Smart Meter Data Based Load Forecasting and Demand Side Management in Distribution Networks With Embedded PV Systems

Khan, Zafar A.; Jayaweera, Dilan

DOI:
10.1109/ACCESS.2019.2962150

Citation for published version (Harvard):
Khan, ZA & Jayaweera, D 2019, 'Smart Meter Data Based Load Forecasting and Demand Side Management in Distribution Networks With Embedded PV Systems', *IEEE Access*, vol. 8, pp. 2631-2644.
https://doi.org/10.1109/ACCESS.2019.2962150

Received December 3, 2019, accepted December 20, 2019, date of publication December 25, 2019, date of current version January 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962150

Smart Meter Data Based Load Forecasting and Demand Side Management in Distribution Networks With Embedded PV Systems

ZAFAR A. KHAN, (Student Member, IEEE), AND DILAN JAYAWEERA, (Senior Member, IEEE)
University of Birmingham, Birmingham B15 2TT, U.K.

Corresponding author: Zafar A. Khan (zafarakhan@ieee.org)

This work was supported by the University of Birmingham and the Mirpur University of Science and Technology, Mirpur A.K.

ABSTRACT With a significant deployment of smart meters across end-user platforms, the dynamic visibility of energy flow among the end-users has been increased significantly. The granular information of smart meters can be used to improve the load forecast accuracy and to influence energy consumption patterns with demand side management (DSM) schemes. This paper addresses the challenges of smart meter data size, complexity, variability and volatility for efficient use in load forecast and DSM. A novel clustering-based approach for analysis of smart meter data, aimed at more accurate and detailed load profiling, reduced profile complexity, improved load forecast accuracy and providing optimal DSM solutions is proposed. The proposed approach utilizes an advanced clustering algorithm to reduce the data size. The approach addresses data complexity, variability and volatility by linearizing the load profiles and minimizing the errors. The validity of the approach is demonstrated on an Irish smart meter dataset and on a simulated solar photovoltaic (PV) data and showed an improved load forecast accuracy, improved DSM solutions, and reduced computational burden. The improvements in the DSM solution are evidenced by a higher cost saving with a higher peak load reduction at the lower level of demand flexibility.

INDEX TERMS Data clustering, energy management, load forecasting, load profiling, solar photovoltaic (PV), smart meter data.

I. INTRODUCTION

The electrical power system is undergoing significant changes to provide a secure, reliable, affordable and low carbon electricity. The driver towards decarbonization of energy industry is leading towards increasing integration of intermittent renewable energy sources (RES) such as solar photovoltaic (PV) and wind power generation. Increasing penetration of intermittent RES has changed the paradigm of conventional power system from a centralized generation system to a decentralized generation system having the meshed distribution networks that are no longer passive circuits supplying loads but active systems with power flow and voltages determined by generation as well as load [1]. The benefits from the RES penetration are promising, however, it brings unprecedented uncertainties in the power system. Apart from the uncertainties due to intermittent RES, demand side, particularly transportation and heating, is also going through a transformation that adds to these uncertainties. Increasing integration of loads such as electric vehicles and heat pumps into distribution system exacerbates the uncertainties in a power system with varied loading patterns. To manage these uncertainties, new sources of flexibility are required to manage secure, reliable and affordable supply of electricity.

A significant amount of flexibility can be achieved by managing the energy at the end-user side. A common term used to represent use of such flexibility is demand side management (DSM). DSM is the portfolio of measures taken at the consumer's side to alter a consumer's energy consumption pattern in response to a pricing signal or any other incentive [2]. DSM is deemed to be more successful and can provide faster power reduction in comparison with the generator ramping [3]. The underlying principle behind implementation of DSM within the context of smart grids is to improve the system efficiency, security, reliability and sustainability [4], [5]. This is achieved by enabling the RES integration into the system while utilizing the maximum capacity of the existing infrastructure [4], [5]. Efficient DSM can potentially

The associate editor coordinating the review of this manuscript and approving it for publication was Arup Kumar Goswami.

optimize the utilization of existing infrastructure and support deferral of construction of new infrastructures for generation, transmission and distribution networks [5].

According to Faruqui *et al.* [6], five to eight percent of the installed generation capacity in Europe only handles load peak, which occurs only one percent of time. Reduction in this peak can potentially save huge financial investment and optimize the system efficiency. The latent demand flexibility at the consumer end can be exploited to operate in a secure and reliable manner using DSM. Although, DSM seems an appealing option to enhance power system flexibility, direct involvement of large number of consumers makes scalability of such approach a major challenge. Moreover, implementation of DSM requires prior knowledge about the expected load profile of the consumer and sophisticated coordination between the consumer and utility. For a distribution network with thousands of consumers, managing the demand at individual consumer level is a challenging task. The smart meter data provides the opportunity to manage the demand, however size of smart meter data, high dimensionality and heterogeneity of the load profiles [7] pose great computational challenges to DSM at consumer level.

In order to address the non-linearity of DSM optimization using linear techniques, a piece-wise linear cost function is introduced by [8]. Similarly, [9] presented a new linearized formulation for optimization of DSM as well as carbon-emissions. Only a few researchers have tried to linearize the profiles and conventional methods to linearize the inherent non-linearity of the load profiles increase the computational burden with larger number of segments [10], [11]. The non-linearity of profiles does not only limit DSM solutions, but also poses challenges for load forecasting and new techniques are required to address the non-linearity, variability and volatility of the smart meter data and load profiles to benefit load forecasting and DSM applications including others.

This paper presents a novel holistic DSM approach by considering all aspects of the DSM by developing a systematic approach to deal with the size, complexity, variability and volatility of the smart meter data. It also proposes an advanced and effective approach to load forecasting using smart meter data. This is achieved by advancing k-means clustering algorithm and refining the profiles with the aim of generating alternative profiles using smart meter data and PV generation profiles. The alternative load and generation profiles transpose the N-dimensional non-linear data functions into a concatenation of continuous differentiable linear functions that reduces the data complexity while preserving the data accuracy in the application. These profiles are used for load forecasting and then in DSM with the presence of RES. DSM is applied by mapping clusters using a new cluster selection index, which incorporates the effect of forecasting accuracy in the cluster selection.

Proposed DSM application algorithm systematically controls the outcome of pricing signal by quantifying the response of each cluster till the desired results are achieved. The novelty of the approach originates from the innovative

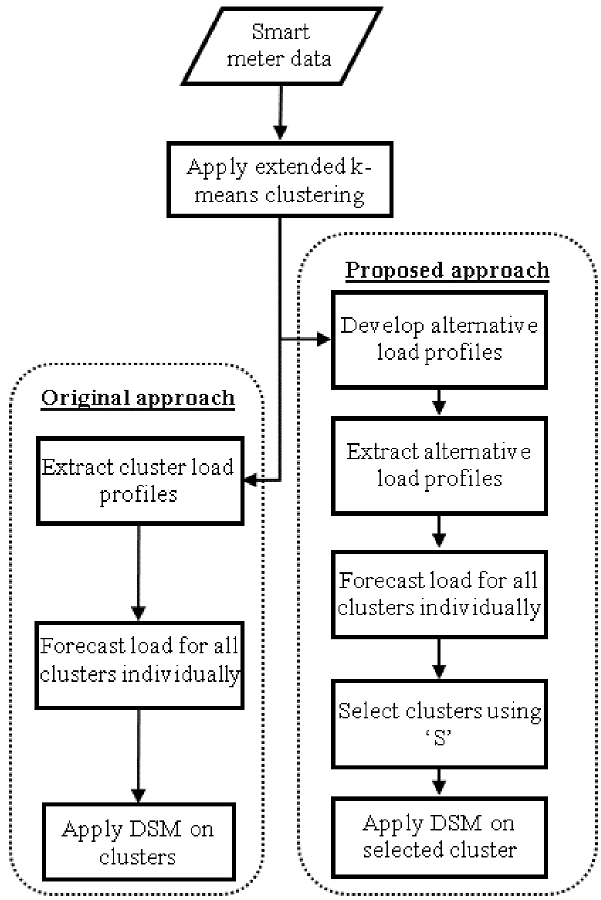


FIGURE 1. Original and Proposed approach for DSM application.

build-up of extended k-mean clustering algorithm, alternative profiling for the use in load forecasting, new cluster selection index, controlling feature of the pricing signal until the desired outcome is achieved, and micro level DSM application.

The rest of the paper is organized as follows: Section 2 delineates the methodology adopted for development of alternative profiling, forecasting and DSM. Section 3 presents the numerical application, results and analysis. Finally, section 4 concludes the findings.

II. METHODOLOGY

The methodology for the proposed DSM approach is composed of multiple stages. An essential pre-requisite for the DSM is the knowledge of load demand that requires to be managed. This is achieved by forecasting load using smart meter data. However, forecasting individual consumer load on large scale is not possible. As can be seen from Figure 1, DSM application using smart meter data involves reducing data using data clustering to extract cluster load profiles. These profiles are used to forecast the load which helps in determining the need for DSM for peak reduction. However, the approach proposed in this paper differs from this convention approach as it includes additional steps of developing the alternative load profiles and after forecasting the load using alternative load profiles, and selection of appropriate clusters for DSM application to benefit system operator and consumer. Detailed methodology adopted for each step is given below.

