Smart Grid Consumer Behavioral Model

Gaurav Kumar.
Under the guidance of Prof. S. A. Khaparde
Department of Electrical Engineering
Indian Institute of Technology Bombay
Email: 140070039@iitb.ac.in, sak@ee.iitb.ac.in

***Abstract*—Constructing behavior model and profiling consumers are important for power system planning and operation. This helps the utility and the consumer for the demand response program which is the main focus of the analysis. Authors attempt to cluster sample data using two traditional clustering techniques, k-means and Expectation Maximization (EM) clustering. Clustering was carried out using MATLAB software. 3-D surface plots are given to check the volatility of data sets. Results are compared using Performance index and conclusions are drawn.**

***Index Terms*—k-means, EM method, 3-D surface plot, Performance Index.**

A. Introduction

With the advent of Smart Meter Acquisition System, tons of smart meter data are being collected in the real time stamp. The data collected can play an important role to the distributed network operators (DNOs) in helping them understand the various energy profiles of the different categories of consumerss. Hence the analysis of the data plays an important role. For the analysis the smart meter data is clustered to classify the consumers based on their energy utilization pattern. By the means of clustering of the smart meter data lots of insights can be gained to understand the low voltage networks [1]. Better clustering corresponds to choosing the right features. Clustering of data has been helpful in

1) Better Tariffs : Checking the load profiles of the consumers which help in better identification of the hurtful households in turn helping them in cost mitigation and choosing better tariffs [1].
2) Planning : Helps in the decision making of the load curtailment by better load forecasting [2] [3].
3) Operation : Identification of suitable consumers for the demand response program, energy demand balance [1], energy profile modeling and management [4][5].
4) Asset Management

Smart meter data recorded can be classified based on various clustering algorithms [4]. The algorithms like the k-means,k-medoids,Expectation Maximization(EM) [4], hierarchical clustering algorithm, self organization map clustering algorithm [6], fuzzy clustering[7], dynamic clustering [8] are used in analyzing of the household data for different applications. The comparison of various algorithms can be found in [6]. Selection of features is one of the major challenges in clustering algorithm. Feature selection varies with different application [4]. In the work reported, features are selected keeping demand response as the main objective to be achieved which eases the decision making for the DNOs. Different household data has been clustered to identify the behaviour of different users which indeed can be used by DNOs for the demand response program. In this work k-means and Expectation Maximization (EM) algorithms are used. A brief methodology of these algorithms is presented in the Section I, followed by the sample system selection and feature selection in Section II. Results are briefed in Section III. Comparison of the results using both the algorithms is done in Section IV. Conclusion is given in Section V and finally future work is stated in Section VI.

I. CLUSTERING ALGORITHM

The two methods used for the analysis of the sample system. First method is the k-means clustering algorithm and second method is the Expectation Maximization (EM) algorithm. The brief description of both the methods are given below

A. k-means algorithm

The k-means is the widely used method because of its ease of implementation and capacity to handle large data set [9]. The methodology followed by the k-means in clustering of the data is [9] :
Input : With number k and database containing n objects
Output : Set of k-clusters minimizing the squared-error criterion.

1) Step 1: Choose k objects arbitrarily as the initial centroid of the cluster.
2) Step 2: Reassign each object to the cluster to which its most similar based on the calculation of the mean value of the objects in the cluster.
3) Step 3: Update the cluster mean, do it till the value of the mean calculated for the cluster remains constant.

B. EM algorithm

The EM algorithm stems from Gaussian Mixture model (GMM)[9]. GMM method improves the density of the given data set to model the distribution of the data set by modeling it as a function of probability density of a single-density estimation method with multiple Gaussian probability density function. The methodology followed is [9] :
Input : k: Cluster number, data set, Stopping tolerance
Output : A set of k-clusters with weight that maximizes the log-likelihood function.

