Smart Households' Available Aggregated Capacity Day-ahead Forecast Model for Load Aggregators under Incentive-based Demand Response Program

Biao Xiang, Kangping Li, Xinxin Ge, Fei Wang
Department of Electrical Engineering
North China Electric Power University
Baoding, China
feiwang@ncepu.edu.cn

Jingang Lai
E.ON Energy Research Center
RWTH Aachen University
Aachen, Germany
jingang.lai@eonerc.rwth-aachen.de

Payman Dehghanian
Department of Electrical and Computer Engineering
George Washington University
Washington DC, USA
payman@gwu.edu

***Abstract*—The rapid development of smart grid and smart appliances helps those smart households (SHs) more actively participate in the incentive-based demand response (IBDR) programs. As the agent facilitating the SHs' participation in the IBDR program, load aggregators (LAs) need to comprehend the available SHs' demand response (DR) capacity when trading with the system operator in the day-ahead market. This paper proposes a forecasting model aiming to aid LAs forecast the available aggregated SHs' DR capacity in the day-ahead market. Firstly, a home energy management system (HEMS) is implemented to perform an optimal scheduling for SHs and to model the customers' responsive behavior in the IBDR program; secondly, a customer baseline load (CBL) estimation method is applied to quantify the SHs' aggregated DR capacity during DR days; thirdly, several features which may have significant impacts on the aggregated DR capacity are extracted; finally, a support vector machine (SVM) based forecast model is proposed to forecast the aggregated SHs' DR capacity in the day-ahead market.**

Keywords—Smart household (SH), load aggregator (LA), home energy management system (HEMS), aggregated capacity, incentive-based demand response (IBDR).

NOMENCLATURE

A. Indices

t	Timeslot
k	DR day k in the test datasets
i	The i^{th} customer that participate in the IBDR program.

B. Sets

$\boldsymbol{T}$	The set of timeslots
$\boldsymbol{T_{DR}}$	The set of timeslots during the DR time duration
$\boldsymbol{K}$	The set of DR days
$\boldsymbol{I}$	The set of the actual number of customers that participate in the IBDR program.

C. Parameters and variables

λ	Daily fixed electricity price
Δd	The amount of electricity consumption reduction
d^{cont}	Contracted electricity consumption reduction
Inc	Monetary reward
$s(\mu)$	Logical variable
$req(P_{t,i,k})$	Household's initial required electricity consumption at the time t
$P_{t,i,k}^{baseline}$	Customer baseline load at the time t
$P_{t,i,k}$	Customer actual load at the time t
Δt	Time resolution (15minutes)
t_{start}	The start time of DR event
t_{end}	End time of DR event
$P_{t,i,k}^{shift}$	Electricity consumption of shiftable appliance at the time t
$req(P_{t,i,k}^{shift})$	The initial required electricity consumption of shiftable appliance at time t
$p_{i,k}^{shift}$	Rated power of the shiftable appliance
$\tau_{t,i,k}^{shift}$	Running time of the shiftable appliance from the time $t-1$ to t
ΔT	Shifting time of the shiftable appliance
$P_{t,i,k}^{Ine}$	Electricity consumption of inelasticity appliance at the time t
$req(P_{t,i,k}^{Ine})$	The initial required electricity consumption of inelasticity appliance at the time t
$P_{t,i,k}^{air}$	Electricity consumption of air conditioner at the time t
p_i^{air}	Rated power of AC
θ_{room}^{t}	Room temperature at the time t
C	Equivalent heat capacity
R	Equivalent thermal resistance
Q_i	Equivalent heat rate
θ_0	Ambient temperature
θ_{set}	Temperature set-point air conditioner
θ_{db}	Temperature dead bandwidth
$\Delta\tau_{on,i,k}$	The time duration when the AC is on
$\Delta\tau_{off,i,k}$	The time duration when the AC is off
α	$\alpha=-1$ when in summer, $\alpha=1$ when in winter
η	Energy efficiency ratio of the AC
$\tau_{t,i,k}^{air}$	The time duration when the AC is on from time $t-1$ to t
$f_{i,k}$	Individual customer DR capacity during the DR event.
f_k^{agg}	Actual aggregated DR capacity value in DR day k
$\hat{f}_k^{agg}$	Forecasted aggregated DR capacity value in DR day k

D. Abbreviation

SH	Smart household
CBL	Customer baseline load

CBE	Customer baseline energy
HEMS	Home energy management system
DR	Demand response
IBDR	Incentive-based demand response
RMSE	Root mean square error
MAE	Mean absolute error
APE	Absolute percent error
MAPE	Mean absolute percent error
MIC	Maximal information coefficient

I. Introduction

Demand response (DR) is a tariff or program established to motivate changes in electricity consumption by the end-use customers in response to variations in electricity price over time [1], or to offer incentive payments designed to induce lower electricity use at times of high market prices or when the grid reliability is jeopardized [2]. It utilizes the flexible demand-side resources to maintain the system operation reliability and significantly helps address the prevailing stochasticity in massive penetration of renewable energy resources [3]-[6].

Residential customers are responsible for a considerable proportion of electrical energy consumption among all end-users and represent a significant potential for DR. However, only energy-intensive industrial and commercial end-users were traditionally permitted to participate in the incentive-based DR (IBDR) programs, primarily supported by two main reasons: (i) individual residential customers possess a limited DR capacity which is unable to reach the capacity threshold of IBDR in the electricity market, and (ii) residential end-users exhibit heterogeneous characteristics [7], [8], making them a challenge for the system operator to operate and manage. The emergence of load aggregators (LAs) [9] offers a solution to this challenge since the LAs could act as an agent for the residential customers to participate in the market [10]-[11]. Hence, the system operator only need to trade with the LAs instead of a large number of individual residential customers.

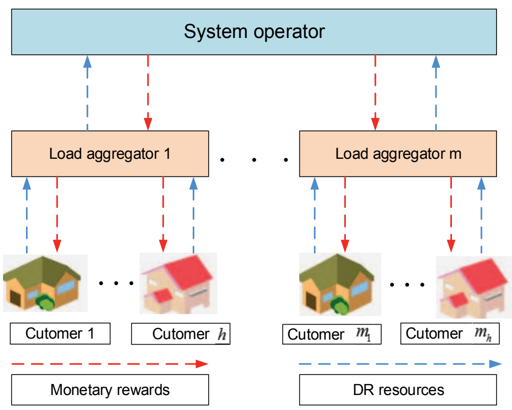


Fig. 1. Different roles in the IBDR program in the day-ahead market

In an IBDR program, the LA procures the DR resources from residential customers by giving monetary rewards to them and then provides these DR resources to the system operator who calls for DR services as shown in Fig. 1. There is some essential information that the LAs needs to figure out when participating in the day-ahead market trading, one of which is how much DR resources, which is also called as the "available aggregated DR capacity" that the LAs is able to acquire from customers in this paper. An accurate quantification and estimation of the available aggregated DR capacity in the IBDR program prior to the DR event is significantly critical in the decision-making process when the LAs is trading in the day-ahead market. However, quantification and estimation of the aggregated DR capacity prior to DR event is usually a complex challenge since residential customers are not only affected by DR signals, but their behavior [12], [13] can also play a significant role with decisions that may present gaps between knowledge, the perception of weather conditions, and attitudes, which are difficult to model and forecast specifically. Therefore, how to understand and model the customers' responsiveness reasonably and thereby accurately quantifying and estimating available aggregated DR capacity is an unavoidable issue for LAs' participation in IBDR program in day-ahead market.