A. SMART METER DATA CLUSTERING

Rapid development of technology has enabled penetration of smart meters at individual consumer levels. However, as the number of consumers increases, the amount of data generated by smart meters tends to increase exponentially. Mathematically, smart meter data of a consumer can be presented as in (1);

$$y_c = [y_0(x), y_1(x), \ldots, y_m(x)] \tag{1}$$

where y_c is load profile of a single consumer consisting of 'm' time series record. 'y_m' is the energy consumption magnitude at the time step 'x'. If 'f' is the vector-valued function of x and its derivatives, then mathematically 'y' can be represented by a system of ordinary differential equations of order n and m dimensions with a_m as m^{th} coefficient and (m') as m^{th} derivative of x such that;

$$\begin{aligned} y_0 &= f_0(x) = a_1x_0 + a_2x_0^2 + \ldots + a_nx_0^n \\ y_1 &= f_1(x) = a_1x_1' + a_2x_1^{2'} + \ldots + a_nx_1^{n'} \\ &\vdots \\ y_m &= f_m(x) = a_1x_m^{m'} + a_2x_m^{2m'} + \ldots + a_nx_m^{n(m')} \end{aligned} \tag{2}$$

From (2), complexity of smart meter data functions becomes apparent and handling such data with thousands of consumers is a challenging task. This paper adopts Squared Euclidian distance as the similarity measure and for two data points $X(x_1, x_2, \ldots, x_t)$ and $Y(y_1, y_2, \ldots, x_t)$, the squared Euclidian distance can be given as in (3) [12];

$$d_{squec}(x_i, y_j) = \sum_{i=1}^{t} (x_i - y_i)^2 \tag{3}$$

A modified form of k-means clustering namely extended k-means clustering, which combines speed of k-means and divisive approach used in hierarchical clustering, was developed in our previous work [13], is used to cluster the smart meter data. The detailed steps of algorithm are given in Table 1 [13]. A flow chart of the algorithm is also given in Figure 2 [13].

The key advantage of using extended k-means clustering include compact and distinctive clusters having high intra-cluster pattern similarity at lower aggregation level. This algorithm explores the deeper level of the data to extract similar patterns that can benefit in improved forecast accuracy with reduced uncertainty and more accurate DSM planning.

Clusters extracted from the smart meter data are used to develop typical cluster load profiles which are further developed into alternative profiles. Detailed methodology adopted for load profiling is presented below.

TABLE 1. Algorithm : Extended k-means clustering.

1.	Let k=2, Initialize the k-means clustering with 2 clusters
2.	Check stopping criterion 1 for cluster of child node 1
3.	If stopping criterion 1 is true, save the cluster as output cluster and break the cluster out of loop
4.	If false, check stopping criterion 2
5.	If stopping criterion 2 is true, save the node of cluster which resulted in cluster satisfying criterion 2 as output cluster and break the cluster out of loop
6.	If false, go to step 7
7.	Repeat steps 2-5 on remaining cluster child node clusters
8.	If stopping criteria 1 & 2 are false, apply k-means clustering with k=2 on all the child nodes/clusters
9.	Repeat the steps 2-5 on all nodes of child clusters
10.	Repeat the steps 2-8 at each stage until criteria 1 & 2 are true for all clusters
11.	Save all the terminal clusters as final clusters

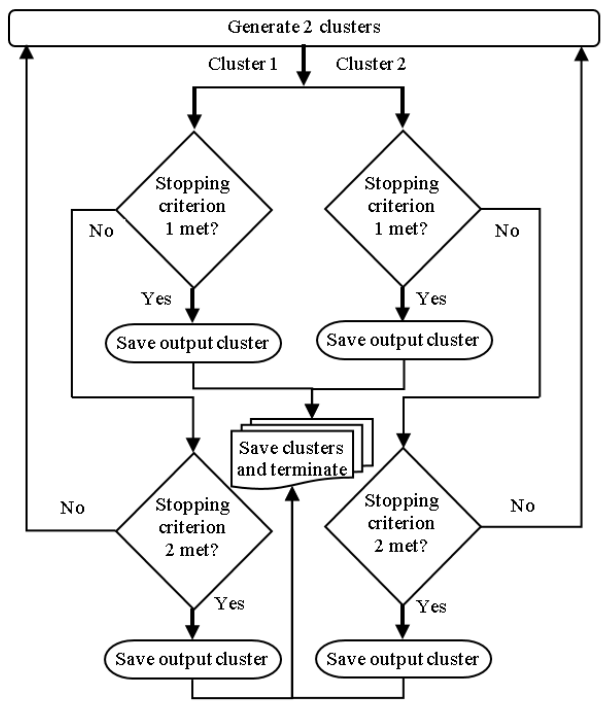


FIGURE 2. Extended k-means clustering algorithm flow chart [13].

B. ALTERNATIVE LOAD PROFILING

The clusters resulting from extended k-means clustering algorithm are used to extract representative profile which are averaged profiles of the cluster. The resulting raw profiles are often considered as the final representative profiles for the cluster. The raw cluster profiles tend to be non-linear, volatile and complex. The complexity of modelling these profiles is apparent in (2). Different studies have used linearization approaches including the Douglas-Peucker algorithm, by selecting typical days instead of the full year [14], [15] or by using a stepwise approximation of load duration curve [16] to reduce the complexity and non-linearity of the profiles.

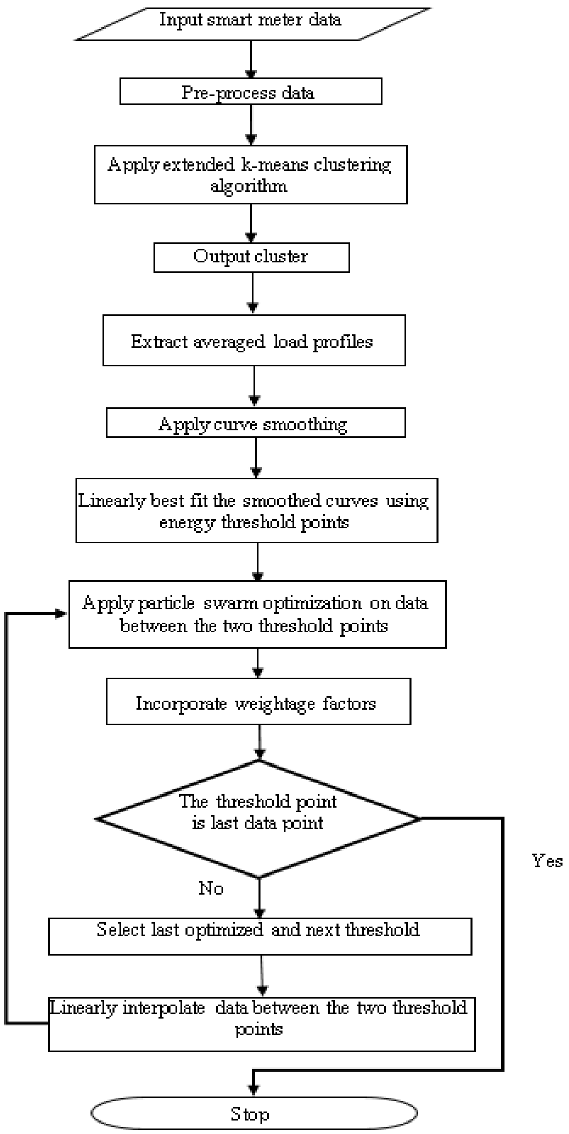


FIGURE 3. Alternative profile development process.

Review of literature shows that no published study has been conducted in power systems using smart meter data to address these issues except in [13]. Therefore, a uniform and systematic profiling approach is used to convert these complex profiles into simplistic profiles called alternative profiles [13].

To develop the alternative profiles, the raw profiles are smoothed by applying moving average smoothing and then systematically linearized around the energy threshold points using Taylor series linearization process [13]. The alternative linearized profiles are optimized to enhance accuracy of representation by incorporating weighting factors for each individual pattern (linear pattern) in the profiles. These weighting factors are determined using particle swarm optimization (PSO) with objective function to minimize the difference of energy consumed by alternative and raw profiles during the interval of each pattern [13]. Thus, m-dimensional

TABLE 2. Comparison between Raw and alternate profiles.

RAW PROFILES	ALTERNATE PROFILES
Highly non-linear profiles	Concatination of linear profiles
Raw smart meter data	Accurate as real smart meter data but in an alternative representation
Require higher order non-linear differential equations for the mathematical representation	Requires linear equations for mathematical representaiton
Difficult to predict due to high variablility and volatility	Easier to predict due to reduced variability and volatility
Possibility for convexity is low due to high non-linearity in energy optimization applications	Possibility for convexity is high due to increased linearity
Challenging to incorporate in feasibility studies	Convenient for feasibility studies
Higher time demand for stochastic simulations in a Monte Carlo Simulation environment	Reduced time demand in a Monte Carlo Simulation environment
High complexity in stochastic modelling in a smart power system	Reduced complexity in stochastic modelling applications in a smart power system

raw profiles are linearized into concatenation of r linear functions and can be given as in (4) [13];

$$y_i = \sum_{i=0}^{r} w_i(f(a_i) + f'(a_i)(x - a_i)) \tag{4}$$

where, y_i represents load profile, w_i represents the weighting factor for pattern i and a_i symbolises the threshold point or operating point for the pattern i around which the curve is linearized. f′(a_i) represents the derivative of a_i. The detailed process of alternative profile development is given in Figure 3 [13]. These profiles are alternative to the raw profiles which are complex, non-linear and variable. Their reduced complexity and variability can benefit in many power system applications where linear processing of non-linear data is required. These profiles can particularly benefit energy optimization applications where increased convexity of the optimization function is required to achieve global optima. A comparison between the raw profiles and alternative profiles is presented in Table 2.

