Methodologies for Customer Baseline Load Estimation and their Implications

Shubhasmita Pati, Satish J. Ranade, Olga Lavrova
Klipsch School of Electrical & Computer Engineering
New Mexico State University, Las Cruces, NM 88003, USA
{shubha, sranade , olavrova}@nmsu.edu

***Abstract*—Demand response (DR) has gained increasing importance with the grid modernization and evolution of the smart grid. Customer baseline load (CBL) informs the expected consumption of the consumers and is an important factor in DR management. To develop forecasting methods for CBL that work satisfactorily is a major challenges faced by independent system operators (ISOs) and utilities. Consequently, it is critical that these methods are evaluated based on their performance. The major objective of this paper is to empirically evaluate the performance of CBL estimation from historical observations, for residential customers. In addition to some well established methods, we have used an artificial neural network regression for estimating CBL. We use the CER smart meter data from 6435 residential and small industrial customers collected between July 2009 to Dec 2010, for this evaluation. We performed the error analysis of different methods using both the accuracy and bias metrics.**

***Index Terms*—Customer Baseline Load, Load Forecasting, Demand Response**

I. Introduction

The consumers around the globe expect an sustained and on-demand supply of electricity, at all the time. In recent years, the consumer consumption patterns have changed and are expected to change further with introduction of new loads such as, intelligent electronic equipment and electric vehicles. Meeting the dynamically changing customer demands is challenging for the grid operators as electricity cannot be efficiently stored in substantial quantities. Operators must organize the generation and transmission grid so that demand and supply exactly match, every moment of the day, every day of the year, in every location. Most electric utilities serve a variety of customers: residential, commercial and industrial. Each class of customer have different consumption needs, resulting in a differing load profile for each class. Accurate forecasts of such consumption are necessary for suppliers, financial institutions and other participants in electric energy generation, transmission, distribution and trading. These forecasting issues have become more convoluted with the penetration of distributed energy resources (DERs). The relatively recent emergence of the smart grid technologies attempts to address these challenges and facilitates new paradigms under the umbrella of demand and response (DR) issues.

978-1-7281-4436-8/20/$31.00 ©2020 IEEE

A. Demand and response in the smart grid

With the advent of *smart grid* technologies, the concept of power delivery transcends to efficient energy delivery. This ensures optimized generation, transmission, and consumption over conventional electricity systems. The assimilation of DERs necessitates the bidirectional information exchange between consumers and the grid operators. The incentives associated with roof-top solar panels transform a large number of residential consumers to potential producers. This motivates users participation in energy management and savings through the DR mechanism. A key mission of the smart grid is to manage DR for reducing peak electricity load and facilitate consumers to participate in energy-saving programs [1]. The duplex information flow through communication infrastructure enables the control and scheduling of deferable or controllable loads such as residential air-conditioners and electric heaters, to optimize energy consumption and costs. An alternative approach for DR management is to dynamically vary tariff or incentives based on consumption. Essentially, the DR management programs integrates the varying demand with the market dynamics.

In the US, the Federal Energy Regulatory Commissions (FERC) 745 order mandates that the DER owners have the right to participate in the energy market [2]. The demand reduction is treated as a supply source according to this order. In addition, FERC instructs the independent system operators (ISOs) and regional transmission operators (RTOs) to remunerate the residential DER producers with locational marginal price. However, appropriate compensation for DER producers for demand reduction is not yet standardized [3], [4]. The FERC order 745 and its ratification by the Supreme Court of the United States, forces the DR programs to heavily depend on accurate forecasting of the energy consumption profiles of customers, to decide on how the providers can fix tariff or compensate the customers [5], [6].

B. Customer Baseline Load

The *predicted* consumption level, or the amount of electricity that a customer would have consumed in the absence of any DR program, is defined as the Customer Baseline load (CBL) [7]. This is different from the load forecasting methods which predict the actual load to be consumed by a customer irrespective of the DR scheme. The CBL can be used to determine the incentive payment in DR programs is a preferred approach for many operators [6]. The payment settlement in Peak Time Rebate (PTR) programs are regulated by the estimated CBL. The erroneous estimation of CBL would adversely affect the performance of a DR program. Thus, effectiveness of DR, necessitates accurate forecasting of CBL [8]. Some of the well-established CBL calculation methods have been adopted by utilities for PTR programs targeted at large industrial customers [9]. Because this program is not offered to the residential customers, the evaluation of CBL calculation methods for these customers have gained limited attention [10].

