

Clustering Analysis of Residential Loads

by

Kambiz Karimi

B.S., Kabul University, 2011

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Electrical and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Major Professor
Dr. Anil Pahwa

Abstract

Understanding electricity consumer behavior at different times of the year and throughout the day is very important for utilities. Though electricity consumers pay a fixed predetermined amount of money for using electric energy, the market wholesale prices vary hourly during the day. This analysis is intended to see overall behavior of consumers in different seasons of the year and compare them with the market wholesale prices. Specifically, coincidence of peaks in the loads with peak of market wholesale price is analyzed.

This analysis used data from 101 homes in Austin, TX, which are gathered and stored by Pecan Street Inc. These data were used to first determine the average seasonal load profiles of all houses. Secondly, the houses were categorized into three clusters based on similarities in the load profiles using k-means clustering method. Finally, the average seasonal profiles of each cluster with the wholesale market prices which was taken from Electric Reliability Council of Texas (ERCOT) were compared.

The data obtained for the houses were in 15-min intervals so they were first changed to average hourly profiles. All the data were then used to determine average seasonal profiles for each house in each season (winter, spring, summer and fall). We decided to set three levels of clusters). All houses were then categorized into one of these three clusters using k-means clustering. Similarly electricity prices taken from ERCOT, which were also on 15-min basis, were changed to hourly averages and then to seasonal averages.

Through clustering analysis we found that a low percent of the consumers did not change their pattern of electricity usage while the majority of the users changed their electricity usage pattern once from one season to another. This change in usage patterns mostly depends on level of income, type of heating and cooling systems used, and other electric appliances used.

Comparing the ERCOT prices with the average seasonal electricity profiles of each cluster we found that winter and spring seasons are critical for utilities and the ERCOT price peaks in the morning while the peak loads occur in the evening. In summer and fall, on the other hand, ERCOT price and load demand peak at almost the same time with one or two hour difference. This analysis can help utilities and other authorities make better electricity usage policies so they could shift some of the load from the time of peak to other times.

Table of Contents

List of Figures	iv
List of Tables	v
Acknowledgements	vi
Chapter 1 - Introduction.....	1
Chapter 2 - Methodology	3
2.1-Data:	3
2.2- K means clustering	3
Chapter 3 - Data Processing and Observations.....	6
3.1-Average Seasonal Curves:	6
3.1.1 Averaging data on hourly basis.....	7
3.1.2 Averaging hourly data over season.....	8
Chapter 4 - Clustering analysis of seasonal data	11
4.1- Normalized average seasonal curves.....	11
4.2- Clustering of all houses in each season	13
Chapter 5 - Conclusion and Recommendation	26
References.....	27

List of Figures

Figure 3.1 15 minute based bar plot for house # 30 on Dec/1/2013.	6
Figure 3.2 Plot of hourly averaged load data	7
Figure 3.3 Box and whisker plot of seasonal demand in winter.....	8
Figure 3.4 Box and whisker plot of seasonal demand load in spring.....	9
Figure 3.5 Box and whisker plot of seasonal demand load in summer.	9
Figure 3.6 Box and whisker plot for of seasonal demand load in fall.....	10
Figure 4.1a Power demand profiles for House 30 averaged over summer season. b- Same demand normalized.....	12
Figure 4.2a, b and c Power demand profiles of houses included in each cluster in winter.	13
Figure 4.3 Plot of cluster centers and ERCOT SPP real time price average in winter.	14
Figure 4.4a, b and c Power demand profile of houses included in each cluster in spring.	16
Figure 4.5 Plot of cluster centers and ERCOT SPP real time price average in spring.	17
Figure 4.6a and b Power demand profile of houses included in each cluster in summer.....	18
Figure 4.7 Plot of cluster centers and ERCOT SPP real time price average in summer.....	19
Figure 4.8a, b and c Power demand profile of houses included in each cluster in fall.	21
Figure 4.9 Plot of cluster centers and ERCOT SPP real time price average in fall.	22

List of Tables

<i>Table 1 Clustering result for winter season.....</i>	15
<i>Table 2 Clustering result for spring season.....</i>	18
<i>Table 3 Clustering results for summer season.....</i>	20
<i>Table 4 Clustering result for fall season.....</i>	23
<i>Table 5 Clustering result for all houses in all seasons.(continues on next page)</i>	24

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Dr. Anil Pahwa for his continuous support of my MS studies and related project, for his encouragement, motivation and guidance. I would also like to recognize my committee members Dr. Sanjoy Das and Dr. Ruth Miller for their support and thoughtful comments. And Dr. Andrew Rys, ECE department graduate program director, for his continuous support.

My special appreciation goes to the Fulbright program, Department of State, my Fulbright advisor Tom Koerber and all other individuals associated to this prestigious program for granting me Fulbright Scholarship and supporting through last two years to complete my MS degree in the US.

Last but not the least; I thank my parents, my brothers and sister, my fiancé and my friends for being there for me and supporting me.

Chapter 1 - Introduction

Electric energy and its vast usage in all aspects of life has become a critical part of the current life style on earth. From lighting to communication, transportation systems to education and military uses, the demand for more electricity is increasing with passing time. Though there is enough information collected on electricity generation and transmission, there are very little data on end users.

Consumer behavior related to electricity usage can differ throughout the day and many factors such as time of the year, weather, level of income, can affect this behavior. The new technologies of smart metering have enabled utilities and research institutes to gather consumer load data on 1-minute, 15-minute or 1-hour basis. These data can be analyzed to find consumer behavior on temporal basis and find the driving factors behind these behaviors. Such analysis could help utilities manage the loads during peak times and shift them to other times to not coincide with market wholesale price peaks.

To analyze large number of consumers' load data and their behavior throughout the year, we can categorize them into distinctive groups based on their pattern of electricity use and see which group of houses uses the most electricity during peak market prices. In order to implement this grouping process on load data, clustering algorithms are proving to be one of the best ways.

Clustering is a grouping process that, using the similar features of the data, categorizes them into distinctive groups. This technic works best for quantitative data; therefore is a good tool for analyzing electric consumer data. Some work has already been done using this technic to analyze load patterns. For example, Chicco used clustering technics to analyze the behavior of non-residential loads at medium voltage level. He analyzed 400 load patterns during a weekday in Italy. The load data he used was in 15-minute intervals and he normalized it to the minimum and maximum value of the loads. He found that methods that find outliers in the data are the most effective for creating typical profiles of consumers [4].

Rasanen et al. using an efficient methodology based on self-organizing maps (SOM) and clustering methods (K-means and hierarchical clustering) analyzed the load patterns of 3989 small customers in Savo, Finland. They grouped the users into different clusters based on their electricity usage based on the hourly data they had obtained for these houses. New load curves were created for each cluster of users and these load curves were compared to the old ones that

utility companies had before using the real time data of the houses. The results from the new load curves were much more accurate and estimated near to the real demands of the houses [5].

Similar analysis was done by Kim et al. to study the typical load profiles of 3183 high voltage consumers in South Korea. They took the data from Korea Electric Power Corporation (KEP-CO). They performed three methods of clustering; hierarchical, k-mean and fuzzy c-means, on the data to divide them into clusters and create load profiles for each group. The results showed that hierarchical clustering was optimal approach to ensure consistent processing time. Also k-means clustering technique was found to be the most efficient option to minimize the mean absolute error (MAE) between the typical load profile and the real load profile [6].

Also, Kwat et al. formed a set of hourly electricity use profiles from 220,000 homes in California. Using clustering techniques (adaptive k-means and hierarchical clustering) they classified them into certain groups based on their level of electricity use. The purpose of this classification was to target the high energy users during peak hours for demand response (DR) and energy efficiency (ER) policies [7].

Panapakidis et al. using a k-means clustering on 150 consumers in Greece found out that other survey data such as electricity cost, household appliances can be effective factors in classifying houses into different clusters [8].

Using the experience from other works done in applying clustering methods for classifying load profiles; and based on the original work of Rhodes, J.D. et al [10], in which they performed clustering analysis on residential loads based on different seasons of the year, we performed similar analysis on residential load data obtained from residents of Austin, TX. We used clustering algorithm to categorize all houses into distinctive clusters (groups) based on their load usage pattern. This clustering approach is performed for all four seasons separately. Each cluster center, which is average of that group, is then compared with the market wholesale prices to see how close or far does the demand peaks and market wholesale price peak occur. This analysis helps utilities identify which houses use the most expensive electricity, and therefore find a solution on how to shift these houses' loads to times when the wholesale market prices are not very high.

Chapter 2 - Methodology

2.1-Data:

In this analysis load data of 101 houses were used. These houses are located in Austin, Texas and are volunteers with Pecan Street [2] for acquiring electrical and water data for various analysis purposes. Pecan Street Inc. is an organization researching electrical and water systems and has more than 1300 volunteers around the country. The data acquired from these volunteers regarding their general electric use, solar PV panel performance, electric vehicle charging and other valuable survey reports make up Pecan Street's Dataport [2]. Dataport makes terabytes of electric load data easily accessible, manageable and usable to researchers from around the globe.

Load data is available in 1-min and 15-min basis for each house for different years. Because we do not need very specific details of data in 1-min basis, we used 15-min basis data for the purpose of this analysis. Also we chose the year 2014 because it had better data than the previous years with less data record errors and more details.

We also used market wholesale price data in this analysis for Austin area. This data was obtained from ERCOT, which is responsible for recording the market wholesale prices for Texas. The data is recorded on 15-minute basis and is available for public use.

2.2- K means clustering

Clustering is the process of partitioning data into small number of groups called clusters. Different features of the data are observed and those with similar features are categorized onto one cluster. K-means algorithm measures the distance between the data point and the cluster centers and minimizes it [3]. K-means algorithm can be performed through following steps:

- 1- Select number of clusters.
- 2- Select center of the clusters
- 3- Measure distance between each data point and the cluster centers to assign the data point to the cluster with the least distance from it.
- 4- Change the position of the cluster center to the mean of the data points that belong to that cluster.
- 5- Repeat steps 3 and 4 until the algorithm converges.