C. ALTERNATIVE PV GENERATION PROFILING

Consideration of embedded PV power integration at the consumer level, reflected in the individual clusters, enables a more futuristic and realistic smart grid scenario of integration of PV. The PV profiles are simulated using the weather data generated by UKCP09 weather generator [17]. This enables the study to have a more realistic PV profiles simulated using weather data which reflects intermittency of PV. The PV profiles are also non-linear in nature and are converted into alternative PV generation profiles using method described above [18]. The raw PV data are incorporated in the raw load profiles and alternative PV power generation data are in the alternative load profiles. Different levels of PV power penetrations are used by scaling the PV to 10%, 20% and 30% of average load of individual clusters. The raw and alternative

PV generation profiles are used in each cluster to ascertain the impact of PV power penetration on forecasting and DSM for both raw and alternative profiles at the cluster level. Once the representative profiles for each cluster are developed, these profiles are used to forecast load at the cluster level which can benefit in the planning of DSM.

D. LOAD FORECASTING

The smart meter data provides an opportunity to forecast the load at different hierarchical system levels, however, the smaller the system, more the variability in profiles and hence results more uncertainty. Load forecast error at a lower aggregation level is usually high due to many contributing factors such as consumers' energy consumption behavior, high variations in load, variety and volatility of complex load profiles which is characterized by complex mathematical formulations. According to [19], the main difficulty in load forecasting is interpretation of the random nature of demand and its representation in equations of the model.

In this research, load forecasting is performed using Feed Forward artificial neural network (ANN) and adopting multiple linear regression (MLR). Comparison of a statistical and artificial intelligence techniques-based forecasting results can be used to validate the impact of simpler mathematical model applied in alternative profiles. The prediction variables used in this study includes temperature as the weather variable and, day of the week and hour of the day as calendar variable. The load variables included 24 hour lagged load, load at same hour from previous week and average load of previous 24 hours. The forecast accuracy is commonly evaluated using mean absolute percentage error (MAPE) [20] and thus, the same metric is used in this research.

Different scenarios are considered for load forecasting. These scenarios include forecasting using raw profiles, forecasting using alternative profiles without any PV penetration, forecasting using raw and alternative profiles with 10%, 20% and 30% PV penetration for each cluster. The PV profiles are embedded in the load profiles and net profiles are used to forecast the load. Moreover, in addition to the forecast variables described above, forecast considering PV included PV generation as a forecast variable. The variables used to show PV generation are simulated values of the PV generation. Training data used for both raw and alternative profiles is half-hourly energy consumption records of 504 days i.e. 24,192 half-hourly records with the measuring unit as kWh. The forecast horizon considered for this study is for a one week, i.e. 336 data points.

E. CONSUMER SELECTION AND DEMAND SIDE MANAGEMENT (DSM)

The primary objective of the DSM is the reduction of system peak load and the operational cost. Due to its effectiveness, the load shifting is most commonly used and widely applied energy management technique in current electrical distribution networks [5]. This technique shifts the load from peak hours to off-peak hours by benefiting from the time independence of the load.

With development of smart loads, the deferrable/flexible loads can contribute in load shifting by automatically responding to the utility signals. The impact of consumer participation in load shifting can be quantified with variable demand flexibility. One of the challenges in DSM implementation is determining the flexibility potential of the consumers at individual household level. However, flexibility potential of an individual household can be quantified only by using the detailed information about the energy consumption at appliance level [21], [22]. The smart meters do not necessarily provide appliance level data and the proposed study uses smart meter gathered data. Therefore, using the smart meter data requires a different approach to ascertain the impacts of flexibility potential of consumers rather than using detailed information at individual household level. Thus, this study considers an alternative approach for DSM application by considering different levels of flexibility to ascertain the impact of flexibility. The study incorporated load shifting with different levels of demand flexibility ranging from 10% to 90% in steps of 5% for each class of consumers or consumers in a cluster. A comparison of cost and saving due to DSM application has been made between raw forecasted profile and alternative forecasted profile, both with and without integration of PV, to assess the efficacy of alternative profiles in DSM.

1) CONSUMER SELECTION

An important aspect in DSM application using the smart meter data at distribution system level is the selection of the appropriate consumers to reduce the system peak. The appositeness of the consumers for participation in DSM is dictated by the cluster load profiles. Clusters can be selected using different indices based on energy consumption for example based on overall maximum energy consumption, high-energy consumption during peak hours, etc. These criteria for selection of clusters will select clusters having significant impact on the system load. On the other hand, such criteria can also result in selection of clusters with large number of consumers and consequently either ineffective response to demand signals or consumer discomfort with undesired affect can arise. Moreover, higher forecast error at a lower aggregation level of load can also cause uncertainty in the system profile. Selection of the clusters should cause minimum disruption to the consumers and avoid customer discomfort, which is one of the underlying rules applied in the proposed approach.

The objective of cluster selection should enable DSM to achieve the maximum cost saving with the minimum consumer disruption and reduction in peak with increased certainty. To achieve this, an algorithm (given in Figure 4) is developed for DSM application. The algorithm selects the clusters based on an index, which considers energy density in each cluster during the peak load hours and incorporates the forecast uncertainty. The peak hours considered in this case are system load peak hours which are the hours in which the system load exceeds 85% of the maximum load. The peak is result of combined load by all clusters as given in (5);

$$P_T = \sum_k PC_k \tag{5}$$

where, P_T represents the total energy consumed by the system during peak hours and PC_k represents the energy consumed by cluster k during peak hours. Average energy consumed by each member of cluster k during peak hours is quantified as in (6);

$$EC_k = PC_k/N_k \tag{6}$$

EC_k represents the average energy consumed by members of cluster k during peak hours. N_k is number of consumers in cluster k. Finally, the cluster selection index S is calculated using (7).

$$S = EC_k + EC_k \times (MAPE/100) \tag{7}$$

All of clusters are arranged in descending order of value of cluster selection index S for DSM application. Initially a single cluster with highest value of S is applied with DSM, and the final number of clusters selected for DSM is decided by considering the demand flexibility and required percentage reduction in energy peak. This percentage is determined by the variable α. The variable α will vary according to the value of required demand flexibility. With lower demand flexibility, higher value of α will be needed to get the required reduction in peak whereas, if the demand flexibility is higher, lower value of α can be used to get higher peak reduction. The index S helps in selection of clusters with high energy density, low number of consumers and reduces the uncertainty.

Thus, the algorithm proposes an index that allows the minimum consumer disruption to achieve required DSM results by selecting appropriate clusters. The algorithm also proposes a control mechanism which quantifies the response from each consumer class considering their demand flexibility potential and iteratively considers best clusters to achieve the desired flexibility in system load. Depending on the flexibility of each class, the desired flexibility in the system can be achieved by different number of clusters/classes while avoiding the undesired scenarios such as rebound effect or over generation etc.

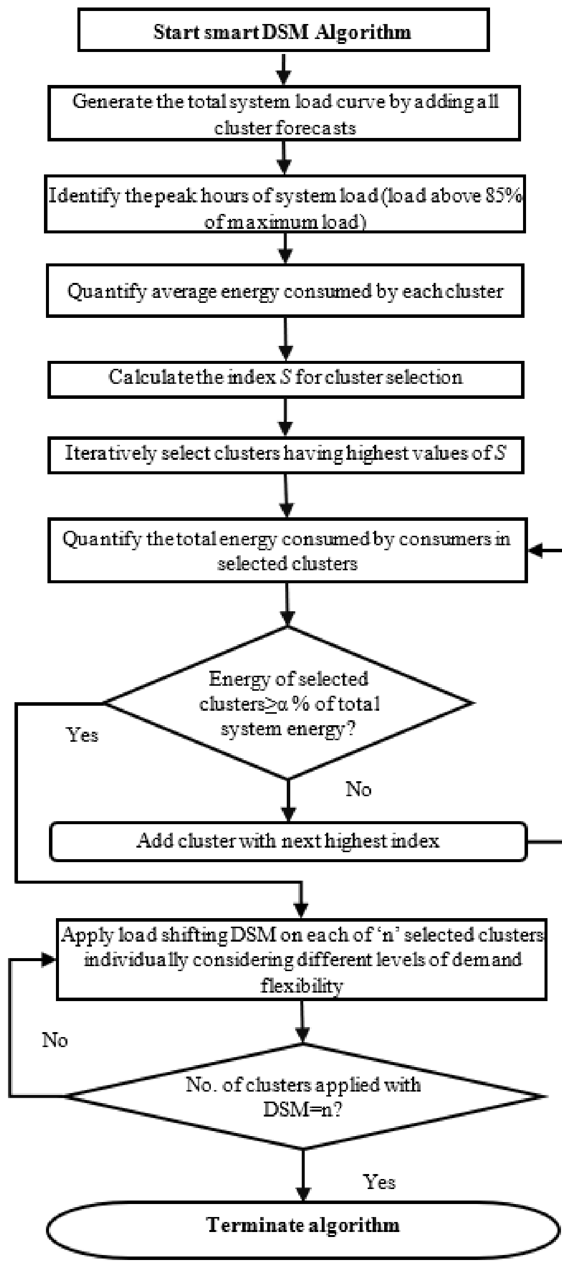


FIGURE 4. Smart DSM and consumer selection algorithm.