In this paper, the CBL estimation methods are extensively evaluated for residential customers similar to [7]. In addition to the widely used methods for industrial CBL estimation, we also evaluate a method based on neural network regression. The Section II discusses the methods in detail, Section III describes the evaluation metrics, and Section IV summarizes the results, followed by the conclusion in Section V.

II. Methods

We consider three established kinds of CBL estimation methods that are popular among ISOs [11]. These are *XofY*, *exponential moving average*, and the *regression* methods. Many attempts have been made in the literature to enhance the performance of these methods by either modifying the tunable parameters or by refine the input data [12]. The *XofY methods* have been divided into three categories, namely *HighXofY*, *LowXofY* and *MidXofY*. In the *moving avarage* category we consider the *exponential moving average* method and for regression, we chose the *artificial neural network (ANN)*.

A. HighXofY *Method*

We define the non-DR days as the days for which DR events are absent. *HighXofY* method selects Y non-DR days excluding weekends and holidays. Let C be the set of all customers, and T be the set of hourly time-stamps in a day. X number of days are selected from the chosen Y non-DR days according to a specific criteria based on consumption. The baseline load for is defined as the average consumption of the X days. The *HighXofY* CBL of customer $n \in C$ for time-stamp $t \in T$ for day d is expressed as:

$$B_n^{HighXofY}(d,t) = \frac{1}{X} \sum_{d \in High(X,Y,d)} l_n(d,t) \qquad \text{(II-A)}$$

Here, $l_n(d,t)$ is the load consumed by n^{th} customer on d^{th} day at time t. The New York ISO (NYISO) uses this method for CBL calculation with parameters $X = 5$ and $Y = 10$, and we have adopted the same parameters in this study.

B. LowXofY *Method*

This method does the opposite of *HighXofY*. It chooses X number of lowest consumption days out of Y days to calculate the baseline. The *LowXofY* CBL of customer $n \in C$ for time-stamp $t \in T$ on day d is defined as follows:

$$B_n^{LowXofY}(d,t) = \frac{1}{X} \sum_{d \in Low(X,Y,d)} l_n(d,t) \qquad \text{(II-B)}$$

For the simulations in this paper, we have chosen X and Y as 4 and 5 based on [6].

C. MidXofY *Method*

This method rejects some of the low and high consumption days and the remaining X consumption days in between, are used for estimating the CBL.
Let $X \leq Y$, $(Y - X)\,\%\,2 = 0$, and $Z = (Y - X)/2$. In this method, the Z-lowest and Z-highest consumption days are excluded and, the remaining X days are used in the CBL calculation. The *MidXofY* CBL of customer $n \in C$ for time-slot $t \in T$ on day d is expressed as follows:

$$B_n^{MidXofY}(d,t) = \frac{1}{X} \sum_{d \in Mid(X,Y,d)} l_n(d,t) \qquad \text{(II-C)}$$

We chose $X = 4$ and $Y = 6$, for the simulations in this paper.

D. *Exponential Moving Average Method*

The historical consumption pattern of consumers are given due importance in this method. The initial average load of a consumer is computed and the exponential moving average is computed continuously, with the arrival of new observations. The estimated baseline is expressed as:

$$S_n(d_\tau, t) = \frac{1}{\tau} \sum_{j=1}^{\tau} l_n(d_j, t) \qquad \text{(II-D)}$$

Here, τ is a constant with a value between 1 and k. It represents the number of days considered for determining the initial average load for consumer $n \in C$ for time-stamp $t \in T$.

The exponential moving average baseline for j^{th} day, with $\tau \leq j \leq k$ is expressed as:

$$S_n(d_j, t) = \delta . S_n(d_{j-1}, t) + (1 - \delta) l_n(d_j, t) \qquad \text{(II-D)}$$

where, δ is a constant with a value between 0 and 1. It can be observed that the relative importance of the load consumption for each day, in the calculation of exponential moving average baseline, decreases exponentially with time. The CBL for customer $n \in C$ for day d at time-slot $t \in T$ is estimated as:

$$B_n^{exp}(d,t) = S_n(d_k, t) \qquad \text{(II-D)}$$

One of the limitations of this method is its inability to estimate baseline for days earlier than $d_{\tau+1}$. To alleviate this, the DR programs ensure that data for sufficient number of days are taken into account while calculating the initial average load. This methodology is employed by the NewEngland ISO (ISONE). For a new consumer in the DR program, they estimate the initial baseline load as the hourly average of the preceding five business days, rejecting holidays and irrelevant event days. This method of initial baseline calculation is also known as *CBL6* and is expressed mathematically as:

$$S_6 = \frac{\sum_{t=1}^{5}}{5} kWh_{n,h} \qquad \text{(II-D)}$$

After calculating *CBL6*, the CBL for the new consumer is calculated according to equation II-D, with $\delta = 0.9$. The CBL

