the entire dataset) are considered as the test weeks. The reason for selecting the last weeks of the dataset as the 'test set' is because the same can be applied to any new user where the historical information of that user can be used as the training set in order to predict the upcoming values. Figure 3.3 shows the power consumption profiles of user '19' and user '25' forming training and test sets on Wednesdays pertaining to the case study.



(a) User '19'



(b) User '25'

Figure 3.3. Power consumption profiles of user '19' and user '25' forming training and test Sets on Wednesdays

The symbolic representation of the entire training set consisting of the power consumption values of K users at T metering intervals on all days of the week for W^{train} weeks is given as $p_k(t, d, w)$ with $k \in \mathcal{K}$, $t \in \mathcal{T}$, $d \in \mathcal{D}$, and $w \in \mathcal{W}^{\text{train}}$ whereas the test set can be represented as $p_k(t, d, \mathrm{j})$ with $k \in \mathcal{K}$, $t \in \mathcal{T}$, $d \in \mathcal{D}$, and $\mathrm{j} \in \mathcal{W}^{\text{test}}$ where,

$$\mathcal{K} = \{1, 2, ..., K\}, \text{ with } K = 75,$$

$$\mathcal{T} = \{1, 2, ..., T\}, \text{ with } T = 48,$$

$$\mathcal{D} = \{1, 2, 3, 4, 5, 6, 7\},$$

$$\mathcal{W}^{\text{train}} = \{1, 2, ..., W^{\text{train}}\}, \text{ with } W^{\text{train}} = 55,$$

$$\mathcal{W}^{\text{test}} = \{W^{\text{train}} + 1, W^{\text{train}} + 2, ..., W^{\text{test}}\}, \text{ with } W^{\text{test}} = 75.$$

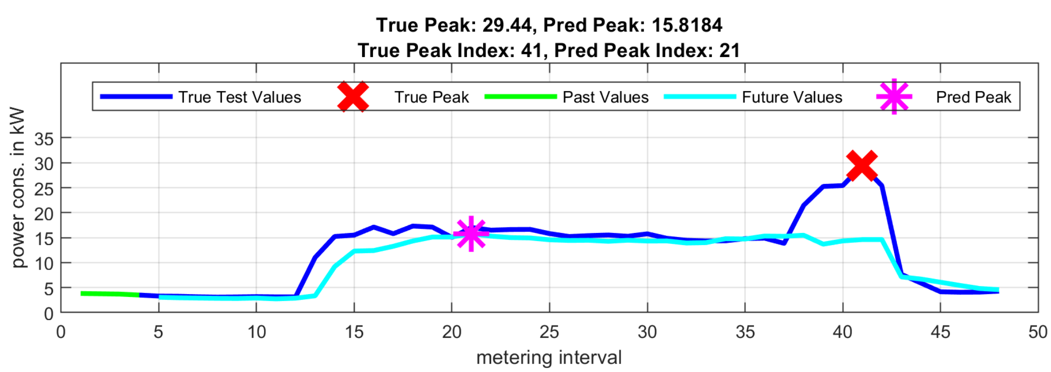
For our case study, $k = \{19, 25\}$, $d = \{3\}$ with all other variables as mentioned above unless otherwise stated.

3.2.1 Offline Average Model

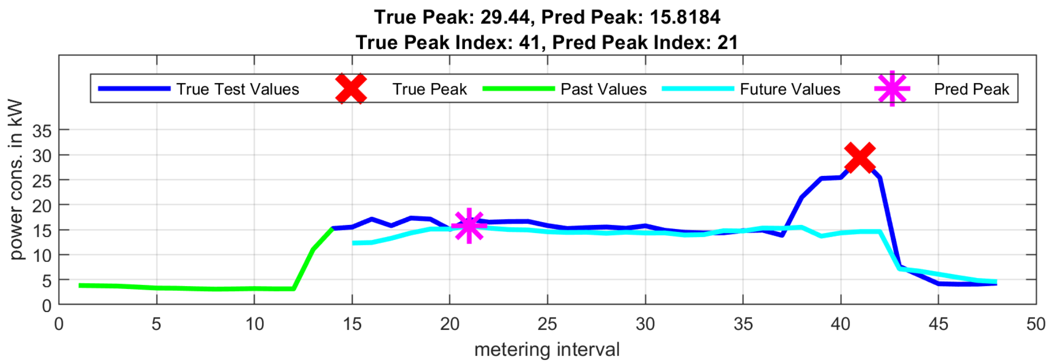
The basic model of forecasting the power consumption values is the average model. In offline average model, the forecasting of the power consumption values for the upcoming metering intervals during the day is calculated by taking the average of all training data corresponding to respective upcoming metering intervals on different days of the week for different users. Mathematically, it can be expressed as

$$\hat{p}_k(\tau, t, d, \mathrm{j}) = \begin{cases} p_k(t, d, \mathrm{j}), & \forall t < \tau, \\ \frac{1}{W^{\text{train}}} \sum\limits_{w=1}^{W^{\text{train}}} p_k(t, d, w), & \forall t \geq \tau. \end{cases} \tag{3.1}$$

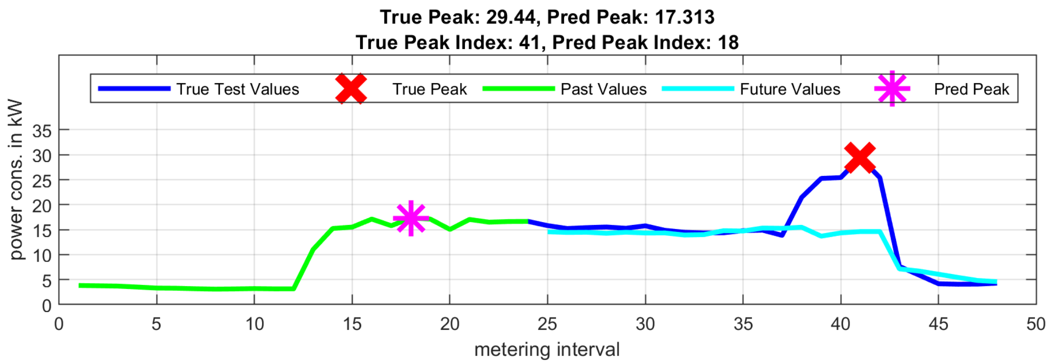
In the above sets of equations, τ represents the metering interval at which the prediction is done for the rest of the day. For example, figure 3.4 (a) shows the predicted power consumption profile of user 25 on Wednesday of week 65 at $\tau = 5$, where the forecasting is done for the metering intervals greater than or equal to $\tau = 5$. Moreover, the green curve indicates the power consumption values occured in the past, the cyan curve indicates the predicted power consumption values and the blue curve indicates the actual values for the upcoming metering intervals for this user. It also shows the actual peak value with $\times$ and the predicted peak with $*$. Figure 3.4 shows the predicted power consumption profiles for user 25 on Wednesday of week 65 at different values of τ.



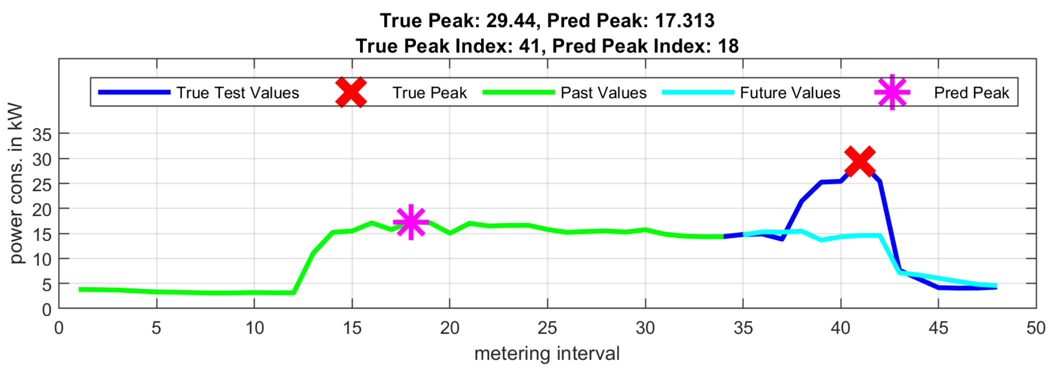
(a) $\tau = 5$



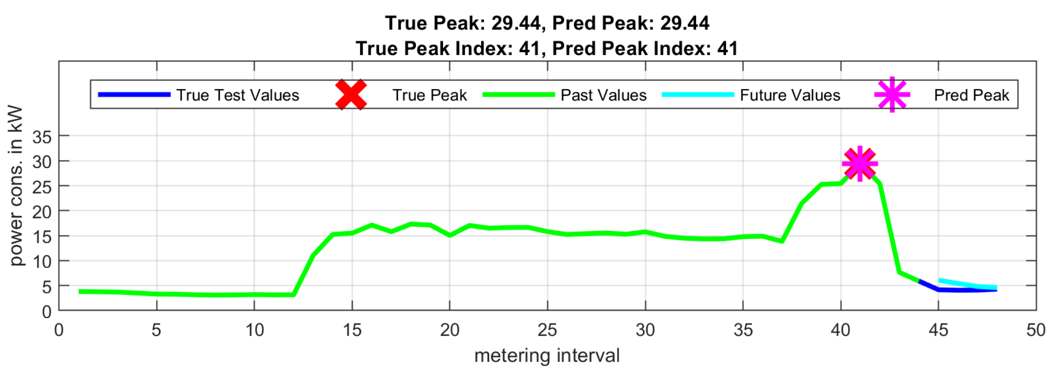
(b) $\tau = 15$



(c) $\tau = 25$



(d) $\tau = 35$



(e) $\tau = 45$

Figure 3.4. User 25's predicted power consumption profile on Wednesday of week 65 using offline average model

The metric calculations for every model discussed in this chapter are based on three objectives (a) how good the model is predicting the power consumption values for the upcoming metering intervals, (b) how good the model is predicting the peak values, and (c) how good the model is predicting the metering interval corresponding to the peak power consumption value. The metric calculations corresponding to each objective is termed as the Mean Absolute Percentage Error (total), Mean Absolute Percentage Error (peak), and Mean Absolute Error (peaktime), where the first two are expressed in % whereas the last one is expressed in minutes. The notation used for these three metric values are $\mathrm{MAPE}_k^{\mathrm{total}}(d)$, $\mathrm{MAPE}_k^{\mathrm{peak}}(d)$, and $\mathrm{MAE}_k^{\mathrm{peaktime}}(d)$ respectively. Mathematically, these values can be expressed as

$$\mathrm{MAPE}_k^{\mathrm{total}}(d)(\text{in }\%) = \left(\frac{2 \times 100}{|\mathcal{W}^{\mathrm{test}}| \cdot T^2 \cdot (T+1)}\right) \sum_{\mathrm{j} \in \mathcal{W}^{\mathrm{test}}} \sum_{\tau=1}^{T} \sum_{t=\tau}^{T} \frac{|p_k(t,d,\mathrm{j}) - \hat{p}_k(\tau,t,d,\mathrm{j})|}{p_k(t,d,\mathrm{j})}, \quad (3.2)$$

$$\mathrm{MAPE}_k^{\mathrm{peak}}(d)(\text{in }\%) = \left(\frac{1 \times 100}{|\mathcal{W}^{\mathrm{test}}| \cdot T}\right) \sum_{\mathrm{j} \in \mathcal{W}^{\mathrm{test}}} \sum_{\tau=1}^{T} \frac{|p_k^{\mathrm{peak}}(d,\mathrm{j}) - \hat{p}_k^{\mathrm{peak}}(\tau,d,\mathrm{j})|}{p_k(d,\mathrm{j})}, \quad (3.3)$$

$$\mathrm{MAE}_k^{\mathrm{peaktime}}(d)(\text{in mins}) = \left(\frac{\Delta}{|\mathcal{W}^{\mathrm{test}}| \cdot T}\right) \sum_{\mathrm{j} \in \mathcal{W}^{\mathrm{test}}} \sum_{\tau=1}^{T} |t_k^{\mathrm{peak}}(d,\mathrm{j}) - \hat{t}_k^{\mathrm{peak}}(\tau,d,\mathrm{j})|, \quad (3.4)$$

where $|\mathcal{W}^{\mathrm{test}}|$ represents the number of elements present in the test set, $p_k^{\mathrm{peak}}(d,\mathrm{j})$ represents the actual peak power consumption value, $\hat{p}_k^{\mathrm{peak}}(d,\mathrm{j})$ represents the predicted peak power consumption value, $t_k^{\mathrm{peak}}(d,\mathrm{j})$ represents the metering interval corresponding to the actual peak power consumption value and $\hat{t}_k^{\mathrm{peak}}(d,\mathrm{j})$ represents the metering interval corresponding to the predicted peak power consumption value. Mathematically, these parameters can be given as

$$p_k^{\mathrm{peak}}(d,\mathrm{j}) = \max_{t \in \mathcal{T}} p_k(t,d,\mathrm{j}), \quad (3.5)$$

$$\hat{p}_k^{\mathrm{peak}}(\tau,d,\mathrm{j}) = \max_{t \in \mathcal{T}} \hat{p}_k(\tau,t,d,\mathrm{j}), \quad (3.6)$$

$$t_k^{\text{peak}}(d, \mathrm{j}) = \arg\max_{t \in \mathcal{T}} p_k(t, d, \mathrm{j}), \tag{3.7}$$

$$\hat{t}_k^{\text{peak}}(\tau, d, \mathrm{j}) = \arg\max_{t \in \mathcal{T}} \hat{p}_k(\tau, t, d, \mathrm{j}). \tag{3.8}$$

Using the equations mentioned in 3.2 - 3.4, the metric values obtained for the case study using the offline average model are given in table 3.1.