2) DSM OPTIMIZATION

Once a cluster is selected, it is applied with DSM. The objective of DSM application is set to acquire maximum saving for the consumer by minimizing the energy consumption during peak hours in response to energy price.

DSM optimization problem is solved using linear programming. The optimization problem can be mathematically formulated as in (8);

$$Min.C = \sum_{i=1}^{k}\sum_{j=1}^{24} P_{(i,j)} \times Cd_j \tag{8}$$

where, C represents the total cost of energy consumed during 24 hours, $P(i,j)$ represent the load demand of cluster i at time j. The price of electricity at time 'j' is given by Cd_j. The constraints considered for the minimization optimization problem are given below.

Subject to:

Equality constraint:

$$\sum_{i=1}^{k}\sum_{j=1}^{24} Pnew_{(i,j)} = \sum_{i=1}^{k}\sum_{j=1}^{24} Pold_{(i,j)} \tag{9}$$

It should be noted that the constraints used in this study are energy constraints, however they are based on the flexibility in terms of load demand only.

Inequality constraints:

Demand flexibility lower bound;

$$Pnew_{(i,j)} \geq Pold_{(i,j)} \times df_{l} \tag{10}$$

where *Pnew and Pold* refer to energy after and before DSM, df_l is the demand flexibility ranging from 0.90 to 0.05 in decremental steps of 0.05. This sets the lower bound according the required demand flexibility.

For off-peak hours;

$$Pnew_j \geq Pold_j \tag{11}$$

Demand flexibility upper bound level;

$$Pnew_{(i,j)} \leq Pold_{(i,j)} \times df_{u} \tag{12}$$

where df_u is the demand flexibility ranging from 1.10 to 1.90 in steps of 0.05. This sets the upper bound according the assumed demand flexibility.

For peak hours,

$$Pnew_j \leq Pold_j \tag{13}$$

An additional constraint that is considered for Case II limits the upper bound to 95% i.e.

$$Pnew_j \leq 0.95 \times MaxPold \tag{14}$$

where *MaxPold* is the maximum value of the *Pold*.

The optimization objective for the PV integrated scenarios can be formulated as;

$$Min.C = \sum_{i=1}^{k}\sum_{j=1}^{24} Pres_{i,j} \times Cd_j \tag{15}$$

where $Pres_{i,j}$ is the energy after renewable energy integration.

$$Pres_{i,j} = (P_{(i,j)} - PV_{(i,j)}) \tag{16}$$

where, $PV_{i,j}$ is the power input from the PV generation for cluster i at time j and can be calculated as in (16).

$$PV = \left(\frac{\sum_{j=1}^{24} P_{(i,j)}}{24}\right) \times \beta \times PV_R \tag{17}$$

PV_R is the real simulated PV generation (scaled from 0-1.0) and β determines the level of PV penetration i.e. for 10% PV penetration β will be 0.10, for 20% value of β will be 0.20 and 0.30 for 30%.

III. NUMERICAL APPLICATION OF NOVEL APPROACH

For the case study, smart meter data of more than 5,000 homes and small businesses from ISSDA (Ireland) [23] with 30 minutes data resolution are used. After data pre-processing i.e. removal of missing and erroneous values, the data is clustered. The clusters are used to extract the raw load profiles, which are further modelled into alternative profiles. A comparison of raw and linear profiles is shown in Figure 5. It can be seen from the Figure 5 that the raw profile is highly non-linear, volatile and variable in nature and modelling and forecasting such patterns are a complex task. Whereas, alternative profile is concatenation of linear profiles and is linear in nature which makes it much simpler and easier to model and forecast. The average variance of energy in alternate profiles is less than 2%, which shows the high accuracy of representation by alternative profiles. These profiles represent groups of consumers and can be used to forecast load demand as given in proceeding section.

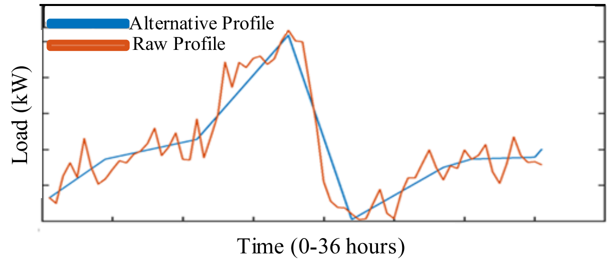


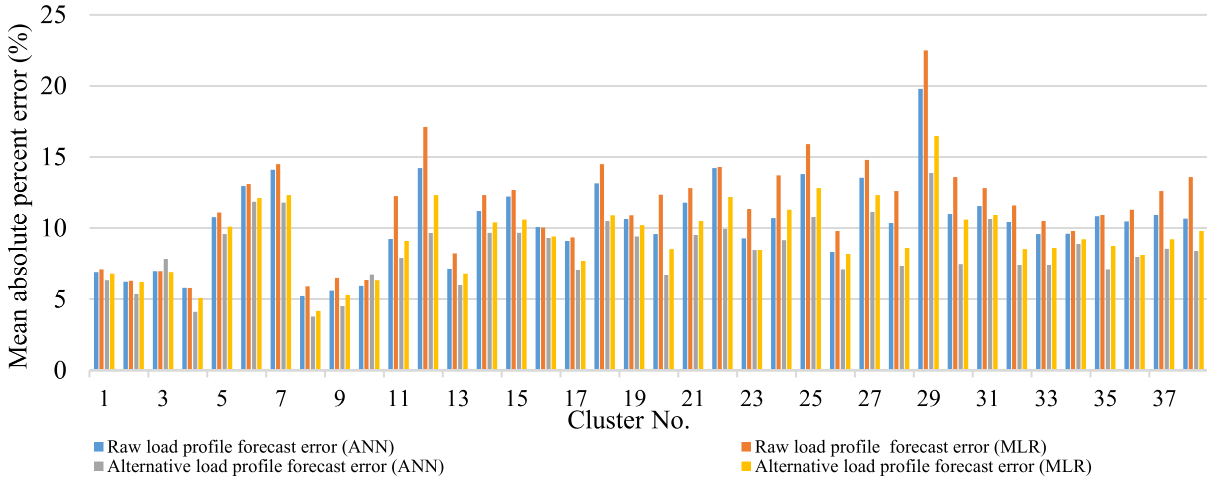
FIGURE 5. A comparison of raw and alternative load profiles.

A. FORECASTING

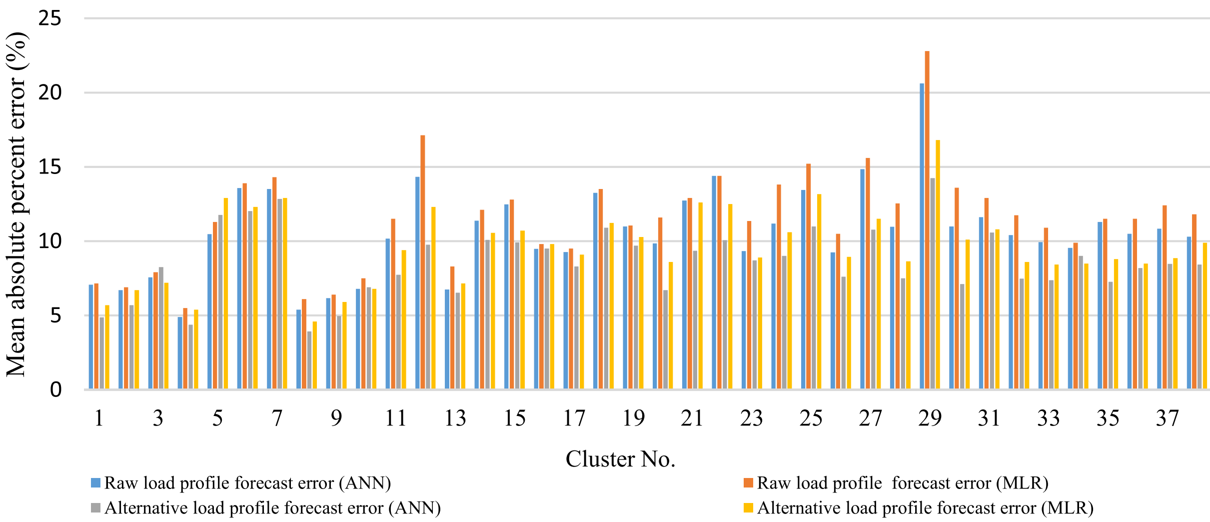
As discussed in methodology, the raw and alternative load profiles extracted and developed from the smart meter data clusters are used to forecast the load demand on short term horizon. Raw profiles forecast is evaluated against the raw future profiles, whereas, the accuracy of forecast for alternative profiles is assessed against the alternative future profiles. As the variables used for both types of profiles are the same for both ANN and MLR techniques, the accuracy of the model can be compared for performance against both models.

Comparison of forecast error of both profiles for different clusters is given in Figure 6. For ANN model, from the MAPE of the clusters in Figure 6 (a) and Figure 6 (b), it becomes evident that the forecast accuracy of the alternative profiles is higher than that of the raw profiles, with exception of two clusters. The clusters with high number of consumers have comparatively smoother raw profiles, which can result in higher forecast accuracy for raw profiles. Moreover, the comparison between the raw and alternative MLR forecast models shows that although alternative MLR forecast performs better as compared to raw MLR forecast. However, overall results suggest that MLR model does not necessarily perform better than that of ANN. The results suggest that in case of the clusters where raw profile forecast using ANN is more accurate than alternative profiles, the MLR model generate better alternative forecast.