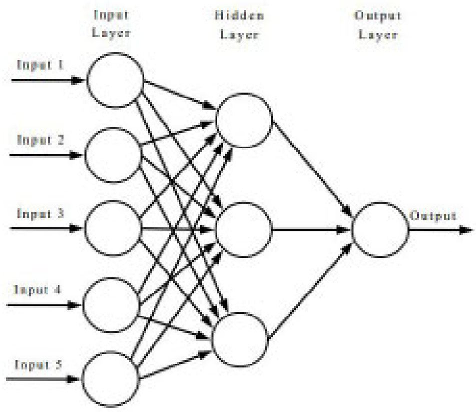

Fig. 1. Feed-forward Neural Network structure

is estimated for each day, excluding weekends, holidays, and event days. The choice of $\delta = 0.9$ implies that 90% of the importance is given to the previous day's CBL and 10% of the importance is given for the current day consumption, for estimating CBL.

E. Neural network Regression Method

This method uses multiple regressions based on neural network to calculate the baseline. The structure of a feed-forward neural network (NN) is shown in figure II-E. The network diagram shown has three layers; input, hidden, and output layer. Each layer consists of one or more nodes connected by links or weights. The feed-forward NNs allow signals to travel one way only; from input to output. The raw data fed into the network, is processed by the input nodes. The hidden node activity is determined by the output of the input units, the weights between the input and hidden layer nodes, and the non-linear activation function inside the hidden layer nodes. The output depends on the hidden nodes output, the weights between hidden and output nodes, and the activation function of the output nodes. The NN model is simply a nonlinear function from a set of input to a set of output variables, controlled by a set of adjustable parameters or weights [13]. The weights are adjusted by a process of training. The NN are trained by minimizing the error between a set of known input and and output data, in an iterative process. Once trained, the NNs can be used for prediction by applying new input data.

From the context of determining CBL, the NN is fed with an input from energy consumed from past few days and the desired output is to estimate the CBL for the current day. The error from the predicted and actual value are fed back to the network and weights are adjusted recursively to improve the prediction of following day.

Let a set of days upto the current day d is, $D(\infty, d) = d_1, ..., d_P$. The CBL for customer $i \in C$ on day d for time-slot $t \in T$ is expressed as:

$$B_n^{NN}(d,t) = \sum_{m=1}^{M} W_m^H \left(\sum_{j=1}^{P} W_{jm}^I l_n(d_j,t) + W_0^I \right) + W_0^H \qquad \text{(II-E)}$$

Here, W_m^H denotes the weight between mth the hidden layer node and the output layer node. W_{jm}^I denotes the weight between j^{th} input layer node and m^{th} the hidden layer node. The bias corresponding to Input layer and hidden layer are represented as W_0^I and W_0^H. The weights and biases were determined through a Levenberg-Marquardt algorithm. The weights and biases were initialized randomly. The number of hidden layer weights controls the non-linearity of the NN model and is a determining factor for fitness of the regression. We chose the number of hidden layer as 10 based on a hit and trial approach. The number of input nodes were chosen as 5 which implies the load consumption of past five business days, were fed as input to the NN.

III. Evaluations

The smart meter technology provides utilities a large amount of data at measurement rates of typically half hour or less. Thus, the smart meter infrastructure promises efficient execution of DR schemes for residential customers by providing the grid with near real-time energy consumption. This also provides the customers with an opportunity to either manage the load or enroll into DR incentive programs [14], [15]. This motivated us to consider the smart meter data to evaluate the estimation of CBL.

A. The CER dataset

The Irish Commission for Energy Regulation (CER) established a Smart Metering Project Phase 1 in late 2007 with the objective of setting up and running smart metering trials [16]. As part of this phase, a number of smart metering technology trials were conducted and the associated data is published to support research and analysis. For the simulations in this paper, we considered smart meter data collected between July 2009 to December 2010, for 6435 residential consumers. We considered the benchmark data from customers' consumption in the traditional fixed rate tariff environment.