Table 3.1. MAPE values using offline average model

Metric values	**User 19**	**User 25**
$\text{MAPE}_k^{\text{total}}(d)$	23.52 %	50.10 %
$\text{MAPE}_k^{\text{peak}}(d)$	11.39 %	20.34 %
$\text{MAE}_k^{\text{peaktime}}(d)$	305.22 mins ≈10.17 intervals	400.94 mins ≈13.36 intervals

Offline average model can be used as a base model which can be used to compare against all the other models that will be discussed in the forthcoming subsections.

3.2.2 Online Average Model

In online average model, the idea is to include the precedent test weeks to the training set in order to calculate the average power consumption values corresponding to respective upcoming metering intervals on different days of the week for different users. Mathematically, it can be expressed as

$$\hat{p}_k(\tau, t, d, \mathrm{j}) = \begin{cases} p_k(t, d, \mathrm{j}), & \forall t < \tau, \mathrm{j} \in \mathcal{W}^{\text{test}}, \\ \frac{1}{W^{\text{train}}} \sum\limits_{w=1}^{W^{\text{train}}} p_k(t, d, w), & \forall t \geq \tau, \mathrm{j} = W^{\text{train}} + 1, \\ \frac{1}{\mathrm{j}-1} \sum\limits_{w=1}^{\mathrm{j}-1} p_k(t, d, w), & \forall t \geq \tau, \forall \mathrm{j} > W^{\text{train}} + 1. \end{cases} \tag{3.9}$$

Using the equations mentioned in 3.2 - 3.4, the metric values obtained by the online average model for the case study are given in table 3.2.

Table 3.2. MAPE values using online average model

Metric values	User 19	User 25
$\text{MAPE}_k^{\text{total}}(d)$	23.10 %	50.78 %
$\text{MAPE}_k^{\text{peak}}(d)$	11.36 %	19.94 %
$\text{MAE}_k^{\text{peaktime}}(d)$	304.47 mins ≈10 intervals	316.62 mins ≈10 intervals

By comparing the metric values of the offline and online average models, it can be seen that there is a modest improvement on the total prediction of the power consumption values for upcoming metering intervals and on the peak power consumption values for both users. However, there is a drastic improvement in predicting the metering interval corresponding to the peak power consumption value for user 25. The improvement of this model can be attributed to the strong correlation of the power consumption values of the preceding weeks which further helps the average values across all metering intervals to change accordingly.

3.2.3 First Order Auto Regression Model (Conventional Method)

Auto regression models are quite common for the studies of time series analysis. In general, an auto regressive model is when a value from a time series is regressed on previous values from that same time series. Specifically, a first order auto regressive model (or AR1) of a stationary time series y_t can be given as

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t. \tag{3.10}$$

But the time series for the most of the users in our dataset being non–stationary as discussed in the section 2.4, the idea is to regress the model for every metering interval. Thereby, the modified first order auto regression model used can be expressed as

$$p_k(t,d,w) \sim p_k(t-1,d,w), \qquad t > 1, \tag{3.11}$$

$$\hat{p}_k(t,d,w) = \alpha_k(t,d)p_k(t-1,d,w) + \beta_k(t,d), \qquad t > 1, \tag{3.12}$$

where $w \in \mathcal{W}^{\text{train}}$.

In order to calculate the model parameters $\alpha_k(t,d)$ and $\beta_k(t,d)$, a squared loss objective function formed for every metering interval can be used and minimized as shown in the following expression:

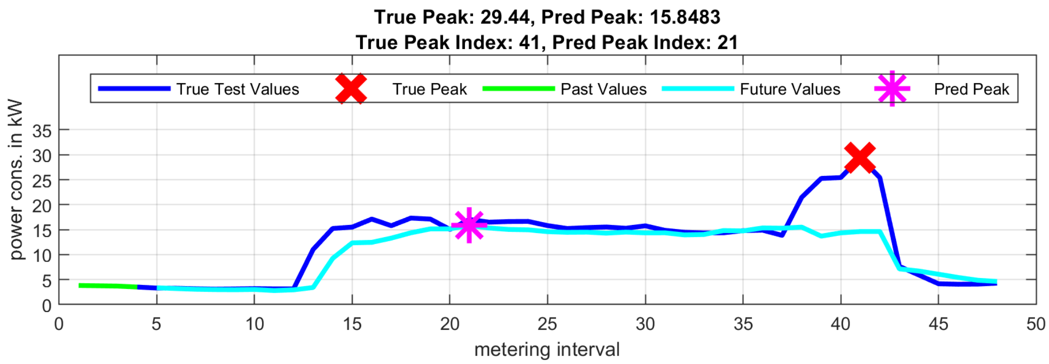
$$\min_{\alpha_k(t,d),\beta_k(t,d)} \sum_{w=1}^{W^{\text{train}}} \Big(p_k(t,d,w) - \hat{p}_k(t,d,w)\Big)^2.$$

Based on the model parameters obtained, the predicted power consumption values are

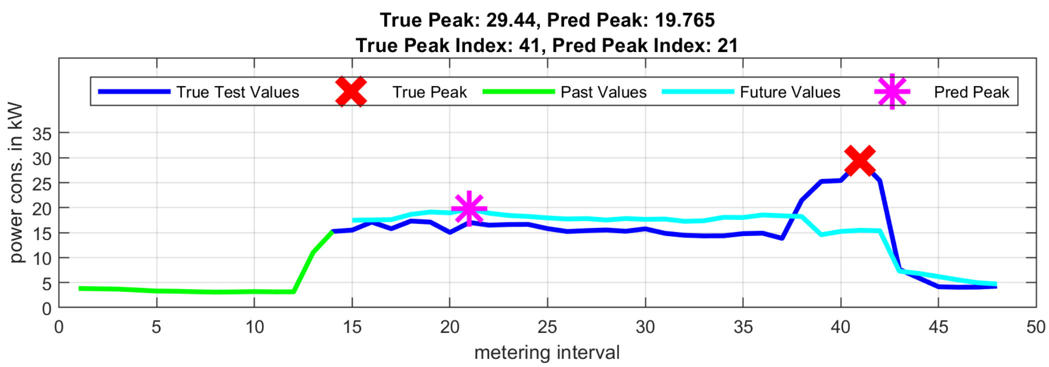
$$\hat{p}_k(\tau,t,d,\mathrm{j}) = \begin{cases} \frac{1}{W^{\text{train}}} \sum\limits_{w=1}^{W^{\text{train}}} p_k(t,d,w), & \tau = 1, \mathrm{j} = W^{\text{train}} + 1, \\ \frac{1}{\mathrm{j}-1} \sum\limits_{w=1}^{\mathrm{j}-1} p_k(t,d,w), & \tau = 1, \forall \mathrm{j} > W^{\text{train}} + 1, \\ p_k(t,d,\mathrm{j}), & \forall t < \tau, \forall \mathrm{j} \in \mathcal{W}^{\text{test}}, \\ \alpha_k(t,d)p_k(t-1,d,\mathrm{j}) + \beta_k(t,d), & \forall t \geq \tau, \forall \mathrm{j} \in \mathcal{W}^{\text{test}}. \end{cases} \tag{3.13}$$

Essentially, the above set of equation means that we predict the power consumption values using an online average model at the beginning of the day ($\tau = 1$) and thereby use the model parameters for predicting the power consumption values for the upcoming metering intervals based on the precedent power consumption values. It should also be noted that the precedent power consumption values of the metering interval at that metering interval are equal to their actual power consumption values.

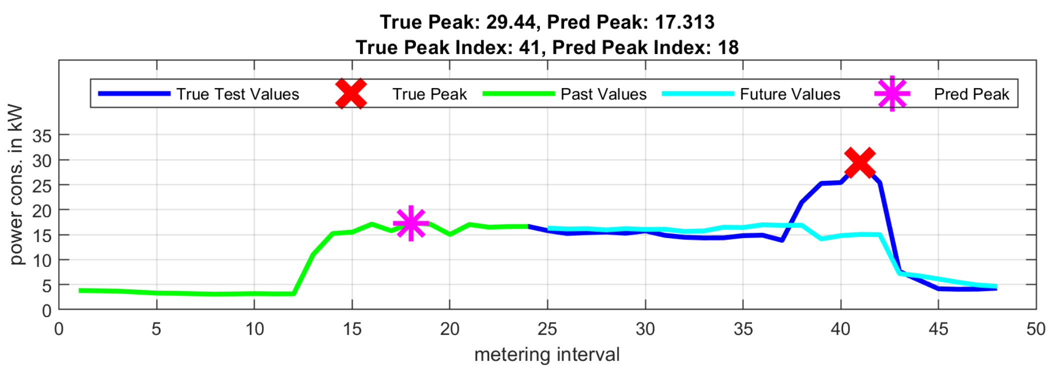
Figure 3.5 shows the same example of user 25 on Wednesday of week 65 as mentioned in the offline average model for different τ values for first order auto regression model. The main difference that can be observed by comparing the two models is the variation of the power consumption values for the upcoming metering intervals at different τ values. In offline average model, the power consumption values for the upcoming metering intervals do not change with different values of τ but in this model, the power consumption values for the upcoming metering intervals do change with different values of τ making this model more adaptive to the updated power consumption values within the day.