The impact of PV penetration is determined at 10%, 20% and 30% of the average cluster load. The PV profiles are embedded into the cluster profiles to generate a net profile. MAPE of the forecasts for raw and alternative profiles for 10% PV power integrated profiles is given in Figure 6 (b). Overall increase in MAPE has been observed due to integration of PV. In some cases, the MAPE of alternative profiles has increased closer to raw profile MAPE, however, the overall alternative profile forecast outperforms the raw profile forecast and in an overall



(a) MAPE for Raw and Alternative Profiles without PV Integration



(b) MAPE for Raw and Alternative profiles with 10% PV

FIGURE 6. Forecast error for clusters with and without PV for raw and alternative profiles.

increase in MAPE for raw profile based forecast is observed. Extended study with 20% and 30% PV power integration shows that the forecast error behaves non-linearly yet the alternative profile-based forecasts perform better than raw profiles.

To see the overall impact of the consolidated forecast for the system, the raw and alternative forecasts of all clusters are added to generate a system profile and the MAPE is compared. The MAPE of load forecast for the system using raw profiles for the DSM day is noted as 5.06% and for alternative profiles 4.64%. This shows that although the alternative profiles perform significantly better at cluster level due to high variability at lower aggregation level, the difference in MAPE starts to decrease with aggregation of load at higher level.

Another important aspect of the forecasting observed during the neural network training is that alternative profiles took shorter duration to train as compared to the raw profiles. The forecasting time for MLR remained same for both profiles. The improved forecast accuracy and lower computational time shows that alternative profiles are better alternative for forecasting at cluster level. Therefore, the results show that the proposed forecasting approach provides improved forecasting accuracy due to reduced complexity, volatility and variability of the profiles which in turn reduces the intricacy of load modeling. Reduced complexity in load modeling benefits in mapping the relationship between the variables and load resulting in increased forecast accuracy.

B. CLUSTER SELECTION

For DSM applications, the forecasted load of one day is considered for each cluster's raw and alternative profiles. The clusters are selected using the cluster selection index 'S' based on system load profile of the day of the DSM application.

TABLE 3. Impact of MAPE on cluster selection.

Cluster No.	1	2	3	4	5	6	7	8
S with MAPE %	134.36	72.52	58.48	**43.33**	**41.33**	35.88	35.33	33.81
S without MAPE %	126.36	69.64	54.24	**35.92**	**36.27**	32.10	34.04	30.83

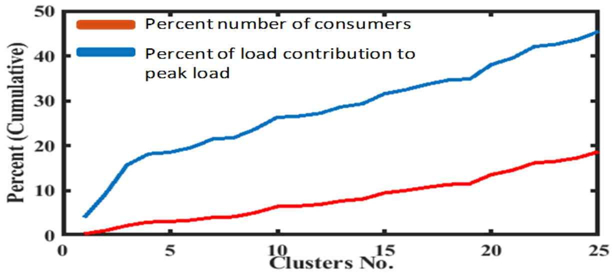


FIGURE 7. **Cumulative percentage of consumer numbers against load contribution.**

The cluster selection algorithm resulted in selection of clusters with lower number of consumers and higher level of peak hours energy consumption. This is reflected in Figure 7 where impact of cluster selection is represented in terms of percentage of peak load contribution against the percentage of consumers. Figure 7 shows an exponential rise in the load consumption (blue line) during the peak hours with first three clusters selected by S. The first three clusters having less than 3% of total consumers consist of more than 18% of the total energy.

Moreover, Table 3 shows the impact of including MAPE in cluster selection. With the inclusion of MAPE in cluster selection, clusters with higher MAPE or uncertainty are prioritized over clusters with lesser uncertainty.

Without considering the MAPE in cluster selection, cluster 5 is selected as it has higher value of 'S' as compared to cluster 4 (Table 3). On the other hand, if MAPE is included in cluster selection, cluster 4 is prioritized over cluster 5. The benefit of inclusion of MAPE can be observed from Figure 8 where cluster 4 presents better shaped cluster for DSM participation as compared to cluster 5. Although the cluster 4 has comparatively higher number of consumers with almost same level of energy density per consumer, the coinciding peak of the cluster 4 with system peak makes it more suitable for DSM. Similar cases are also observed in other cases such as cluster 6 and 7 where better load shape is prioritized only due to incorporation of the forecast uncertainty. Thus, inclusion of MAPE helps in selecting better clusters.

Impact of cluster selection is shown in Figure 9 where the original system profile (blue line) and system profile after removing the load of 10 clusters selected (red line) using cluster selection index are plotted. From the Figure 9, it can be clearly seen that the selected clusters closely follow the shape of system profile which results in a uniform reduction in the system profile, except the from 09:00 hours to 1400 hours. This indicates that the usual trend in the system load is partly

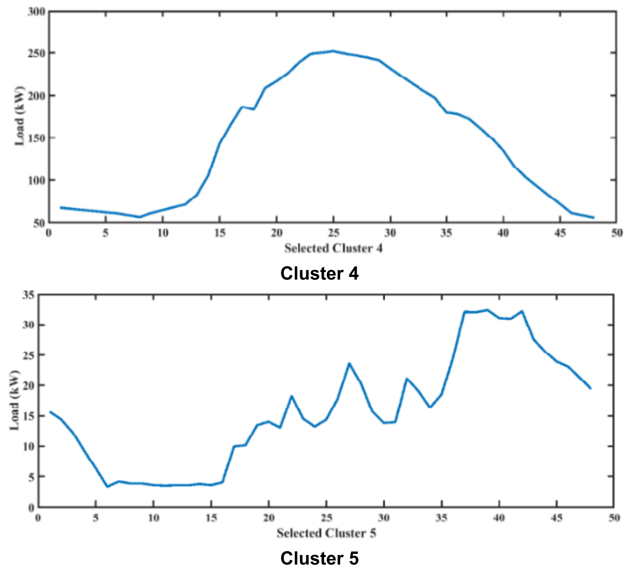


FIGURE 8. **Clusters selected with and without considering MAPE.**

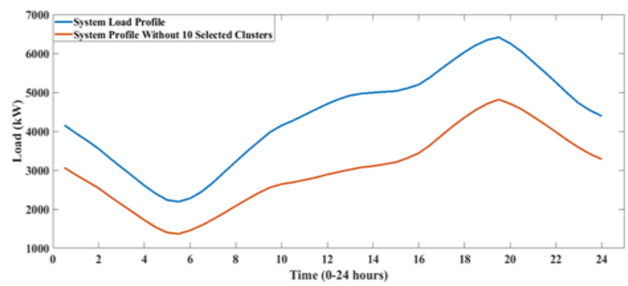


FIGURE 9. **Plot of system load profile and without 10 selected clusters.**

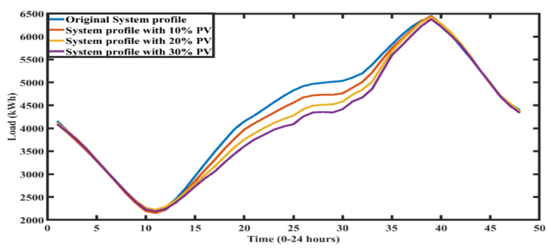


FIGURE 10. **Forecasted original and PV embedded system profiles.**

set by these selected clusters and as the peak of the selected clusters coincides with the system peak, variation in load profiles of these selected clusters can be potentially beneficial for use in DSM. Moreover, impact of embedded PV generation can be observed from Figure 10. The net load profiles after incorporating the PV generation are given in Figure 10 for no PV, 10% PV, 20% PV and 30% PV power penetration. The forecasted net profiles show the intermittency and impact of forecast error for different levels of PV penetration. Higher the PV power penetration, the higher the forecast error due to higher level of intermittency and potentially higher level of forecast error. Despite the forecast error, the overall load

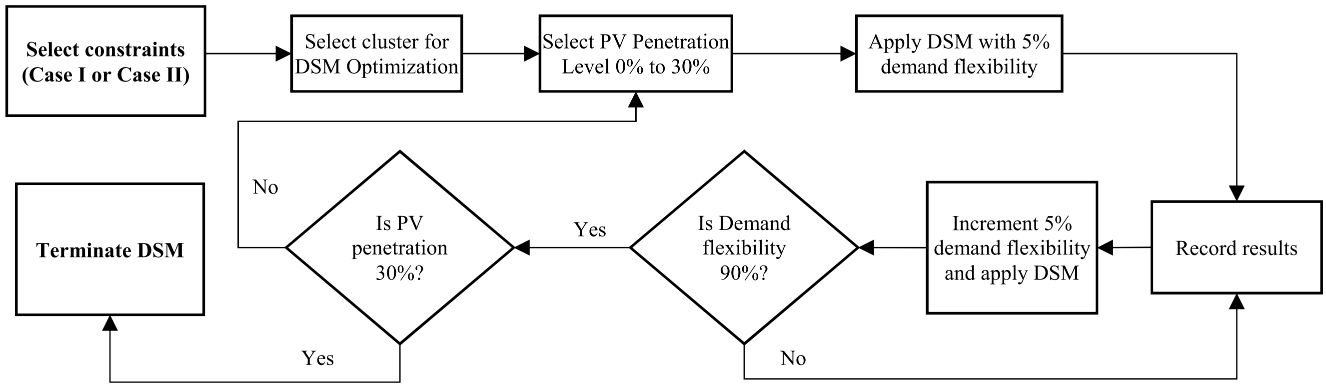


FIGURE 11. Selection of Cases (constraints) and, PV integration and demand flexibility level Scenarios for DSM applications.

profile trend of net load profiles is similar to the profile without PV power integration. The net profiles provide a more realistic scenario where distribution network observed the net load after incorporation of embedded PV power generation.

