European Commission

JRC CONFERENCE AND WORKSHOP REPORT

Proceeding of the 11th International Conference on Energy Efficiency in Domestic Appliances and Lighting (EEDAL'21)

Bertoldi, Paolo (Eds.)

2023

EEDAL
Energy Efficiency in Domestic Appliances and Lighting

Joint Research Centre

Contact information
Name: Bertoldi Paolo
Address: European Commission, Joint Research Centre, Via Enrico Fermi 2749 - 21027 Ispra (VA), Italy
Email: paolo.bertoldi@ec.europa.eu
Tel.: +39 0332789299

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC132721

PDF ISBN 978-92-76-99908-9 doi:10.2760/356891 KJ-07-23-093-EN-N

Luxembourg: Publications Office of the European Union, 2023

How to cite this report: *Proceeding of the 11th International Conference on Energy Efficiency in Domestic Appliances and Lighting (EEDAL '21)*, Bertoldi, P. (ed.), Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/356891, JRC132721.

Verification of Demand Response: the customer baseline load in small/medium customers

Antonio Gabaldon[1], Ana Garcia-Garre[1], Ramon Ruiz-Molina[2], Carlos Alvarez[3], MariaCarmen Ruiz-Abellón[4], Luis Alfredo Fernandez-Jimenez[5] and Antonio Guillamon[4]

1 Electrical Engineering Area, U. Politécnica de Cartagena, Cartagena, Spain
2 MIWenergia, Murcia, Spain
3 Institute for Energy Engineering, U. Politècnica de Valencia, Valencia, Spain
4 Dpt. of Applied Mathematics and Statistics, U. Politécnica de Cartagena, Cartagena, Spain
5 Dpt. Of Electrical Engineering, Universidad de La Rioja, Logroño, Spain

Abstract

The development of Demand Response (DR) is a basic step to achieve an increase of the flexibility in Power Systems, in the short and medium, term to balance the volatility of the new generation mix foreseen in the horizon 2020. At the same time, it is necessary to deploy tools to evaluate the performance of DR policies to obtain precise economic feedback for all the actors. This should increase the engagement of new resources from the demand-side. The verification of DR involves a right estimation of the customers' steady-state load without control: the customer baseline load (CBL). The aim of this paper is to compare the accuracy of the traditional and simple methods based on historical data to calculate CBLs with a specific Neural Network based method and, with both methods test the significance of adjustment coefficients in the increase of the accuracy or results. To develop this proposal, a demand database from a SME customer in the south east of Spain is analysed. Results show that it is possible to improve the performance of CBLs without increasing their complexity, which enables the removal of some technical barriers of more complex baseline approaches.

1. Introduction

A main concern for energy actors and authorities is the development of the portfolio of Demand Response (DR) on an equal footing with respect to conventional Supply-Side resources and integrate in this portfolio new resources (Distributed Energy Resources, DER) such as Energy Storage Systems (ESS) and Renewable Energy Systems (RES). For instance, the article 17 of the EU Directive 2019/944 establishes that "Member States shall allow final customers, including those offering DR through aggregation, to participate alongside producers in a non-discriminatory manner in all electricity markets" [1]. This issue includes the payment for the resource's performance, but the flexibility must be measured and verified in an easy and right way.

The achievement of this objective requires accurate and understandable economic flows: customers should receive credit according to the flexibility they provide, which needs an accurate evaluation of the changes in demand that occurs after DR performs. A forecast of demand (demand and generation, in the case of "prosumers") considering loads and other DER resources is needed. The physical behavior of loads and customers can change due to several parameters: weather, type of day, end-use shares, gaming possibilities and, specifically, the frequency of activation of DR events. Aggregators, Load Serving Entities (LSE) and System Operators (DSOs or TSOs) should estimate the "steady-state" of loads of their customers without DR (that is, the so-called Customer Baseline Load, CBL) with respect to available Smart Meter (SM) measurements after DR.

Several attributes are theoretically required for a good CBL: accuracy, simplicity, replicability and integrity. The accuracy of both the baseline estimation and the achieved flexibility is important to avoid erroneous incentives or penalties for DR and, in this way, promoting and encouraging customer participation. The accuracy is a main concern because the interest in DR may be reduced on the premise that management and enabling technology costs (e.g., smart thermostats and control) are usually high, and this casts a doubt on the recovery of the investment made in those enabling technologies. These issues have been a determining factor (a barrier, as stated in [2]) for the reduced participation of customers in implicit DR (price-response). For instance, in New York ISO there was some historical participation since 2006 [3] and a considerable interest in Demand Side Flexibility (DSF) to price, but during 2017 there was not any active price participation in the NY. Other power systems have a similar experience and the estimated revenue for economic DR (implicit DR) are quite small in comparison with capacity payments for explicit DR. For example, these figures have been reported by PJM in the USA [4].

Moreover, a CBL methodology must be robust to avoid some manipulation attempts of specific customers or entities, more interested in gaming than in DR. Forced changes in patterns to alter revenues should be detected by CBL to ensure DR integrity and a fair and correct revenue. Besides, CBL methodology should be simple to be understanded by customers, and this methodology should consider the idiosyncrasy of customers and markets where DR is deployed. If CBL is too complex it becomes a barrier, it could lead to a lack of interest by aggregators and customers. CBL should be also replicable, in the sense that we need some degree of standardization in CBL, to avoid the development of specific methodologies for each customer, country, product and market. Therefore, it is significant that regulators could provide different methods to compute CBL and agree with the other parties (customer and aggregator) a specific method. This seems a possibility to increase the customer engagement in DR policies and mitigate gaming. This option is usually deployed by operators and utilities (e.g., the notification of DR mechanisms, in France [5]).