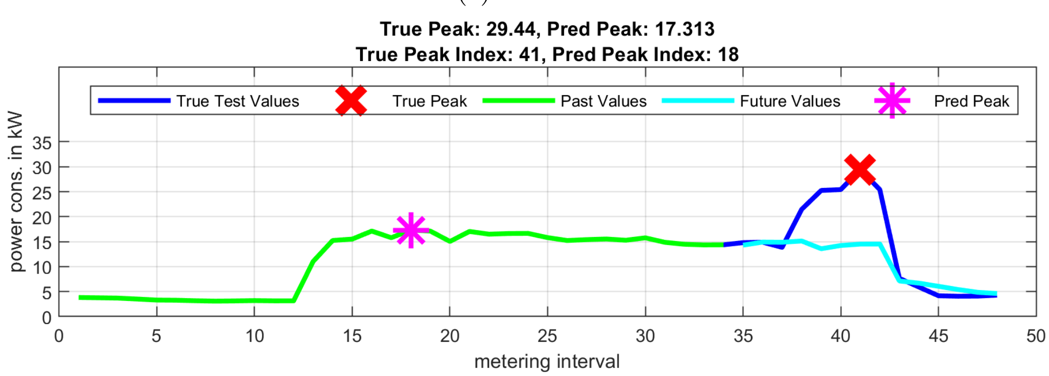
(a) $\tau = 5$



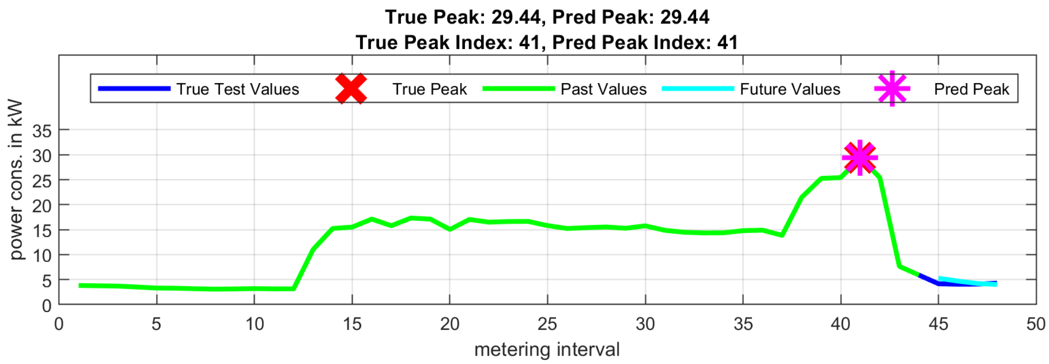
(b) $\tau = 15$



(c) $\tau = 25$



(d) $\tau = 35$



(e) $\tau = 45$

Figure 3.5. User 25's predicted power consumption profile on Wednesday of week 65 using first order auto regression model (conventional method)

Again, to verify how well the model is behaving, we can use the same metric calculations mentioned in eqns. 3.2 - 3.4. The metric values for the first order auto regression model are shown in the table 3.3.

Table 3.3. MAPE values using first order auto regression model (conventional method)

Metric values	**User 19**	**User 25**
$\text{MAPE}_k^{\text{total}}(d)$	22.37 %	40.82 %
$\text{MAPE}_k^{\text{peak}}(d)$	10.17 %	19.35 %
$\text{MAE}_k^{\text{peaktime}}(d)$	298.10 mins ≈9.93 intervals	306.72 mins ≈10.22 intervals

Comparing this model to the other two models, the values across all parameters are improved. The main improvement can be observed in the $\text{MAPE}_k^{\text{total}}(d)$ value of user 25. This means that the first order models predicts the power consumption values of the upcoming metering intervals with a better accuracy compared to the average models. Our main focus being the peak values, there is a slight improvement in the $\text{MAPE}_k^{\text{peak}}(d)$ values for both users.

3.2.4 First Order Auto Regression Model (Asymmetric Penalty Method)

In the previous model, first order auto regression model using a conventional method is discussed. Since the improvement on the peak values when compared to the average models was little, asymmetric penalty method of the same first order auto regression model is discussed in this section. The model being the same as that of the conventional method, it can be expressed as

$$p_k(t,d,w) \sim p_k(t-1,d,w), \qquad t > 1, \tag{3.14}$$

$$\hat{p}_k(t,d,w) = \alpha_k(t,d)p_k(t-1,d,w) + \beta_k(t,d), \qquad t > 1, \tag{3.15}$$

where $w \in \mathcal{W}^{\text{train}}$.

The main difference in between the conventional method and the asymmetric method lies in the way the model parameters $\alpha_k(t,d)$ and $\beta_k(t,d)$ are obtained. Instead of using a squared loss function, this method uses a dead band loss function with a hyper-parameter $\gamma_k(d)$ which is termed as the penalty. This loss function can be given as

$$\min_{\alpha_k(t,d),\beta_k(t,d)} \sum_{w=1}^{W^{\text{train}}} \Big(\hat{p}_k(t,d,w), p_k(t,d,w)\Big)_+ + \gamma_k(d)\Big(p_k(t,d,w), \hat{p}_k(t,d,w)\Big)_+,$$

where

$$(x,y)_+ = \max\{0,(x-y)\}. \tag{3.16}$$

In the above expression, $\Big(\hat{p}_k(t,d,w), p_k(t,d,w)\Big)_+$ represents the overestimating component whereas $\Big(p_k(t,d,w), \hat{p}_k(t,d,w)\Big)_+$ represents the underestimating component of the loss function. Depending on the values of the penalty used, minimizing the loss function will furnish us with model parameters which either overestimates or underestimates power consumption values for the upcoming metering intervals. By this we mean, if the value of the penalty is greater than 1 ($\gamma_k(d) > 1$), minimizing the loss function will provide us with the model parameters which overestimates the power consumption values for upcoming metering intervals whereas if the value of the penalty is between 0 and 1 ($0 < \gamma_k(d) < 1$) will correspond to the model parameters which underestimates the power consumption values for upcoming metering intervals. If the value of the penalty is equal to 1 ($\gamma_k(d) = 1$), the loss function will simply transform to the absolute loss function that can be expressed as

$$\min_{\alpha_k(t,d),\beta_k(t,d)} \sum_{w=1}^{W^{\text{train}}} |p_k(t,d,w) - \hat{p}_k(t,d,w)|.$$

To obtain the best value for the penalty $\gamma_k(d)$ that provide can us with the model parameters $\alpha_k(t,d)$ and $\beta_k(t,d)$ such that there can be an improvement in predicting the peak values, following four steps are used for this method.

Step 1: Split the training data into pure training set and cross validation set

First, the training set is further is divided into pure training set and the cross validation set. Symbolically, $p_k(t,d,w)$ with $w \in \mathcal{W}^{\text{pure}}$ represents the pure training set and $p_k(t,d,w)$ with $w \in \mathcal{W}^{\text{cv}}$ represents the cross validation set such that $\mathcal{W}^{\text{pure}} \cup \mathcal{W}^{\text{cv}} = \mathcal{W}^{\text{train}}$. It should be noted that the division to obtain the pure training set and cross validation set from the training set is random.

Step 2: Train the model using the pure training set for different values of the hyper-parameter $\gamma_k(d) \in \mathcal{G} = [g_{min}, g_{max}]$

After the training set is split into pure training set and cross validation set, first order auto regression model is built on the pure training set for different values of the hyper-parameter in the range $[g_{min}, g_{max}]$. This can be represented as

$$p_k(t,d,w) \sim p_k(t-1,d,w), \qquad t > 1, \tag{3.17}$$

$$\hat{p}_k(t,d,w) = \alpha_{k,\gamma_k(d)}(t,d)p_k(t-1,d,w) + \beta_{k,\gamma_k(d)}(t,d), \qquad t > 1, \tag{3.18}$$

where $w \in \mathcal{W}^{\text{pure}}$. The model parameters $\alpha_{k,\gamma_k(d)}(t,d)$ and $\beta_{k,\gamma_k(d)}(t,d)$ for different values of hyper-parameter $\gamma_k(d)$ can be obtained by minimizing the deadband loss function expressed as

$$\min_{\alpha_{k,\gamma_k(d)}(t,d),\beta_{k,\gamma_k(d)}(t,d)} \sum_{w=1}^{W^{\text{train}}} \Big(\hat{p}_k(t,d,w), p_k(t,d,w)\Big)_+ + \gamma_k(d)\Big(p_k(t,d,w), \hat{p}_k(t,d,w)\Big)_+.$$

Step 3: Based on the model parameters obtained, predict on cross validation set

After obtaining the model parameters for different values of the hyper-parameter, power consumption values can be predicted on the weeks corresponding to the cross validation set as follows

$$\hat{p}_k(\tau, t, d, c) = \begin{cases} \frac{1}{|\mathcal{W}^{\text{pure}}|} \sum_{w \in \mathcal{W}^{\text{pure}}} p_k(t, d, w), & \tau = 1, c \in \mathcal{W}^{\text{cv}}, \\ p_k(t, d, c), & \forall t < \tau, c \in \mathcal{W}^{\text{cv}}, \\ \alpha_{k,\gamma_k(d)}(t, d) p_k(t-1, d, c) + \beta_{k,\gamma_k(d)}(t, d), & \forall t \geq \tau, c \in \mathcal{W}^{\text{cv}}. \end{cases} \tag{3.19}$$

Predicting the power consumption values on the cross validation set is quite similar to that of the first order auto regression model using conventional method. The power consumption values for all the metering intervals at the beginning of the day i.e at $\tau = 1$ is given by the offline average model of the pure training set and the power consumption values for the upcoming metering intervals are given by the model parameters obtained for different values of hyper-parameter. Again, it should be noted that the precedent power consumption values of a metering interval at that metering interval can be given as the actual power consumption values.

Step 4: Using metric calculations, choose the best value of $\gamma_k(d)$

Based on the prediction on the cross validation set using different model parameters corresponding to different values of the hyper-parameter as mentioned in step 3, a metric calculation used for the peak values (similar to eqn 3.3) can be used for finding the best hyper-parameter value, which is shown in eqn. 3.20.

$$\text{MAPE}_k^{\text{peak}}(d)(\text{in } \%) = \left(\frac{1 \times 100}{|\mathcal{W}^{\text{cv}}| \cdot T} \right) \sum_{c \in \mathcal{W}^{\text{cv}}} \sum_{\tau=1}^{T} \frac{|p_k^{\text{peak}}(d, c) - \hat{p}_k^{\text{peak}}(\tau, d, c)|}{p_k(d, c)} \tag{3.20}$$

where $|\mathcal{W}^{\text{cv}}|$ represents the number of elements present in the cross validation set. The best value of the hyper-parameter is the one that corresponds to the lowest $\text{MAPE}_k^{\text{peak}}(d)$ on the cross validation set. This can be represented as

$$\gamma_k^{best}(d) = \arg \min_{\gamma_k(d) \in \mathcal{G}} \text{MAPE}_{k,\gamma_k(d)}^{\text{peak}}(d). \tag{3.21}$$

The values of the hyper-parameter $\gamma_k(d)$ explored for the case study involving user 19 and user 25 on Wednesdays are in the interval of [0,4] which resulted in the $\text{MAPE}_k^{\text{peak}}(d)$ values on the cross validation set for different values of $\gamma_k(d)$ shown in figure 3.6.

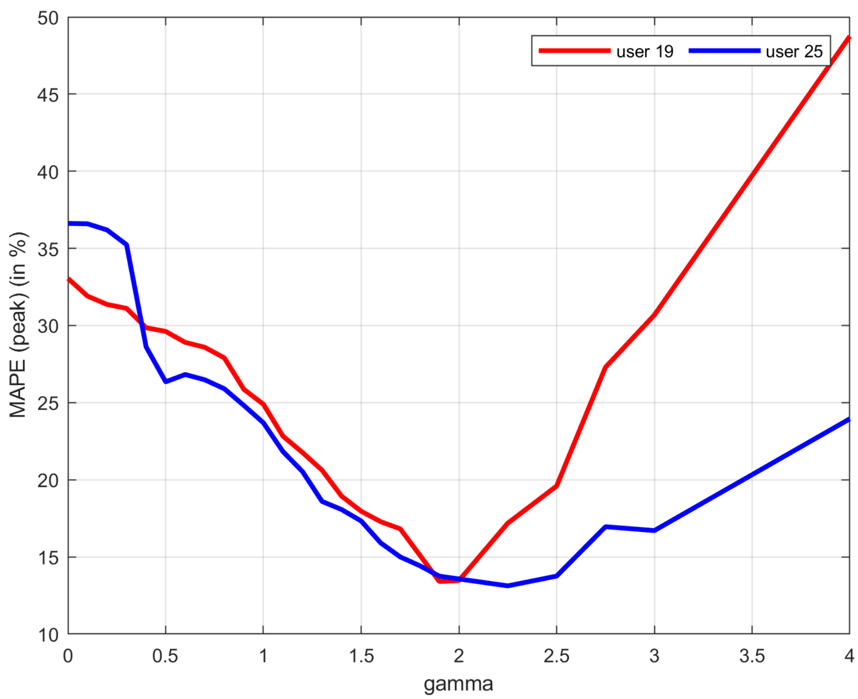


Figure 3.6. Variation of MAPE (peak) values on the cross validation set with respect to different values of the hyper-parameter for user 19 and user 25 on Wednesdays

The values of the hyper-parameter corresponding to the lowest MAPE (peak) values on the cross validation set for user 19 and user 25 on Wednesdays are 1.9 and 2.25 respectively. These values being greater than 1 suggest overestimation for predicting the peak values with a better accuracy.