C. DEMAND SIDE MANAGEMENT APPLICATION

Upon selection of the clusters, load shifting is applied to shift the load from peak hours to off-peak hours. This reduces the peak of the load and increases savings for the consumers as well as utility. Time-of-Use tariff (ToU) is used to reduce the peak of load by load shifting [24], which is taken from the office of gas and electricity markets (Ofgem) in the United Kingdome [25]. Raw forecasted profiles require processing of the constraints to make the non-linear boundaries linear, whereas linearity of alternative forecasted profiles increases the convexity of the profiles. To evaluate the benefits of proposed DSM approach, an efficient linear optimization technique, linear programming is used to optimize the DSM solution.

Figure 11 shows the flow diagram for selection of different scenarios for each Case study i.e. Case I or Case II. The optimization constraints given in Case I consider the upper bound equal to the original peak of the cluster and in Case II, the peak is clipped to 95% of the total peak load. After selection of the Case, the selected cluster with highest value of s is retrieved and required level of PV generation is incorporated in the cluster profile. The profile after incorporation of PV profile is optimized for cost minimization with demand flexibility ranging from 5% to 90% in steps of 5%. In next step, the PV penetration is increased to 10% and

same process is repeated for 20% and 30% PV penetration. After 30% PV generation, the next selected cluster is applied with the same procedure. In total 21,760 simulations are carried out for each case (for both approaches) with combination of seventeen demand flexibility scenarios, four PV integration and application of DSM on up to 20 individual clusters. Results of both cases are discussed below.

1) CASE I

In the first case, the forecasted load is applied with load shifting using cost minimization objective functions given in (8) and (15). Different scenarios considered for DSM in this case include load forecast without PV and with 10%, 20% and 30% PV integrated load forecast. The load shifting is applied to both raw and alternate forecasted profiles to ascertain the impacts of the profiling approaches on the optimization of DSM. Figure 12 shows results of 10 clusters participating in the DSM for cost saving maximization. The savings and peak reduction shown in the Figure 12 reflect the system level savings and peak reduction.

It can be clearly seen from the Figure 12 (a) that in all scenarios, the cost saving using the forecasted alternative profiles is higher as compared to the raw profile forecasts including all scenarios of PV integration. It is pertinent to mention that savings of up to 2.31%, 4.06% and 5.98% are already achieved by integration of 10%, 20% and 30% PV respectively. These savings are not considered in this comparison. The savings for PV integrated profiles in Figure 12 are exclusively due to the DSM application so that a fair comparison of peak reduction and savings can be carried out for all scenarios.

The cost savings come with the load shifting from peak to off-peak hours. As the clusters are individually optimized and they are oblivious to other clusters' load, potentially a second peak or rebound effect can occur. Therefore, as described in the methodology, a constant quantification of the demand response will be beneficial in avoiding any undesired results. In case I, it can be seen from Figure 12 (b) that initially peak reduction is achieved, however, with increase in demand flexibility the peak reduction percentage reduces and beyond 25% demand flexibility, the peak reduction starts to decline for alternative profiles. All scenarios of alternative profiles achieve the maximum reduction in peak at 25% demand flexibility, whereas for raw profiles, some profiles even require demand flexibility of up to 80% to achieve the maximum reduction in peak. The maximum peak reduction achieved by raw profiles at higher levels of demand flexibility i.e. 80%, is lower than that of alternative profiles at 25% demand flexibility. Thus, alternative profiles present better optimization solution with higher cost saving at all levels of demand flexibility and higher peak reduction at lower level of demand flexibility.

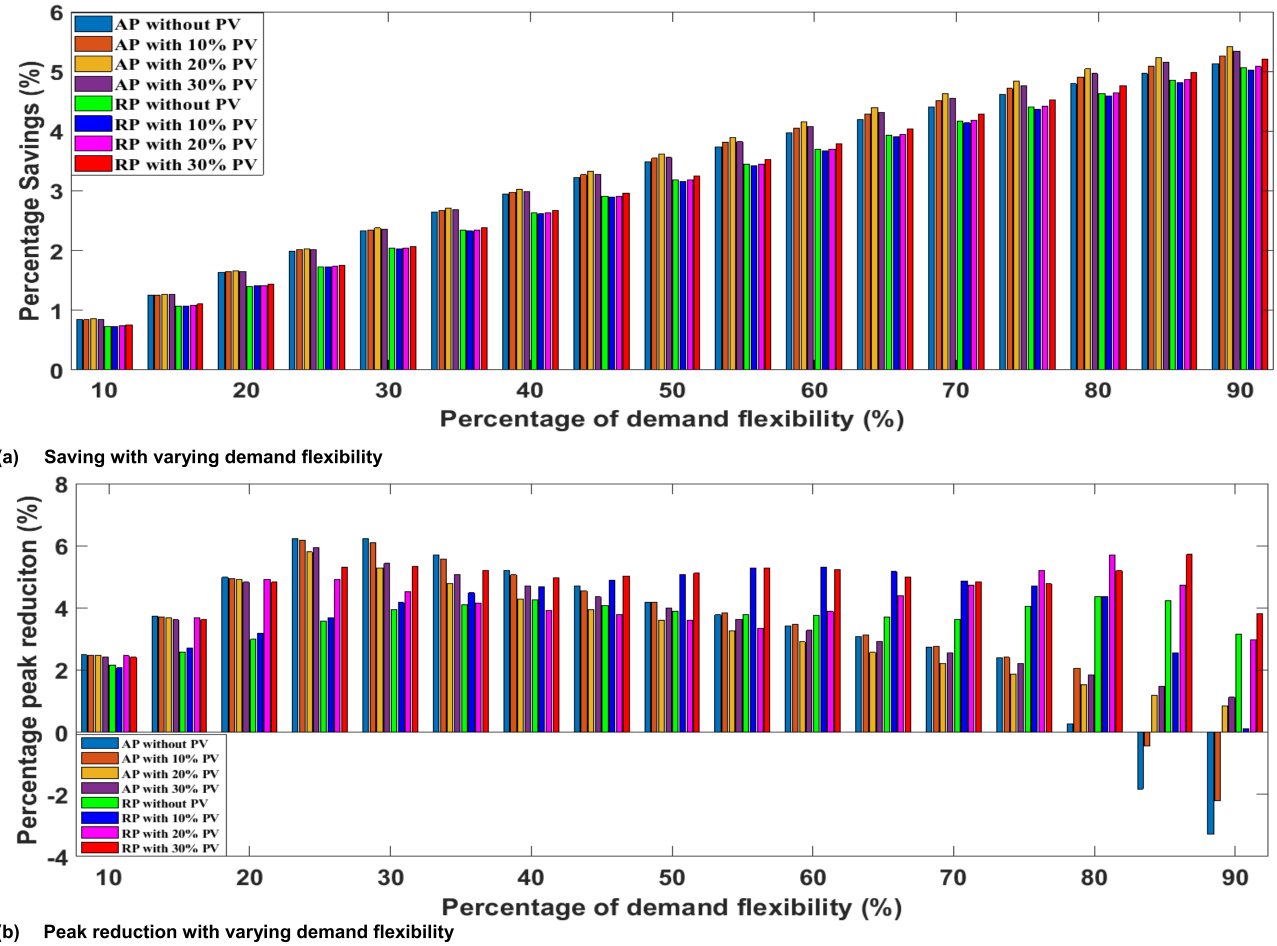


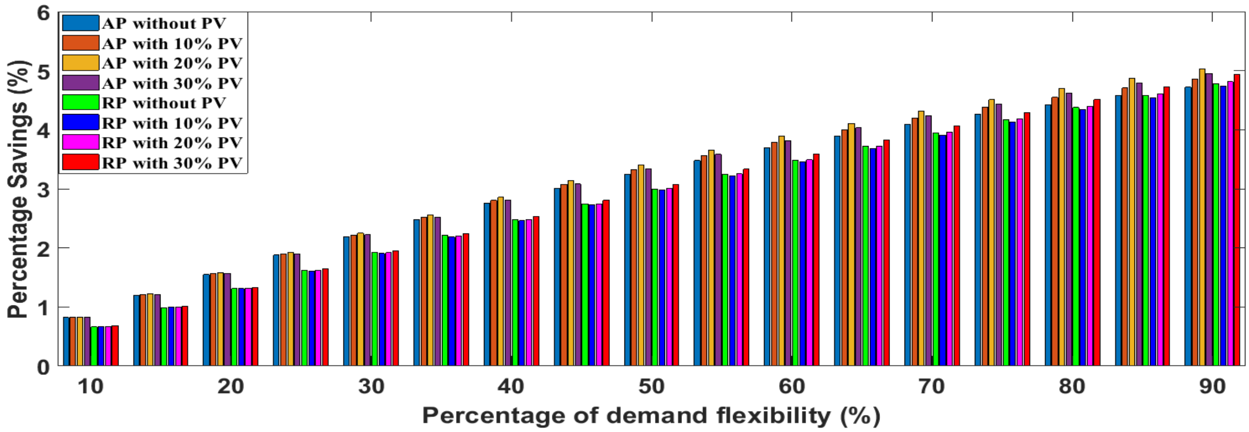
FIGURE 12. Savings and peak reduction without peak clipping (AP: Alternative Profile, RP: Raw Profile).