1) Step 1: Expectation step: For each data set record x, compute the membership probability of x in each cluster h = 1,. . . ,k.
2) Step 2: Maximization step: Updating the probability weight of the mixture model.
3) Step 3: Stopping criteria: If stopping criteria is satisfied then stop, else set j = j +1 and return Step 1.

By using the above two methods,clustering is performed on the following sample system.

II. SAMPLE SYSTEM

The sample system contains the energy data of 5567 London households that participated in the UK Power Networks led by the Low carbon London project [10]. For the analysis, mixture of the data of different classes of people is considered. The three different classes of people involved in the analysis are affluent achievers (lavishly living people, rich people), comfortable communities (middle-class), financially stretched (lower middle class) [11]. In total 186 consumers are considered for the analysis. The data collected is at an interval of 30 minutes, so 48 time stamps a day.

Figure 1 gives the 3-Dimensional surface plot of total power consumption per half hour(kW) over a year for all the 186 consumers. From the figure, different colors indicate different power consumption range, red color being the highest. The plot gives an idea that data is highly volatile and clustering of these data is quite a challenging task.

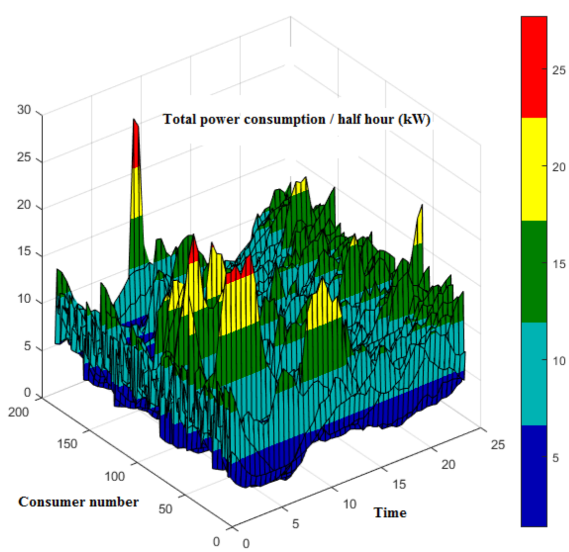

Figure 1: 3-D surface plot of total power consumption of all consumers

A. Feature Selection

After the availability of the data, selection of the features play a crucial role. The features are selected according to the end goal of the analysis. The procedure involved in the selection of features for the analysis is:

1) Normalization of the data

$$a = \sum_{t=1}^{48} l(t) \text{ and } s(t) = \frac{l(t)}{a} \quad (1)$$

where,
a is the daily total consumption,
$s(t)$ is the normalized data profile.

2) A representative curve of the normalized power data value for individual consumer has been plotted by adding all 365 days data of a particular time stamp. So a total of 48 time stamps in a year.
3) By using the representative curve of each consumer, three features have been extracted. They are as follows:
   a) Feature 1 (F1)- The peak power consumption : The maximum consumption of the each consumer is calculated. A total of 186 peak points are obtained.
   b) Feature 2 (F2)- Ratio of peak to off-peak : Ratio of peak power consumption to off peak power consumption is calculated for each consumer. Total of 186 ratios are obtained.
   c) Feature 3 (F3)- Ramp rate : For the calculation of ramp rate, 00:00hrs is considered as 1^{st} instant, 11:30hrs as 23^{rd} instant and 23:30hrs as 48^{th} instant. Formulae for the calculation of the ramp rate is:

$$\frac{P(i) - P(i-4)}{4}, \text{ if } i > 4$$

$$\frac{P(i) - P(44+i)}{4}, \text{ if } i \leq 4$$

   where,
   i is the time instant of maximum power consumption,
   $P(k)$ is the power consumption at the k^{th} time instant where $k = 1, 2, \ldots 48$.

From the sample system, the features like the peak power consumption, ratio of maximum to minimun power consumption and ramp rate are extracted. These features help the utility to take certain decision for the demand response program.

III. RESULTS

The extracted features are clustered by using k-means and EM algorithm in MATLAB. Two features were considered at the time of clustering.