To choose the right number of clusters the easiest way is to have an estimated range of number of clusters and perform trial and error to see how many of clusters give the best results. Also many computational tools such as Matlab and R can perform analysis to find the optimum number of clusters [9]. In this analysis we selected the number of clusters to be 3 based on electricity usage patterns.

This algorithm can get stuck in local minima so initializing the cluster centers is important. We can choose any data point as cluster center as we want or we can initialize randomly. If the algorithm is stuck in local minima we need to reinitialize the cluster center to get the optimum results. The algorithm is performed as follows:

1- Initialize randomly C_k , $k = 1, \dots, K$

2- Repeat

a. For each $X(n)$:

$$C_k \leftarrow \arg \min_{C_j} \|X(n) - C_j\|$$

$$\mu_k(n) \leftarrow 1, \forall j \neq k, \mu_k(n) \leftarrow 0$$

b. For each C_k :

$$C_k \leftarrow \frac{\sum_n \mu_k(n) X(n)}{\sum_n \mu_k(n)}$$

End

Here $X(n)$ is the n th data point, C_j is the j th cluster center, μ is an n by k matrix which can have either values of 1 or 0. Value of 1 in this matrix means that the data point $X(n)$ is included in k th cluster and value of 0 means that the data point is not included in the k th cluster.

Clustering categorizes data points into clusters. In this analysis we want to assign each house into a cluster, therefore each house is considered as a data point. Each house as a data point has 24 numbers each showing the load of the house during different hours of the day for a specific season.

As the number of clusters has already been chosen to be three we directly go to the second step of the clustering algorithm described above and randomly choose three different houses as the cluster centers for each of the four seasons separately.

To apply the third step of the algorithm on the load data we find the difference between the loads of each hour of the houses to the cluster centers respectively. Then add the amounts we found from this subtraction separately for each cluster center and see which of the added amounts is the least; we will then assign the house to that cluster.

For example, let us assume House # 20, 40 and 60 are randomly selected as the cluster centers for the three clusters in winter. We will start from House # 1 and subtract the amount of load of this house at 1:00am from the amount of load of center of cluster 1 (i.e. House # 20), and do this subtraction for 2:00am and the rest of the load data till 12:00 at night. We add the values obtained from this subtraction. This final added value shows the total load difference from House # 1 to center of cluster 1. We repeat this process for House # 1 with center of cluster 2 and 3 and find these load differences. Now we see which load difference is the lowest. If House # 1 had the lowest load difference with center of cluster 1 then we assign it to that cluster; or if this house had the lowest load difference with center of cluster 2 or 3 we will assign it to one of those clusters respectively. We do this process for all houses and assign each house to one of the clusters. At the end of this step we will have all the houses assigned to one of the three clusters.

The fourth step is changing the cluster centers to the average of each cluster. This step can be done on this data by averaging the load on hourly basis of all houses assigned to each cluster. We add the loads of all houses in the same cluster at 1:00am and divide it by the number of houses. We repeat it for the rest of 23 hours of the day. This way new cluster centers are assigned for each cluster.

We now apply step three with new cluster centers, assign all houses to one of the three clusters again based on their load difference from the cluster centers. We find the average of the clusters and assign them as the new cluster centers.

We continue this process until the algorithm has converged. This means that no more houses change their cluster position and the cluster center does not change their value anymore. We apply this process in all four seasons to get clusters for all seasons.

Chapter 3 - Data Processing and Observations

The analysis for this report was done to create average seasonal curves from 15-min data given for each house. The second step was to determine number of clusters and their associated houses. Afterwards a comparison was made between the peak market prices and peak demand loads of each cluster center.

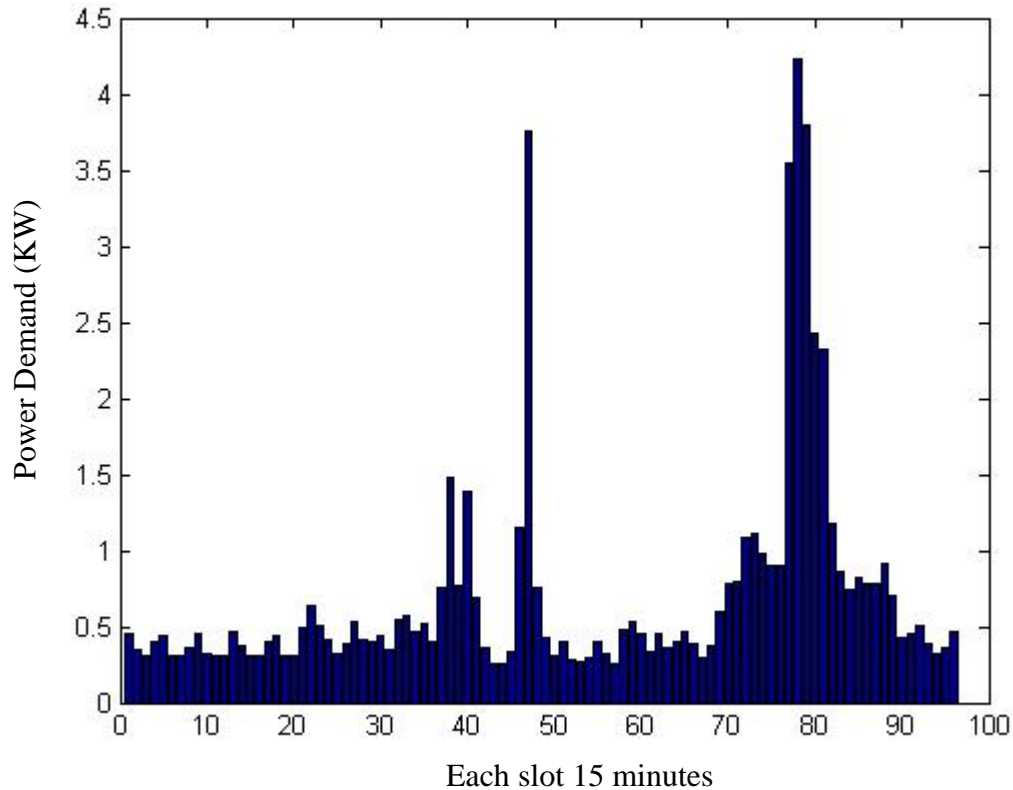


Figure 3.1 15 minute based bar plot for house # 30 on Dec/1/2013.

3.1-Average Seasonal Curves:

In this analysis four seasons of the year are considered, which are Spring: March-May; Summer: June - August; Fall: September – November and Winter: December – February. The load data used in this analysis is 15-minute based demand load data collected by Pecan Street Inc. [2] through their project in Austin, Texas. The data is from 12/1/2013 to 11/30/2014 which was downloaded through pgadmin and SQL from the Dataport (the main data servers of Pecan Street).

3.1.1 Averaging data on hourly basis

A sample 15-minute based demand load data is plotted for demonstration in figure 3.1 above. This plot is for 12/1/2013 for house # 30. As can be seen the peak demand has occurred near 80th 15-min time interval which is around 8:00pm. Also a demand spike is seen around 12:30pm which could be due to cooking and dishwashing.

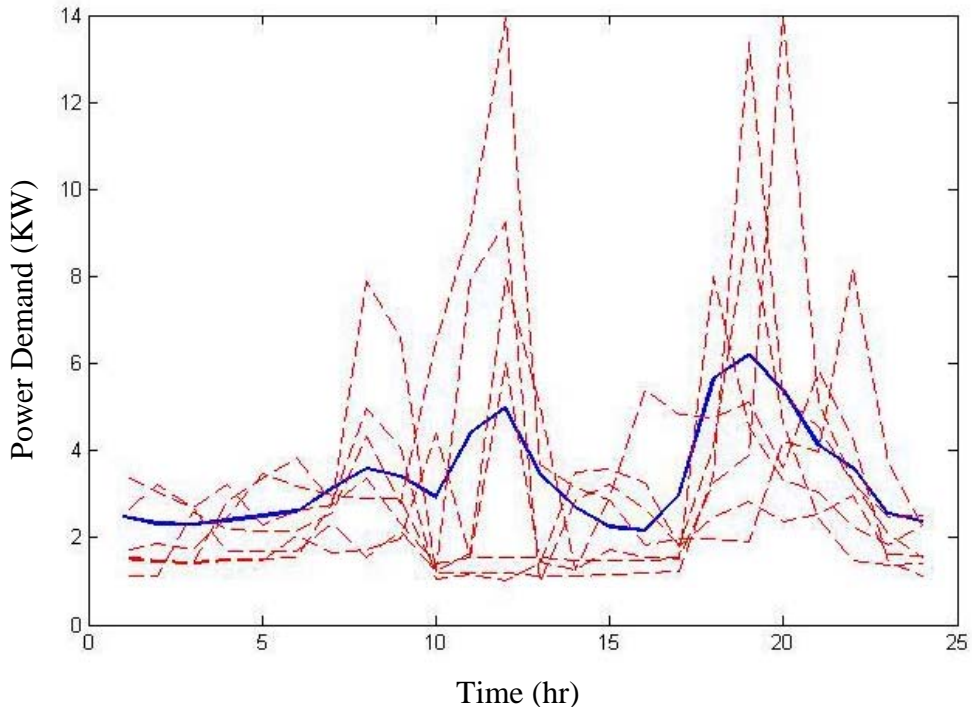


Figure 3.2 Plot of hourly averaged load data

The raw 15-minuta data was averaged over the hour to obtain hour based demand using equation (1) below:

$$L_{jh} = \sum P_{jm} / 4 \tag{1}$$

Where j is the house number and h is the hour of day and m is the 15-minute interval and P is the demand on 15-minute basis.

Figure 3.2 above shows a plot of hourly averaged demand for House 30 for ten days (in red) Dec 1 to Dec 10 2013. The plot also contains an average demand profile plot for the entire month of December for the same house. As can be seen from the plot different days have different load demands peaking up to 14kW on some days, while their average over a month gives an in between demand showing the most common time of peak load by this house.