The key driver for higher saving and peak reduction using alternative profiles is linearity of the profiles. Due to the linearity of the profiles, the convexity of objective function and constraints increases which makes the boundaries linear and provides freedom to find global optima. On the other hand, the raw profiles tend to reduce the peak by shifting load from the peak hours but fails to get optimum solution with linear optimization techniques. From the above, it can be deduced that to maximize the benefits for system operator i.e. reduction in peak, the demand flexibility of 25% is sufficient using alternative profiles. Whereas, due to non-linearity of raw profiles, the maximum reduction in peak is achieved at higher levels of demand flexibility and peak reduction of raw profiles even at higher levels of demand flexibility is less than that presented by alternative profiles. The benefits for consumers are presented in terms of cost savings and higher savings are achieved using alternative profiles as compared to raw profiles.

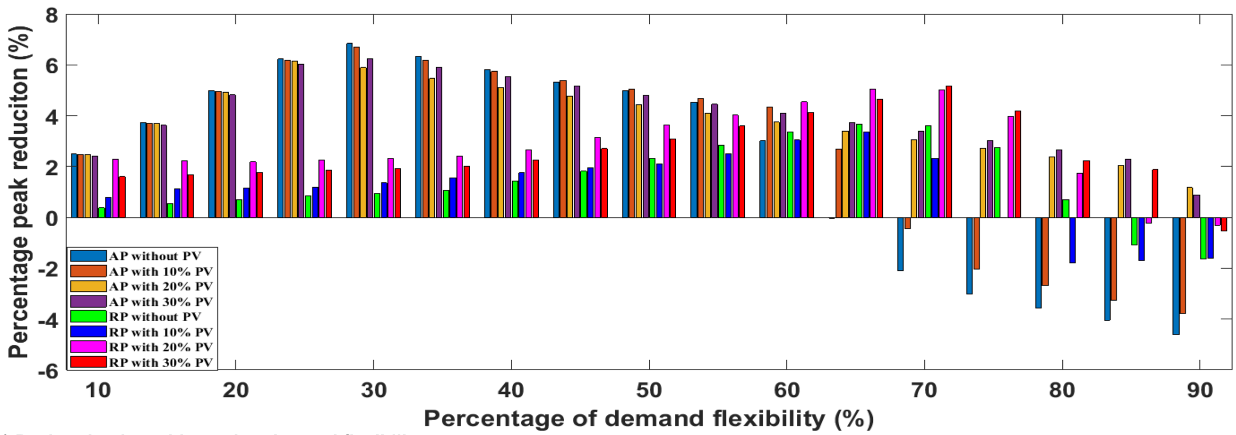
2) CASE II

Additional constraint of 5% peak shaving is considered for Case II. The additional constraint is implemented by limiting the maximum magnitude of the load to 95% of the maximum of the peak load of cluster. Different scenarios described in Case I i.e. varying levels of demand flexibility (5%-90%), varying levels of PV penetration (0%-30%) and different number of clusters (5-20), are simulated with this additional constraint.

Figure 13 shows the results of cost saving in all scenarios with 10 clusters only. The savings show similar behavior as Case I but savings are lesser than that of Case I. The reduction in savings is due to additional constraint that limits the savings. Peak reduction achieved in this case shows better performance by alternative profiles. Maximum peak reduction by alternative profiles is attained at 30% demand flexibility whereas the peak reduction for raw profiles remains relatively low. The maximum peak reduction for raw profiles is attained between 65-70% demand flexibility. An interesting factor observed in this Case is pronounced rebound effect for both raw and alternative profiles. With reduced margin for peak to grow during the peak hours, new peak emerges in off-peak hours. Restricting the peak of individual clusters to 95% of the original peak does not stop the rebound effect. The rebound occurs due to coincidental peaks of the clusters



(a) Saving with varying demand flexibility



(b) Peak reduction with varying demand flexibility

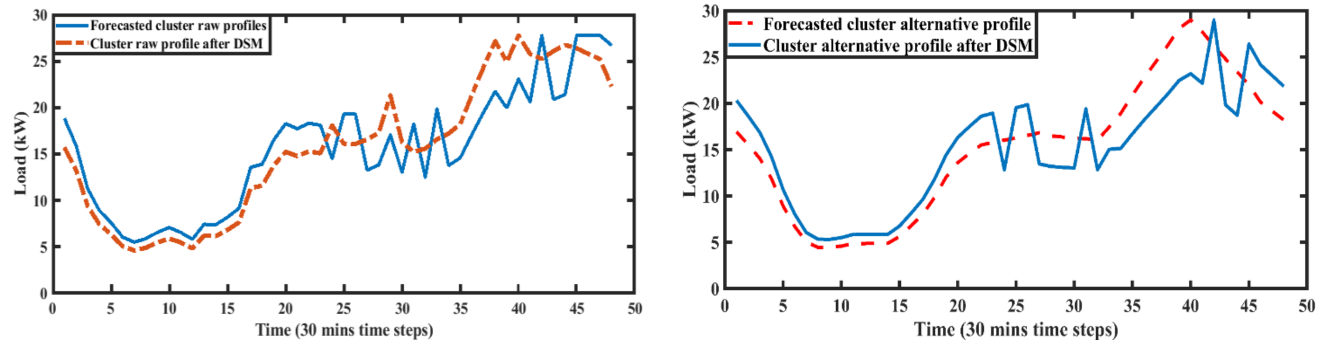
FIGURE 13. Savings and peak reduction with peak clipping (AP: Alternative Profile, RP: Raw Profile).

applied with DSM. This shows that selection of appropriate level of demand flexibility and number of clusters is essential to avoid rebound of peak. This appropriate level of demand flexibility can also be interpreted as coordinated DSM approach where different consumers cooperate with grid but by communicating with each other to minimize the peak.

From Case II, it becomes clear again that linearity of the profiles significantly influences the optimization results. The complexity of raw profiles does not provide appealing peak reduction solution when compared with the solution provided by raw profile. This is again due to non-linearity of the profiles that is further exacerbated by additional constraint of peak shaving. The additional constraint limits the boundaries of the solution thus reducing the feasible region, which is highly non-linear in case of raw profiles. This non-linearity and reduced freedom can lead the linear programing solution for raw profiles to suboptimal solutions. This is evident from the non-linear behavior of peak reduction with demand flexibility in both cases for raw profiles (Figure 13 (b)). On the other hand, peak reduction in alternative profiles, shows a comparatively linear and uniform increasing and decreasing trend. Extended studies with higher levels of load using 15 and 20 clusters also showed that the alternative profiles provide better DSM solutions with high cost savings and high peak reduction at lower demand flexibility level. Impacts of the optimization of the cluster and objective function as is evident from Figure 14. Thus, a comparatively uniform solution with less variability is achieved using alternative profiles at individual cluster level which tends to provide a smoother profile. Figure 14 shows the pre and post DSM system profile for both original and alternative approaches. Visualizing system profiles provides impact of non-linearity of individual clusters on the entire system load profiles is discussed in proceeding section.

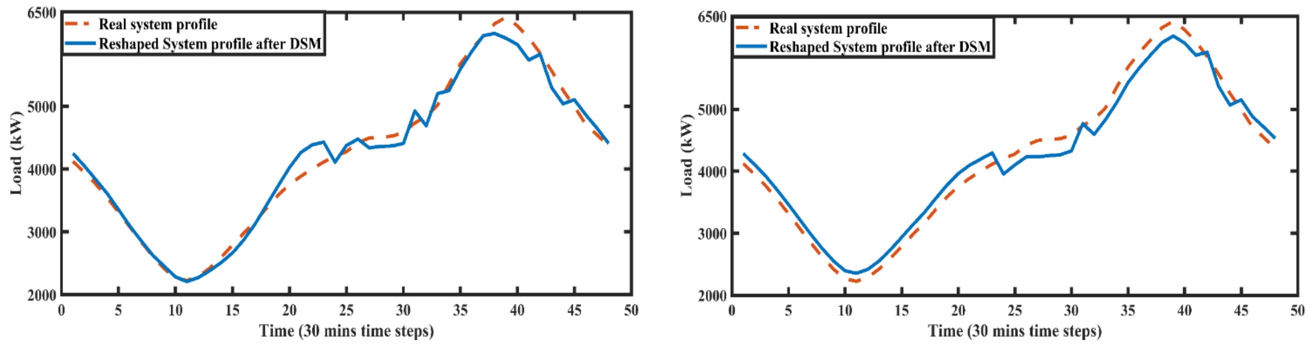
3) IMPACTS OF DSM ON SYSTEM LOAD PROFILES

The results for both cases and the extended studies indicate that being refined, smooth and linear, alternative profiles provide better cost saving as well as a peak reduction. The financial benefits from the linear nature of alternative profiles come in the shape of higher forecast accuracy and convex data functions. Therefore, alternate approach for DSM i.e. using the alternative profiles, provides benefits for both electricity



(a) Raw profile for single cluster before and after DSM (b) Alternative profile for single cluster before and after DSM

FIGURE 14. Raw vs alternate profile after DSM (Cluster Level, Case II, 15% demand flexibility, 10% PV penetration).



(a) Raw profile before and after DSM (b) Alternative profile before and after DSM

FIGURE 15. Reshaped system profiles using raw and alternative profiles (Case II, 10 clusters, 15% demand flexibility, 10% PV).

consumer and network operator in terms of higher savings and peak reduction respectively. However, additional benefit of DSM at cluster level and system level for network operator can be observed from Figure 14 and Figure 15 respectively. From Figure 14 (a) it can be seen that the variability and non-linearity of the raw profile of a single cluster results in DSM solution with high variability. The alternative profiles require lower number variations in the system profile to minimize the objective function as is evident from Figure 14 (b). Thus, a comparatively uniform solution with less variability is achieved using alternative profiles at individual cluster level which tends to provide a smoother profile. Figure 15 shows the pre and post DSM system profile for both original and alternative approaches. Visualizing system profiles provides impact of non-linearity of individual clusters on the entire system profile. The system profile before DSM tends to be smooth as the load aggregation process cancels the noise in the load data. However, DSM application introduces variations in the load, which will reduce the smoothness of the system profile.

As can be seen from Figure 15 (a), the raw system profile lacks smoothness and uniformity in re-shaped profiles. A uniform growth in load during off-peak hours can be seen from Figure 15 (b) which is not the case for raw profile Figure 15 (a). The solution for DSM should tends to reduce the peak load but, in some cases, where load growth is expected to fill the valley, load level for raw profiles even falls below the original load level. This happens due to the non-linearity of the profiles, which reduces the uniformity, and smoothness of the re-shaped system profile. The alternative profiles on the other hand are more convex in nature, their constraints tend to be linear, and thus the solution achieved using alternative profiles is uniform and smoother as compared to raw profiles. Therefore, alternative profiles produce a better resolution without using computationally intensive and complex non-linear modelling to get an improved solution.

IV. CONCLUSION

A novel approach is proposed for the end user energy demand management through the effective load forecasting using smart meter data in a power distribution network. The case studies have evidenced the ability of the proposed approach to select appropriate cluster for DSM application and the effective energy management. Consumer can increase profit opportunities in terms of higher cost saving and for utility and higher peak reduction can be achieved with lower demand flexibility.

Multiple sensitivity analyses suggest that the higher demand flexibility can potentially result in considerable savings, however, chances of rebound effect could also be increased with demand flexibility. Thus, the selected clusters should be iteratively applied with DSM and their demand response should be quantified to avoid any undesired outcome. The alternative profiles can benefit with higher savings and peak reduction at much lower level of demand flexibility as compared to raw profiles.

The proposed approach is an alternative and efficient energy management approach incorporating effective load forecasting with smart meter data. It provides dynamic visibility of energy flows between the end users and the low carbon power generation more effectively, opening a pathway for efficient use of low carbon power generation and energy demand management for the planning and operation of modern power distribution systems.

REFERENCES

[1] C. Ramsay, G. Strbac, A. Badelin, and C. Srikandam, "Impact analysis of increasing (intermittent) RES and DG penetration in the electricity system," Tech. Rep., 2007.

[2] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Trans. Ind. Informat.*, vol. 7, no. 3, pp. 381–388, Aug. 2011.

[3] O. Ma, N. Alkadi, P. Cappers, P. Denholm, J. Dudley, S. Goli, M. Hummon, S. Kiliccote, J. MacDonald, N. Matson, D. Olsen, C. Rose, M. D. Sohn, M. Starke, B. Kirby, and M. O'Malley, "Demand response for ancillary services," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1988–1995, Dec. 2013.

[4] G. Strbac, "Demand side management: Benefits and challenges," *Energy Policy*, vol. 36, pp. 4419–4426, Dec. 2008.

[5] T. Logenthiran, D. Srinivasan, and T. Z. Shun, "Demand side management in smart grid using heuristic optimization," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1244–1252, Sep. 2012.

[6] A. Faruqui, D. Harris, and R. Hledik, "Unlocking the €53 billion savings from smart meters in the EU: How increasing the adoption of dynamic tariffs could make or break the EU's smart grid investment," *Energy Policy*, vol. 38, pp. 6222–6231, Oct. 2010.

[7] B. D. Pitt and D. S. Kirschen, "Application of data mining techniques to load profiling," in *Proc. 21st 1999 IEEE Int. Conf. Power Ind. Comput. Appl. (PICA)*, May 1999, pp. 131–136.

[8] M. Kia, S. M. M. Reza, D. E. Abedini, and H. S. Hamid, "Simultaneous implementation of optimal demand response and security constrained unit commitment," in *Proc. 16th Electr. Power Distrib. Conf.*, 2011, pp. 1–5.

[9] M. Mollahassani-Pour, A. Abdollahi, and M. Rashidinejad, "Investigation of market-based demand response impacts on security-constrained preventive maintenance scheduling," *IEEE Syst. J.*, vol. 9, no. 4, pp. 1496–1506, Dec. 2015.

[10] F. Domínguez-Muñoz, J. M. Cejudo-López, A. Carrillo-Andrés, and M. Gallardo-Salazar, "Selection of typical demand days for CHP optimization," *Energy Buildings*, vol. 43, pp. 3036–3043, 2011.

[11] W. Ko, J.-K. Park, M.-K. Kim, and J.-H. Heo, "A multi-energy system expansion planning method using a linearized load-energy curve: A case study in South Korea," *Energies*, vol. 10, p. 1663, 2017.

[12] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*. Philadelphia, PA, USA: SIAM, 2007.

[13] Z. A. Khan, D. Jayaweera, and M. S. Alvarez-Alvarado, "A novel approach for load profiling in smart power grids using smart meter data," *Electr. Power Syst. Res.*, vol. 165, pp. 191–198, 2018.

[14] A. Shahmohammadi, M. Moradi-Dalvand, H. Ghasemi, and M. S. Ghazizadeh, "Optimal design of multicarrier energy systems considering reliability constraints," *IEEE Trans. Power Del.*, vol. 30, no. 2, pp. 878–886, Apr. 2015.

[15] E. A. M. Ceseña, T. Capuder, and P. Mancarella, "Flexible distributed multienergy generation system expansion planning under uncertainty," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 348–357, Jan. 2016.

[16] X. Zhang, M. Shahidehpour, A. Alabdulwahab, and A. Abusorrah, "Optimal expansion planning of energy hub with multiple energy infrastructures," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2302–2311, Sep. 2015.

[17] P. Jones, C. Kilsby, C. Harpham, V. Glenis, and A. Burton, "UK climate projections science report: Projections of future daily climate for the UK from the weather generator," Univ. Newcastle, Newcastle upon Tyne, U.K., Tech. Rep., 2009.

[18] H. Gunduz, Z. A. Khan, A. Altamimi, and D. Jayaweera, "An innovative methodology for load and generation modelling in a reliability assessment with PV and smart meter readings," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2018, pp. 1–5.

[19] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret, and J. Massana, "A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1460–1495, 3rd Quart., 2014.

[20] A. Dedinec, S. Filiposka, A. Dedinec, and L. Kocarev, "Deep belief network based electricity load forecasting: An analysis of macedonian case," *Energy*, vol. 115, pp. 1688–1700, Nov. 2016.

[21] R. Sharifi, S. H. Fathi, and V. Vahidinasab, "Customer baseline load models for residential sector in a smart-grid environment," *Energy Rep.*, vol. 2, pp. 74–81, Nov. 2016.

[22] N. Blaauwbroek, P. H. Nguyen, M. J. Konsman, H. Shi, R. I. G. Kamphuis, and W. L. Kling, "Decentralized resource allocation and load scheduling for multicommodity smart energy systems," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1506–1514, Oct. 2015.

[23] CER Electricity, ISSDA. (2012). *CER Smart Meter Customer Behaviour Trials Data*. Accessed: Nov. 20, 2015. [Online]. Available: https://www.ucd.ie/issda

[24] J. Torriti, "Price-based demand side management: Assessing the impacts of time-of-use tariffs on residential electricity demand and peak shifting in Northern Italy," *Energy*, vol. 44, pp. 576–583, 2012.

[25] OFGEM. (Aug. 23, 2018). *OFGEM Making a Positive Difference for Energy Consumers*. [Online]. Available: https://www.ofgem.gov.uk/

ZAFAR A. KHAN received the Ph.D. degree from the University of Birmingham, U.K. He is currently a Researcher with the University of Birmingham, and also an Assistant Professor with the Mirpur University of Science and Technology (MUST), Mirpur. He has wide experiences in energy industry, research, and teaching.

DILAN JAYAWEERA (Senior Member, IEEE) is currently a Senior Lecturer in electrical power systems with the Department of Electronic, Electrical and Systems Engineering (EESE), University of Birmingham. He has wide experiences in energy industry, research, and teaching. He has authored a significant number of research articles in scientific journals, conference proceedings, and book chapters in the fields of power system security, reliability, active distribution network operation, smart grids, smart asset management, and risks in power systems. He is a Chartered Engineer in U.K., a Chartered Professional Engineer in Australia, a Fellow at Engineers, Australia. He is an Editor for the IEEE TRANSACTIONS ON POWER SYSTEMS and *Power Engineering Letters*.