A. Features F1 & F2

Firstly, peak power consumption and ratio of maximum to minimun power consumption is considered. Results using k-means method and EM are given below:

3 clusters are obtained from k-means algorithm as shown in Figure 2. Cluster 1 contains 13 consumers, cluster 2 contains 97 consumers and cluster 3 contains 76 consumers. This implies that there are 13 consumers(Cluster 1) whose peak power consumption is high followed by cluster 3,cluster 2 respectively.

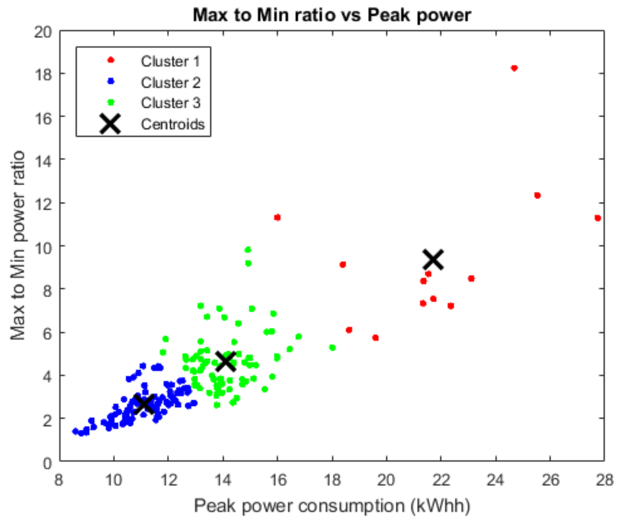

Figure 2: maximum to minimun power ratio vs Peak power consumption (k-means)

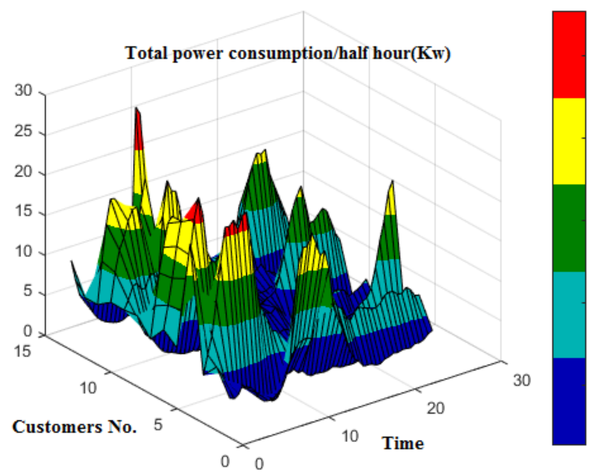

Figure 3: 3-D surface plot of total power consumption per half hour of all consumers in cluster1(k-means)

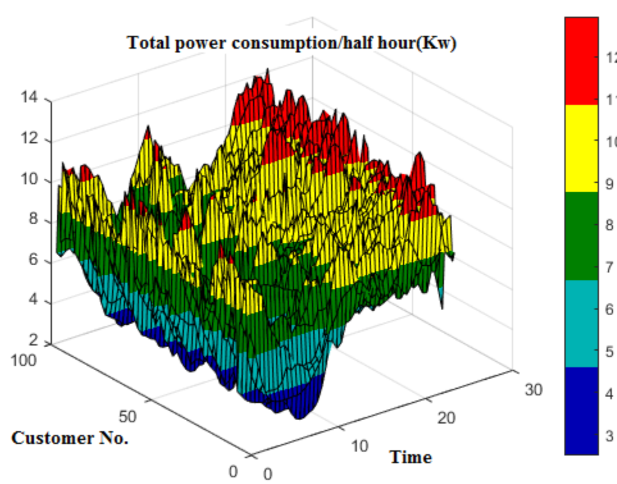

Figure 4: 3-D surface plot of total power consumption per half hour of all consumers in cluster2(k-means)