Finally, the model parameters corresponding to the best value of the hyper-parameter can be used for predicting on the test set which can be given as

$$\hat{p}_k(\tau, t, d, \mathrm{j}) = \begin{cases} \frac{1}{W^{\text{train}}} \sum\limits_{w=1}^{W^{\text{train}}} p_k(t, d, w), & \tau = 1, \mathrm{j} = W^{\text{train}} + 1, \\ \frac{1}{\mathrm{j}-1} \sum\limits_{w=1}^{\mathrm{j}-1} p_k(t, d, w), & \tau = 1, \forall \mathrm{j} \geq W^{\text{train}} + 1, \\ p_k(t, d, \mathrm{j}), & \forall t < \tau, \forall \mathrm{j} \in \mathcal{W}^{\text{test}}, \\ \alpha_{k,\gamma_k^{best}(d)}(t, d) p_k(t-1, d, \mathrm{j}) + \beta_{k,\gamma_k^{best}(d)}(t, d), & \forall t \geq \tau, \forall \mathrm{j} \in \mathcal{W}^{\text{test}}. \end{cases} \tag{3.22}$$

Using the best value of the hyper-parameter, the metric values for both users using this method is shown in table 3.4.

Table 3.4. MAPE values using first order auto regression model (asymmetric penalty method)

Metric values	**User 19**	**User 25**
$\text{MAPE}_k^{\text{total}}(d)$	33.98 %	62.58 %
$\text{MAPE}_k^{\text{peak}}(d)$	9.83 %	13.44 %
$\text{MAE}_k^{\text{peaktime}}(d)$	291.41 mins ≈9.71 intervals	155.5 mins ≈5.18 intervals

It can be seen that the error on the peak values ($\text{MAPE}_k^{\text{peak}}(d)$) do have a significant improvement compared to all the other three methods. However, this improvement comes with a larger error on the total prediction ($\text{MAPE}_k^{\text{total}}(d)$) of the power consumption values. Also, there is a drastic improvement in the prediction of the metering interval corresponding to the peak power consumption value using this model.

3.2.5 Artificial Neural Network

Artificial Neural Networks are more advanced compared to the models previously discussed. ANNs are comprised of a node layers, containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, connects to another and has an associated weight and threshold. A double layered artificial neural network is shown in the figure 3.7.

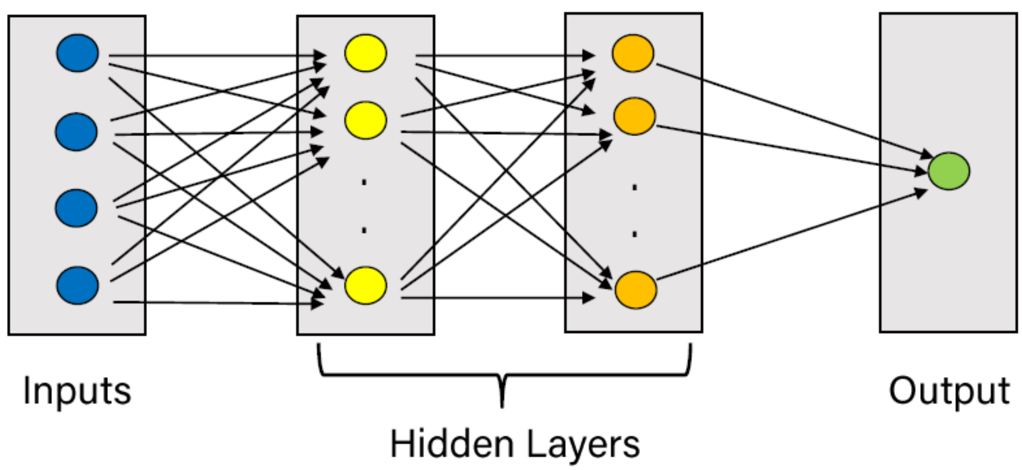
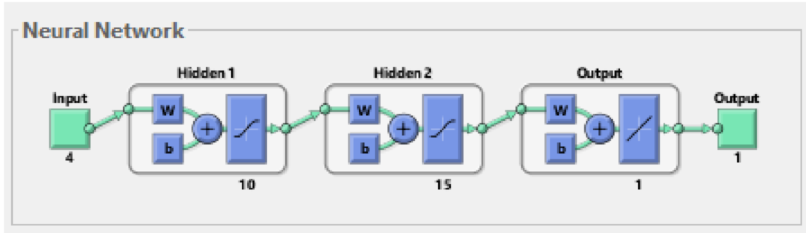


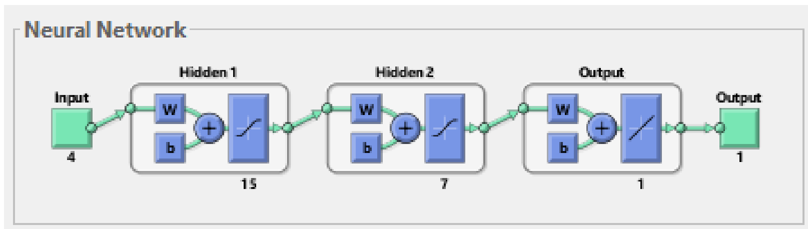
Figure 3.7. Double layered feedforward artificial neural network

A network similar to the figure 3.7 is used for our case study. The four training inputs fed to the network are the metering interval t which is defined a categorical variable, the power consumption value at the previous metering interval $p_k(t-1,d,w)$, power consumption value of the same metering interval on the previous day $p_k(t,d-1,w)$, and the power consumption value of the same metering interval on the previous week $p_k(t,d,w-1)$. The training output is the actual power consumption value on the metering interval $p_k(t,d,w)$, where $t \in \mathcal{T} \setminus \{1\}$ and $w \in \mathcal{W}^{\text{train}} \setminus \{1,2\}$.

The double layered artificial neural network can shown in figure can be build in MATLAB using *'feedforwardnet'* function which is shown in figure 3.8 for for the case study consisting user 19 and user 25 on Wednesdays.



(a) User '19'



(b) User '25'

Figure 3.8. Double layered artificial neural network used for predicting the peak power consumption values for user '19' and user '25' on Wednesday using indirect approach

Based on the trained network, the metric values for the case study are given in table 3.5.

Table 3.5. MAPE values using double layered artificial neural network

Metric values	**User 19**	**User 25**
$\text{MAPE}_k^{\text{total}}(d)$	19.90 %	37.43 %
$\text{MAPE}_k^{\text{peak}}(d)$	9.65 %	12.25 %
$\text{MAE}_k^{\text{peaktime}}(d)$	283.72 mins ≈9.46 intervals	150.83 mins ≈5 intervals

The above table suggests that with the right number of layers and the number of neurons in each layer, the performance of ANN can be better compared to other models as the errors corresponding to the entire power consumption values, peak values and the metering intervals corresponding to the peak values are lesser compared to that of other models.

3.3 Summary

The highlights of this chapter are the following:

1. The two-step process of predicting the power consumption values for the entire time series and then identifying the peak power consumption value can be used for learning the peaks for different users on different days of the week.

2. Highlights on the case study of user '19' and user '25' on Wednesdays.

    - All models (offline average model, online average model, first order auto regression model using conventional method, fist order auto regression model using asymmetric penalty method, and artificial neural network) provide a decent prediction on the peak values as the mean errors are less than 20%.
    - Online average model performs better compared to that of the offline average model.
    - First order auto regression models work better compared to the average models.
    - Overestimating the power consumption values at every metering interval helps in the better prediction on the peak values but affects the prediction on the entire time series.
    - Artificial neural network using the appropriate features can be helpful for complex relationships between the inputs and the output.

4. MACHINE LEARNING FOR PEAKS DATA

In the previous chapter, forecasting the peak power consumption values of user '19' (belonging to cluster 1) and user 25 (belonging to cluster 5) on Wednesdays were discussed using the indirect approach. In this chapter, direct approach for forecasting the peak power consumption values of the same users on Wednesdays is discussed.

4.1 Actual data pertaining to case study

Before discussing the machine learning algorithms used to forecast the peak values directly, it is important to visualize the input data which in our case study is the peak power consumption values of user '19' and user '25' on Wednesdays across different weeks. The same is plotted in figure 4.1. It can be observed that the actual peak power consumption values are not steady, rather fluctuate a lot around their corresponding mean peak values shown by the dashed lines (**- -**) which have the values close to 10kW and 20kW for user '19' and for user '25' respectively.

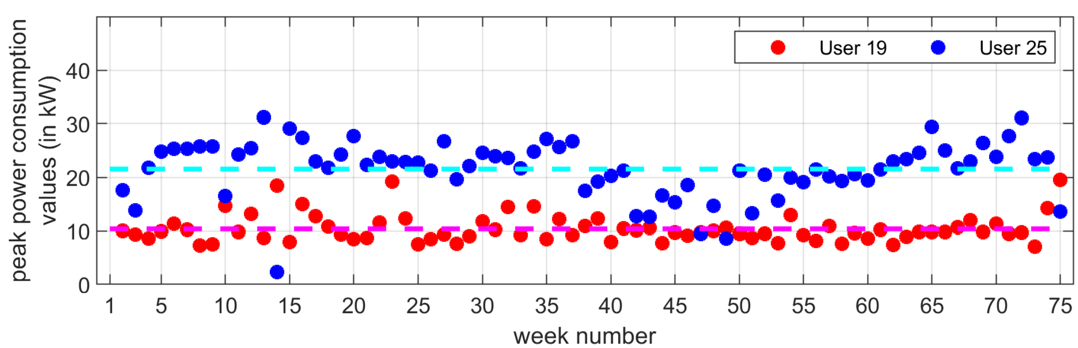
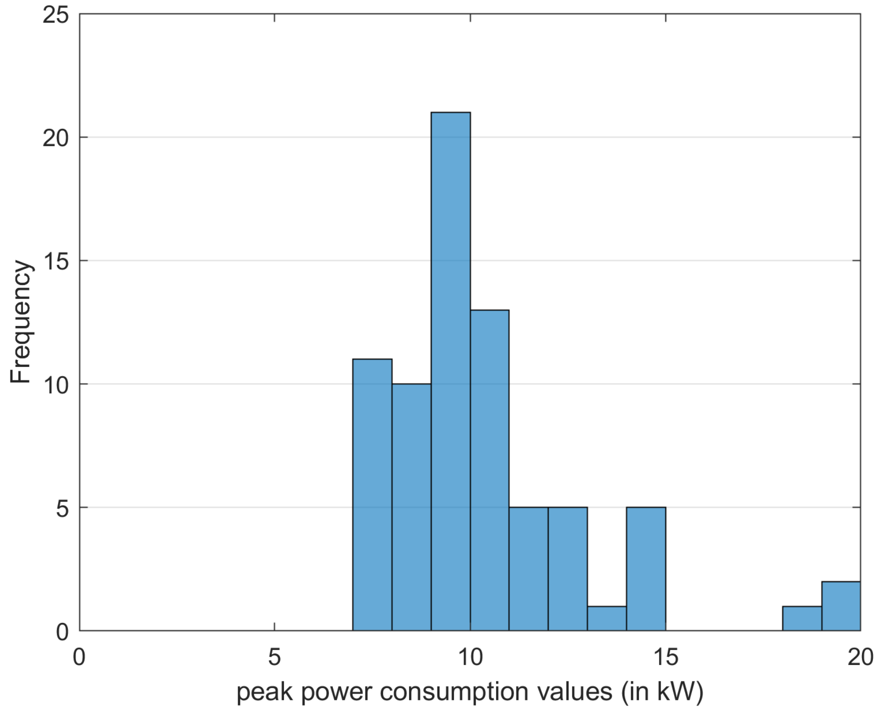
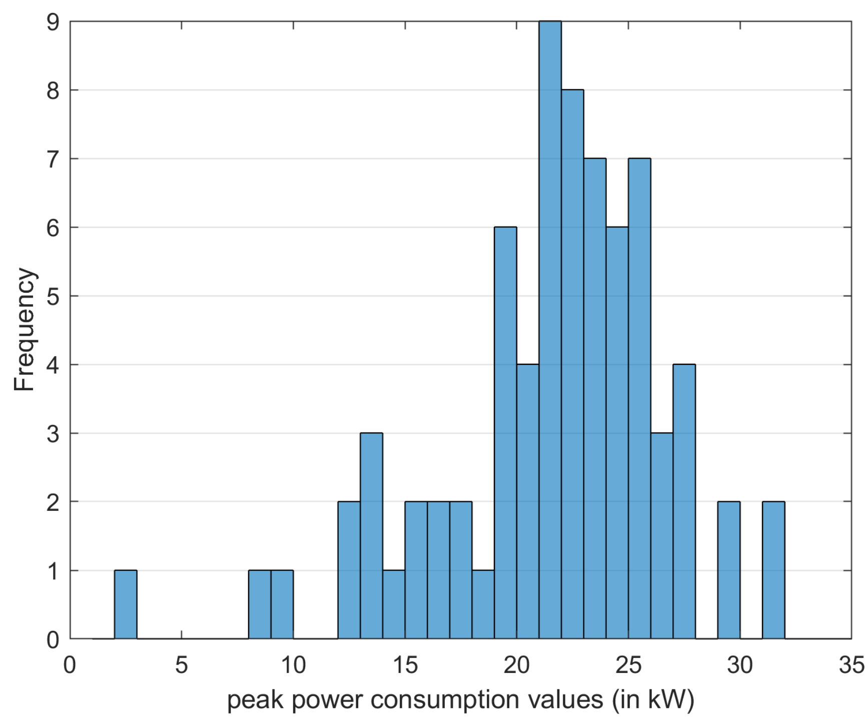