B. Performance metrics

To evaluate the error performance of the CBL methods, we use two metrics, *accuracy* and *bias* [6].

The Mean Absolute Error (MAE) is used to quantify the accuracy of methods. The MAE of CBL accuracy is defined as:

$$\alpha = \frac{\sum_{n \in C} \sum_{d \in D} \sum_{t \in T} |b_n(d,t) - l_n(d,t)|}{|C||D||T|} \qquad \text{III-B}$$

where, C is the set of all customers, D is the set of all days in the dataset, and T is the set of hourly time-slots in a day. *Lower* value of the MAE indicates *higher* the accuracy.

The estimation bias plays a significant role in assessing the performance of an estimator. The bias associated with each of the CBL estimation methods is expressed in equation (III-B).

$$\beta = \frac{\sum_{i \in C} \sum_{d \in D} \sum_{t \in T} (b_n(d,t) - l_n(d,t))}{|C||D||T|} \qquad \text{III-B}$$

The CBL methods with positive bias overestimate the customers' actual consumption and vice-versa.

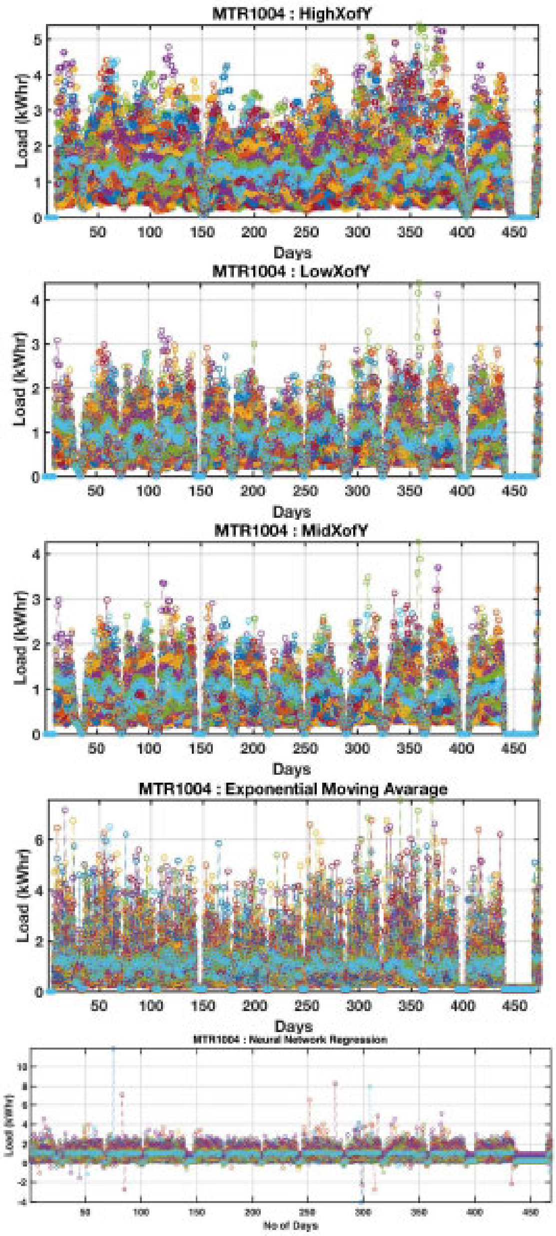

Fig. 2. CBL estimates for one randomly chosen customer. Different colors represent different half-hourly time Windows.

IV. Results

CBLs calculated for a one customer chosen randomly with different methods described in Section II are shown in figure 2. Although the figures can't be interpreted to extract relevant information regarding performance, they demonstrate the difference in CBL estimate patterns. The NN regression showed more outliers in the estimate, compared to other methods. The individual error metrics computed for the same customer is shown in figure 3. The MAE is the lowest for *LowXofY* or method and correspondingly has highest accuracy, however has a negative bias. This implies that the CBL may underestimate the actual consumption. The MAE as a function of bias for the same customer is shown in figure 4. The X-axis in this figure represents accuracy or *MAE kWh/h* and

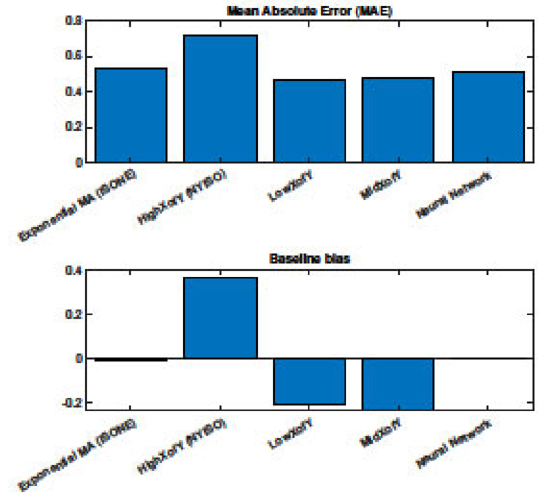

Fig. 3. Errors and biases for one customer.