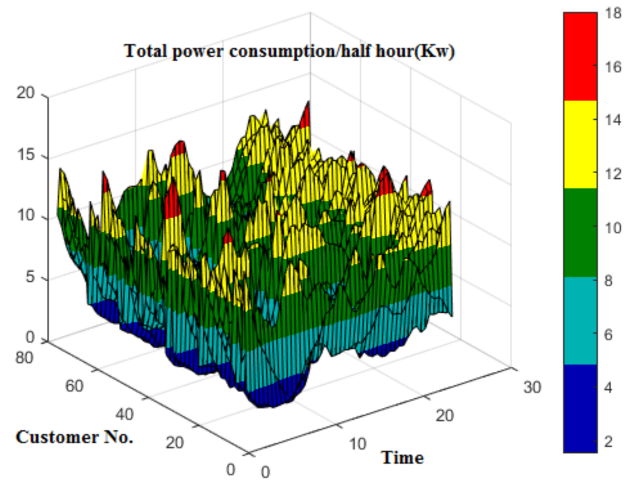

Figure 5: 3-D surface plot of total power consumption per half hour of all consumers in cluster3(k-means)

Figure-3,Figure-4 & Figure-5 represent 3-D surface plots for different clusters obtained using k-means algorithm considering features F1 & F2.

3 clusters are obtained from EM algorithm as shown in Figure 6. Cluster 1 contains 88 consumers, cluster 2 contains 81 consumers and cluster 3 contains 17 consumers. This implies that there are 17 consumers(cluster 3) whose peak power consumption is high followed by cluster 1,cluster 2 respectively.

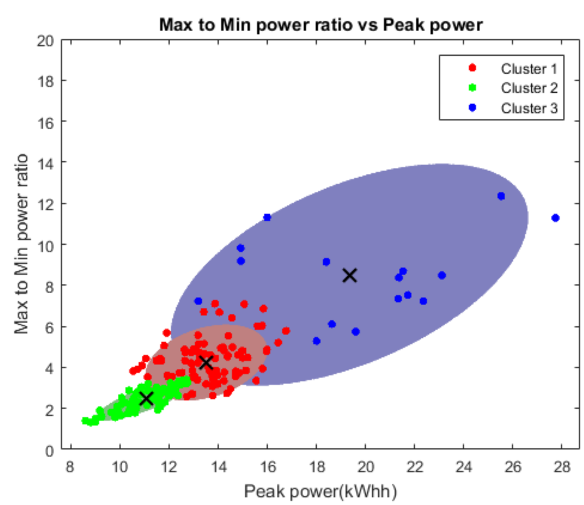

Figure 6: Maximum to minimun power ratio vs Peak power consumption (EM)

From the above considered features, if the ratio of maximum to minimun power consumption is high this implies that these particular set of consumers are good candidates for the demand response program. The utility can use this data for peak shaving.

B. Features F1 & F3

Secondly, peak power consumption and the ramp rate are chosen for the clustering algorithm. Results by using the k-means method and EM are given below:

3 clusters are obtained from k-means algorithm as shown in Figure 7. Cluster 1 contains 84 consumers, cluster 2 contains 89 consumers and cluster 3 contains 13 consumers. This implies that there are 13 consumers(cluster 3) whose peak power consumption is high followed by cluster 1,cluster 2 respectively.

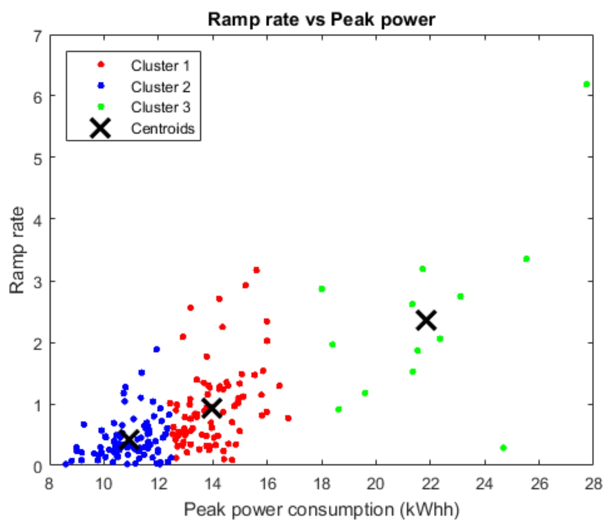

Figure 7: Ramp rate vs Peak power consumption (k-means)