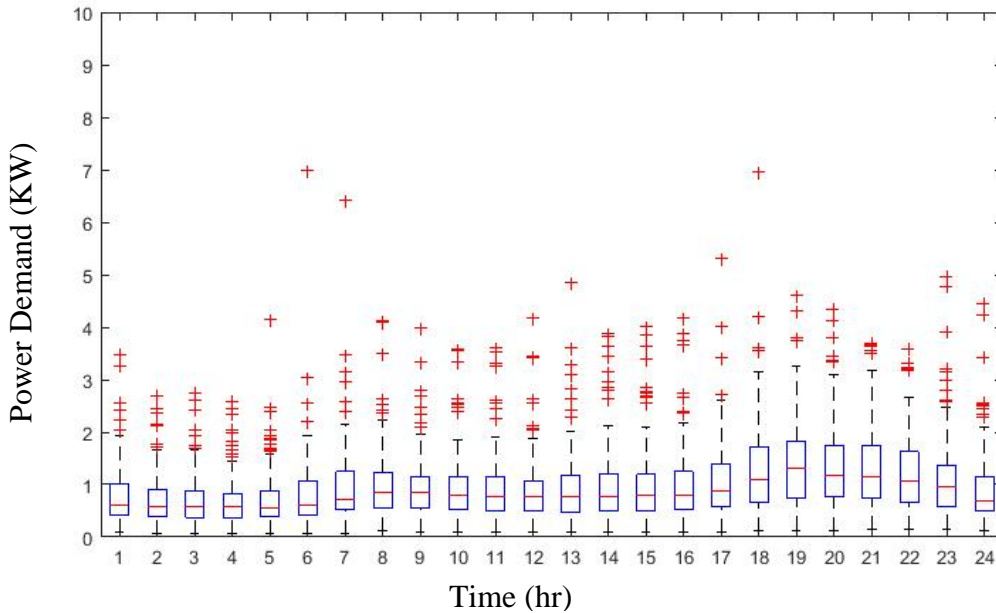


Figure 3.3 Box and whisker plot of seasonal demand in winter.

3.1.2 Averaging hourly data over season

Next step is to average the load data over four seasons. The average demand profiles are calculated using equation (2) as following:

$$L_{s,h} = \sum L_{jh} / N \tag{2}$$

Where “s” indicates the season, j is the hour of the day and h is the house number. N here is the number of days in that season. The summation is done over all days included in that season. Each season; winter, spring, summer and fall, have the following number of days in them respectively 90, 92, 92 and 91 which totals 365. Using the same method ERCOT’s price data is also averaged over each season to compare with the demand profiles.

Figures 3.3 through 6 are box and whisker plots of all houses’ demands averaged over seasons. The box indicates the interquartile range and the red line inside the boxes are the medians. The lower whisker is the lowest demand during that hour and the upper whisker is 1.5 times the interquartile range from the upper edge of the box. All + signs in red show the outliers. Outliers are the data that are higher than 1.5 times the interquartile range from the third quartile.

The box plot of demand load in winter, Figure 3.3, shows that the average demand for these 101 houses during winter is not very high. The highest the whisker reaches is at 19:00 when it reaches about 3KW. There are some outliers of course but the majority of the houses do

not peak higher than 3KW. The reason is that most of the houses have gas furnace for heating purposes in winter, therefore very few houses use electricity for heating.

As the weather gets warmer the electricity demand increases. This can be seen in figures 3.4 through 3.6. In figure 3.4, the spring season, the highest whisker reaches around 4KW. Though outliers still exist, most of consumers do not use more than 4KW at peak which occur at 19:00.

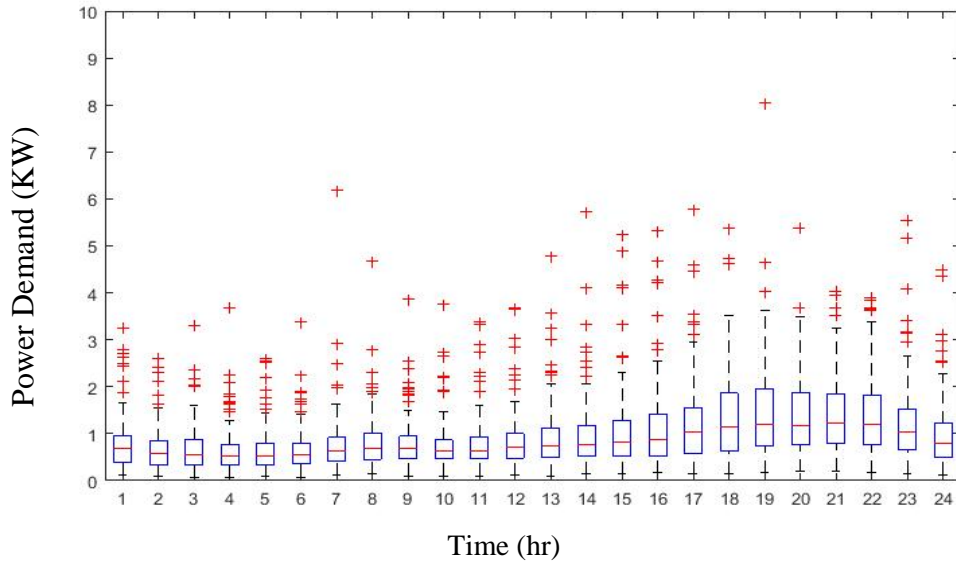


Figure 3.4 Box and whisker plot of seasonal demand load in spring.

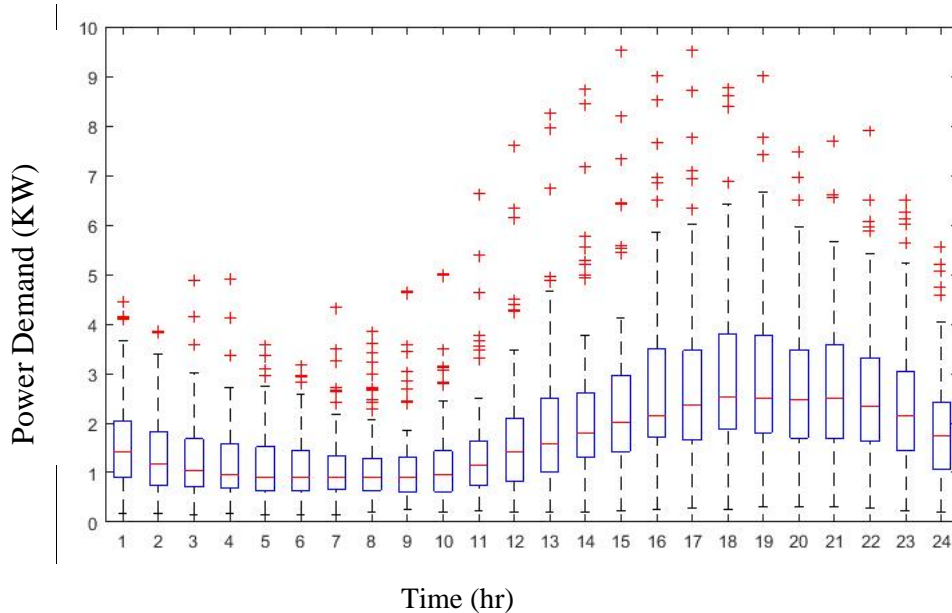


Figure 3.5 Box and whisker plot of seasonal demand load in summer.

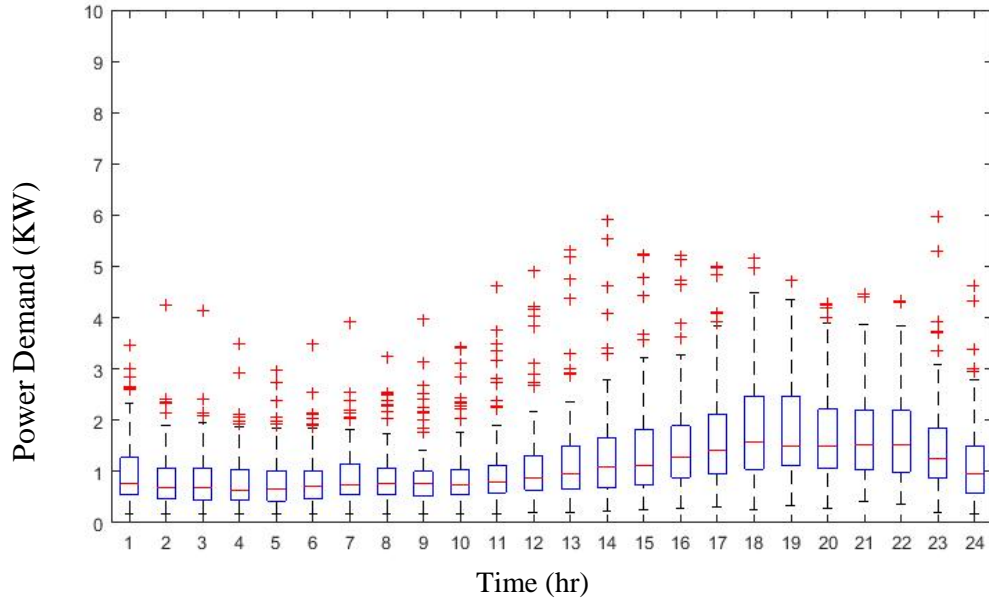


Figure 3.6 Box and whisker plot for of seasonal demand load in fall.

In summer consumers use AC central or wall mounted systems for cooling purposes. All these cooling systems use electricity therefore the demand increases. From figure 3.5 we see that the peak load occurs at 19:00 which is around 7KW. There are houses that uses up to 10KW, but the majority limit their use by 7KW at max.

The use of AC continues till mid fall so the average demand load in fall stays higher than winter but lower than summer due to decrease in temperature and less use of cooling systems compared to summer. Figure 3.6 shows that behavior of these 101 houses during fall 2014.