There exists a broad literature and research efforts concerning the residential customers' responsiveness under DR programs from different perspectives [14]. Majority focus on price-based DR (PBDR) programs [15] with many related experiments [16]. Woo et al. [17], [18] study the impacts of time-of-use (TOU) price and critical peak price (CPP) on the residential users' power and energy consumption patterns according to a PBDR pilot study in British Columbia, Canada, and found that implementation of certain control strategies on users would further help in significant reducing the power and energy consumptions. Paterakis et al. [19] acquire the load datasets through a home energy management system (HEMS) and via the optimal scheduling of appliances for smart households (SHs): considering an hourly varying price tariff scheme, they study the effects of the PBDR program on load forecasting. Compared with the PBDR program, there are a very few studies focusing on residential customers' responsiveness and DR capacity under IBDR schemes. Vallés et al. [20] present an empirical methodology to obtain a full characterization of residential consumers' flexibility in response to economic incentives in a Spanish location and model individual observed flexibility, which provides a concise parametric representation of customers that allows a straightforward application of classification methods to partition the sample of customers into categories of similar flexibility. However, it fails to evaluate the flexibility in aggregation level and is not able to considering about the impact that customers' responsiveness may vary from one DR event to another in different DR days. Considering accurate information of the amount of the aggregated DR capacity is crucial for LAs in the day-ahead market trading, this paper proposes a forecasting model aiming to provide an accurate forecast of the aggregated DR capacity for the LAs under IBDR program in the day-ahead market. The contributions of this paper are summarized as follows:

- A machine learning mechanism is proposed to forecast the aggregated SHs' DR capacity under the IBDR programs in the day-ahead markets.
- We comprehensively identify and study several factors that may have significant influences on the aggregated DR capacity forecasts, ranging from the internal parameters affecting the customers' electricity consumption and external DR signals.
- Real-life daily load dataset from the Pecan Street experiments in Austin, TX is adapted to verify the effectiveness of the proposed approach.

The remainder of this paper is organized as follows. In Section II, the problem description and a basic idea of the proposed framework are explained. Section III presents how to model the customers' responsiveness and quantify the aggregated DR capacity. In Section IV, a forecasting model is proposed and mathematically elaborated. Numerical case study is carried out in Section V. Finally, the conclusion is given in Section VI.

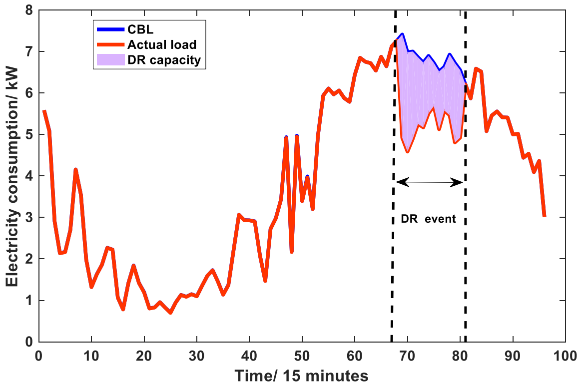


Fig. 2. Schematic diagram of the DR capacity

II. Problem Statement and Proposed Framework

A. Problem Statement

1) DR capacity

In this paper, the DR capacity is defined as the ability of the customers to adjust consumption with respect to normal consumption habits in response to an external signal with an economic incentive. And only the downward signals in the peak electricity usage time are considered in this paper. The DR capacity can be assessed through the customer baseline load (CBL) [21] and the actual loads as demonstrated by an individual customer in Fig. 2.

2) Mathematical description

Let $\boldsymbol{T}=\{t\,|\,1,2,...,T\}$ be the set of timeslots in a day and $\boldsymbol{T}_{DR}=\{t\,|\,t_{start}...t_{end}\}$ be the set of timeslots during the DR duration time in a given DR day, where $\boldsymbol{T}_{DR}\subseteq\boldsymbol{T}$ and the time resolution is Δt=15min . Furthermore, let $\boldsymbol{I}=\{i\,|\,1,2,...,I\}$ be the set of customers that participate in the IBDR program and $\boldsymbol{K}=\{k\,|\,1,2,...,K\}$ be the set of the DR days. For a customer i , let $f_{i,k}$ denote the DR capacity during the DR time duration in the DR day k, and f_k^{agg} represents the aggregated DR capacity of all customers who participate in the IBDR program in the DR day k. They could be mathematically expressed in (1) and (2), respectively.

$$f_{i,k}=\sum\nolimits_{t\in\boldsymbol{T}_{DR}}(P_{t,i,k}^{baseline}-P_{t,i,k})\cdot\Delta t \tag{1}$$

$$f_k^{agg}=\sum_{i=1}^{I}f_{i,k} \tag{2}$$

Where $P_{t,i,k}^{baseline}$ and $P_{t,i,k}$ denote the CBL and actual load respectively at timeslot t during the DR duration time $t\in\boldsymbol{T}_{DR}$. This paper aims to formulate a forecasting model in order to forecast the aggregated DR capacity f_k^{agg} that the LAs could harness in response to a certain DR signal during the IBDR program in the day-ahead market.

B. Basic Idea and the Proposed Framework

There are two main challenges concerning the forecast of the aggregated DR capacity. The first challenge is to model the customers' responsiveness to DR signals in IBDR program and to acquire the customers' load data because the actual residential customers' load data in IBDR programs is typically private and not readily accessible. The second challenge that follows the first one concerns with the accurate forecast of the aggregated DR capacity contingent to the availability of models that capture the customers' responsiveness to DR signals.