Figure 4.1. Peak values of user '19' and user '25' on Wednesdays

To perceive the above data numerically, histogram of the peak power consumption values is shown in figure 4.2. About 30% (21 out of 75) of the peak values for user '19' are in the range of 9kW to 10kW and the rest of 70% values are widely scattered in the range of 7kW to 15kW. The spread of the peak values for user '25' is comparatively larger than that of the user '19'. Merely, 10% (9 out of 75) of the peak values lie between 21kW to 22kW. Most of the other peak values are fall in the range of 19kW to 28kW.

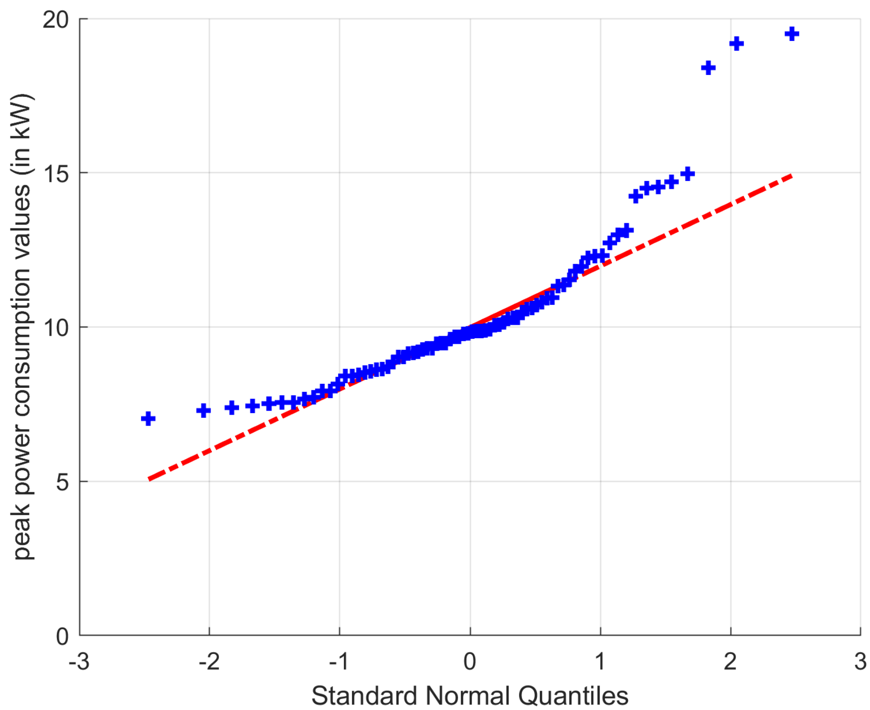


(a) User '19'

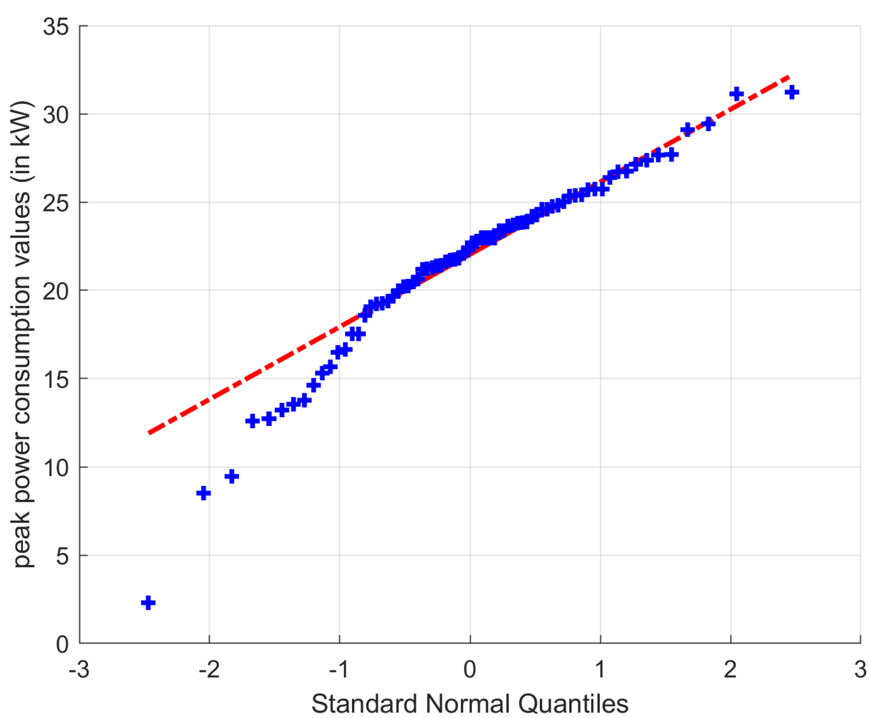


(b) User '25'

Figure 4.2. Histogram of peak values of user '19' and user '25' on Wednesdays



(a) User '19'



(b) User '25'

Figure 4.3. Comparison of peak values of user '19' and user '25' on Wednesdays against normal distribution

The figure 4.3 represents the comparison of the peak power consumption values of user '19' and user '25' against the standard normal distribution. Looking at these figures, it can be claimed that the peak values do not follow normal or Gaussian distribution. The exact probability distribution functions of these peak values are shown in figure 4.4. It can be inferred that these distribution are heavy tailed i.e. tails are not exponentially bounded. Theoretically, heavy tailed distribution are classified into three classes: the fat-tailed distributions, the long-tailed distributions and the sub-exponential distributions. It can be seen that the peak values of user '19' belong to the sub-exponential class whereas the peak values of user '25' are of the long-tailed class.

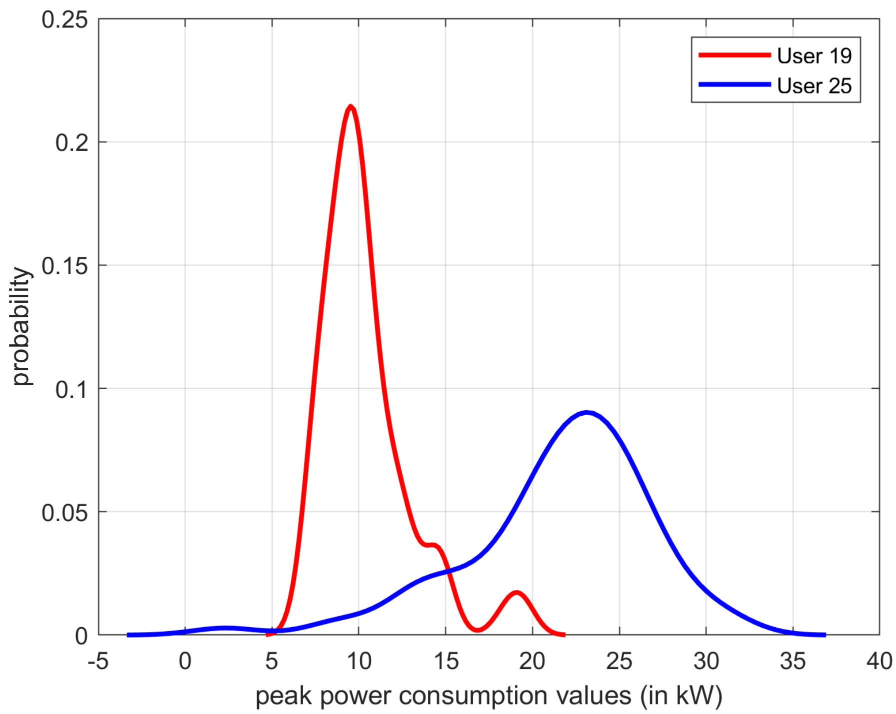


Figure 4.4. Probability Distribution Functions (PDF) of peak values for user '19' and user '25' on Wednesdays

Forecasting the values which follow heavy tailed distribution is not trivial. Essentially, it means that the dataset has a significant number of outliers that requires the application of advanced machine learning techniques to capture them using different features for better prediction.

4.2 Machine Learning Models

In this section, different machine learning models used to forecast the peak power consumption values directly is discussed. The study is performed on the same typical users i.e. user '19' from cluster 1 and user '25' from cluster 5 on Wednesdays. Figure 4.5 shows the values corresponding to their training and test sets. For each user, the first 55 Wednesdays have been considered as the training set whereas the last 20 Wednesdays as the test set.

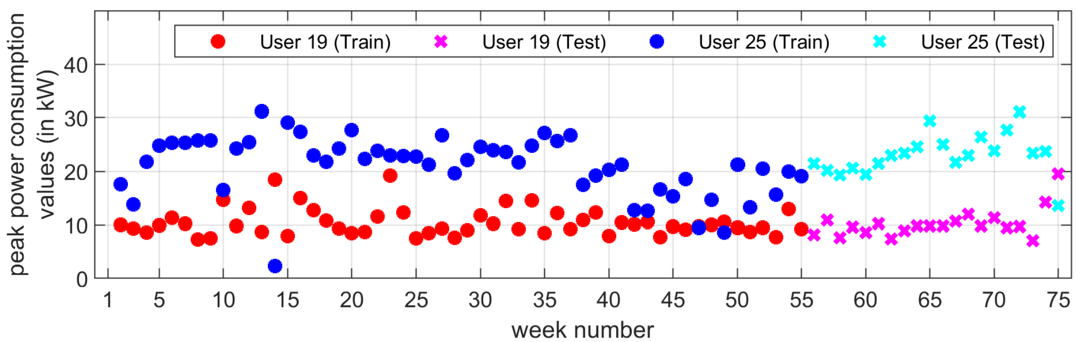


Figure 4.5. Peak values forming training and test sets for user '19' and user '25' on Wednesdays

The symbolic representation of the entire training set consisting of the peak power consumption values of K users on all days of the week for W^{train} weeks is given as $p_k^{\text{peak}}(d, w)$ with $k \in \mathcal{K}$, $d \in \mathcal{D}$, and $w \in \mathcal{W}^{\text{train}}$ whereas the representation of the entire test set consisting of the peak power consumption values for the same number of users on all days of the week for W^{test} weeks is given as $p_k^{\text{peak}}(d, \text{j})$ with $k \in \mathcal{K}$, $d \in \mathcal{D}$, and $\text{j} \in \mathcal{W}^{\text{test}}$ where

$$\mathcal{K} = \{1, 2, ..., K\}, \text{ with } K = 75,$$
$$\mathcal{D} = \{1, 2, 3, 4, 5, 6, 7\},$$
$$\mathcal{W}^{\text{train}} = \{1, 2, ..., W^{\text{train}}\}, \text{ with } W^{\text{train}} = 55,$$
$$\mathcal{W}^{\text{test}} = \{56, 57, ..., W^{\text{test}}\}, \text{ with } W^{\text{test}} = 75.$$

For the case study, $k = \{19, 25\}$, $d = \{3\}$ with all other variables as mentioned above unless otherwise stated.