Chapter 4 - Clustering analysis of seasonal data

The next step after creating average curves for each house in all four seasons was to divide all the houses into distinctive groups using k-means clustering. This algorithm classifies all houses into one of the clusters by measuring their sum square distance to each cluster's center and putting them into the cluster with the lowest sum square distance.

For this algorithm the number of clusters should be predefined. J.D. Rhodes et al. performed clustering analysis with two cluster centers, one for high power users and the other low power users. We decided to perform this analysis using three clusters [10].

This analysis is performed for each season and we obtain three distinctive clusters for each season. Houses classified to each cluster are identified at the end of the analysis and cluster centers are calculated for the purpose of observance and comparison with each other and the ERCOT SPP prices [1].

Using the average electricity use of each cluster in each hour and the average seasonal ERCOT SPP Price we can calculate the average cost of electricity for each cluster using equation (3) below:

$$C_a = \sum_h E_{a,h} * SPP_h / E_a \quad (3)$$

Where C_a is the average cost of electricity use (\$/kWh) of cluster "a". $E_{a,h}$ is average electricity use for cluster a in hour "h". SPP_h is the ERCOT market price at hour "h" and E_a is the total energy used in entire 24 hours by cluster "a".

4.1- Normalized average seasonal curves

Though electric utility customers pay a fixed amount of money for every unit of electric energy they use, utilities need to deal with the changing market wholesale price of electricity throughout the day. Utilities would benefit the most if the peak load and peak price do not occur at the same time.

This clustering analysis is therefore more concerned with the peak values of the loads rather their true value. Hence all load data is normalized for the purpose of this analysis. The data of all houses which were recorded on 15-min intervals was averaged over each hour. Then

the average seasonal profiles were created for each house by averaging the loads each hour of the day over all days of the season. At this point all seasonal curves were divided by their peak values to obtain normalized average seasonal curves. This method of analyzing normalized load data has also been used by Rhodes, J.D. He also used normalized data to analyze residential houses behavior in different times of the year [10].

ERCOT prices were recorded and obtained on 15-minute basis. So they were also transformed to seasonal averages through the same process. The seasonal averages of each house were divided by their respective peak value to normalize them. Figure 4.1 below shows the demand of House # 30 averaged over summer based on its real values and normalized values. We notice that the overall shape of the load profile of the house does not change, only the y-axis measure changes.

Figure 3.1 to 3.6 are plots of load data with their true values in kW while figures 4.2 to 4.9 are all normalized data plots. In figures 3.3 to 3.6 we see that the average values of the demands differ from season to season and there are many high users in each season. The outliers go as high as twice or three times the average users. On the other hand in figures 4.2a-c we notice that all houses at some point reach the same value of 1 which does not mean that all the houses have the same peak rather means that each individual house reaches its own peak demand.

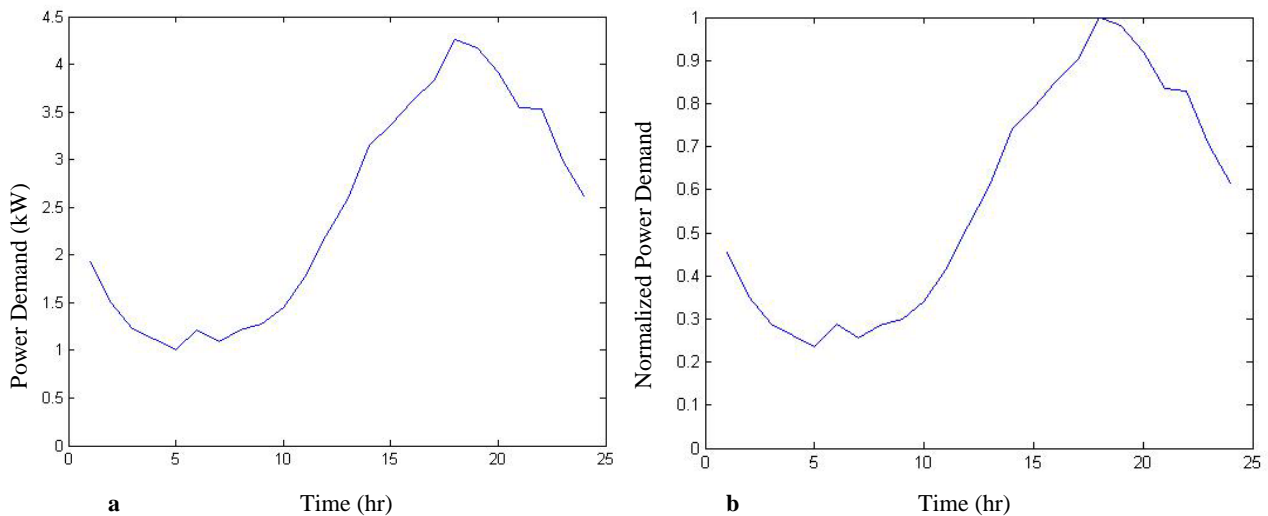


Figure 4.1a Power demand profiles for House # 30 averaged over summer season. b- Same demand normalized

4.2- Clustering of all houses in each season

As the number of clusters was already decided to be three representing three different groups of users, all houses were grouped into one of these clusters in each season. Figures 4.2-4.9 show electricity profiles of members of each cluster in each season and their respective cluster centers plots together with ERCOT SPP average prices.

Figure 4.2a-c show profiles of members of all three clusters in winter season. Members of each cluster are shown in Table 1. Cluster 1 with mean variance of 0.019 contains the most members, 43 houses. Cluster 2 and 3 with mean variances of 0.029 and 0.034 has 31 houses 27 houses respectively. From figure 4.2a we see that houses in cluster 1 starts mostly between 0.3 and 0.5 and continue to decrease until around 7:00am and then gradually

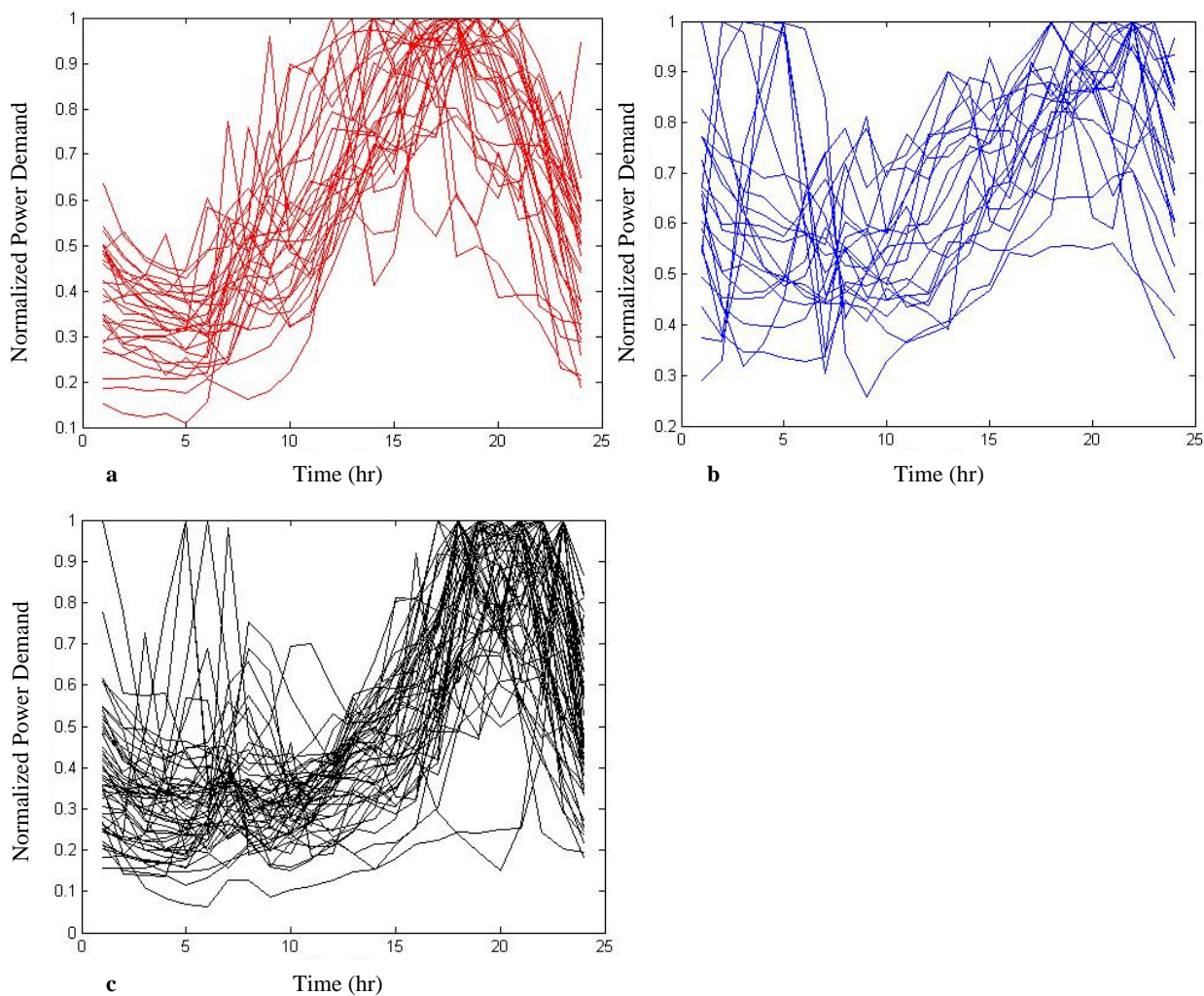


Figure 4.2a, b and c Power demand profiles of houses included in each cluster in winter.