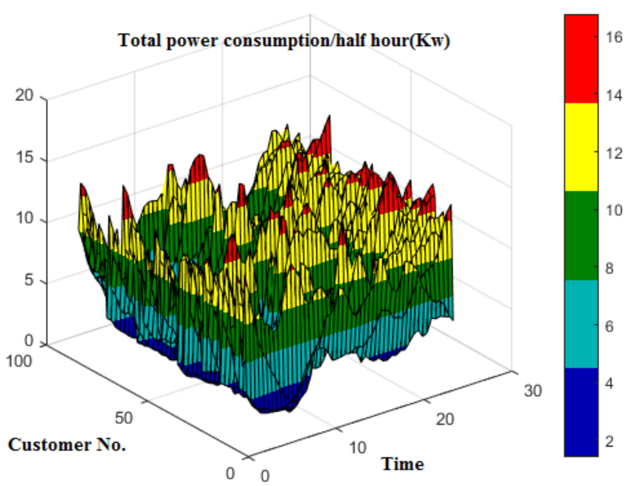

Figure 8: 3-D surface plot of total power consumption per half hour of all consumers in cluster1

Figure-8,Figure-9 & Figure-10 are 3-D surface plots for different clusters obtained using k-means algorithm considering features F1 & F3.

3 clusters are obtained from k-means algorithm as shown in Figure 11. Cluster 1 contains 57 consumers, cluster 2 contains 20 consumers and cluster 3 contains 109 consumers. This implies that there are 20 consumers (cluster 2) whose peak power consumption is high followed by cluster 1,cluster 3 respectively.

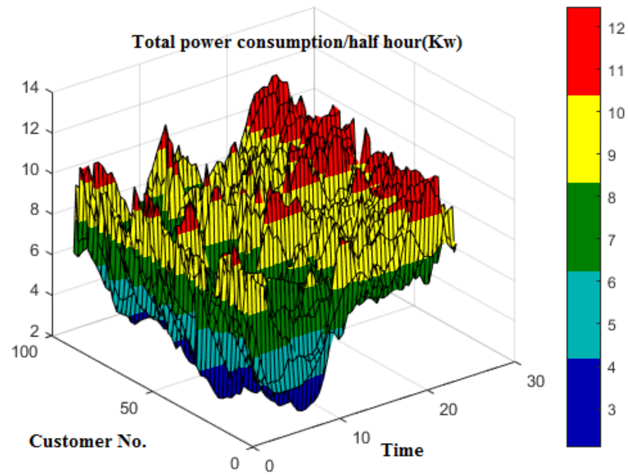

Figure 9: 3-D surface plot of total power consumption per half hour of all consumers in cluster2

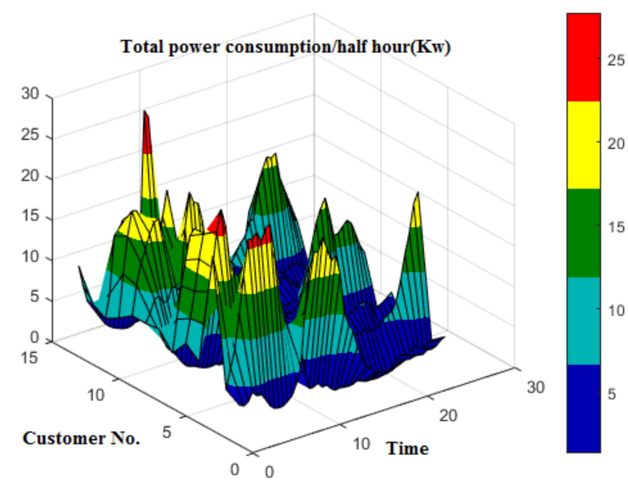

Figure 10: 3-D surface plot of total power consumption per half hour of all consumers in cluster3