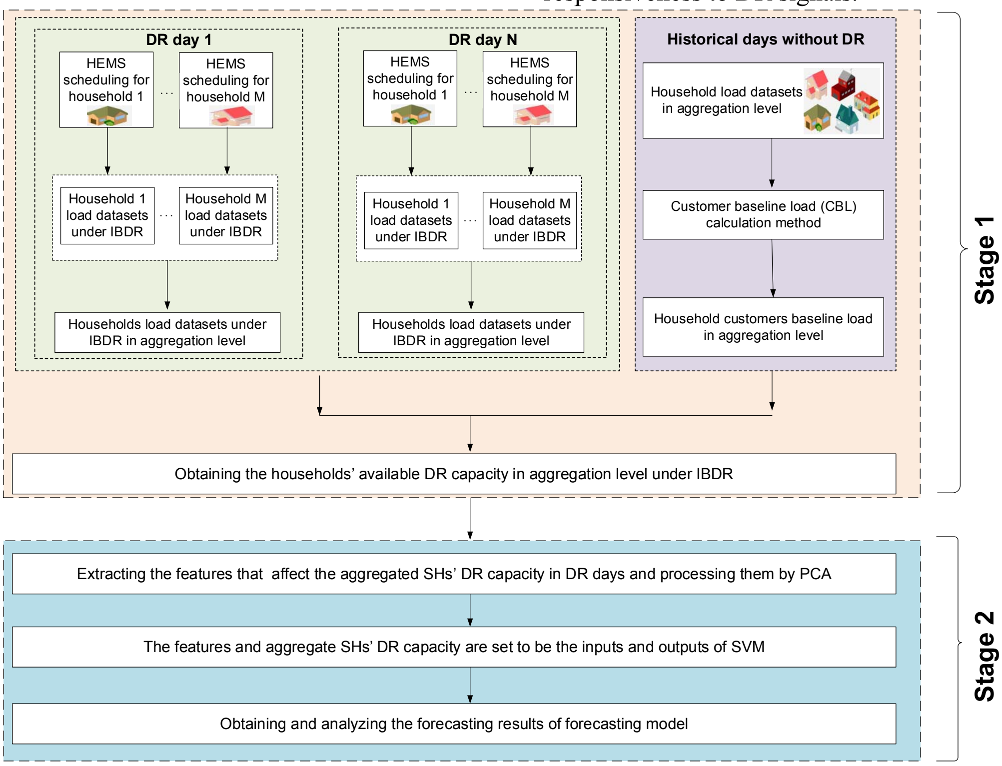


Fig. 3. Architecture of the proposed framework for aggregated DR capacity forecast

We proposed two basic ideas to address the issues listed above: (i) we focus on SHs and a model of HEMS [22], [23] is applied to perform an optimal appliance scheduling and accordingly model the customers' responsiveness during the IBDR program to minimize the individual customer's electricity cost; (ii) In order to accurately forecast the aggregated SHs' DR capacity, we analyse the characteristics that influence the aggregated SHs' DR capacity and extract 9 features as the input to the forecasting model. In addition, a machine learning approach is employed as the forecasting engine for the aggregated DR capacity forecasting. The overall architecture of the proposed framework is illustrated in Fig. 3.

III. Modeling Customers' Responsiveness

A. CBL Estimation

The CBL, which refers to the amount of electricity consumption that would have been consumed by the participants in the absence of the DR event, is introduced to quantitatively assess the DR capacity. The difference between the CBL and the actual load is regarded as the DR capacity. Generally, a CBL estimation method should be simple enough for all stakeholders to understand, assess, and implement. Thus, even though more sophisticated machine learning mechanisms could make the CBL estimation a higher accuracy, we insist on a user-friendly easy-to-understand average CBL estimation technique: $High X of Y$. Additional details on $High X of Y$ are available in [24].

B. HEMS Model for Smart Households Operation

HEMS has been widely deployed in the energy sector for several years. With the deployment of advanced metering infrastructure and communication devices, load control through the HEMS has become a reality. The HEMS could perform an optimal scheduling of the electrical appliances primarily to shift the demand during DR events by considering several factors, such as DR signals, load profiles, and customers' comfort to reduce the individual's energy cost. The HEMS architecture adopted in this paper is a tool to model the customers' electricity consumption behavior under the IBDR program. Fig. 4 shows a schematic diagram of the smart household equipped with a HEMS.

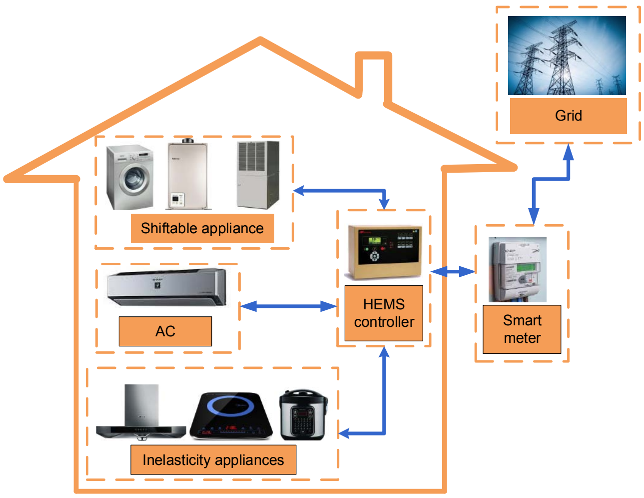


Fig. 4. Schematic diagram of the smart household

1) Objective Function

The objective is to minimize the individual customer's total daily cost of electricity consumption under the IBDR program. The objective function is reflected in equation (3), λ is the daily fixed electricity price. For a customer i in a DR day k, $P_{t,i,k}$ is the customer's total electricity consumption at time t. Δd represents the electricity consumption reduction (individual customer's DR capacity) during the DR event which is estimated through the actual load and the CBL shown in equation (5), d^{cont} denotes the contracted electricity consumption reduction between the individual customer and the LAs, Inc is the monetary reward for unit electricity consumption reduction an Pen is the monetary penalty for unit electricity consumption if individual customer's reduction does not reach to a pre-specified contracted level. $s(\mu)$, reflected in equation (4), is a logical variable that is used to judge whether the electricity consumption reaches to the contracted level, thus knowing if the customer needs to pay for the penalty. Equation (6) is used to judge if the customer will pay less compared to the circumstance without DR, the $req(P_{t,i,k})$ denotes the initial required electricity consumption of customer i at time t if there is no DR event. The main reason to add this equation is that the possibility of underestimation of CBL which could cause the underestimation of Δd, even the negative value of Δd. If the customers discover that unavailability to reduce the cost compared with the initial situation without participation in DR in day-ahead trading with LA, they will refuse to participate in that IBDR program. Equation (7) enforces the power balance constraints, where the total electricity consumption is made up of all shiftable loads $P_{t,i,k}^{shift}$, air conditioning (AC) $P_{t,i,k}^{air}$ and other inelastic loads $P_{t,i,k}^{lne}$.

$$Minimize\ F_1 = \sum_{t=1}^{T} \lambda \cdot P_{t,i,k} \cdot \Delta t \quad -Inc \cdot \Delta d + Pen \cdot (d^{cont} - \Delta d) \cdot s(\mu) \tag{3}$$

$$s(\mu) = \begin{cases} 1 & \mu > 0 \\ 0 & \mu \le 0 \end{cases} \qquad \mu = d^{cont} - \Delta d \tag{4}$$

$$\Delta d = \sum_{t \in T_{DR}} (P_{t,i,k}^{baseline} - P_{t,i,k}) \cdot \Delta t \tag{5}$$

$$F_2 = F_1 - \lambda \cdot \sum_{t=1}^{T} req(P_{t,i,k}) \cdot \Delta t \tag{6}$$

$$P_{t,i,k} = P_{t,i,k}^{shift} + P_{t,i,k}^{air} + P_{t,i,k}^{other} \tag{7}$$

2) Modeling the Shiftable Appliances

The shiftable appliances are considered as elastic loads, where the electricity consumption could be shifted forward or postponed at a certain time interval and will not greatly affect the customers' comfort. When modeling the shiftable appliances in the IBDR program, the optimized amount of shiftable loads should meet the initial requirements as enforced in equation (8). ΔT is the shifting interval under a certain incentive Inc at which the customer would like to modify the usage of the shiftable appliances. The power of the shiftable appliance at time t is reflected in equation (9), where p_i^{shift} is the rated power of shiftable appliance and $\tau_{t,i}^{shift}$ denotes the running time of appliance from time t-1 to t, and must be no more than Δt.