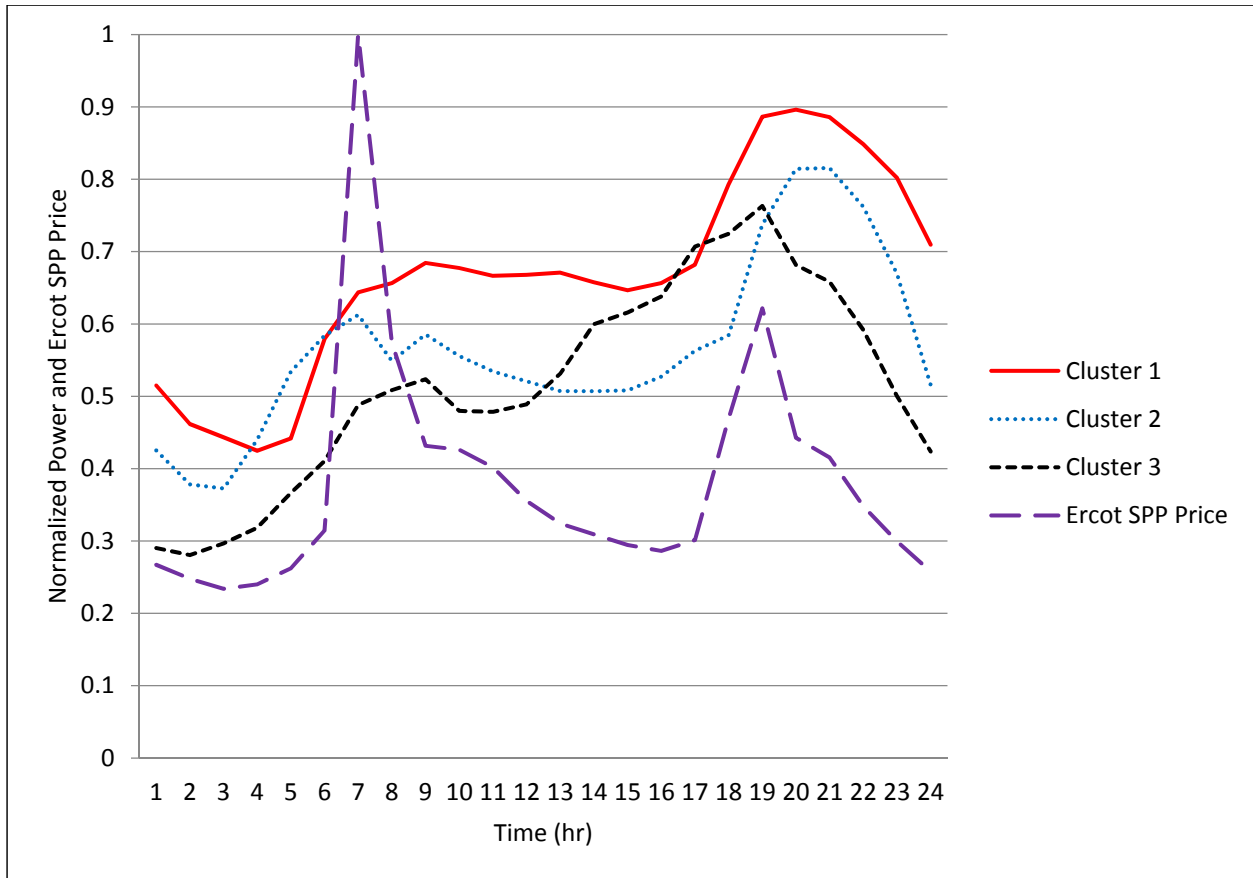


Figure 4.3 Plot of cluster centers and ERCOT SPP real time price average in winter.

increase till they hit their peak mostly between 3:00 pm and 8:00pm and then decrease. From Figure 4.3 we see that the average peak of the cluster happens at 7:00pm.

In contrast to cluster 1, we see that cluster 2 shows totally different electricity usage behavior. Houses included in cluster 2 as show in Figure 4.2b start off with wide range of usage levels from 0.4 to 0.8 and some of them have their peak usage around 5:00am. After 8:00am their usage gradually decreases until 3:00pm when the increase in electricity usage again starts till they peak between 6:00pm to 10:00pm. The cluster average peak occurs at 9:00pm.

Cluster 3 has a similar start as cluster 1 with some houses peak between 5:00 to 8:00 am. Cluster 3's usage also gradually increases until they peak between 5:00pm and 10:00pm, while their average usage peak occurs at 7:00pm.

Figure 4.3 above is a plot of cluster centers and ERCOT SPP price average over winter season. The ERCOT price peaks at 7:00am. At this time cluster 1 is using more energy than clusters 2 and 3. However, because the ERCOT price peak and the clusters' peak do not occur at

the same time, utilities need not worry about the peak market price. There seems a local peak around 7:00pm in ERCOT prices, but it is not very high.

Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
1	6	2	58	72	74
4	13	3	60	73	77
5	15	8	65	75	78
7	18	9	66	79	84
11	21	10	70	81	90
12	23	19	76	88	
14	24	22	80	89	
16	25	26	82	95	
17	30	28	83	99	
20	36	29	85		
27	37	33	86		
31	41	38	87		
32	46	40	91		
34	48	42	92		
35	49	45	93		
39	50	47	94		
43	62	52	96		
44	63	54	97		
51	64	57	98		
53	68	59	100		
55	69	61	101		
56	71	67			

Table 1 Clustering result for winter season

The profiles of spring season as shown in Figures 4.4 and 4.5 start at higher values compared to winter. Cluster 1 has more power usage during entire day while its peak demand is at 9:00pm and its mean variance is 0.029. Cluster 2 with mean variance of 0.037 has some high morning users and peaks at 10:00pm. Cluster 3, having a mean variance of 0.04 exhibits a high usage pattern around 8:00am and between 4:00pm to 8:00pm and its peak demand occur at 6:00pm.

In Figure 4.5 we see that the ERCOT SPP peak occurs at 7:00am. At this time cluster 1 is at 63% of its peak value while cluster 2 and 3 are at 47% and 45% of their peak values respectively.

There also occurs a local peak in ERCOT SPP at 10:00pm which is 80% the peak price. At this time cluster 1 is at its peak demand and cluster 2 and 3 are at 89% and 83% of their peak loads respectively.

Compared to winter the local peak of ERCOT peak price in spring is much higher. In winter it was just 60% of the peak price while in spring it is 80% of the peak price and it occurs near the peak demands of clusters which is not good for the utilities.

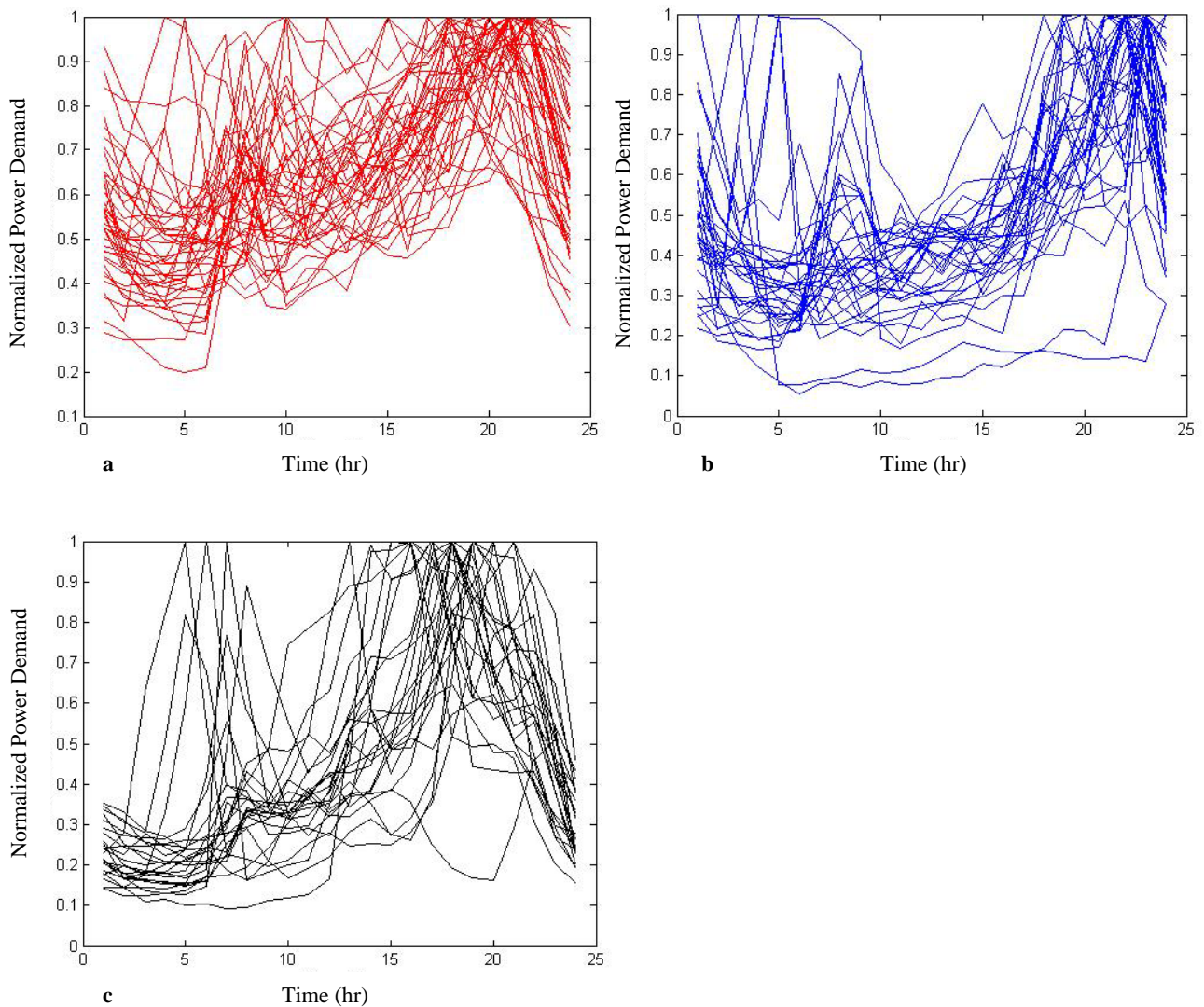


Figure 4.4a, b and c Power demand profile of houses included in each cluster in spring.