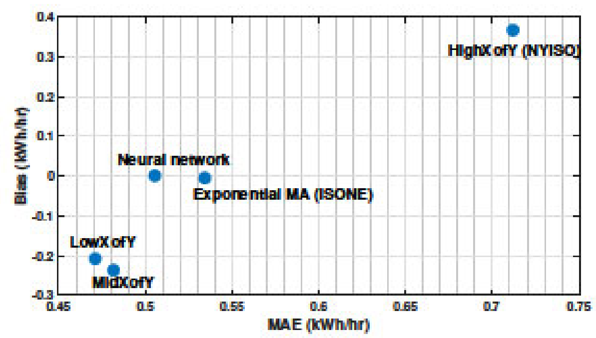

Fig. 4. Error performance for one customer.

the Y-axis represents the estimation *bias* in *kWh/h*. Methods above and below the zero bias line or the X-axis show positive and negative bias, respectively. This indicates that the Neural Network has minimum bias while having comparable MAE with other methods. The bias of exponential moving average method is also relatively small compared to other methods. However, the NN has the lowest MAE and the least bias, implying it to be the the most preferred approach. The NN is a data driven method and the convergence of the weights plays a significant role in the accuracy of prediction. As the weights are initialized randomly, the converged weights could have some variation. The NN results reported in this manuscript is the average of 100 independent runs. However, this result could not be generalized for all the customers. The performance of the methods largely relied on the data consuption pattern of the customer. To quantify which method results in minimum accuracy for how many customers, we derived a statistics. This is summarized in figure 5. To our surprise, We found that, the *MidXofY* method results in highest accuracy for most of the consumers, followed by exponential moving average and NN methods. This is confusing, because there is no one method can be suggested that works best for all the customers. To resolve this, a cumulative evaluation of CBL performance of all the customers was carried out.

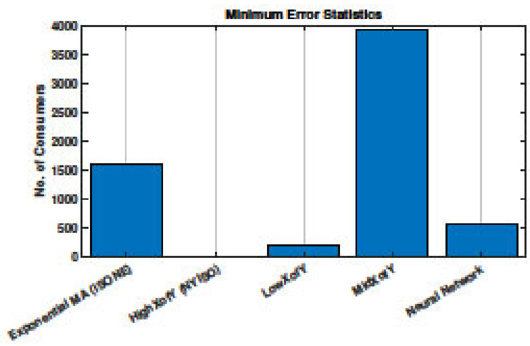

Fig. 5. Minimum error statistics for all the customers

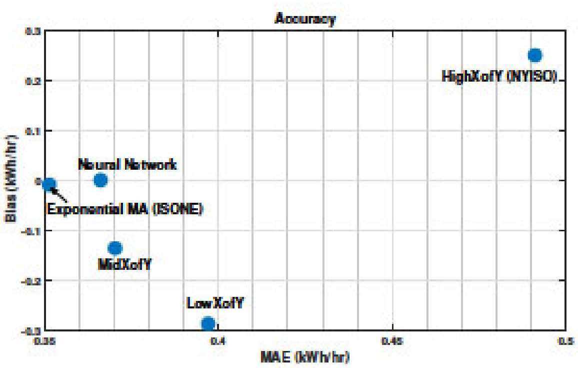

Fig. 6. Cumulative error performance for all the customers.

This mean that, although one method might not be the most suitable one for one customer, cumulatively one method can be recommended for a set of customers for a utility to maximize accuracy and minimize bias. The CBL for all the customers cumulatively taken together, was calculated using each estimation method and the error metrics for all the methods were computed. The cumulative errors for different methods are shown in figure 6. It was observed that the exponential moving average has the minimum MAE and the highest accuracy among all the CBL estimation methods. Moreover, the NN shows comparable performance for residential customers with lesser bias than exponential moving average, but with slightly lower accuracy.

V. Conclusion

We presented a comparative performance assessment of methods to calculate CBLs for residential customers. Our finding indicated that, not one method can be attributed as the most efficient method for CBL estimation and the efficiency of methods are largely dependent on individual customer consumption patterns. However, when the consumption patterns of a group of customers are aggregated to estimate the CBL, we found that both the NN and exponential moving average methods show superior performance. If the bias is taken into consideration the NN regression emerges as the most efficient method. It should be noted that, this conclusion is derived from a subset of the the CER smart meter data, which might not lead to generalization. Similar evaluations can be performed over historical data-set of an ISO, to decide the CBL estimation method that would be preferable. In addition, the financial impact of the accuracy and bias of each CBL estimation method depends on the type of PTR program. In summary, along with accuracy and bias, PTR program and operational issues will drive the decisions for the utilities to opt for a particular CBL calculation method.

The future work will focus on integrating the weather variables and conduct the experiments by considering the PTR programs.

Acknowledgment

This work was supported by the Electric Utility Management Program (EUMP) and the National Science Foundation (NSF) under grant HRD-1345232.

References

[1] L. Hatton, P. Charpentier, and E. Matzner-Løber, "Statistical estimation of the residential baseline," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1752–1759, 2016.

[2] FERC, "Demand response compensation in organized wholesale energy markets," *Docket RM10-17-000*, vol. Order 745, Mar. 2011.

[3] W. Hogan, "Demand response pricing in organized wholesale markets," *ISO/RTO Council Comments on Demand Response Compensation in Organized Wholesale Energy Markets*, vol. 13, 2010.

[4] P. Siano, "Demand response and smart grids—a survey," *Renewable and sustainable energy reviews*, vol. 30, pp. 461–478, 2014.

[5] K. J. Williams, "Federal energy regulatory commission v. electric power supply association," *Public Land and Resources Law Review*, no. 6, p. 30, 2013.

[6] S. Mohajeryami, M. Doostan, A. Asadinejad, and P. Schwarz, "Error analysis of customer baseline load (cbl) calculation methods for residential customers," *IEEE Transactions on Industry Applications*, vol. 53, no. 1, pp. 5–14, 2017.

[7] S. Mohajeryami, I. N. Moghaddam, M. Doostan, B. Vatani, and P. Schwarz, "A novel economic model for price-based demand response," *Electric Power Systems Research*, vol. 135, pp. 1–9, 2016.

[8] S.-C. Chan, K. M. Tsui, H. Wu, Y. Hou, Y.-C. Wu, and F. F. Wu, "Load/price forecasting and managing demand response for smart grids: Methodologies and challenges," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 68–85, 2012.

[9] Working Group 2 Measurement and Evaluation Committee, *Demand Response program evaluation – Program Year 2004 Sub-metering summary report*. Quantum Consulting Inc., 2004.

[10] J. Jazaeri, T. Alpcan, R. Gordon, M. Brandao, T. Hoban, and C. Seeling, "Baseline methodologies for small scale residential demand response," in *2016 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia)*, Nov 2016, pp. 747–752.

[11] T. K. Wijaya, M. Vasirani, and K. Aberer, "When bias matters: An economic assessment of demand response baselines for residential customers," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1755–1763, July 2014.

[12] T. Chen, J. M. Lehr, O. Lavrova, and M. Martínez-Ramón, "Distribution feeder-level day-ahead peak load forecasting methods and comparative study," *IET Generation, Transmission Distribution*, vol. 12, no. 13, pp. 3270–3278, 2018.

[13] C. M. Bishop, *Pattern recognition and machine learning*. Springer Science+ Business Media, 2006.

[14] B. Kim and O. Lavrova, "Two hierarchy (home and local) smart grid optimization by using demand response scheduling," in *2013 IEEE PES Conference on Innovative Smart Grid Technologies (ISGT Latin America)*, April 2013, pp. 1–8.

[15] Byungchul Kim and O. Lavrova, "Optimal power flow and energy-sharing among multi-agent smart buildings in the smart grid," in *2013 IEEE Energytech*, May 2013, pp. 1–5.

[16] Commission for Energy Regulation (CER), "CER smart metering project - electricity customer behaviour trial, 2009-2010 [dataset]. 1st edition. irish social science data archive. SN: 0012-00." www.ucd.ie/issda/CER-electricity, (2012).