$$\sum_{t=t_{start}-\Delta T}^{t_{end}+\Delta T} (P_{t,i,k}^{shift}) = \sum_{t=t_{start}-\Delta T}^{t_{end}+\Delta T} req(P_{t,i,k}^{shift}) \tag{8}$$

$$P_{t,i,k}^{shift} = p_i^{shfit} \cdot \tau_{t,i,k}^{shift} / \Delta t \tag{9}$$

$$0 \le \tau_{t,i,k}^{shift} \le \Delta t \tag{10}$$

3) Modeling the Air Conditioner

Air conditioner (AC) is an ideal appliance for the DR program as it does not need to be completely switched off during the DR event. The electricity usage could be modified by changing its temperature set points in an acceptable range without greatly impact on the user's comfort zone that much. This paper adopted a set of simplified equivalent thermal parameters (ETP) [25]-[27] in order to model the AC units. When the AC is turned on, then the following constraint is enforced:

$$\theta_{room}^{\tau+1} = \theta_0^{\tau} + QR - (\theta_0^{\tau} + QR - \theta_{room}^{\tau}) e^{-\Delta\tau/RC} \tag{11}$$

When the AC is turned off, we will then have

$$\theta_{room}^{\tau+1} = \theta_0^{\tau} - (\theta_0^{\tau} - \theta_{room}^{\tau}) e^{-\Delta\tau/RC} \tag{12}$$

Where τ represents the resolution time of one minute. The AC unit can only work at its rated power p_i^{air} (on) or 0 (off) otherwise, which makes the indoor temperature change periodically within the range $[\theta_{\min}, \theta_{\max}]$ ($\theta_{\min} = \theta_{set} - \theta_{db}$, $\theta_{\min} = \theta_{set} + \theta_{db}$, where θ_{set} is the temperature set-point and θ_{db} is the temperature dead bandwidth). Therefore, one is able to acquire the time duration of switching on and off, which are reflected in equations (13) and (14) (this paper takes AC's cooling mode as an example). θ_1 is set to $\theta_{\max}$, θ_2 is set to $\theta_{\min}$, and α is set to -1 in summer when the AC's cooling ability is needed, while the circumstance is just the reverse in the winter when the heating ability is activated.

$$\Delta\tau_{on,i,k} = RC \ln\left(\frac{\theta_0^{\tau,k} - \theta_1 + Q_i R}{\theta_0^{\tau,k} - \theta_2 + Q_i R}\right) \tag{13}$$

$$\Delta\tau_{off,i,k} = RC \ln\left(\frac{\theta_0^{\tau,k} - \theta_2}{\theta_0^{\tau,k} - \theta_1}\right) \tag{14}$$

$$Q_i = \alpha\eta p_i^{air} \tag{15}$$

$$P_{t,i,k}^{air} = p_i^{air} \cdot \tau_{t,i,k}^{air} / \Delta t \qquad \tau_{t,i,k}^{air} \in \Delta\tau_{on,i,k} \tag{16}$$

$$0 \le \tau_{t,i,k}^{air} \le \Delta t \tag{17}$$

4) Modeling the Inelastic Appliances

In this paper, the appliances that greatly affect the customers' daily habits and have little potential to reduce the electricity usage are regarded as inelastic appliances. In such cases, any modification on the electricity consumption different from its regular usage habit could result in a significant violation of the customers' comfort zone. Thus, inelastic appliances are kept to their initial required electricity consumption and are not considered variable during the optimal operation of HEMS, as enforced in equation (18).

$$P_{t,i,k}^{Ine} = req(P_{t,i,k}^{Ine}) \tag{18}$$

IV. Forecasting Model

A. Forecast Engine: Overall Structure

Having modeled the customers' responsiveness to the DR signals and acquiring the load data under the IBDR program, one is then able to estimate the aggregated DR capacity in the DR days. This section elaborates how the forecasting model is constructed to forecast the aggregated DR capacity as illustrated in Fig. 5. The forecasting model includes three main stages: (i) extraction of the features that could capture the characteristics of aggregated DR capacity as input pairs to be processed by the PCA in order to avoid noise and redundant information; (ii) adoption of a proper forecast engine; (iii) evaluation of the forecast performance by accuracy metrics.

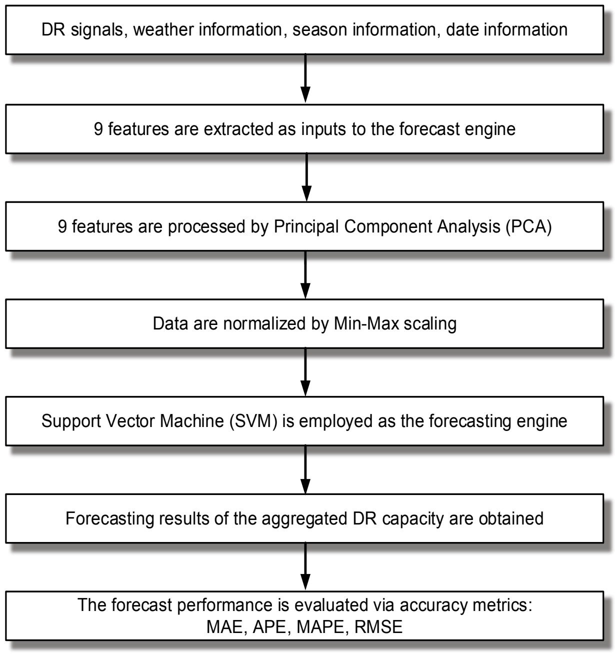


Fig. 5. Overall forecast structure

B. Feature Extraction

While daily electricity consumption is conditioned mostly by daily activities and weather patterns, DR signals and particularly the economic incentives affect the ability and the willingness of consumers to be responsive in the IBDR program, thereby affecting the DR capacity. Compared with forecast of the DR capacity for individual customers, forecast of the aggregated DR capacity does not primarily focus on the individual customer's household features such as house size, type of dwelling, the number of occupants in the house, etc., since individual customer's electricity usage characteristics will not greatly influence the aggregated DR capacity. Hence, only the factors with significant impacts on the evolution principle of aggregated DR capacity in different DR days should be considered. The extracted features in this research are mainly categorized in the following two groups: (i) features that influence the daily aggregated electricity consumption: these features mainly include the weather conditions, season and date information. More specifically, they are the highest and lowest temperature, as well as season, weekday and weekend labels in the upcoming DR day; (ii) features that decide how much electricity consumption the customers are willing to reduce on the basis of their daily electricity usage: these features mainly include the DR signals and related information such as monetary reward, customer baseline energy (CBE, i.e., the initial electricity consumption under CBL during the DR event duration), DR start time and duration in the upcoming DR day.

C. Principal Component Analysis (PCA)

Principal component analysis (PCA) is a mathematical transformation approach that converts a given set of related variables into another set of unrelated variables by orthogonal transformation. The main role of PCA is to reduce noise and redundant data (i.e., dimensionality reduction) while preserving all critical information in the original dataset as much as possible. In this paper, PCA is utilized to process the datasets and analyze the extracted features.

D. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a statistical learning approach that can be applied to solve problems in nonlinear regression and forecasting. Unlike the classical neural networks, SVM formulates the statistical learning problem as quadratic programming with linear constraints through nonlinear kernels, offering a high generalization ability and solution sparsity. In addition, SVM has better computational performance with a promising convergence. In this paper, SVM is used for day-ahead aggregated DR capacity forecast engine.

E. Performance Metrics for Model Accuracy Evaluations

To assess the forecast performance of the proposed model, different error metrics are employed as the benchmark. The most widely used metrics are the mean absolute error (MAE), the absolute percent error (APE), the mean absolute percent error (MAPE) and the root mean square error (RMSE), where the lower the error, the better the prediction performance.

$$MAE = \frac{1}{K}\sum_{k=1}^{K}\left|\hat{f}_k^{agg} - f_k^{agg}\right| \tag{19}$$

$$RMSE = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(\hat{f}_k^{agg} - f_k^{agg})^2} \tag{20}$$

$$APE = \left|\frac{\hat{f}_k^{agg} - f_k^{agg}}{\hat{f}_k^{agg}}\right| \tag{21}$$

$$MAPE = \frac{1}{K}\sum_{k=1}^{K}\left|\frac{\hat{f}_k^{agg} - f_k^{agg}}{\hat{f}_k^{agg}}\right| \tag{22}$$

Where f_k^{agg} and $\hat{f}_k^{agg}$ denote the actual and forecasted aggregated DR capacity values at the DR day k.

V. Case Study

A. Studied Dataset: Assumptions and Sepcifications

The data used in this research is a real-world dataset from Pecan Street experiment in Austin, TX [28]. The dataset contains minute-resolution electricity consumption data from 500 homes (both the home-level and individually monitored appliance circuits). In order to consider the seasonal and yearly effect, this paper chooses a two-year load data with a 1-minute interval from Jan.1st, 2015 to Dec.31st, 2016. The data set is trimmed by removing the customers with missing load data, and finally, 170 customers with two-year load data are characterized for the analysis.

B. Experimental Settings

1) Selection of DR Days and DR Signal Settings

Since the dataset is on the residential customers' daily load profiles with no DR event, this paper sets 65 DR days artificially and assumes having an HEMS performing optimal households' appliance operation for each customer during DR events.

TABLE I
PARAMETERS IN THE DR EVENTS

Base Fixed Tariff	0.3$/kWh		
Monetary Reward	0.3$/kWh	0.4$/kWh	0.5$/kWh
DR Event Time	12:00-14:00 17:00-19:00	12:00-15:00 17:00-20:00	13:00-15:00 18:00-20:00

The DR signals are mainly related to DR start time, duration and monetary rewards. As tabulated in Table I, there are 3 different types of monetary reward and 6 different types of DR event times. In each DR event, monetary reward and DR event time are randomly selected accordingly, thereby leading to varying DR signals during different DR days.

2) Settings on Different Customer Types

In this paper, two types of customers are studied, the sensitivity of which to the monetary rewards is different; specifically, one type pays more attention to the comfort level and willing to modify the electricity usage less (Type 1), the other is willing to change the consumption habits to pursue cheaper electricity bills in the IBDR program (Type 2). Customer Type 1 and 2 are set to 7 different proportions of all customers as presented in Table II.

TABLE II
DISTRIBUTION OF TWO CUSTOMER TYPES

Customer Distribution	Customer Type 1	Customer Type 2
Distribution 1	0.2	0.8
Distribution 2	0.3	0.7
Distribution 3	0.4	0.6
Distribution 4	0.5	0.5
Distribution 5	0.6	0.4
Distribution 6	0.7	0.3
Distribution7	0.8	0.2

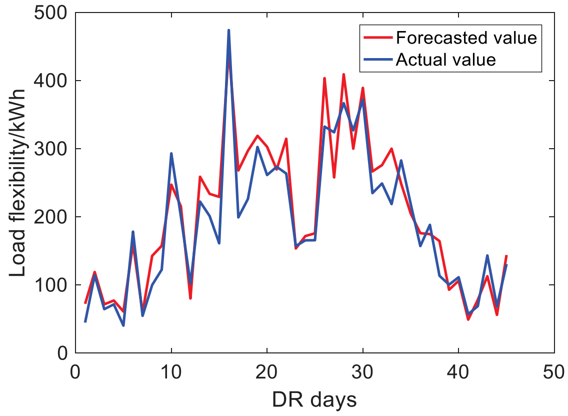
C. Numerical Results and Analysis

TABLE III
COMPARISON OF MAE, MAPE, AND RMSE UNDER DIFFERENT DISTRIBUTION OF CUSTOMERS

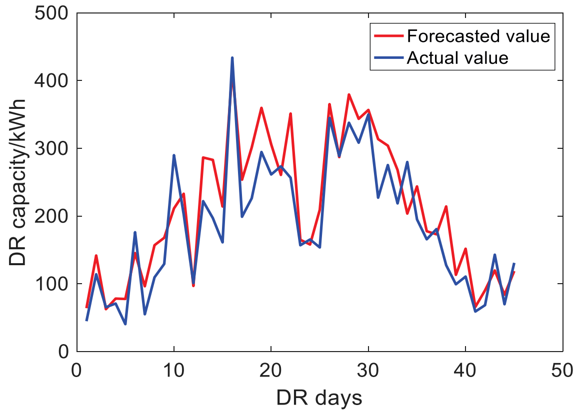
Customer Distribution	MAE	MAPE	RMSE
Distribution 1	28.1653	0.1477	35.4016
Distribution 2	27.3701	0.1417	34.8597
Distribution 3	26.9967	0.1374	35.1978
Distribution 4	25.8484	0.1323	34.0790
Distribution 5	28.9893	0.1426	38.4267
Distribution 6	29.5047	0.1419	39.2675
Distribution 7	30.5032	0.1437	40.7109

Different types of customers may have different sensitivity to DR signals. LAs, however, are usually not able to figure out details on that of every single customer. What the LAs need to understand is the circumstance in aggregation level when participating in the day-ahead market. Therefore, the LAs seek a universal forecast of the aggregated DR capacity irrespective of the distribution of customers. Table III presents the forecast accuracy of the proposed approach with regards to MAE, MAPE, and RMSE metrics under different distributions of customers, where one can see the similarity of the metrics for forecasts under different distributions of customers. Furthermore, the accuracy and stability of the forecasts are verified in Table III through the low MAPE metric on 7 different distributions.