4.2.1 Offline Average Model

One of the most basic models to start the discussion with on forecasting the peaks directly is by taking the average of all the values pertaining to the training set. This model is referred as the offline average model. Mathematically, the predicted peak power consumption values using this model can be expressed as

$$\hat{p}_k^{\text{peak}}(d, \mathrm{j}) = \frac{1}{W^{\text{train}}} \sum_{w=1}^{W^{\text{train}}} p_k^{\text{peak}}(d, w); \forall \mathrm{j} \in \mathcal{W}^{\text{test}}. \tag{4.1}$$

The metric calculations for the models used in the process of directly predicting the peaks can be expressed as

$$\text{MAPE}_k^{\text{peak}}(d)(\text{in } \%) = \frac{1 \times 100}{|\mathcal{W}^{\text{test}}|} \sum_{\mathrm{j} \in \mathcal{W}^{\text{test}}} \frac{|p_k^{\text{peak}}(d, \mathrm{j}) - \hat{p}_k^{\text{peak}}(d, \mathrm{j})|}{p_k^{\text{peak}}(d, \mathrm{j})}, \tag{4.2}$$

where $|\mathcal{W}^{\text{test}}|$ represents the number of the elements in the test set.

With the offline average as the model, MAPE values obtained corresponding to the case study are given in table 4.1.

Table 4.1. MAPE values using offline average model

Model	**User 19**	**User 25**
Offline Average	17.39%	14.16%

4.2.2 Online Average Model

With a slight modification in the way the average is calculated, the results can be improved by appending the test weeks which would have already occured to the training set and then calculating the average of the appended training data. This model is referred as the online average model. Mathematically, the model can be expressed as

$$\hat{p}_k^{\text{peak}}(d, \text{j}) = \begin{cases} \frac{1}{W^{\text{train}}} \sum\limits_{w=1}^{W^{\text{train}}} p_k^{\text{peak}}(d, w), & \text{j} = W^{\text{train}} + 1, \\ \frac{1}{\text{j}-1} \sum\limits_{w=1}^{\text{j}-1} p_k^{\text{peak}}(d, w), & \forall \text{j} \geq W^{\text{test}} + 1. \end{cases} \tag{4.3}$$

With the online average as the model, MAPE values obtained using the eqn. 4.6 corresponding to the case study are given in table 4.2.

Table 4.2. MAPE values of online average model

Model	User 19	User 25
Online Average	16.50%	13.49%

By comparing the MAPE values from table 4.1 and 4.2, it can be noticed that the online average model performs better as compared to the offline average as the MAPE values are smaller for online average model. This is because the peak power consumption values for any user on different days of the week are highly correlated to their immediate previous values as compared to the values that would have been recorded long back. With this idea, the next models that are going to be discussed will contain the most related features, so that the future values can be forecasting with a better accuracy.

4.2.3 Simple Linear Regression (SLR) Model using previous week's peak power consumption value as the feature

One of the most common techniques used for forecasting is simple linear regression. The idea in this technique is to find a single exploratory variable (or feature as otherwise known) which has a linear relationship with the response variable. From online average, it is realised that using the immediate previous week's peak power consumption value definitely have the chance in improving the forecasting. But using previous week's power consumption value as a feature directly will not help as it would essentially mean that the upcoming values will always be either increasing or decreasing depending on the slope obtained using this model. Hence, some kind of transformation is needed. In the model shown below, the value online average of the previous weeks' peak power consumption value is subtracted from the previous week's peak power consumption value for the same day of the week and is modelled using simple linear regression. Mathematically, it can be expressed as

$$p_k^{\text{peak}}(d,w) - \bar{p}_k^{\text{peak}}(d,w) \sim p_k^{\text{peak}}(d,w-1) - \bar{p}_k^{\text{peak}}(d,w-1), \tag{4.4}$$

$$\hat{p}_k^{\text{peak}}(d,w) - \bar{p}_k^{\text{peak}}(d,w) = \alpha_k(d) p_k^{\text{peak}}(d,w-1) - \bar{p}_k^{\text{peak}}(d,w-1) + \beta_k(d), \tag{4.5}$$

where $p_k^{\text{peak}}(d,w)$ is the true peak power consumption value present in the training set, $\hat{p}_k^{\text{peak}}(d,w)$ is the predicted peak power consumption value corresponding to the true value present in the training set, $\bar{p}_k^{\text{peak}}(d,w)$ is the online average component which can be given as

$$\bar{p}_k^{\text{peak}}(d,w) = \left(\frac{1}{w-1}\right)\left\{\sum_{m=1}^{w-1} p_k^{\text{peak}}(d,m)\right\}. \tag{4.6}$$

For the model to be properly defined, the peak power consumption values corresponding to the training week 1 and 2 should not be considered while training the model i.e. $w \in \mathcal{W}^{\text{train}} \setminus \{1,2\}$.

The model parameters $\alpha_k(d)$ and $\beta_k(d)$ can be found by minimizing the total loss function given by the following expression:

$$\min_{\alpha_k(d),\beta_k(d)} \sum_{w=3}^{W^{\text{train}}} \left(p_k^{\text{peak}}(d,w) - \hat{p}_k^{\text{peak}}(d,w)\right)^2. \tag{4.7}$$

Based on the parameters obtained for the model, the predicted peak power consumption values on the test set can be expressed as

$$\hat{p}_k^{\text{peak}}(d, \text{j}) = \alpha_k(d) p_k^{\text{peak}}(d, \text{j}-1) - \bar{p}_k^{\text{peak}}(d, \text{j}-1) + \beta_k(d) + \bar{p}_k^{\text{peak}}(d, \text{j}), \forall \text{j} \in \mathcal{W}^{\text{test}}. \tag{4.8}$$

With the SLR model with previous week's peak power consumption values as the feature, MAPE values obtained corresponding to the case study are given in table 4.3.

Table 4.3. MAPE values of SLR model using previous week's peak power consumption value

Model	User 19	User 25
SLR using previous week's peak value	15.75%	12.91%

4.2.4 Simple Linear Regression (SLR) Model using previous day's peak power consumption value as the feature

When compared to the online average model, SLR model with previous week's peak power consumption value as a feature improves the prediction for the future values. Continuing the exploration with an idea similar to using the previous week's peak power consumption value, this time previous day's peak power consumption value can be used as a feature. But it should be noted that the users belonging to two dominating clusters, cluster 1 and cluster 5 have the different power consumption profiles (refer figure 2.8). So, the models for the users belonging to cluster 1 and cluster 5 will be different although the feature used will be the same (previous day's peak power consumption value) which are given in the following set of equations.

Case 1: For users belonging to cluster 1

Subcase 1: When $d \in \{2, 3, 4, 5, 6, 7\}$,

$$p_k^{\text{peak}}(d, w) - \bar{p}_k^{\text{peak}}(d, w) \sim p_k^{\text{peak}}(d-1, w) - \bar{p}_k^{\text{peak}}(d-1, w), \tag{4.9}$$

$$\hat{p}_k^{\text{peak}}(d, w) - \bar{p}_k^{\text{peak}}(d, w) = \alpha_k(d) p_k^{\text{peak}}(d-1, w) - \bar{p}_k^{\text{peak}}(d-1, w) + \beta_k(d). \tag{4.10}$$

Subcase 2: When $d = 1$,

$$p_k^{\text{peak}}(d, w) - \bar{p}_k^{\text{peak}}(d, w) \sim p_k^{\text{peak}}(7, w-1) - \bar{p}_k^{\text{peak}}(7, w-1), \tag{4.11}$$

$$\hat{p}_k^{\text{peak}}(d, w) - \bar{p}_k^{\text{peak}}(d, w) = \alpha_k(d) p_k^{\text{peak}}(7, w-1) - \bar{p}_k^{\text{peak}}(7, w-1) + \beta_k(d). \tag{4.12}$$

Case 2: For users belonging to cluster 5

Subcase 1: When $d \in \{3, 4, 5, 6\}$,

$$p_k^{\text{peak}}(d, w) - \bar{p}_k^{\text{peak}}(d, w) \sim p_k^{\text{peak}}(d-1, w) - \bar{p}_k^{\text{peak}}(d-1, w), \tag{4.13}$$

$$\hat{p}_k^{\text{peak}}(d, w) - \bar{p}_k^{\text{peak}}(d, w) = \alpha_k(d) p_k^{\text{peak}}(d-1, w) - \bar{p}_k^{\text{peak}}(d-1, w) + \beta_k(d). \tag{4.14}$$

Subcase 2: When $d = 2$,

$$p_k^{\text{peak}}(d, w) - \bar{p}_k^{\text{peak}}(d, w) \sim p_k^{\text{peak}}(6, w-1) - \bar{p}_k^{\text{peak}}(6, w-1), \tag{4.15}$$

$$\hat{p}_k^{\text{peak}}(d, w) - \bar{p}_k^{\text{peak}}(d, w) = \alpha_k(d) p_k^{\text{peak}}(6, w-1) - \bar{p}_k^{\text{peak}}(6, w-1) + \beta_k(d). \tag{4.16}$$

The users belonging to cluster 5 barely consume any power on Sundays and Mondays, so the online average model during these days of the week can furnish us with a good prediction. For the model to be properly defined, the peak power consumption values corresponding to the training week 1 and 2 should not be considered while training the model i.e. $w \in \mathcal{W}^{\text{train}} \setminus \{1, 2\}$.

The model parameters $\alpha_k(d)$ and $\beta_k(d)$ can be found by minimizing the total loss function given by the following expression:

$$\min_{\alpha_k(d), \beta_k(d)} \sum_{w=3}^{W^{\text{train}}} \left(p_k^{\text{peak}}(d, w) - \hat{p}_k^{\text{peak}}(d, w) \right)^2. \tag{4.17}$$

It should also be noted that the online average component is subtracted from the feature and from the response variable for the same reason stated for the previous model. Based on the parameters obtained for this model, the predicted peak power consumption values can be expressed as mentioned in the following equations for users belonging to different clusters.

Case 1: For users belonging to cluster 1

Subcase 1: $d \in \{2, 3, 4, 5, 6, 7\}$,

$$\hat{p}_k^{\text{peak}}(d, \text{j}) = \alpha_k(d) p_k^{\text{peak}}(d-1, \text{j}) - \bar{p}_k^{\text{peak}}(d-1, \text{j}) + \beta_k(d) + \bar{p}_k^{\text{peak}}(d, \text{j}), \forall \text{j} \in \mathcal{W}^{\text{test}}. \quad (4.18)$$

Subcase 2: $d = 1$,

$$\hat{p}_k^{\text{peak}}(d, \text{j}) = \alpha_k(d) p_k^{\text{peak}}(7, \text{j}-1) - \bar{p}_k^{\text{peak}}(7, \text{j}-1) + \beta_k(d) + \bar{p}_k^{\text{peak}}(d, \text{j}); \forall \text{j} \in \mathcal{W}^{\text{test}}. \quad (4.19)$$

Case 2: For users belonging to cluster 5

Subcase 1: When $d \in \{3, 4, 5, 6\}$,

$$\hat{p}_k^{\text{peak}}(d, \text{j}) = \alpha_k(d) p_k^{\text{peak}}(d-1, \text{j}) - \bar{p}_k^{\text{peak}}(d-1, \text{j}) + \beta_k(d) + \bar{p}_k^{\text{peak}}(d, \text{j}); \forall \text{j} \in \mathcal{W}^{\text{test}}. \quad (4.20)$$

Subcase 2: When $d = 2$,

$$\hat{p}_k^{\text{peak}}(d, \text{j}) = \alpha_k(d) p_k^{\text{peak}}(6, \text{j}-1) - \bar{p}_k^{\text{peak}}(6, \text{j}-1) + \beta_k(d) + \bar{p}_k^{\text{peak}}(d, \text{j}); \forall \text{j} \in \mathcal{W}^{\text{test}}. \quad (4.21)$$

With the SLR model using previous day's peak power consumption values as the feature, MAPE values obtained corresponding to the case study are given in table 4.4.

Table 4.4. MAPE values of SLR model using previous day's peak power consumption value

Model	**User 19**	**User 25**
SLR using previous day's peak value	14.95%	12.51%

Comparing with all the other three models, this model performs the best for our case study. However, as the number of features increases, there are better chances for explaining the response variable. So, continuing the exploration in the direction of multiple features, our next step is to discuss the models based on Multiple Linear Regression.

4.2.5 Multiple Linear Regression (MLR) Model using previous seven days' peak power consumption values as the features

Multiple Linear Regression is a widely used method for forecasting in which the number of predicting variables are more than one. In many cases, it is not possible to find a single variable on which the response variable many be dependent on. In these cases, combination of different variables can be useful in developing the model which can potential lead to better performance. For our case study, a model with previous seven days peak power consumption values without the respective online average components are considered as the features. The corresponding mathematical model can be expressed as

$$p_k^{\text{peak}}(d,w) - \bar{p}_k^{\text{peak}}(d,w) \sim \sum_{n=1}^{7} \left(p_k^{\text{peak}}(d_n,w_n) - \bar{p}_k^{\text{peak}}(d_n,w_n) \right), \tag{4.22}$$

$$\hat{p}_k^{\text{peak}}(d,w) - \bar{p}_k^{\text{peak}}(d,w) = \sum_{n=1}^{7} \alpha_{k,n}(d) \left(p_k^{\text{peak}}(d_n,w_n) - \bar{p}_k^{\text{peak}}(d_n,w_n) \right) + \beta_k(d). \tag{4.23}$$

where $n \in \mathcal{D} = \{1,2,3,4,5,6,7\}$, d_n and w_n can be given as

$$d_n = \begin{cases} d-n, & \text{if } d-n > 0, \\ 7+d-n, & \text{if } d-n \leq 0, \end{cases} \tag{4.24}$$

$$w_n = \begin{cases} w, & \text{if } d-n > 0, \\ w-1, & \text{if } d-n \leq 0. \end{cases} \tag{4.25}$$

For the model to be properly defined, the peak power consumption values corresponding to the training week 1 and 2 should not be considered while training the model i.e. $w \in \mathcal{W}^{\text{train}} \setminus \{1,2\}$.

The model parameters $\alpha_{k,n}(d)$ and $\beta_k(d)$ can be found out by minimizing the total loss function given by ther following expression:

$$\min_{\alpha_{k,n}(d),\beta_k(d);n\in\mathcal{D}} \sum_{w=3}^{W^{\text{train}}} \left(p_k^{\text{peak}}(d,w) - \hat{p}_k^{\text{peak}}(d,w) \right)^2. \tag{4.26}$$

Based on the model parameters, the prediction on the peak power consumption values pertaining the test set can be expressed as

$$\hat{p}_k^{\text{peak}}(d, \text{j}) = \sum_{n=1}^{7} \alpha_{k,n}(d)\left(p_k^{\text{peak}}(d_n, \text{j}_n) - \bar{p}_k^{\text{peak}}(d_n, \text{j}_n)\right) + \beta_k(d) + \bar{p}_k^{\text{peak}}(d, \text{j}) \forall \text{j} \in \mathcal{W}^{\text{test}} \quad (4.27)$$

where $n \in \mathcal{D} = \{1, 2, 3, 4, 5, 6, 7\}$, d_n and j_n can be given as

$$d_n = \begin{cases} d - n, & \text{if } d - n > 0, \\ 7 + d - n, & \text{if } d - n \leq 0, \end{cases} \quad (4.28)$$

$$\text{j}_n = \begin{cases} \text{j}, & \text{if } d - n > 0, \\ \text{j} - 1, & \text{if } d - n \leq 0. \end{cases} \quad (4.29)$$

With the MLR model using previous seven day's peak power consumption values as the features, MAPE values obtained corresponding to the case study are given in table 4.5.

Table 4.5. MAPE values of MLR model using previous seven days's peak power consumption values

Model	User 19	User 25
MLR using previous seven days peak power consumption values	13.21%	12.06%

In comparison to the previously discussed models, this model tops the list with the lowest MAPE values. This model has the potential to capture the short time trend of users' power consumption behaviour which is believed as the primary advantage when compared to the other models.

4.2.6 Artificial Neural Network

As discussed in the indirect approach, ANNs can be helpful in building more complex models. In all the previous models discussed, the input features were linearly regressed. But ANNs have the capability of self building nonlinear models usually termed as the activation functions present in each neuron of the hidden layers based on the features inputting to the network. A single layered feedforward artificial neural network is shown in the figure 4.6.

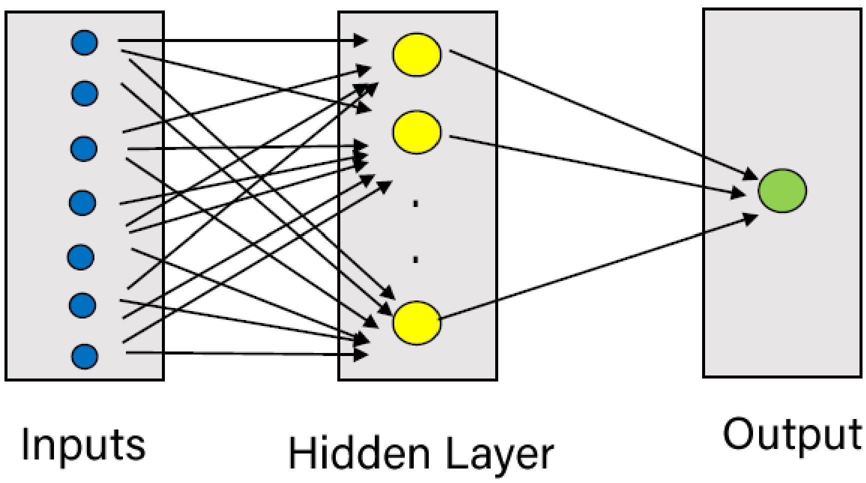


Figure 4.6. Single-layer feedforward artificial neural network

The network that is built for the case study takes the same seven features as discussed for the multiple linear regression. These are the previous seven days' peak power consumption values for each user. However, the input features supplied to the neural network do not have the component of online average. Mathematically, these features for every user on different days can be represented as $p_k^{\text{peak}}(d_n, w_n)$ where $n \in \mathcal{D} = \{1, 2, 3, 4, 5, 6, 7\}$, d_n and w_n can be given as

$$d_n = \begin{cases} d - n, & \text{if } d - n > 0, \\ 7 + d - n, & \text{if } d - n \leq 0, \end{cases} \tag{4.30}$$

$$w_n = \begin{cases} w, & \text{if } d - n > 0, \\ w - 1, & \text{if } d - n \leq 0. \end{cases} \tag{4.31}$$

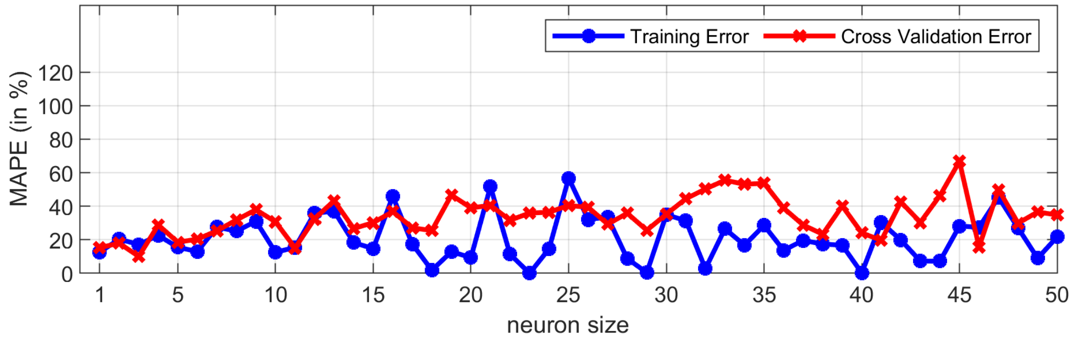
For having a well defined networks for all users on all days of the week, $w \in \mathcal{W}^{\text{train}} \setminus \{1\}$. Any network can be built using the inputs with any number of neuron presents in the hidden layers. But the down side on having too many neurons is over-fitting the network for the training set. This results in a bad performance on the test set. In order to overcome this issue, training set is further split into pure training set and cross validation set. The network is built using the pure training set and then will be used to predict on the cross validation set. By using certain metric calculations, the best number of neurons can be selected and later can be used for the test set. The split of training set is generally in the ratio of 70% and 30%, where 70% of the training set is pure training set and the remaining 30% can be used for cross validation. It should be noted that the split is random for training the network efficiently. Mathematical representation of the split can be expressed as $\mathcal{W}^{\text{train}} = \mathcal{W}^{\text{pure}} \cup \mathcal{W}^{\text{cv}}$. The total number of training weeks (W^{train}) being 55 for our case study, 40 weeks (around 72%) is considered as the pure training set whereas the remaining 15 weeks(around 28%) is considered as the cross validation set.

For building the network for the case study, MATLAB's predefined Neural Network function '*feedforwardnet*' is used. The algorithm used for training the network is Levenberg-Marquardt. The network is trained for a range of number of neurons $r \in \mathcal{R} = \{1, 2, ...R\}$ with $R = 50$ for the case study. For different number of neurons, the MAPE values for both training set and the cross validation set are calculated using the expressions below to identify the optimized number of neurons for different users on different days of the week:

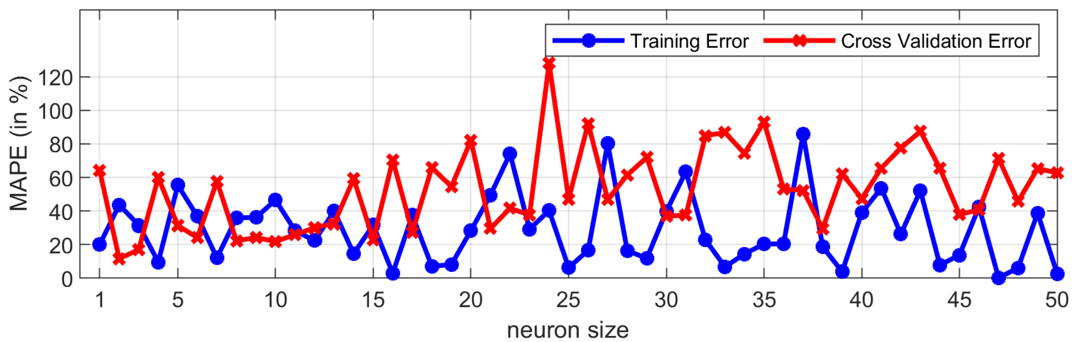
$$\text{MAPE}^{\text{pure}}_{k,r}(d)(\text{in }\%) = \frac{1 \times 100}{|\mathcal{W}^{\text{pure}}|} \sum_{w \in \mathcal{W}^{\text{pure}}} \frac{|p^{\text{peak}}_{k}(d,w) - \hat{p}^{\text{peak}}_{k,r}(d,w)|}{p^{\text{peak}}_{k}(d,w)}, \tag{4.32}$$

$$\text{MAPE}^{\text{cv}}_{k,r}(d)(\text{in }\%) = \frac{1 \times 100}{|\mathcal{W}^{\text{cv}}|} \sum_{w \in \mathcal{W}^{\text{cv}}} \frac{|p^{\text{peak}}_{k}(d,w) - \hat{p}^{\text{peak}}_{k,r}(d,w)|}{p^{\text{peak}}_{k}(d,w)}. \tag{4.33}$$

where $|\mathcal{W}^{\text{pure}}|$ represents the number of elements present in the pure training set and $|\mathcal{W}^{\text{cv}}|$ represents the number of elements present in the cross validation set. The MAPE values of the peaks values corresponding to the training set and to the cross validation set with respect to the neuron size of user '19' and user '25' on Wednesdays are shown in figure 4.7.