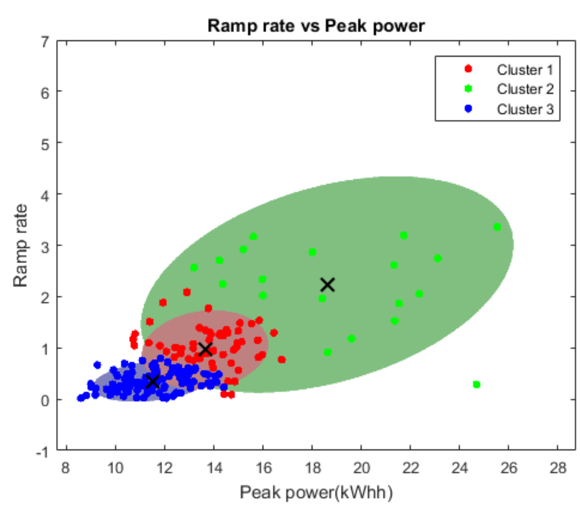

Figure 11: Ramp rate vs Peak power consumption (EM)

From the above considered features, if ramp rate is high this implies that the generator load balancing will be difficult for these particular set of consumers. So, in this regard the utility should be prepared well in advance to meet the load requirements.

IV. COMPARISON OF RESULTS

As seen above the clustering results obtained by using both the methods are slightly different. Both the methods use different techniques to group the consumers. Hence, it becomes difficult to assess which technique gives the best results. In this aspect, a small analysis has been performed to check which clustering result is good. In this analysis, the euclidean distance between the centroids of the three clusters have been calculated. The spread of consumers within the cluster has been evaluated by standard deviation. So, if distance between the centroids is more and spread of clusters is less then the clusters formed are good.

By calculation,

Table 1: Considering features F1 & F2

	Centroid(C_i)	Standard Deviation	$\|\|C_i - C_j\|\|_2$
k-means	21.6916, 9.3705	3.1604, 3.3311	12.5512
	11.1018, 2.6336	1.0223, 0.7224	3.6340
	14.1281, 4.6454	1.1201, 1.3556	8.9182
EM	13.6528, 4.3247	1.3269, 1.0414	3.2919
	10.9645, 2.4250	1.0117, 0.5398	11.3302
	20.1778, 9.0197	4.0200, 3.0879	8.0385

Table 2: Considering features F1 & F3

	Centroid(C_i)	Standard Deviation	$\|\|C_i - C_j\|\|_2$
k-means	13.9878, 0.9373	1.0488, 0.6514	3.0933
	10.9395, 0.4115	0.9137, 0.3395	11.0802
	21.8463, 2.3661	2.8967, 1.4661	7.9873
EM	13.8576, 1.0430	1.4161, 0.3800	5.7428
	19.4288, 2.3481	4.1319, 1.1907	8.1936
	11.5058, 0.3481	1.3403, 0.1885	2.4523

From the Tables 1 & 2, it is observed that the distance between the centroids obtained by the k-means method is more sparse and the spread of the consumers in the clusters is less compared to EM method.Assuming that shape of clusters are elliptical, performance index have been calculated for different clustering algorithm.If Performance Index is high, Clusters formed using one algorithm for the chosen features are more efficient than other clustering algorithm.

Performance index:- It is defined as distance between centroids divided by sum of area of respective clusters. i.e.,

$$PI = \frac{||C_i - C_j||_2}{A_i + A_j} \tag{2}$$

where,
$i, j = 1, 2, 3$ and $i \neq j$
A_i is area of a cluster where $i = 1, 2, 3$
.