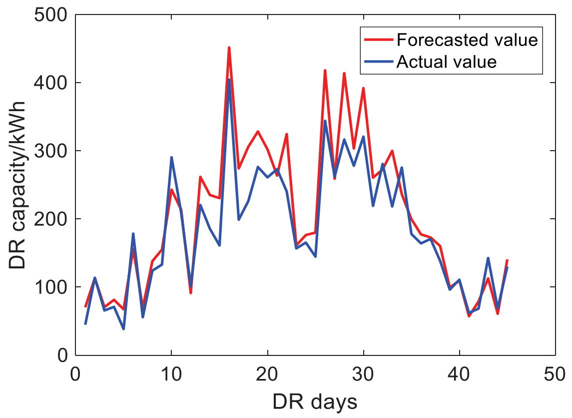
We further provide a comparison of the forecasted and actual values of aggregated DR capacity in figure 6 and a cumulative distribution of the APE in different DR days in Fig. 7 (considering the length and layout of the paper, we only list the figures under distributions 1, 4, and 7). It can be seen intuitively in Fig.6 that the forecasted value could consistently follow the actual trend, reflecting a promising fitting performance between the lines of forecasted and actual value under different distribution of customers. Furthermore, Fig. 7 shows that the proposed forecast model could provide the aggregated DR capacity forecasts with over 90% accuracy for about 40% of the testing dataset and over 80% accuracy for over 70% of the testing dataset



(a) Distribution 1

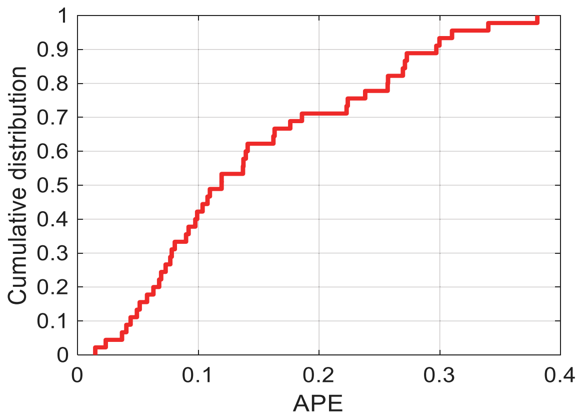


(b) Distribution 4

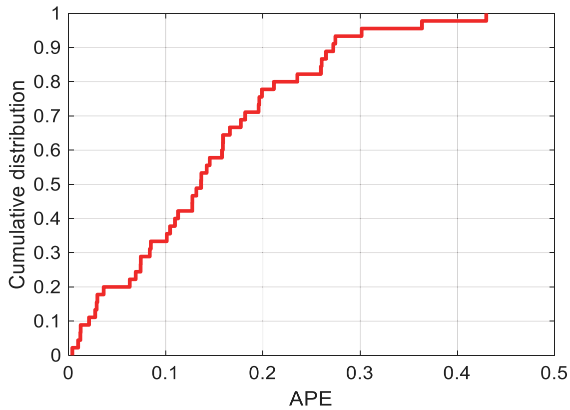


(c) Distribution 7

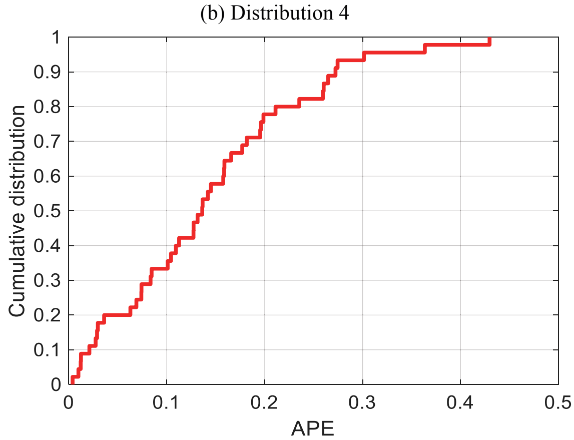
Fig. 6. Forecast results: comparison of the forecasted and actual values of the aggregated DR capacity under 3 different distributions of customers.



(a) Distribution 1



(b) Distribution 4



(c) Distribution 7

Fig. 7. Cumulative distribution of APE corresponding to the forecast results in different DR days under 3 different distributions of customers

D. .Discussions

The main goal in this paper is to pursue an accurate forecast of aggregated DR capacity that the LAs are able to acquire from customers in day-ahead market. Therefore, in addition to the forecast results presented earlier, we will further explore that how the forecast performance is affected by (i) extracted features and (ii) the number of customers.

1) Aggregated DR capacity Forecasting and the Extracted Features

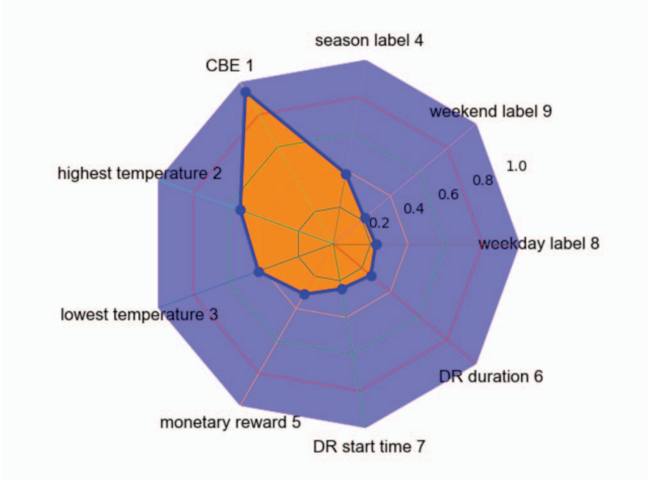


Fig. 8. The MIC between 9 features and the aggregated DR capacity

TABLE IV
COMPARISON OF RESULTS WITH AND WITHOUT PCA PROCESSING ON DIFFERENT FEATURES

	MAE		MAPE		RMSE	
	No-PCA	PCA	No-PCA	PCA	No-PCA	PCA
1	38.9296	38.9296	0.1780	0.1780	51.1309	51.1309
2	30.5665	30.8505	0.1510	0.1484	43.7043	45.0578
3	26.3856	27.0816	0.1349	0.1329	35.5612	36.4180
4	31.7741	27.2996	0.1545	0.1356	41.9170	36.6172
5	35.8724	26.9893	0.1866	0.1345	46.4380	36.1970
6	31.3267	27.2799	0.1844	0.1356	40.6089	36.5772
7	29.3432	26.3768	0.1807	0.1337	36.4692	34.8447
8	31.1051	25.9818	0.1597	0.1336	39.3625	34.1979
9	36.4368	25.8484	0.1838	0.1323	45.0469	34.0790

In order to explore the correlation of the extracted features and the aggregated DR capacity forecasts, the maximal information coefficient (MIC) [29] with promising performance in analyzing sophisticated relationships between variables (i.e., linear and nonlinear, functional and non-functional) is employed as demonstrated in Fig. 8. It can be inferred from Fig. 8 that the CBE, highest temperature, and lowest temperature reveal a higher correlation (i.e., higher MIC) with the aggregated DR capacity forecasting. To further explore the impacts of these features on the forecast results, this paper ranks the features according to the corresponding MIC with the aggregated DR capacity (from 1 to 9 most correlative features) and compares the forecast results with regards to the MAE, MAPE, and RMSE metrics under the conditions of with and without PCA processing. As illustrated in Table IV, if the features are not processed by PCA, the 3 most correlative features are found the best choices as the input to the forecast model, while choosing all 9 features would be preferred when PCA processing is adopted. The reason for such a difference primarily lies in the redundancy of input information. On the one hand, the increase in the number of features could offer more information in forecasting to help improve its accuracy to some extent; on the other hand, additional number of features may result in information redundancy due to the coupling relationships that may exist in several features.