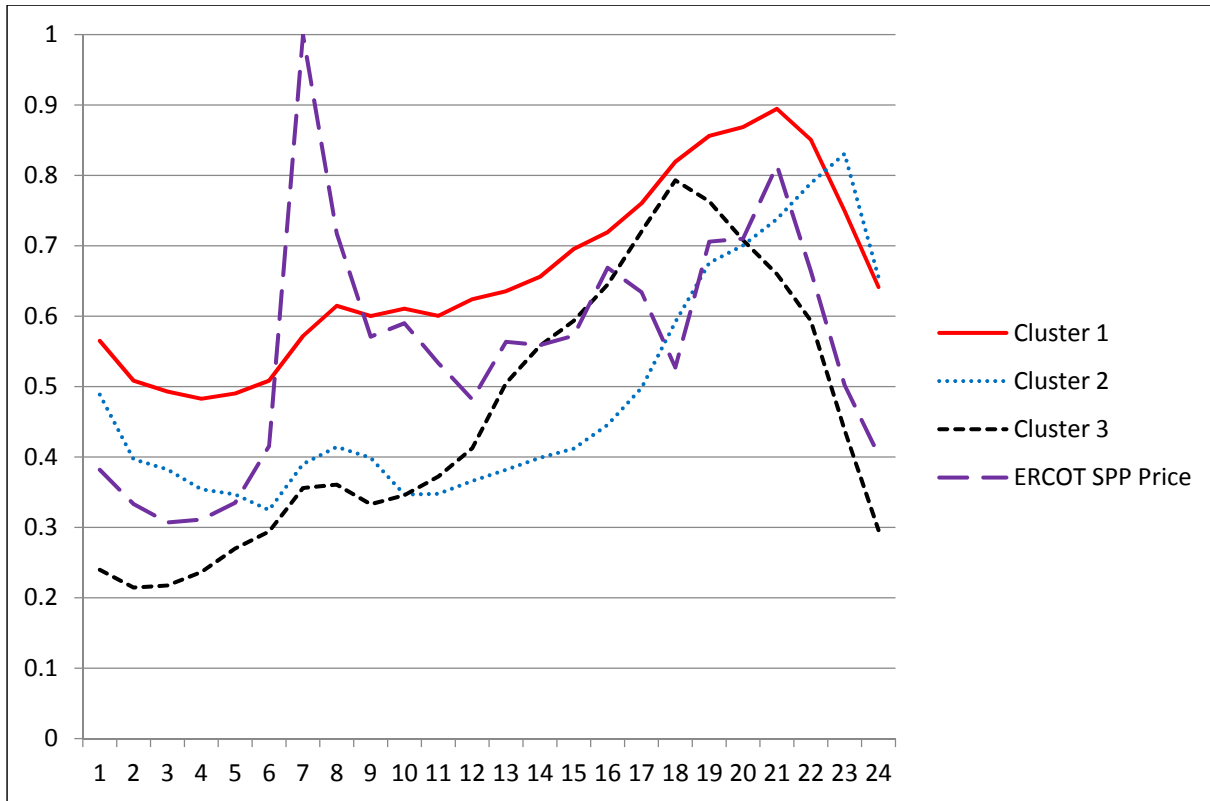


Figure 4.5 Plot of cluster centers and ERCOT SPP real time price average in spring.

Table 2 below is the list of houses per each cluster in spring season. Cluster 1 again contains the most members, 42 houses. Cluster 2 has 35 houses and cluster 3 has only 24 houses.

Comparing this table to Table 1 which showed the cluster members for winter, we see that about 62% of the users did not change their clusters and remained at the same level of usage. 18% of the users switched to a lower level of electricity use and remaining 20% of the users switched to a higher level of electricity use.

Similarly the three clusters for summer season were created and their members' profiles are plotted in Figure 4.6a-c. Figure 4.7 shows the cluster centers for the summer season with the average ERCOT prices and Table 3 is the list of cluster members in this season.

Summer clusters show the highest electricity use of all the four seasons. Cluster 2 has a very high early morning use while clusters 1 and 3 have lower power demand in early morning. Cluster 1 later in the midday increases much faster than the other two clusters and peaks at 6:00pm. Cluster 2 also has a higher peak relative to cluster 3 and peaks at 7:00pm. Cluster 3 on the other hand is of late evening users and overall low electricity users. This cluster's peak demand occurs at 9:00pm.

Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
4	2	1	62	64	74
7	3	6	63	68	78
9	5	8	65	69	98
12	10	11	66	75	
17	14	13	70	79	
18	15	16	71	83	
20	19	22	73	84	
27	21	23	76	87	
32	24	28	77	88	
33	25	29	80	89	
34	26	31	81	94	
38	30	40	82	95	
39	35	42	85	97	
44	36	50	86	99	
45	37	52	90		
49	41	54	91		
51	43	57	92		
55	46	59	93		
56	47	61	96		
58	48	67	100		
60	53	72	101		

Table 2 Clustering result for spring season.

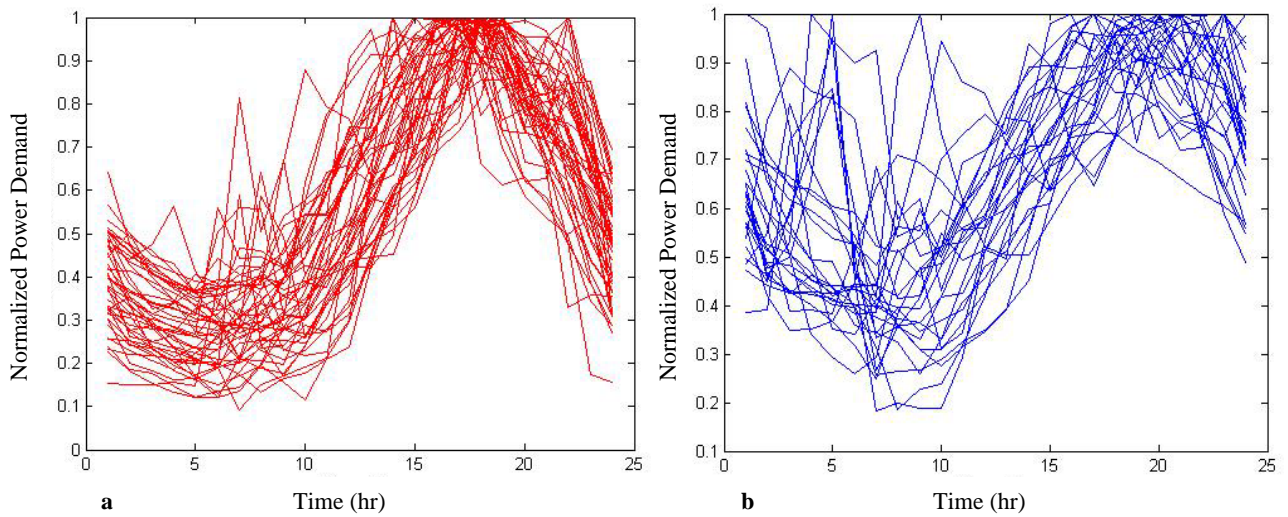


Figure 4.6a and b Power demand profile of houses included in each cluster in summer.

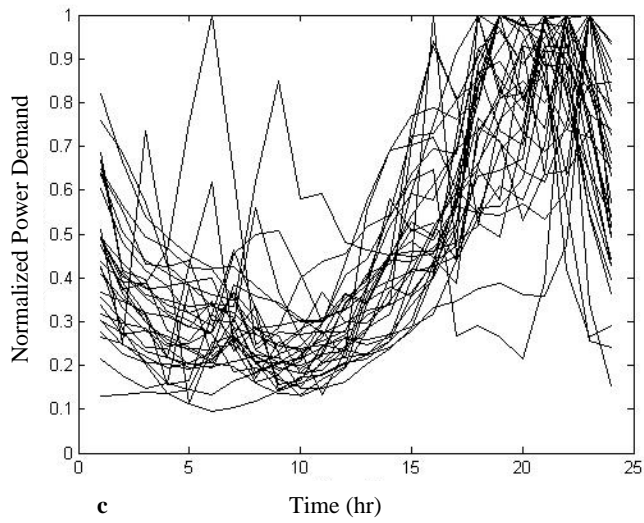


Figure 4.5 c Power demand profile of houses included in each cluster in summer.