Table 3: Considering features F1 & F2

Method	PI_1	PI_2	PI_3	PI
k-means	0.3546	0.5125	0.2357	1.1028
EM	0.5435	0.2783	01855	1.0073

Table 4: Considering features F1 & F3

Method	PI_1	PI_2	PI_3	PI
k-means	0.9912	0.7740	0.5157	2.2809
EM	0.3349	0.5042	0.9870	1.8262

V. CONCLUSION

The clustering of the stochastic smart grid data of a sample system has been performed by using k-means and EM method in MATLAB. Features assumed in the analysis are the peak power consumption, ratio of maximum to minimun power consumption and ramp rate. It is observed that three distinct clusters are obtained by using both the methods. If the ratio of maximum to minimun power consumption is high, it implies that these particular set of consumers are good candidates for the demand response program. The utility can use this data for peak shaving. And, if ramp rate is high, it implies that the generator load balancing will be difficult for these particular set of consumers. So, in this regard the utility should be prepared well in advance to meet the load requirements. This information helps the utility for the demand response program. Coming to the choice of method best suited to perform clustering, a small analysis was done to compare both the methods. The euclidean distance between the centroids and the spread of the consumers were calculated using standard deviation.If the distance between the centroids is more and the spread is less, the clusters formed are good. From Tables 3 and 4, It is observed that the performance index obtained from k-means method is more as compared to EM method. Hence, it can be concluded that k-means clustering method is more efficient than the EM method in this particular analysis. However, due to the volatility of the data it becomes difficult to identify the best clustering method.

VI. FUTURE WORK

The analysis carried out was keeping in mind a particular application that is the demand response program. The work will further be expanded for different applications by dynamic clustering algorithm which identifies the most suitable algorithm and the features. The algorithm is tested by varying the weights given to the features.

REFERENCES

[1] H. Â. Cao, C. Beckel, and T. Staake, “Are domestic load profiles stable over time? an attempt to identify target households for demand side management campaigns”, in *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, Nov. 2013, pp. 4733–4738.

[2] M. Chaouch, “Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves”, *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 411–419, Jan. 2014.

[3] B. Stephen, A. J. Mutanen, S. Galloway, G. Burt, and P. Järventausta, “Enhanced load profiling for residential network customers”, *IEEE Transactions on Power Delivery*, vol. 29, no. 1, pp. 88–96, Feb. 2014.

[4] S. Haben, C. Singleton, and P. Grindrod, “Analysis and clustering of residential customers energy behavioral demand using smart meter data”, *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 136–144, Jan. 2016.

[5] A. Albert and R. Rajagopal, “Smart meter driven segmentation: What your consumption says about you”, *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4019–4030, Nov. 2013.

[6] O. A. Abbas, “Comparison between data clustering algorithms”, *The Internationala Arab Journal of Information Technology*, vol. 5, 8 Jul. 2008.

[7] Y. Sun, W. Gu, J. Lu, and Z. Yang, “Fuzzy clustering algorithm-based classification of daily electrical load patterns”, in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, Aug. 2015, pp. 50–54.

[8] I. Benítez, A. Quijano, J.-L. Díez, and I. Delgado, “Dynamic clustering segmentation applied to load profiles of energy consumption from spanish customers”, *International Journal of Electrical Power & Energy Systems*, vol. 55, no. Supplement C, pp. 437–448, 2014, ISSN: 0142-0615. DOI: https://doi.org/10.1016/j.ijepes.2013.09.022. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0142061513004043.

[9] Jung, Y. Gyu, M. S. Kang, and J. Heo, “Clustering performance comparison using k-means and expectation maximization algorithms”, *Biotechnology, Biotechnological Equipment*, vol. 28, S44–S48, sup1 2014.

[10] *Smartmeter energy consumption data in london households*. [Online]. Available: https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households.

[11] *The acorn user guide*. [Online]. Available: https://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf.