

Approach for smart meter load profiling in Monte Carlo simulation applications

Khan, Zafar; Jayaweera, Dilan

DOI:

[10.1049/iet-gtd.2016.2084](https://doi.org/10.1049/iet-gtd.2016.2084)

License:

Creative Commons: Attribution (CC BY)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Khan, Z & Jayaweera, D 2017, 'Approach for smart meter load profiling in Monte Carlo simulation applications', *IET Generation, Transmission and Distribution*, vol. 11, no. 7, pp. 1856-1864. <https://doi.org/10.1049/iet-gtd.2016.2084>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

Khan, Z.A. and Jayaweera, D. (2017), Approach for smart meter load profiling in Monte Carlo simulation applications. *IET Gener. Transm. Distrib.*, 11: 1856-1864. <https://doi.org/10.1049/iet-gtd.2016.2084>

Final Version of Record available online at:
<http://dx.doi.org/10.1049/iet-gtd.2016.2084>

Checked 8/5/2017

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

New Approach for Smart Meter Load Profiling in Monte Carlo Simulation Applications

¹Zafar A. Khan and ²Dilan Jayaweera

¹email: kxz511@bham.ac.uk

²email: d.jayaweera@bham.ac.uk

Abstract—Smart grids introduce new technological elements into power systems which take prevalent challenges to a new level by shaping parameters of power systems towards a complex regime of uncertainties. Rapid proliferation of advanced metering infrastructure (AMI) and integration of renewable energy sources (RES) in smart grids increase system-wide complexities. This paper proposes an innovative approach to classify the energy consumptions of smart meter customers with typical profiles by processing with multi-layered clustering of energy consumption data of smart consumers extracted from the AMI. There are two stages for the approach of which the first stage analyses the data for intra-cluster similarity of energy consumption patterns and in case the patterns do not have a high intra-cluster similarity, they are fed back for re-clustering with multi-layered clustering process until the clearly identifiable energy patterns with high intra-cluster similarity is obtained. The second stage linearizes the complex energy patterns using interpolant and curve fitting techniques until stabilised profiles are obtained. The paper also proposes a methodology for smart meter load modelling for Monte Carlo (MC) simulation applications to reduce the computing time compared with traditional alternatives. The paper validates the robustness of the approach and provides the corroboration of the method for MC simulation applications in a smart grid environment.

Index Terms-- Data Clustering, k-means, Load Curves, Load Profiles, Monte Carlo Simulation, Smart Grids

I. INTRODUCTION

With the development of smart grids and in response to carbon dioxide (CO₂) reduction targets, penetrations of intermittent RES, particularly wind and solar power have been intensified. Given the decrease in dependency on fossil fuels and high penetration of intermittent RES with highly variable nature of power generation, the security, reliability, and stability of a power system can be affected considerably. Moreover, with changes on generation side, the consumption side is also evolving with smart loads like electric vehicles etc. The AMI provides a new avenue of opportunity to learn from the recorded historical data of both generation and consumption of smart consumers by incorporating it for system studies. However, exploiting large volumes of data extracted from smart meters is a challenging task particularly with stochastic nature of electric loads. Traditional techniques are limited in exploiting this big data at consumer levels for reliable and robust system studies.

With the development of AMI, an abundant amount of consumer load data from smart meters have become available which is typically recorded at 15 or 30 minutes intervals [1]. The increasing aggregated volume and rate of AMI data has enhanced the complexity of data challenges by limiting the flexibility of data analysis and extraction of specific detailed information. The smart meter data can be used to gain an in-depth view of customers' energy consumption behaviour by analysing the consumers' load profiles. However, with a large

number of consumers, it is not possible to analyse consumer load profiles at individual level. Therefore, typical profiles of consumers are required which can represent a standard consumer energy consumption patterns that can further be used for planning and operation of a power system [2]. Extraction of such typical-profiles from voluminous data is a multifaceted challenge. Such challenges can be dealt with using computationally advanced extended data mining techniques such as neural networks, data clustering, genetic algorithm etc.

The electricity consumption data is unlabelled and stochastic in its nature, the application of unsupervised learning using probabilistic techniques including data clustering is more suitable for data mining rather than deterministic techniques. Data clustering enables extraction of the required information from a big data like smart meter data. Data clustering can classify different consumers without any supervision. This classification can be utilized for different purposes including load modelling, load forecasting, load flow analysis, system security and reliability studies etc. Different authors classify clustering techniques in a different way. However, for load profiling, clustering techniques can be divided into direct and indirect clustering [3]. Direct clustering involves direct application of the clustering technique to data and in indirect clustering it is based on a dimension reduction technique. Classification of data for typical load profiling is discussed in detail with clustering techniques in [3]. One of the issues of the indirect clustering technique is that there can be a potential chance of feature loss which might lead to a lower accuracy in representation of the original profile.

However, most of the work in the past regarding energy profiling (using data clustering) utilized grid or transformer level data to extract the load profiles. In these cases particularly the issue of intra-cluster pattern dissimilarity was not shown significantly affecting the overall profile because of the scale of the loads. Authors of reference [4] used k-means clustering for energy demand forecasting based on the consumer behaviour similarities. Li and Wolf [5] clustered high frequency components of the load using hierarchical clustering and for load modelling. Reference [6] carried out a comparative analysis of some data clustering techniques including k-means, hierarchical, k-means and fuzzy c-means clustering and established that amongst the previously mentioned techniques, k-means clustering was superior to others in terms of processing speed and robustness. Reference [7] used k-means for study of category based consumer load profiling with dimension reduction techniques namely principal components analysis and curvilinear component analysis. Therefore, k-means clustering is frequently used as a robust clustering method in different studies, including load profiling. This paper incorporates k-means data clustering using smart meter data. Smart meter consumer data complicates the clustering process with high volume and frequency of data at individual customer level, showing significantly higher levels of dissimilarity in intra-cluster patterns which is often overlooked in other similar studies.

The typical profiles can be used for MC simulation applications for system studies including probabilistic load flow, security, and reliability. MC simulations simulate random numbers for the stochastic variables and different variable parameters of the system models and analyses deterministically at each sample trial to estimate parameters that are intended. As energy consumption data of consumers are stochastic in nature and for such data, probabilistic techniques produce more realistic results than deterministic techniques, hence MC simulation is widely used for power system studies. MC simulation are generally favourable for complex problem estimations [8] and are highly effective for power system reliability and security studies [9]. References [10-15] used MC methods for network security assessment. Carpinelli et al. [16] used the MC simulation in the

probabilistic load flow studies in the presence of intermittent generation including solar and wind power generation. Chen et al. [17] discussed MC methods in detail for the application of probabilistic load flow study.

Previous studies which use data clustering for load profiling, do not address the issue of intra-cluster pattern dissimilarity, which can potentially improve the estimation of parameters in a Monte Carlo simulation platform. The dissimilarities are often overlooked and typical load profiles are generated despite significantly dissimilar patterns. By taking into account this gap, the paper proposes an innovative technique for data abstraction while preserving the original features of data by applying multi-layered clustering methodology which tends to minimize the intra-cluster pattern dissimilarity for the extraction of typical load profiles. The added merits of the approach include the improved accuracy, simplicity and flexibility for the stochastic applications. Further to these merits, this approach not only reduces the data by the application of data clustering but also defines the energy classification which can be used to model energy threshold levels in such a way that it can reduce the processing time of a MC simulation.

II. THE APPROACH

As given by [18] owing to the applicability, proven robustness, higher processing speed, and proven efficiency, a direct clustering technique namely k-means clustering is incorporated in this approach for the analysis of energy consumption data of smart meters in the first stage and then it is extended to extract clear load patterns. Another added advantage of using k-means clustering in this study is the fast processing speed of integrated operation of k-mean clustering and MC simulation. k-means clustering technique is highly reliant on the proximity/distance measure used for the clustering process. Some of the most commonly used proximity measures include Euclidean distance, Squared Euclidean distance, Minkowski distance and Mahalanobis distance. A number proximity measures can be used in k-means clustering, however, Euclidean distance is the most commonly used proximity measure. Euclidean distance computes the distance between a cluster and centroid using the sums of squares of distances. Therefore, this study incorporates, Euclidean distance as proximity measure.

The proposed approach for load profiling involves many processing stages including data clustering, pattern extraction and curve smoothing. Initially in the k-means clustering, the 'ISSDA CER' dataset [19] which comprises of smart meter data from more than 5,000 homes and businesses in Ireland with a resolution of 30 minutes is populated. The dataset X consists of m time series records, described by n attributes $A = \{A_1, A_2, A_3, \dots, A_n\}$, where A_i is the attribute of consumption data observed time T_i . In total seven thousand random samples were extracted from the dataset to develop and validate the methodology. As the data is extracted from a large dataset which consists of six files with each file containing more than two million rows of data, to ensure the randomness of data, data used in this study was extracted from all six files. The extracted data was then checked for missing values by visual inspections. In some cases, days had some missing data entries. The missing data values can possibly be the result of faulty data collection instruments, errors in data transmission or any other problem at sending or receiving end. Further to this, another possibility of missing values can be originated from day light savings [20]. Different possible options for dealing with the missing values include filling the missing values using extrapolation techniques or excluding samples with missing values. In this work, data from such days was removed instead of extrapolating the missing values to ensure preservation of the original features of the data because of uncertainty in data. After removing the days with

missing load values, 6934 out of total 7000 half-hourly energy consumption data samples were finalized for data clustering. The finalized data samples were structured into the shape of $p \times n$ matrix with p number of rows and n number of columns. A plot of the finalized energy consumption data i.e. 6934 samples with time of day ranging from 0-24 hours on x-axis and load in kWh/30 minutes on y-axis is shown below in Fig. 1.

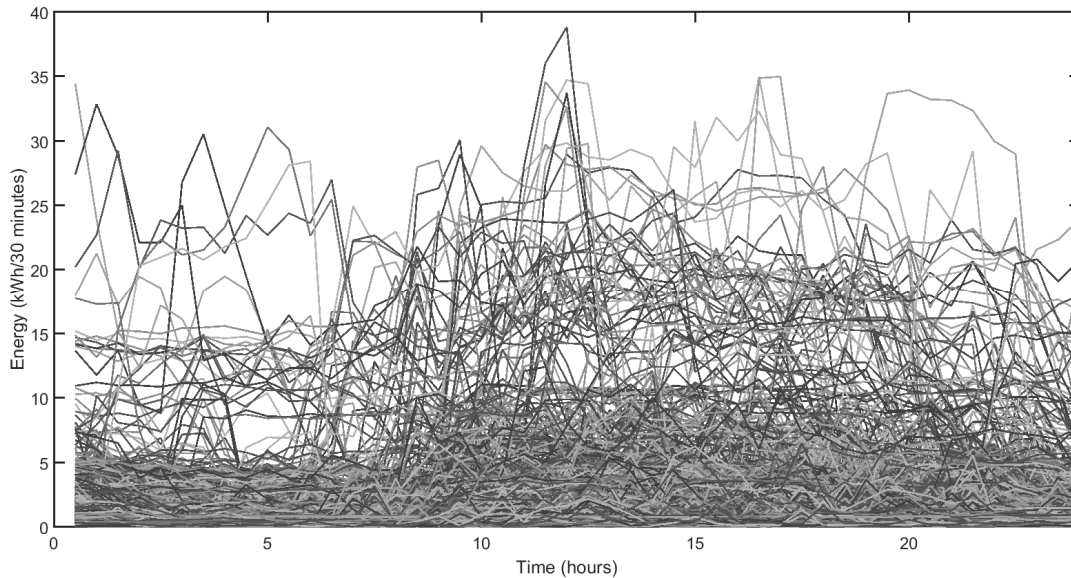


Fig. 1. Energy consumption patterns of 6934 Customers [18]

In Fig. 1, it can be noted that a substantial amount of loads is concentrated at the region below 10 kWh [18]. Extracting specific profiles from such loads is a challenging task as to determine the energy concentrations that are at the dominant levels requires going into the in depth information of the energy consumption. For example, in this case from Fig. 1 due to higher energy concentration in the region below 10 kWh, the load below 10 kWh can easily be perceived as one cluster despite having large variations within the region [18].

To resolve such intricate data challenges this paper introduces an innovative approach. To enhance the reliability, robustness, and effectiveness of the clustering process for MC simulation applications, new approach is developed which is presented in Fig. 2. The approach consists of two parts, part (A) was developed in our previous work [18] and part (B) develops an approach to use the resultant load profiles for MC simulation applications. In this part (A) of the approach, after the data pre-processing, the smart meter data is subjected to k-means clustering as given in steps i to vi.

Let $X = \{x_1, x_2, x_3, \dots, x_m\}$ be the data set with m instances, and let $A_1, A_2, A_3, \dots, A_k$ be the k disjoint clusters of X [18].

- i. Select k cluster centres randomly.
- ii. Calculate the distance between each data point and cluster centres.
- iii. Assign the data point to cluster with minimum distance to its centre
- iv. Re-compute the new centre with:

$$A_i = (1/C_i) \sum_{j=1}^{C_i} x_{ij} \text{ Where ' } C_i \text{ ' represents number of data points in } i^{\text{th}} \text{ cluster.}$$

- v. Re-compute the distance between each data point and new cluster centres.
- vi. If any data point member changes cluster membership, repeat the procedure from iii until no change in cluster membership occurs.

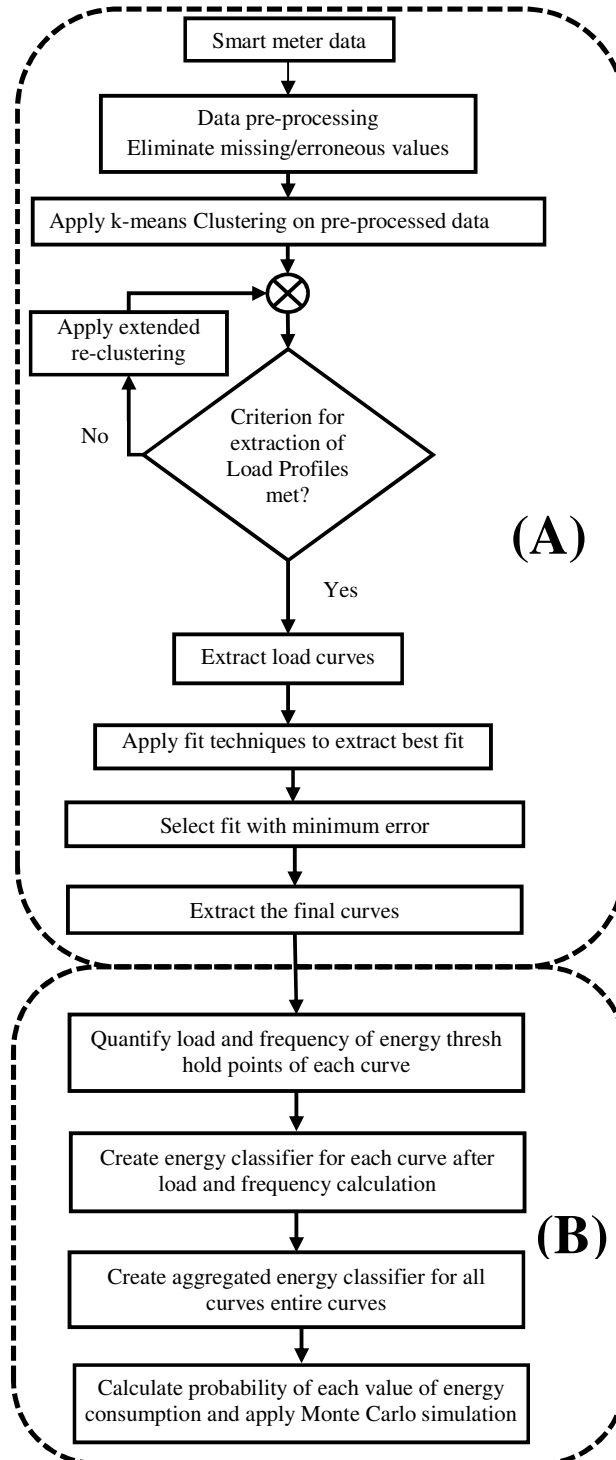


Fig. 2. Extraction of Load Profiles for Monte Carlo simulation (Part (A) was developed in our previous work [18])

Upon completion of the process, the resultant clusters are evaluated for their fitness for the extraction of typical load profiles [18]. If the intra-cluster similarities of energy consumption patterns are such that only clear energy consumption patterns with reasonably small variations from each other are allocated the same cluster, the resultant cluster is considered as decisive cluster for load profile extraction. However, if the cluster members have significant variations in their energy consumption patterns, they are applied with re-clustering procedure in order to minimize the error function [18].

The resultant clusters having higher intra-cluster pattern dissimilarity are applied with re-clustering process. This is achieved by feeding back the clusters containing dissimilar energy consumption patterns to step i of clustering process as input. Each output of the clustering and re-clustering process is checked and fed back for re-clustering until high intra-cluster energy consumption patterns similarity is obtained [18].

Arithmetic mean of the clusters with high intra-cluster pattern similarity is taken to generate a typical load profile [18]. However, the generated profile of the clusters can have substantial variations in it over the duration of curve, which does not suit for the purpose of extraction of typical load profiles. To minimize the variations over the time duration, the resultant curves are smoothed by applying best-fit functions. Depending on the non-linearity level of the resulting curve data, different interpolant and/or the curve fitting techniques are applied as appropriate [18]. They include interpolant and polynomial techniques including Spline interpolant, shape-preserving interpolant, and 4th, 5th, 6th, 7th, 8th, 9th and 10th degree polynomial. In this way, to avoid over-fit or under-fit, different interpolant and degrees of polynomials are applied on different curves for the curve fitting after detailed analysis. To ensure the accuracy of the fitted curve i.e. best-fit curve, after contemplation of different possibilities, the curve with minimum error was obtained. The fitting process is carried out for all the curves obtained from final clustered and final re-clustered data.

The typical load curves extracted using the above approach can be used for load modelling, load flow analysis or power system applications which may incorporate MC simulation platform. However, one of complexities of MC simulation is considerable processing time which can be reduced by classifying the energy threshold levels from the typical load profiles. In this approach, to transcribe the load curve into energy threshold classifier, the extraction of energy consumption values from energy demand curves is done by linearly best fitting the non-linear profiles in such a way that the entire profile is represented with linear set of profiles.

The energy classification process for quantification of the magnitude and frequency of the load at energy threshold points is shown in Figure 3. In the first part of the process, the individual profile is linearized by linearly best fitting the profiles in such a way that entire profile is represented with linear set of profiles. From the linearized profiles, magnitude and frequency of the energy threshold points are extracted in the third step of the process. Finally the energy classifiers are defined which carry the information of the magnitude of load, frequency and classification number according to the individual profile.

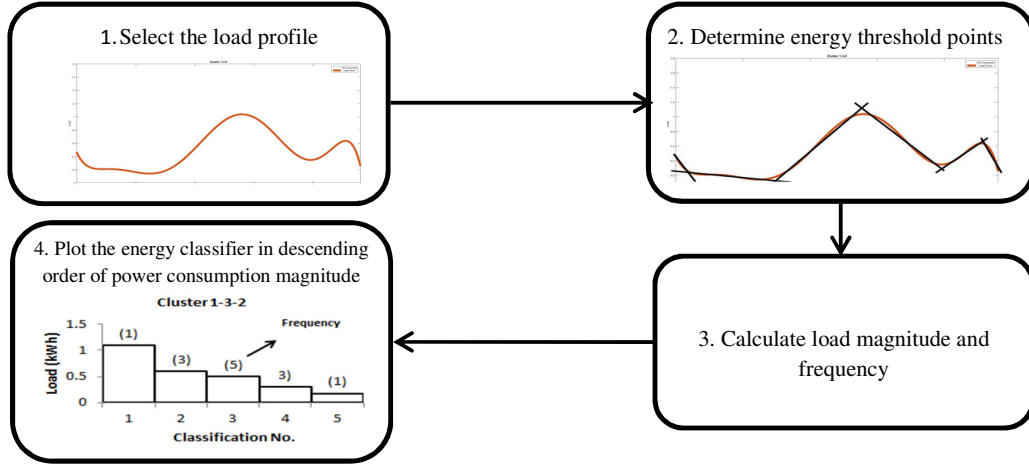


Fig. 3. Energy classification process

The same procedure is adopted for all load curves to classify the energy. Upon completion of the individual energy classification, the individual energy classifiers are aggregated to generate a single energy classifier, which retains the magnitude of load and facilitates to use MC simulation applications.

III. NUMERICAL APPLICATIONS

This section delineates the application of clustering and extraction of load profiles and making the pathways for MC simulation applications. K-means clustering was applied for the entire data population resulted through the data pre-processing discussed in section II. The k-means clustering faces a challenge of selection of the initial number of clusters (k) which are user defined. The selection of initial number of clusters has a significant impact on final clustering solution. Different indices/criteria are given in the literature to estimate the number of clusters such as Calinski–Harabasz Index, Gap Statistic, Silhouette Coefficient and many more [21]. Two different most commonly used techniques namely CalinskiHarabasz and silhouette were used to evaluate the optimal number of clusters. However, it was observed that the optimal number of clusters given by these techniques did not satisfy the intra-cluster energy consumption pattern similarity. Therefore, final number of clusters i.e. ‘k’ was decided through visual analysis. Using visual analysis, different initial numbers of ‘k’ were tested ranging from 2 and up to 20 to ascertain the appropriate number for k [18]. It was observed from the clustering process, that after six clusters with the increase of k, despite having higher intra-cluster similarity of energy consumption patterns, the clusters with lower number of members were more sensitive for change in membership as compared to the clusters with higher memberships and dissimilarity[18]. Therefore, after careful deliberation, it was assessed that for initial clustering, six clusters was the appropriate initial ‘k’ for this particular case. After clustering the total population of consumers, the cluster membership achieved with six initial clusters is shown in Table I.

Table I Cluster membership details [18]

Cluster No.	No. of Members	Total number of energy consumers in %
1	5546	79.98
2	10	0.14
3	88	1.27
4	13	0.19
5	1244	17.94
6	33	0.48

It was observed that out of these six clusters, only two clusters showed clear energy consumption patterns and rest of the clusters had significantly large differences between the consumption patterns. Therefore, the four uncertain clusters out of six were deemed as not being suitable for load profile extraction. The typical k-means clustering approach adopted for load profile extraction would have suggested extracting the energy consumption patterns considering these clusters as final clustering solution. However, in the extended k-means clustering approach, the clusters with significantly dissimilar energy consumption patterns are fed back individually for clustering. The same procedure is adopted for each output if it carries significantly dissimilar patterns until the process produced clearly and identifiable energy consumption patterns.

The final resultant output clusters contain the energy consumption data of all members with significantly similar energy consumption patterns. This data is then utilized to generate a single profile which is referred as a typical load profile/curve of the cluster. Different interpolation and curve fitting techniques are applied on these typical load curves to get the best fit curve which is considered as the load profile that is fully smoothed. Some profiles with low number of consumers were not subjected to interpolation because of higher energy loss. Some of the extracted load curves are given in Fig. 4.

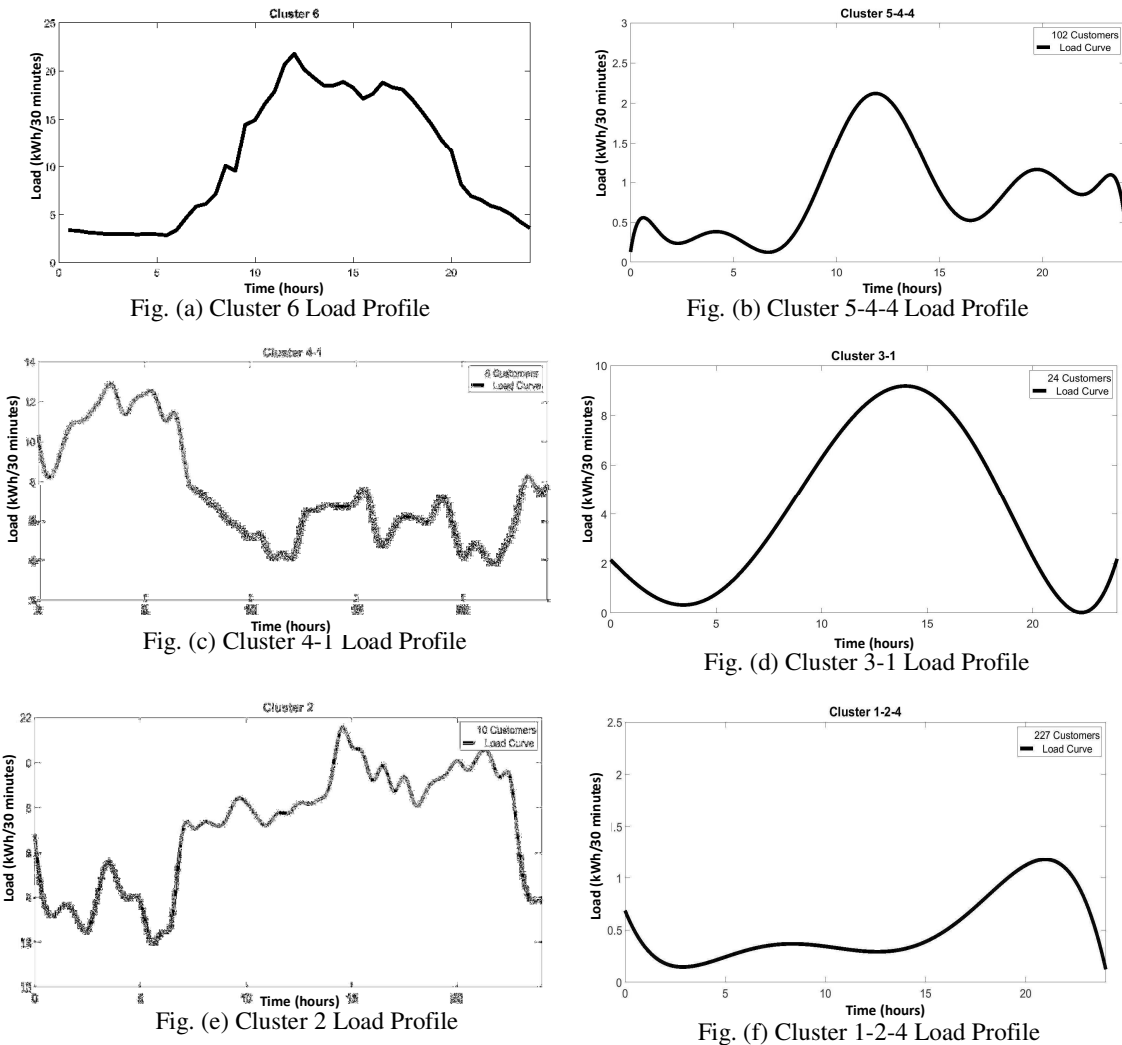


Fig. 4. Cluster load profiles (Horizontal axis gives time of the day in hours and vertical axis gives kWh/30 mins load of smart meter consumers [18])

From the resultant clusters of entire analysis and re-clustering process it was evident that the cluster number 2, 4, 5 and 6 consisted of member with relatively higher energy levels [18]. Cluster 1 and 5 had to be re-clustered further to achieve the load profiles that can be used in MC simulation applications. Fig. 5 shows the load profiles generated for entire population of the consumers containing time duration from 0-24 hours on x-axis and load on y-axis.

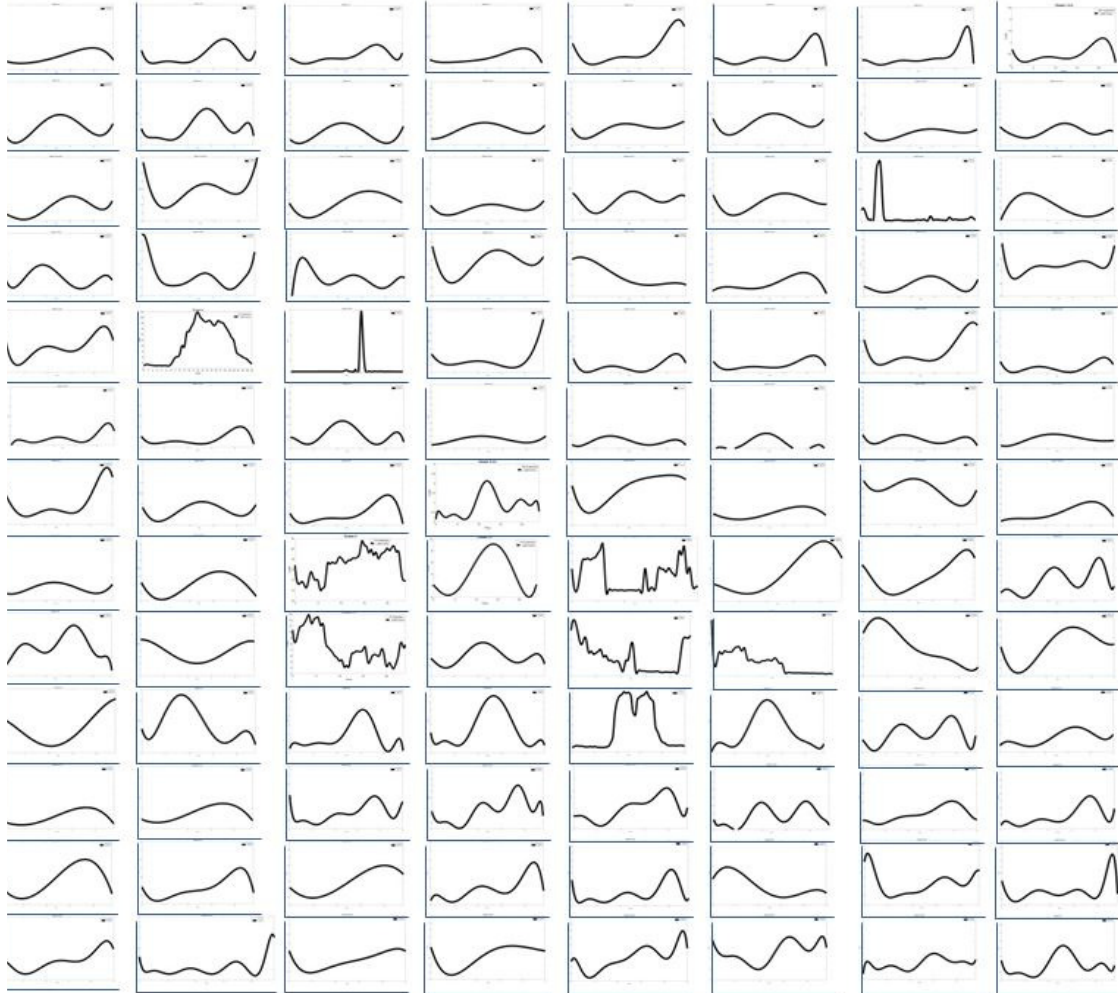


Fig. 5. Load Profiles for all Clusters given in a (13x8) matrix form (Horizontal axis gives time of the day and vertical axis gives kWh/30 mins energy of smart meter consumers [18])

Table II gives the overview of the energy distribution with average percentage load for initial six clusters. From the Table II [18] it is clear that cluster 1 with almost 80% of the total population of consumers, consumes less than 46% of total energy and the rest of the energy, i.e. 54% of energy is consumed by only 20% of the consumers [18].

Table II Overall load distribution in Clusters [18]

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
45.72%	4.18%	9.99%	3.32%	29.44%	7.34%

Using the proposed approach consumers with different loads can be segregated easily and this enables the utility to locate such customers for the inclusion of different programs for load control, including demand side management. Customers in cluster 5 which have less than 20% members of entire population, on average constitute nearly 30% of total energy consumption. Further, the less than 2% of total population which is comprised by clusters 2, 3, 4 and 6, on average constitute almost 25 % of total system load [18]. The clusters with low membership did not require much of re-clustering and load profiles were extracted from them without going into deeper levels of re-clustering.

As for representation of the final load profiles, best fit functions were used, to check the accuracy of the methodology adopted to develop curves, calculations of energy captured by the generated curves were analysed and are given in Table III [18].

Table III % variation of the resulting profile from the actual profile at different times [18]

Cluster	% Energy Captured at 01:00	% Energy Captured at 05:30	% Energy Captured at 10:30	% Energy Captured at 15:30	% Energy Captured at 20:30
C1	+26	+27	+03.4	+33.1	+15.1
C2	-1.5	+06	+01.6	+0.6	+01.1
C3	+07	-5.3	-8.4	+07.3	+03.1
C4	-21.4	-0.02	-16.5	-1.7	+11.8
C5	+16.1	-1.6	-10.1	+2.6	+06
C6	+06.2	+02.1	-4.8	+03.8	+17.8

From Table III, energy calculations at five different time slots suggest that cluster membership has a significant effect on the percentage of energy captured. Clusters with higher number of member tend to show positive variance and on the other hand negative or minimal variance is shown by clusters with lower number of members. Further to this, during the process of curve smoothing, significant deviations from original curves were observed at the starting and ending points of the clusters consisting of small number of members. A degree of accuracy can be achieved in the fitting process by choosing the correct fitting polynomial but it cannot eliminate the deviations entirely. It can be observed from the calculations that the captured energy varies significantly in most cases at the starting point and ending point of the curve. Therefore, it was assumed that clusters with very low number of members, should not be normalized with the best-fit function despite the higher variations in the curve for this particular case [18].

According to the results in Table III, the energy capture calculation of the initial individual clusters as compared to the total energy of system showed that in the process of generating the load profiles using the proposed approach, the approximation taken has resulted a lower loss from the original profiles [18]. Thus, the results evidence the robustness of the proposed approach while reducing the complexity in the energy demand profiles and to making the pathway suitable for MC simulation applications. Transformation of thousands of load curves given in the Fig. 1 into a few given in Fig. 5 provides the network operator more flexibility for analysing individual consumers' load patterns.

To study impact of consumers' load on a smart power system environment using MC simulation, the extracted curves are processed with the steps in the proposed energy classification approach given in Section II. Once the energy classification is obtained for all the profiles, an aggregated energy classifier is designed by aggregating all the energy classifications of profiles and then arranging them according to descending order of their probabilities of occurrences for the convenience of the presentation. Fig. 6 shows the resultant energy classifier.

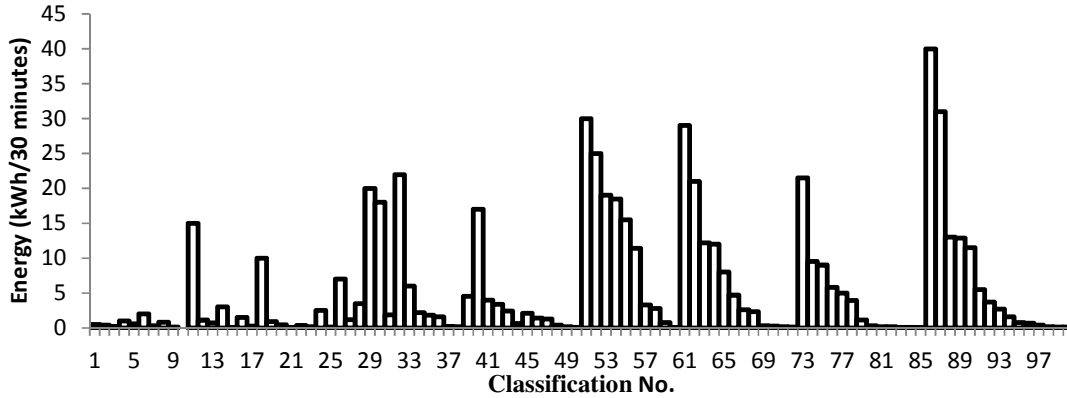


Fig.6. Energy Classification for entire population of Consumers

This classification of energy gives the values of energy threshold points. Values of these energy threshold points can be used for MC simulation applications. The frequency of occurrence of each value of the power consumption can be used to calculate its probability of occurrence. This can be achieved by dividing individual frequency of a specific power value by aggregated frequency of entire classifications. Fig. 7 gives the probability of occurrence of each sample.

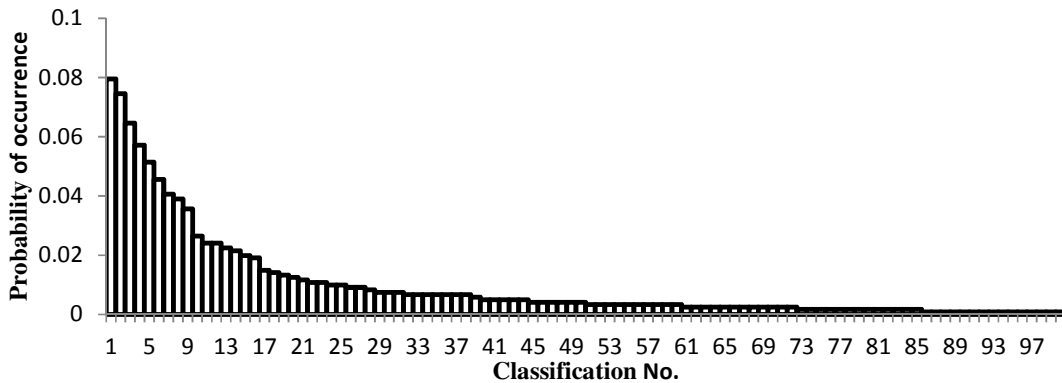


Fig. 7. Probability of occurrence of each sample

The classification number, calculated values of power consumption and frequency are given in Table IV.

In order to show the applicability of the proposed approach for utilization of resulting profiles in MC simulation application, random sampling was applied by incorporating the probabilities of occurrences of individual load demand thresholds in the ultimate load profiles. Initially 10,000 random samples were applied and then it is increased to 20,000. The results showed almost similar probabilities for each power value as given by the

Table IV Calculated Value of Power and Frequency

Class.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Power	0.5	0.4	0.2	1	0.6	2	0.3	0.8	0.1	0	15	1.1	0.7	3	0.05	1.5	0.25	10	0.9	0.45
Freq.	96	90	78	69	62	55	49	47	43	32	29	27	26	24	23	18	17	16	15	
Class.	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Power	0.032	0.35	0.15	2.5	0.12	7	1.2	3.5	20	18	1.9	22	6	2.2	1.8	1.6	0.22	0.17	4.5	17
Freq.	14	13	13	12	12	11	11	10	9	9	9	8	8	8	8	8	8	8	7	6
Class.	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Power	4	3.4	2.4	0.65	2.1	1.4	1.3	0.38	0.18	0.085	30	25	19	18.5	15.5	11.4	3.3	2.8	0.75	0.03
Freq.	6	6	6	6	5	5	5	5	5	4	4	4	4	4	4	4	4	4	4	4
Class.	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Power	29	21	12.2	12	8	4.7	2.6	2.3	0.28	0.26	0.165	0.13	21.5	9.5	9	5.8	5	3.9	1.17	0.27
Freq.	3	3	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2
Class.	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Power	0.155	0.14	0.095	0.07	0.04	40	31	13	12.9	11.5	5.5	3.7	2.7	1.58	0.78	0.67	0.37	0.175	0.135	0.11
Freq.	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

originally calculated probability. Fig. 8 shows a comparison of all three probabilities. Fig. 8 depicts that the probabilities of power consumptions calculated by random sampling are almost close to the calculated probabilities. Further, it is observed that with increase in number of random samples, the difference between calculated and sampled probabilities decreases as can be seen by 10,000 and 20,000 random numbers, which is another indicator to reflect accuracy of the profiles at the convergence levels of Monte Carlo simulation.

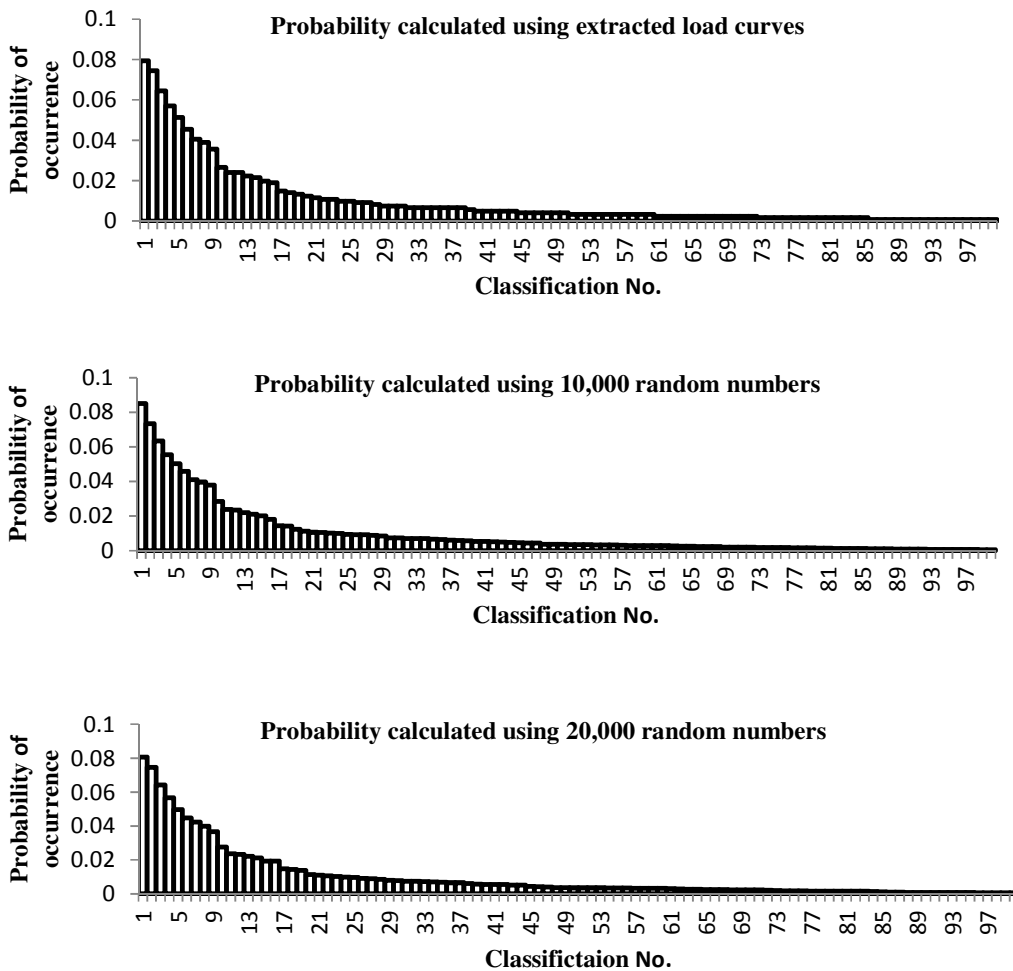


Fig. 8. Comparison of probabilities calculated using load profiles and random sampling

IV. CONCLUSION

Smart planning and operation of a power system requires a vast availability of detailed information on the system loading. Smart meters can potentially provide detailed information of energy consumptions of consumers, however analysing such data on large-scale is complex and challenging. The paper proposed an innovative approach, which abstract the large volume of data to a manageable level and classifies the energy consumption levels. It incorporates multi-layered k-means clustering to linearize complex energy-consumption patterns while ensuring a higher level of intra-cluster similarity.

Model validations proved the robustness of the approach and its extended benefits in the analysis of smart distribution networks. The paper also guides how resulting energy profiles can be used for MC simulation applications.

With the growing scales and volumes of smart meter data, the modelling power consumptions at every node in a smart power system are challenging. The proposed approach can be used as an alternative to mitigate such challenges and complexities without compromising the accuracy.

V. REFERENCES

- [1] F. L. Quilumba-Gudino, "Using Advanced Metering Infrastructure Data For Smart Grid Development," 2014.
- [2] R. Baetens, R. De Coninck, L. Helsen, and D. Saelens, "The impact of load profile on the grid-interaction of building integrated photovoltaic (BIPV) systems in low-energy dwellings," *Journal of Green Building*, vol. 5, pp. 137-147, 2010.
- [3] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, and Y. Zhao, "Load profiling and its application to demand response: A review," *Tsinghua Science and Technology*, vol. 20, pp. 117-129, 2015.
- [4] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *Smart Grid, IEEE Transactions on*, vol. 6, pp. 911-918, 2015.
- [5] Y. Li and P. J. Wolfs, "A hybrid model for residential loads in a distribution system with high PV penetration," *Power Systems, IEEE Transactions on*, vol. 28, pp. 3372-3379, 2013.
- [6] Y.-I. Kim, J.-M. Ko, and S.-H. Choi, "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*, 2011, pp. 1-6.
- [7] I. P. Panapakidis, M. C. Alexiadis, and G. K. Papagiannis, "Enhancing the clustering process in the category model load profiling," *IET Generation, Transmission & Distribution*, vol. 9, pp. 655-665, 2015.
- [8] M. Aien, A. Hajebrahimi, and M. Fotuhi-Firuzabad, "A comprehensive review on uncertainty modeling techniques in power system studies," *Renewable and Sustainable Energy Reviews*, vol. 57, pp. 1077-1089, 5// 2016.
- [9] W. Li, *Reliability assessment of electric power systems using Monte Carlo methods*: Springer Science & Business Media, 2013.
- [10] D. S. Kirschen, D. Jayaweera, D. P. Nedic, and R. N. Allan, "A probabilistic indicator of system stress," *IEEE Transactions on Power Systems*, vol. 19, pp. 1650-1657, 2004.
- [11] D. S. Kirschen and D. Jayaweera, "Comparison of risk-based and deterministic security assessments," *IET Generation, Transmission & Distribution*, vol. 1, pp. 527-533, 2007.
- [12] D. Jayaweera, G. Burt, and J. R. McDonald, "Customer Security Assessment in Distribution Networks With High Penetration of Wind Power," *IEEE Transactions on Power Systems*, vol. 22, pp. 1360-1368, 2007.
- [13] D. Jayaweera and S. Islam, "Steady-state security in distribution networks with large wind farms," *Journal of Modern Power Systems and Clean Energy*, vol. 2, pp. 134-142, 2014.
- [14] D. Jayaweera and S. Islam, "Security of Energy Supply With Change in Weather Conditions and Dynamic Thermal Limits," *IEEE Transactions on Smart Grid*, vol. 5, pp. 2246-2254, 2014.
- [15] D. Jayaweera, "Security Enhancement With Nodal Criticality-Based Integration of Strategic Micro Grids," *IEEE Transactions on Power Systems*, vol. 30, pp. 337-345, 2015.

- [16] G. Carpinelli, P. Caramia, and P. Varilone, "Multi-linear Monte Carlo simulation method for probabilistic load flow of distribution systems with wind and photovoltaic generation systems," *Renewable Energy*, vol. 76, pp. 283-295, 2015.
- [17] P. Chen, Z. Chen, and B. Bak-Jensen, "Probabilistic load flow: A review," in *Electric Utility Deregulation and Restructuring and Power Technologies, 2008. DRPT 2008. Third International Conference on*, 2008, pp. 1586-1591.
- [18] Z. Khan, D. Jayaweera, and H. Gunduz, "Smart meter data taxonomy for demand side management in smart grids," in *Probabilistic Methods Applied to Power Systems (PMAPS), 2016 International Conference on*, 2016, pp. 1-8.
- [19] ISSDA, "Accessed via the Irish Social Science Data Archive ", ed.
- [20] H. Jiawei and M. Kamber, "Data mining: concepts and techniques," *San Francisco, CA, itd: Morgan Kaufmann*, vol. 5, 2001.
- [21] C. C. Aggarwal and C. K. Reddy, *Data clustering: algorithms and applications*: CRC Press, 2013.