2) Aggregated DR capacity Forecasts and the Number of Residential Customers

Different from MAE and RMSE, the MAPE is a relative accuracy performance evaluation metric, taking an average over the APE result in all forecasting DR days and suitable for exploring the impacts of different number of customers on the overall forecast accuracy. Fig. 9 shows the MAPE of the forecast results in all forecasting DR days in the testing datasets under different number and distribution of customers. Note that different proportion of Customer Type1 in Fig. 9 reflects the different distribution of customers as defined in Table I. One can see that, irrespective of the distribution of customers, the MAPE exhibits a descending trend as the number of customers increases, i.e., that higher the number of customer participants in the IBDR program, the higher accurate the forecast is in all forecasting DR days.

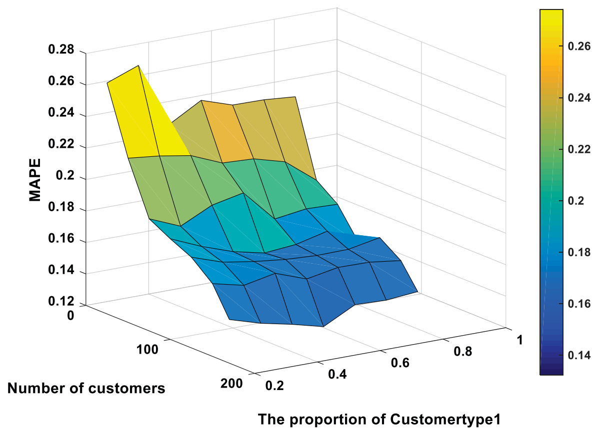


Fig. 9. The MAPE on the forecast results under different number and distributions of customers

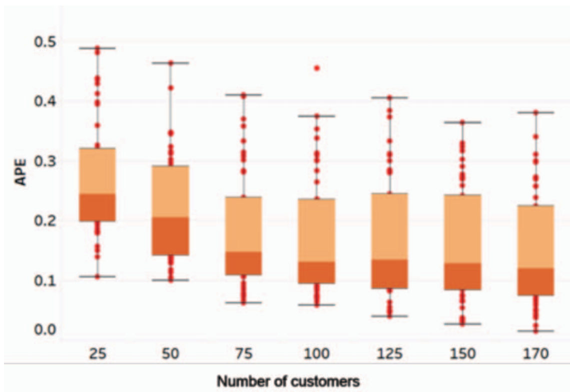


Fig. 10. Distribution of APE on the forecast results in all forecasting DR days under different number of customers in Distribution 1

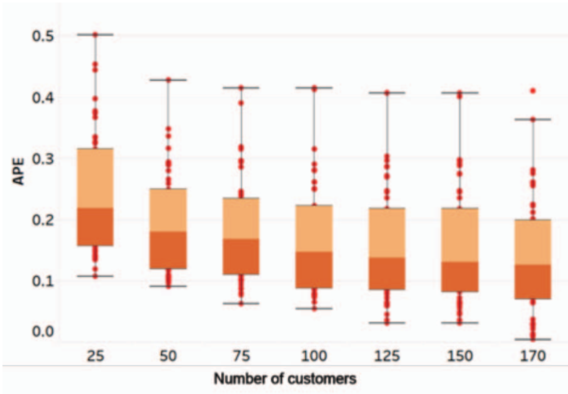


Fig. 11. Distribution of APE on the forecast results in all forecasting DR days under different number of customers in Distribution 4

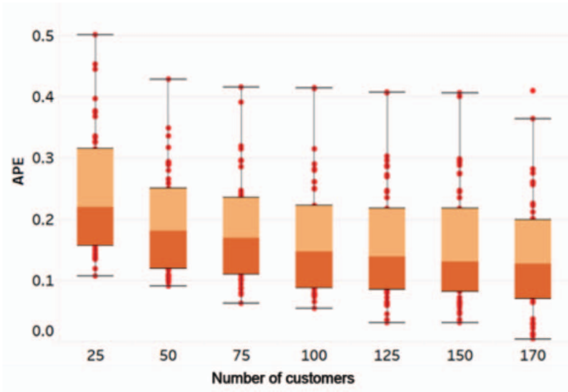


Fig. 12. Distribution of APE on the forecast results in all forecasting DR days under different number of customers in Distribution 7

Furthermore, the box-fit figures which describe the distribution of APE on the forecast results in all forecasting DR days under different number of customers are illustrated in Figs. 10-12 (considering the length and layout of the paper, we only list the results under distribution 1, 4, and 7). One can infer two main observations from these figures: on the one side, with the increase in the number of customers, the box-fit figures demonstrate a lower distribution of APE in all forecasting DR days; on the other side, a descending trend is apparent when the number of customers is small, while it is stable when it is large. This highlights the fact that as more customers participate in the IBDR program, the more accurate forecasts of the aggregated DR capacity could be achieved, thereby reducing the risk for the LAs to participate in the market trading. It could be inferred that if the LAs act on behalf of only a small number of customers to participate in the IBDR program, then the fluctuation in the individual customers' electricity usage might greatly influence the aggregated DR capacity forecasts. However, once a considerable number of customers are willing to choose LAs as their agent to participate in IBDR program, such fluctuations in the electricity usage is limited and a more stable and accurate forecast of the aggregated DR capacity could be realized.

VI. Conclusion

This paper presents a model based on the SVM to forecast the day-ahead aggregated DR capacity that the LAs could obtain from SHs under a IBDR program. The numerical results verifies the promising forecast performance of the proposed model in terms of both stability and accuracy, as well as the universality in different scenarios. Furthermore, this paper shed light on selection of features as a significant step when developing a forecast model. If extraction of the most suitable features is burdensome or not possible, then the PCA is suggested as a reasonable choice for processing the redundant information of the diverse features. In addition, the impacts of the number of customers on forecast results are considered, and it can be implied that if LAs want to pursue an more accurate forecast results and a low-risk participation in day-ahead market trading, then they need to induce more customers to take part in the IBDR program. What's more, considering the increasing interests among customers to install distributed PV systems [30, 31] that will greatly affect their load profiles, future work will explore the impacts caused by additional PV penetrations on the aggregated DR capacity forecasting.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (grant No. 51577067), the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (grant No. LAPS19016), the Fundamental Research Funds for the Central Universities (grant No. 2018QN077), the Science and Technology Project of State Grid Corporation of China (SGHE0000KXJS1800163, kjgw2018-014).

References

[1] F. Wang, K. Li, L. Zhou, H. Ren, J. Contreras, M. Shafie-khah, and J. P. S. Catalão , "Daily pattern prediction based classification modeling approach for day-ahead electricity price forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 105, pp. 529–540, Feb. 2019.

[2] Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them, *U.S. Dept. Energy, Washington, DC, USA, Tech. Rep.*, Feb. 2006.

[3] F. Wang, Z. Zhang, C. Liu, Y. Yu, S. Pang, and N. Duić, "Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting," *Energy Convers. Manag*. vol. 181, pp. 443–462, Feb. 2019.

[4] F. Wang, Z. Zhen, C. Liu, Z. Mi, B.M. Hodge, M. Shafie-khah, and J. P.S. Catalão, "Image phase shift invariance based cloud motion displacement vector calculation method for ultra-short-term solar PV power forecasting," *Energy Convers. Manag*. vol. 157, pp. 123–135, Feb. 2018.