(a) User '19'



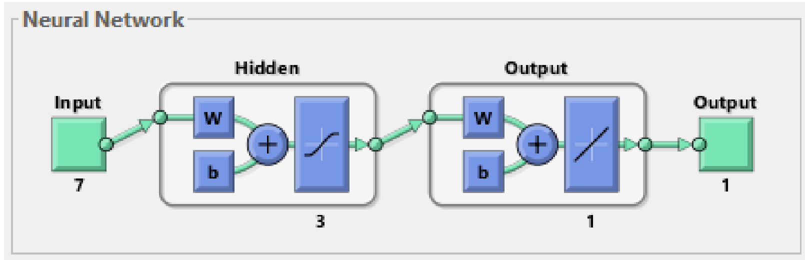
(b) User '25'

Figure 4.7. MAPE values of the peak values constituting the training set and the cross validation set for user '19' and user '25' on Wednesdays using single layered artificial neural network

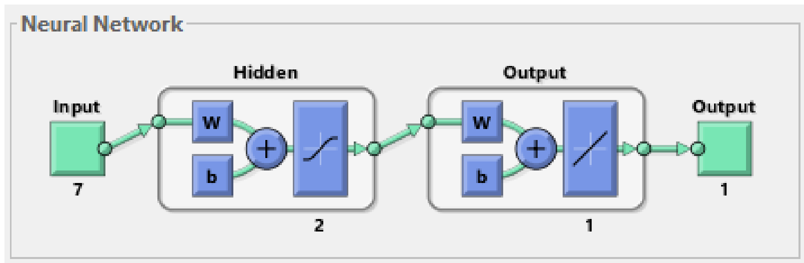
It can be seen that the MAPE values corresponding to the cross validation set increase after a certain neuron size whereas the MAPE values decrease largely because of over-fitting the network for both users. The optimized number of neurons $N_k^{\text{opt}}(d)$ that can be used for building the model for the test set will correspond to the number of neurons that furnish us with the lowest $\text{MAPE}_{k,r}^{\text{cv}}(d)$ for different users on different days of the week. Mathematically, the expression for the optimized number of neurons can be given as

$$N_k^{\text{opt}}(d) = \arg\min_{r \in \mathcal{R}} \{\text{MAPE}_{k,r}^{\text{cv}}(d)\}. \tag{4.34}$$

For the case study, the optimised number of neurons for user '19' is 3 whereas for user '25' is 2. Figure 4.8 shows the single layers artificial neural network with the optimized number of neurons which will be used for the test set.



(a) User ‘19’



(b) User ‘25’

Figure 4.8. Single layered artificial neural network used for predicting the peak power consumption values directly for user ‘19’ and user ‘25’ on Wednesdays using the optimized number of neurons

The MAPE values for the test set using the network formed by the optimized number of neurons are shown in table 4.6.

Table 4.6. MAPE values using single layered artificial neural network

Model	**User 19**	**User 25**
Single layered artificial neural network	11.73%	9.64%

When comparing all the models discussed previously, single layered artificial neural network improves the prediction by choosing a right number of neuron in the hidden layer. By comparing the performance of ANN to the offline average model, the MAPE values have been reduced by 5.66% and 4.52% for user ‘19’ and user ‘25’ on Wednesdays.

4.3 Summary

The highlights of this chapter are the following:

1. Directly predicting the peak power consumption values does provide an alternative for exploration instead of using an indirect approach as discussed in chapter 3 for different users on different days of the week.

2. Highlights on the case study of user '19' and user '25' on Wednesdays.

    - The peak power consumption values does suggest the values to be distributed and possess heavy tailed distribution.
    - Offline and online average models have a decent predictions on the test set.
    - Using features like previous week's or previous day's peak power consumption value improves the model suggesting a high correlation.
    - Using multiple features like peak values corresponding to previous seven days adds cherry on the top.
    - Advanced machine learning algorithms like artificial neural network does capture the nonlinear relationships which supports in explaining the model to a larger extent.

5. CONCLUSION AND FUTURE WORK

In this study, the main objective is to learn about the peak power consumption values of Commercial and Industrial (C&I) users using the real dataset. To pursue the same, we first analyzed the entire dataset in chapter 2. It is found that dividing the dataset into different days of the week helps in reducing large variances for many metering intervals and also helps us in understanding the general trend and correlation among different days of the week. It is also identified that users can be grouped together who have similar power consumption profiles during different days of the week and are divided into five different clusters. The second aspect of the dataset discussed in chapter 2 is the about the peak power consumption values and the metering intervals corresponding to these peak values. The peak values of the C&I users are in the range of 10kW to 40kW. It is found that the mean or the median of the peak values do not provide a good judgement as these values are scattered and follow heavy tailed distribution. The story of the metering intervals corresponding to the peak value is very similar. The mode of the metering intervals corresponding to the peak values do not give a concrete understanding for different users across different clusters.

In chapter 3, indirect approach of predicting the daily peak values is discussed using a case study pertaining to the users belonging to the dominant clusters on typical working day. In this approach, power consumption profile for the entire day is predicted and then the peak value is identified. Five models namely offline average model, online average model, first order auto regression model using conventional and asymmetric penalty methods, and artificial neural network are discussed and evaluated based on Mean Absolute Percentage Error (MAPE) values. Although the error values corresponding to the peaks for all the models is less than 20%, ANNs being advanced and complex perform better compared to other models. It is also observed that selection of the appropriate features play a significant role in developing the models. Features containing the most updated values tend to perform better.

In chapter 4, direct approach of predicting the daily peak values is discussed using the same case study as mentioned in chapter 3. In this approach, the peak power consumption values are directly predicted instead of the two-step process in indirect approach. Six models namely, offline average model, online average model, simple linear regression model using previous week's peak values and previous day's peak value, multiple linear regression model using previous seven day's peak values and artificial neural network with previous seven days' peak values as the input are discussed. Each model is evaluated using the same Mean Absolute Percentage Error (MAPE) values as discussed in chapter 3. The error values for all the models are small and are less than 20%. It is unfair to compare the MAPE values of both approaches simultaneously as the indirect approach uses the values within the day whereas the direct approach predicts the peak values at the beginning of the day. Similar to indirect approach, ANNs perform better as they have the potential to enhance the network with nonlinear features using activation functions.

Future Work

Some of the future work are mentioned as the following

1. Exploring different models with other features like moving average in both indirect and direct approaches can improve the model. Adding features corresponding to the external conditions like ambient temperature, humidity will certainly help in improving the performance of the models. Also, extracting the features from other users' power consumption profiles belonging to the same clusters can be also be studied.

2. Specific to the direct approach, models involving the power consumption values within the day can improve the prediction and will allow a direct comparison between the two approaches. Also, predicting the metering intervals corresponding to predicted peak values using this approach can be helpful.

3. Case study discussed in this research only considers typical users on typical days. Carrying out similar analysis on users belonging to other clusters on different days of the week can be executed.

4. The scope of this work can also be integrated by using the control algorithms for subsidizing the peak values using Distributed Energy Resources (DERs) through smart energy metering systems.

LIST OF REFERENCES

1. A. D. Papalexopoulos and T. C. Hesterberg, 'A regression-based approach to short-term system load forecasting,' Conference Papers Power Industry Computer Application Conference, 1989, pp. 414-423.

2. L. Duan, D. Niu and Z. Gu, 'Long and Medium Term Power Load Forecasting with Multi-Level Recursive Regression Analysis,' 2008 Second International Symposium on Intelligent Information Technology Application, 2008, pp. 514-518.

3. N. Yamaguchi, J. Han, G. Ghatikar, S. Kiliccote, and M. A. Piette, 'Regression models for demand reduction based on cluster analysis of load profiles,' 2009 IEEE PES/IAS Conference on Sustainable Alternative Energy (SAE), 2009, pp. 1-7

4. Rob J Hyndman and George Athanasopoulos, 'Forecasting: Principles and Practice,' 2^{nd} edition, Monash University, Australia.

5. S. N. Dodamani, V. J. Shetty and R. B. Magadum, 'Short term load forecast based on time series analysis: A case study,' 2015 International Conference on Technological Advancements in Power and Energy (TAP Energy), 2015, pp. 299-303.

6. H. Jiang, X. Yu and Y. Yu, 'Application of Support Vector Regression in Power System Short Term Load Forecasting,' 2008 Fourth International Conference on Natural Computation, 2008, pp. 26-30.

7. M. K. Azad, S. Uddin and M. Takruri, 'Support vector regression based electricity peak load forecasting,' 2018 11th International Symposium on Mechatronics and its Applications (ISMA), 2018, pp. 1-5.

8. P. Mukhopadhyay, G. Mitra, S. Banerjee and G. Mukherjee, 'Electricity load forecasting using fuzzy logic: Short term load forecasting factoring weather parameter,' 2017 7th International Conference on Power Systems (ICPS), 2017.

9. S. Singh, S. Hussain and M. A. Bazaz, 'Short term load forecasting using artificial neural network,' 2017 Fourth International Conference on Image Information Processing (ICIIP), 2017, pp. 1-5.

10. G. Chao, Z. Jing-chun, S. Yan-bin and S. Li-ying, 'Application of Dynamic Recurrent Neural Network in Power System Short-Term Load Forecasting,' 2010 International Conference on Computing, Control and Industrial Engineering, 2010, pp. 378-381.

11. Dong-Xiao Niu, Qiang Wanq and Jin-Chao Li, 'Short term load forecasting model using support vector machine based on artificial neural network,' 2005 International Conference on Machine Learning and Cybernetics, 2005, pp. 4260-4265 Vol. 7

12. L. Paull, H. Li and Liuchen Chang, 'The development of a fuzzy neural system for load forecasting,' 2008 Canadian Conference on Electrical and Computer Engineering, 2008, pp. 923-926.

13. A. Y. Saber and A. Al-Shareef, 'Load Forecasting of a Desert: a Computational Intelligence Approach,' 2009 15th International Conference on Intelligent System Applications to Power Systems, 2009, pp. 1-7.

14. S. R. Abbas and M. Arif, 'Electric Load Forecasting Using Support Vector Machines Optimized by Genetic Algorithm,' 2006 IEEE International Multitopic Conference, 2006, pp. 395-399.

15. Commission for Energy Regulation (CER), 'CER smart metering project - electricity customer behaviour trial, 2009-2010,' Tech. Rep., 2012.
Available: www.ucd.ie/issda/CER-electricity

VITA

B Hari Kiran Reddy was born in Bilaspur, Chhattisgarh, India on May 24, 1995. He received Bachelor of Technology degree in Electrical Engineering from National Institute of Technology Agartala, Tripura, India in 2017 where he was awarded with the Institute Gold Medal and the Best Outgoing Student award. From 2017 to 2019, he worked with a team responsible for designing the Extra High Voltage Substations as a Senior Design Engineer for L&T Construction in Chennai, Tamil Nadu, India. He is currently pursuing Master of Science in Electrical and Computer Engineering at Purdue University, West Lafayette, Indiana, USA.