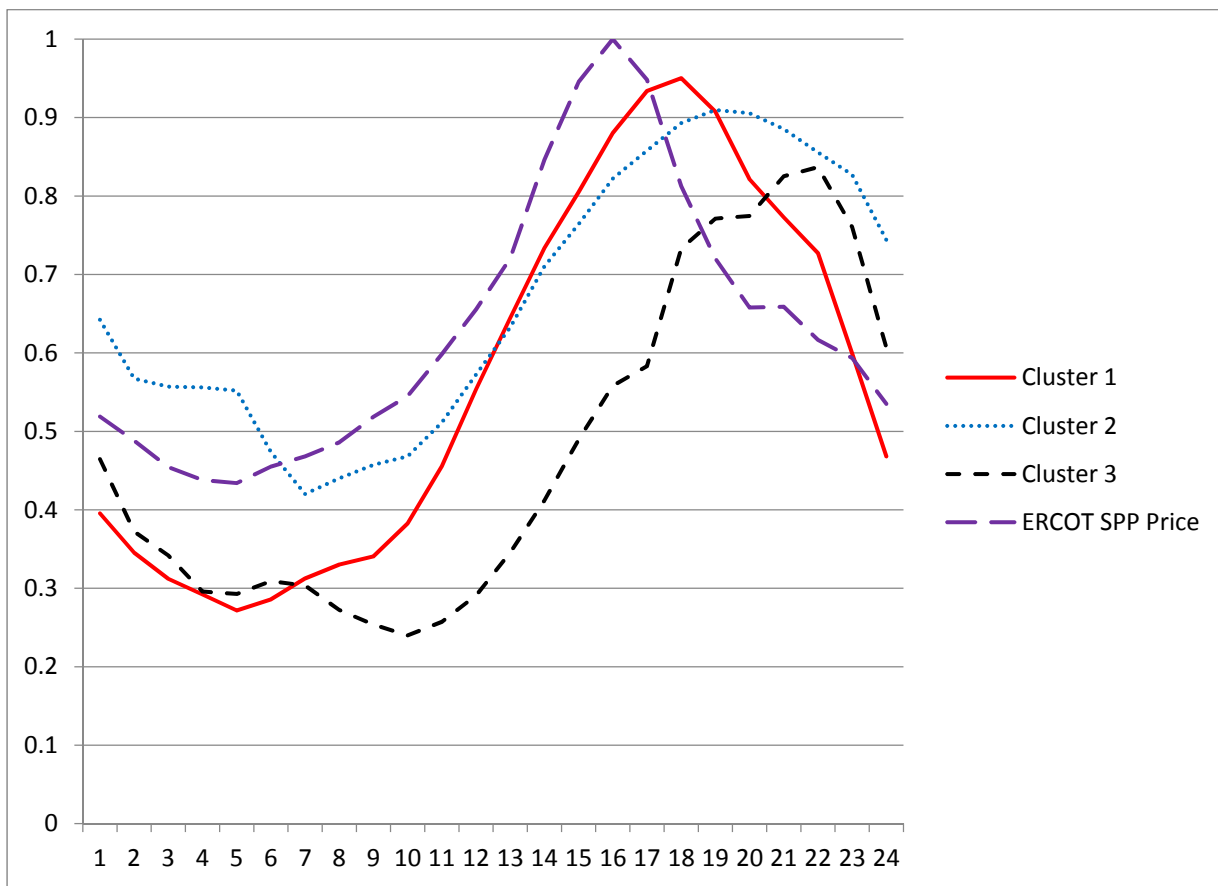


Figure 4.7 Plot of cluster centers and ERCOT SPP real time price average in summer.

This season is very critical for utilities. The high demand of electricity and occurrence of peak load demand and peak prices close to one another implies that utilities need to deal with this issue. In this season ERCOT price peaks at 4:00pm which is very near to cluster 1 and 2 peak demand times.

When market prices get at their peak value in this season, cluster 1 with mean variance of 0.031 is at 93% of its peak value, cluster 2 and 3 with mean variances of 0.043 and 0.036 are in 90% and 67% of their peak values respectively. Also when these clusters hit their peak demands the ERCOT price is at 81%, 72% and 61% of its peak value for each of the clusters 1, 2 and 3 respectively.

Table 3 gives the number of houses that belong to each of the three clusters. Now 45 users are classified to cluster 1. Clusters 2 and 3 respectively have 26 and 30 members.

Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
1	3	2	54	83	87
6	4	5	56	92	88
8	7	15	57	94	89
9	10	21	58		93
11	14	24	60		95
12	19	25	61		97
13	20	35	62		99
16	26	37	65		
17	27	40	69		
18	36	43	74		
22	41	46	77		
23	44	47	78		
28	45	48	80		
29	49	50	82		
30	53	52	85		
31	55	63	86		
32	59	64	90		
33	66	68	91		
34	67	71	96		
38	70	72	98		
39	75	73	100		
42	76	79	101		
51	81	84			

Table 3 Clustering results for summer season.

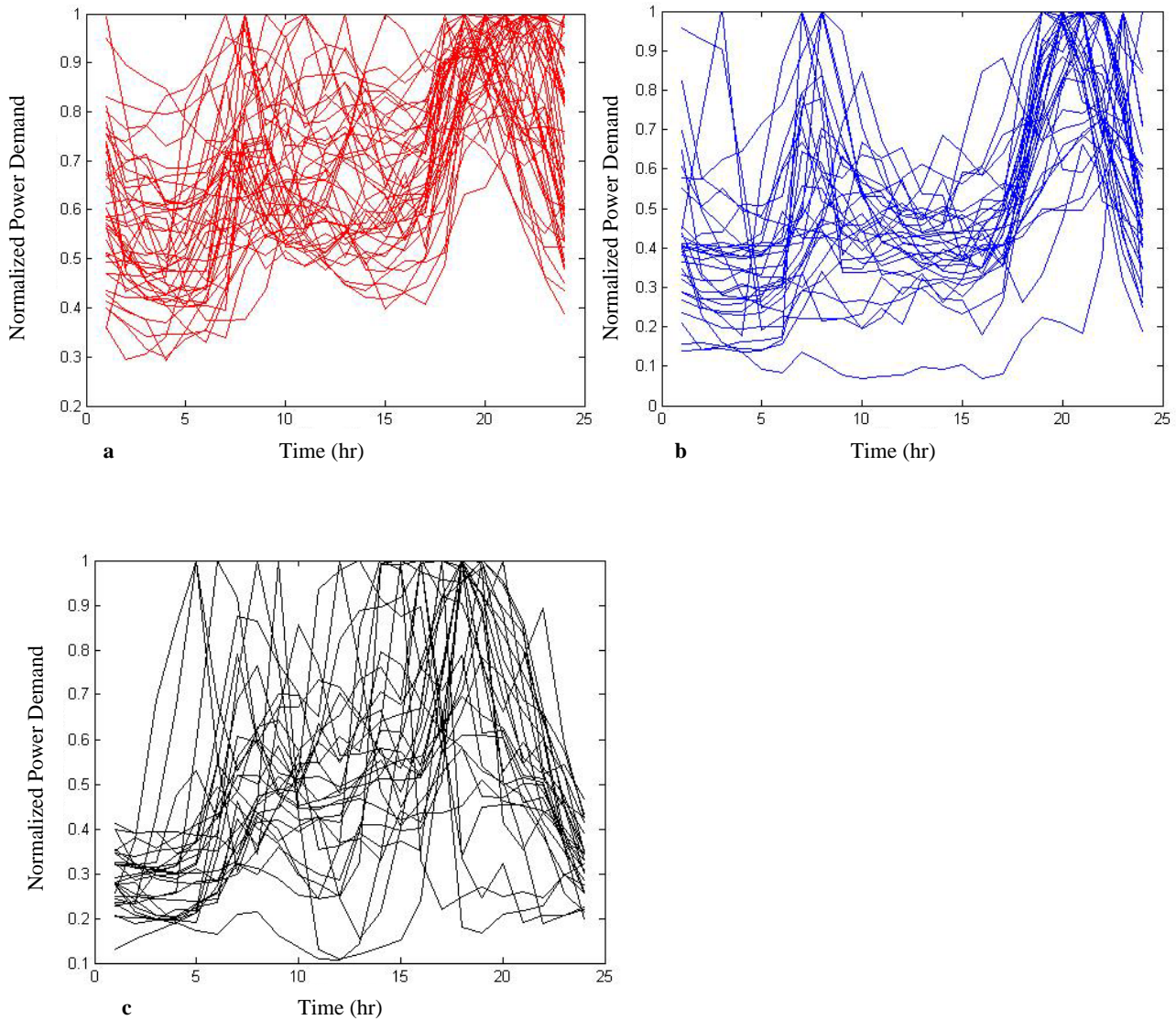


Figure 4.8a, b and c Power demand profile of houses included in each cluster in fall.

Last season for which the clustering analysis was performed is fall season. Figures 4.8a-c show the profile of users classified into each of the three clusters. Figure 4.9 is plot of average ERCOT price and profile of cluster centers for fall season.

The profiles of houses in cluster 1 Figure 4.8a show that this cluster has a local peak at around 8:00am when a number of houses have their peak electricity usage, while most of the houses have their peak demand between 5:00pm to 9:00pm. As a result the peak demand of cluster 1 occurs at 6:00pm. Cluster 2 on the other hand, has a bit higher start and it also has a

local peak at around 7:00am when some of the houses have their peak demand. Cluster 2 has its average peak demand at 7:00pm. Cluster 3 contains mostly those houses that do not have much electricity use before 6:00am, after which their usage increase and houses peak at different times of the day as can be seen in Figure 4.8c. This cluster has its peak demand at 7:00pm.

Fall season is also critical for utilities in terms of occurrence of market price peaks and load peaks at near ranges; Figure 4.9 shows when these peaks occur. ERCOT price peak occurs at 4:00pm, at which time, cluster 1 with mean variance of 0.33 is at 97% of its peak load; clusters 2 and 3 with mean variances of 0.035 and 0.037 are at 85% and 69% of their peak loads respectively.

In terms of cluster peaks if we compare these peaks with ERCOT prices we see that when cluster 1 is at its peak demand, ERCOT price is at 95% of its peak value. Similarly when clusters 2 and 3 are at their peak demands, ERCOT price is at 67% of its peak value. Here we see that cluster 1 is the most expensive electricity user when at its peak. The other two clusters use cheaper electricity because the ERCOT price decreases quickly in 2 to 3 hours from its peak value.