[5] Z. Zhen, S. Pang, F. Wang, K. Li, Z. Li, H. Ren, M. Shafie-khah, and J. P.S. Catalão, "Pattern Classification and PSO Optimal Weights Based Sky Images Cloud Motion Speed Calculation Method for Solar PV Power Forecasting," *IEEE Trans. Ind. Appl.*, vol. 55, no. 4, pp. 3331–3342, Jul. 2019.

[6] F. Wang, Z. Zhen, Z. Mi, H. Sun, S. Su, and G. Yang, "Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting," *Energy Build.*, vol. 86, pp. 427–438, Jan. 2015.

[7] O. G. Santin, L. Itard, and H. Visscher, "The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock," *Energy Build.*, vol. 41, pp. 1223–1232, Nov. 2009.

[8] F. Wang, K. Li, N. Duić, Z. Mi, B.M. Hodge, M. Shafie-khah, and J. P. S Catalão, "Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns," *Energy Convers. Manag.*, vol. 171, April, pp. 839–854, Sep. 2018.

[9] L. Gkatzikis and I. Koutsopoulos, "The Role of Aggregators in Smart Grid Demand," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1247–1257, Jul. 2013.

[10] F. Wang, H. Xu, T. Xu, K. Li, M. Shafie-khah, and J. P. S. Catalão, "The values of market-based demand response on improving power system reliability under extreme circumstances," *Appl. Energy*, vol. 193, pp. 220–231, May. 2017.

[11] K. Li, Q. Mu, F. Wang, Y. Gao, G. Li, M. Shafie-khah, J. P. S Catalão, Y. Yang and J. Ren, "A Business Model Incorporating Harmonic Control as a Value-added Service for Utility-owned Electricity Retailers," *IEEE Trans. Ind. Appl.*, Early Access. DOI: 10.1109/TIA.2019.2922927

[12] F. Wang, L. Liu, Y. Yu, G. Li, J. Li, M. Shafie-khah, and J. P.S. Catalão , "Impact analysis of customized feedback interventions on residential electricity load consumption behavior for demand response," *Energies*, vol. 11, no. 4, pp. 1–22, Apr. 2018.
[13] F. Wang, Y. Yu, X. Wang, H. Ren, M. Shafie-Khah, and J. P. S. Catalão, "Residential electricity consumption level impact factor analysis based on wrapper feature selection and multinomial logistic regression," *Energies*, vol. 11, no. 5, pp. 1–26, May. 2018.
[14] I. A. Sajjad, G. Chicco, and R. Napoli, "Definitions of Demand Flexibility for Aggregate Residential Loads," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2633–2643, Nov. 2016.
[15] Q. Chen, F. Wang, B.M. Hodge, J. Zhang, Z. Li, M. Shafie-Khah, and J. P. S. Catalão., "Dynamic Price Vector Formation Model-Based Automatic Demand Response Strategy for PV-Assisted EV Charging Stations," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2903–2915, Nov. 2017.
[16] J. Schofield, R. Carmichael, S. Tindemans, M. Woolf, M. Bilton, and G. Strbac, "'Residential consumer responsiveness to time-varying pricing - Low Carbon London Learning Lab', Report A3 for the 'Low Carbon London' LCNF project: Imperial College London," *Rep. A3 "Low Carbon London" LCNF Proj. Imp. Coll. London*, pp. 1–77, 2014.
[17] C. K. Woo, I. Horowitz, and I. M. Sulyma, "Relative kW response to residential time-varying pricing in British Columbia," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1852–1860, Dec. 2013.
[18] C. K. Woo, R. Li, A. Shiu, and I. Horowitz, "Residential winter kWh responsiveness under optional time-varying pricing in British Columbia," *Appl. Energy*, vol. 108, pp. 288–297, Aug. 2013.
[19] N. G. Paterakis, A. Tascikaraoglu, O. Erdinc, A. G. Bakirtzis, and J. P. S. Catalao, "Assessment of Demand-Response-Driven Load Pattern Elasticity Using a Combined Approach for Smart Households," *IEEE Trans. Ind. Informatics*, vol. 12, no. 4, pp. 1529–1539, Aug. 2016.
[20] M. Vallés, A. Bello, J. Reneses, and P. Frías, "Probabilistic characterization of electricity consumer responsiveness to economic incentives," *Appl. Energy*, vol. 216, pp. 296–310, Apr. 2018.
[21] F. Wang, K. Li, C. Liu, Z. Mi, M. Shafie-Khah, and J. P. S. Catalao, "Synchronous pattern matching principle-based residential demand response baseline estimation: Mechanism analysis and approach description," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6972–6985, Nov. 2018.
[22] F. Wang, L. Zhou, H. Ren, X. Liu, S. Talari, M. Shafie-khah, and J. P. S. Catalão, "Multi-objective optimization model of source-load-storage synergetic dispatch for building energy system based on TOU price demand response," *IEEE Trans. Ind. Appl.*, vol. 54, no. 2, pp. 1017-1028, Mar. 2018.
[23] M. Shafie-Khah and P. Siano, "A stochastic home energy management system considering satisfaction cost and response fatigue," *IEEE Trans. Ind. Informatics*, vol. 14, no. 2, pp. 629–638, Feb. 2018.
[24] T. K. Wijaya, M. Vasirani and K. Aberer, "When Bias Matters: An Economic Assessment of Demand Response Baselines for Residential Customers," *IEEE Trans. Smart Grid,* vol. 5, no. 4, pp. 1755-1763, Jul. 2014.
[25] Y. Bao, M. Hu, Y. Hong, P. Chen, and J. Ju, "Accuracy analysis and improvement of the state-queuing model for the thermostatically controlled loads," *IET Gener, Transm, Distrib.*, vol. 11, no. 5, pp. 1303–1310, Mar. 2017.
[26] C. H. Wai, M. Beaudin, Z. Hamidreza, A. Schellenberg, Ning Lu , "Cooling Devices in Demand Response : A Comparison of Control Methods," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 249–260, Jan. 2015.
[27] N. Lu, "An evaluation of the HVAC load potential for providing load balancing service," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1263–1270, Sep. 2012.
[28] Pecan Street, "Real energy. real customers. in real time." *http://www. pecanstreet.org/energy/*, 2012
[29] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, Turnbaugh, Peter J, E. S. Lander, M. Mitzenmacher, P. C. Sabeti, "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518-1524, Dec. 2011.
[30] F. Wang, K. Li, X. Wang, L. Jiang, J. Ren, Z. Mi, M. Shafie-khah, and J. P. S. Catalão, "A Distributed PV System Capacity Estimation Approach Based on Support Vector Machine with Customer Net Load Curve Features," *Energies*, vol. 11, no. 7, pp. 1750, Jul. 2018.
[31] J. Lai, X. Lu, F. Wang, P. Dehghanian, and R. Tang, "Broadcast Gossip Algorithms for Distributed Peer-to-Peer Control in AC Microgrids," *IEEE Trans. Ind. Appl.*, vol. 55, no. 3, pp. 2241-2251, Feb. 2019, DOI: 10.1109/TIA.2019.2898367.