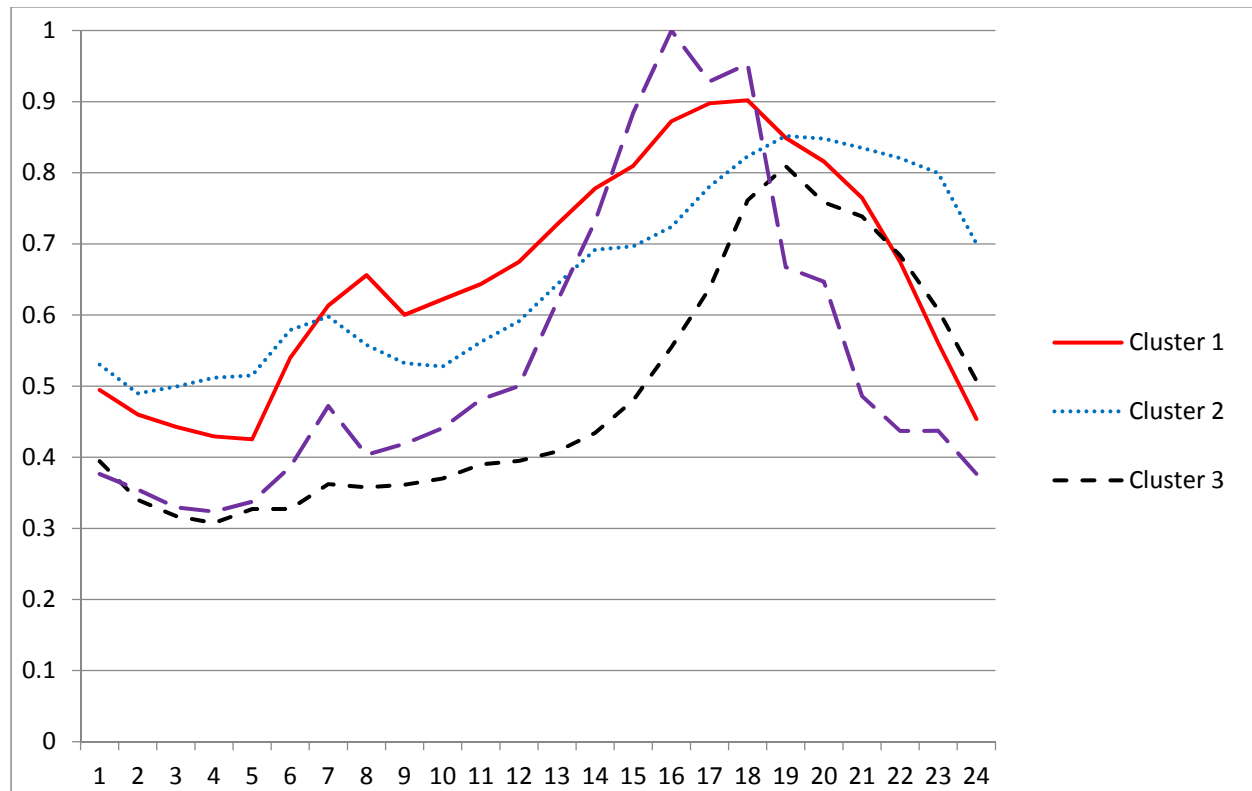


Figure 4.9 Plot of cluster centers and ERCOT SPP real time price average in fall.

Table 4 below is the list of houses included in each cluster in fall season. We can notice from this table that the number of low electricity users have increase a lot compared to all the other seasons, which is good for the utility. Cluster 1 in fall contains 29 houses, cluster 2 contains 20 houses and cluster 3 contains 52 houses, which is near double compared to the summer season. 68% of the houses in fall season stayed in the same cluster as they were in the summer. 27% of the houses changed their cluster to a lower level of electricity usage and only 5% of the houses changed their electricity use to a higher level compared to their summer usage level.

Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
1	4	2	96		48
8	7	3	98		49
9	20	5	101		50
16	26	6			51
17	27	10			52
22	44	11			53
31	45	12			58
32	55	13			63
33	59	14			64
34	62	15			66
38	67	18			68
39	70	19			72
41	75	21			73
42	81	23			74
54	82	24			76
56	90	25			78
57	92	28			79
60	94	29			83
61	97	30			84
65	100	35			86
69		36			87
71		37			88
77		40			89
80		43			93
85		46			95
91		47			99

Table 4 Clustering result for fall season

Table 5 below is the list of all houses with their cluster placements in all four seasons. This table gives us a good way of understanding the general behavior of houses in different seasons. We can see from this table that only 15% of the houses remained in the same cluster all throughout the year while 85% changed their electricity usage behavior based on each season. 63% of the houses were classified in two different clusters for the four seasons, meaning that they had been either in high or medium cluster or in low and medium cluster in different seasons.

House No.	Winter	Spring	Summer	Fall	House No.	Winter	Spring	Summer	Fall
1	1	3	1	1	29	3	3	1	3
2	3	2	3	3	30	2	2	1	3
3	3	2	2	3	31	1	3	1	1
4	1	1	3	2	32	1	1	1	1
5	1	2	3	3	33	3	1	1	1
6	2	3	1	3	34	1	1	1	1
7	1	1	2	2	35	1	2	3	3
8	3	3	1	1	36	2	2	2	3
9	3	1	1	1	37	2	2	3	3
10	3	2	2	3	38	3	1	1	1
11	1	3	1	3	39	1	1	1	1
12	1	1	1	3	40	3	3	3	3
13	2	3	1	3	41	2	2	2	1
14	1	2	2	3	42	3	3	1	1
15	2	2	3	3	43	1	2	3	3
16	1	3	1	1	44	1	1	2	2
17	1	1	1	1	45	3	1	2	2
18	2	1	1	3	46	2	2	3	3
19	3	2	2	3	47	3	2	3	3
20	1	1	2	2	48	2	2	3	3
21	2	2	3	3	49	2	1	2	3
22	3	3	1	1	50	2	3	3	3
23	2	3	1	3	51	1	1	1	3
24	2	2	3	3	52	3	3	3	3
25	2	2	3	3	53	1	2	2	3
26	3	2	2	2	54	3	3	1	1
27	1	1	2	2	55	1	1	2	2
28	3	3	1	3	56	1	1	1	1

Table 5 Clustering result for all houses in all seasons. (Continues on next page)

House No.	Winter	Spring	Summer	Fall	House No.	Winter	Spring	Summer	Fall
57	3	3	1	1	80	1	1	1	1
58	1	1	1	3	81	2	1	2	2
59	3	3	2	2	82	1	1	1	2
60	1	1	1	1	83	1	2	2	3
61	3	3	1	1	84	3	2	3	3
62	2	1	1	2	85	1	1	1	1
63	2	1	3	3	86	1	1	1	3
64	2	2	3	3	87	1	2	3	3
65	1	1	1	1	88	2	2	3	3
66	1	1	2	3	89	2	2	3	3
67	3	3	2	2	90	3	1	1	2
68	2	2	3	3	91	1	1	1	1
69	2	2	1	1	92	1	1	2	2
70	1	1	2	2	93	1	1	3	3
71	2	1	3	1	94	1	2	2	2
72	2	3	3	3	95	2	2	3	3
73	2	1	3	3	96	1	1	1	1
74	3	3	1	3	97	1	2	3	2
75	2	2	2	2	98	1	3	1	1
76	1	1	2	3	99	2	2	3	3
77	3	1	1	1	100	1	1	1	2
78	3	3	1	3	101	1	1	1	1
79	2	2	3	3					

Table 5 Clustering result for all houses in all seasons. (Continued from previous page)

Chapter 5 - Conclusion and Recommendation

This report is based on analysis performed on load data of 101 homes in Austin, TX. The data was first hourly and then seasonally averaged for each house. The seasonal average demand profiles were then normalized to observe the behavior of these houses based on their peak demands. K-mean clustering algorithm was deployed to categorize each house to one of the three clusters based on their pattern of electricity usage. And finally we compared the ERCOT price profiles with the average seasonal cluster centers.

We found that there are three types of users in Austin, TX based on their electricity usage patterns. A low percentage of users maintained their clusters throughout the year while majority of the users changed their cluster once. We concluded from this that electricity usage behavior do not stay the same but rather change from season to season. This change can be due to level of income, PV systems usage, type of heating and cooling systems, number of different electrical appliances and some other factors.

Also we found that winter and spring are not critical seasons for the utilities because the peak market prices and peak demand do not occur at the same time. While summer and fall are critical seasons in terms of market wholesale price and peak demand for utilities as they occur close to each other with an hour or two in between.

The results found in this analysis are much similar to the results found by Rhodes, J.D. et al. The difference in our analysis is that we used three clusters which reduces the variance in clusters allowing the medium valued house profiles be grouped into a third cluster, the medium cluster. The work of Rhodes, J.D. et al had only two clusters in which many houses were classified into one of the two clusters while they were quite different in their pattern, also the survey data analysis done by them cannot provide accurate information as the variance in the clusters are much higher.

This analysis can be used by utilities to reduce electricity use during peak times in summer and fall. Utilities can use similar analysis to identify the users in the highest cluster of electricity usage and find some real world solution on how to get these users to use less electricity at price peak times. Offering incentives to these users for shifting their load from the peak hours to other times could be one of such solutions.

References

- [1] Electric Reliability Council of Texas, <http://www.ercot.com/about/profile>
- [2] What is Pecan Street; 2014. <<http://www.pecanstreet.org/about/what-ispecan-street-inc/>>.
- [3] Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., Fofou, Z. and Bouras, A. (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *Emerging Topics in Computing, IEEE Transactions on*, 2(3), 267-279.
- [4] Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, 42(1), 68-80.
- [5] Räsänen, T., Voukantsis, D., Niska, H., Karatzas, K., and Kolehmainen, M. (2010). Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Applied Energy*, 87(11), 3538-3545.
- [6] Kim, Y. I., Ko, J. M., and Choi, S. H. (2011, April). Methods for generating TLPs (typical load profiles) for smart grid-based energy programs. In *Computational Intelligence Applications In Smart Grid (CIASG), 2011 IEEE Symposium on* (pp. 1-6). IEEE.
- [7] Kwac, J., Flora, J., and Rajagopal, R. (2014). Household energy consumption segmentation using hourly data. *Smart Grid, IEEE Transactions on*, 5(1), 420-430.
- [8] Panapakidis, I. P., Alexiadis, M. C., and Papagiannis, G. K. (2012, May). Electricity customer characterization based on different representative load curves. In *European Energy Market (EEM), 2012 9th International Conference on the* (pp. 1-8). IEEE.
- [9] Package NBClust; 2015. <<http://cran.r-project.org/web/packages/NbClust/NbClust.pdf>>.

[10] Rhodes, J. D., Cole, W. J., Upshaw, C. R., Edgar, T. F., and Webber, M. E. (2014). Clustering analysis of residential electricity demand profiles. *Applied Energy*, 135, 461-471.