The rest of the paper is organized as follows. Section 2 deals with the literature review of CBLs. In Section 3, different CBLs are revisited, and their adjustments are introduced. Section 4 outlines the case study (a commercial customer) through two different methodologies and the results obtained for the case study when the proposed methods are applied. Section 5 presents the conclusions.

2. Literature Review

The increase in renewable generation, and the objective of decarbonisation by 2030-2050 in most of countries around the world, will increase the interest of DR policies in wholesale, retail and future local markets. The measurement and verification of DR is a main concern for the effective engagement of new DR/DSF resources and, consequently, the interest in the development of CBL calculation methodologies has gained momentum since the last decade. At the same time, the figure of aggregators has increased its importance in the European Union. Some examples are France and Spain, where the aggregation is (or will be) more complex due to higher imposed thresholds to responsive demand capacity (around 1 MW) with respect to power systems in the USA (100 kW). These limits can be a problem for customers and aggregators and require the verification of the response (flexibility) for the qualification of responsive resources. This is another complex requirement for small and medium customer segments.

Most of methods described in the literature [6], [7], involve two steps in the definition of CBLs. First, the definition of the base profile and, second, an adjustment method to refine the initial estimation of demand, especially when the load is sensitive to external inputs, for instance: temperature, humidity or solar radiation in weather sensitive loads (air conditioning, heating, water heating, food storage…).

The success of DR policies since 2005 means that payments have increased [8], for example from $50 million to $500 million in 2021 (PJM [4]), so more precision is required in the evaluation of the economic side of DR. The interest and importance of this issue in the USA, Australia or Europe arise from a review of different reports and projects dealing with baseline as a topic in their activities. Research laboratories [9], [10], research consortia [11], [12], power system operators [6], [7], aggregators [13], utilities [14], [15], and energy and environmental agencies [16], have defined and analysed different types of baselines, their metrics, and have proposed some methods to improve their accuracy. This proliferation of methodologies, sometimes for specific customers and systems, has made more complex the management of DR [17] and the participation of small and medium customers (e.g., SMEs) which must be aggregated. The standardization and simplification of CBL methodologies arise as a necessity to remove DR barriers [2]. These attempts start in 2009 by the North American Energy Standard Board which proposed a series of definitions for CBLs to improve the harmonization of CBL methodologies. Later, US authorities assumed these definitions [18], [19]. For instance, the US ISO/RTO council periodically summarizes a table [20] that lists the description, measurement and verification parameters for DR programs across different ISOs.

Literature also depicts methods for different customer segments and regions, and some of them establish comparisons between these baselines. Lawrence Berkeley reported in [9] some methods for non-residential buildings. This research confirmed that base profile benefits its performance with adjustments before the DR period, but they recommend the use of different models for different groups of loads (due to the different weather sensitivity of loads in each segment). In [15] conEdison reported that simple baselines (e.g., High3of5 or Mid5of10) usually perform well for different segments, but when more sophisticated methods are used in a customer segment (e.g., regression analysis for a customer segment), it appears inherently inaccurate for other individual customers and days. This is especially significant for small customers, mainly for residential consumers. Similar results are reported in [14] by San Diego Gas & Electric. Authors conclude that any method is close to being accurate for individual customers on individual event days. In conclusion, more complex baselines only provide marginal improvements in accuracy but at a higher computational cost.

In [21], authors present different methods to evaluate the base profile of demand, from white-box to black-box models, but without any estimation of relative accuracy. Nevertheless, authors state that the accuracy of black-box methods depends on the training procedure, which must be repeated anytime a physical change of the system occurs, which makes more complex the aggregation. In [22] authors propose the so called “control group” approach, i.e. the clustering of customers onto different groups according to consumption patterns to reduce the randomness of individual demands and improve the performance of CBL. This method also presents some drawbacks [23] because it may be hard to uniquely define the best control group that properly captures the customer behavior of the DR group of participants. Basically, the necessary classification of demand into homogeneous, quasi-homogeneous or heterogeneous groups is a conclusion previously established in DR planning and management [24], [25]. This fact is important because it demonstrates that DR operation and verification should share common methodologies and procedures, and this can simplify DR, especially when aggregation of customers is needed.

It is interesting to note that short-term load forecasting (STLF) shares common methodologies with CBL methods because both provide demand forecasts in the short-term. STLF comprises multiple methodologies, for example, linear regression and Artificial Neural Networks (ANN) which can also be used to calculate CBLs [26], [27]. Some other machine learning methods such as Support Vector Regression (SVR) and Support Vector Machine (SVM) have been employed to forecast demand in [28]–[30]. Hybrid parameter optimization [28] and ant colony optimization [29] have been proposed to find the optimal parameters for SVR, whereas SVM with simulated annealing has been presented in [30]. The efficiency of ensemble methods based on regression trees, such as random forest or boosting, have been analysed in [31]. Nevertheless, classical methods like ARIMA models still perform well for demand forecasts. For this reason, hybrid models that combine two or more different methodologies (ARIMA, SVM or ANN) outlined good results. For instance, SVM and ARIMA are proposed in [32], [33], and the combination of ANN and SVM has been developed in [34]. A machine learning approach to disaggregate load and photovoltaic (PV) generation from net load data is analysed in [35] to obtain CBLs in prosumers. Authors conclude that reducing errors in the PV output power estimation can improve the CBLs performance. In [23], authors use Gaussian Process regression for machine learning because they state that the drawback of deterministic methods lies in their failure in capturing the dynamics of complex user behaviors, particularly important for small to medium consumers with more variability.

3. Methodology

3.1. Overall methodology

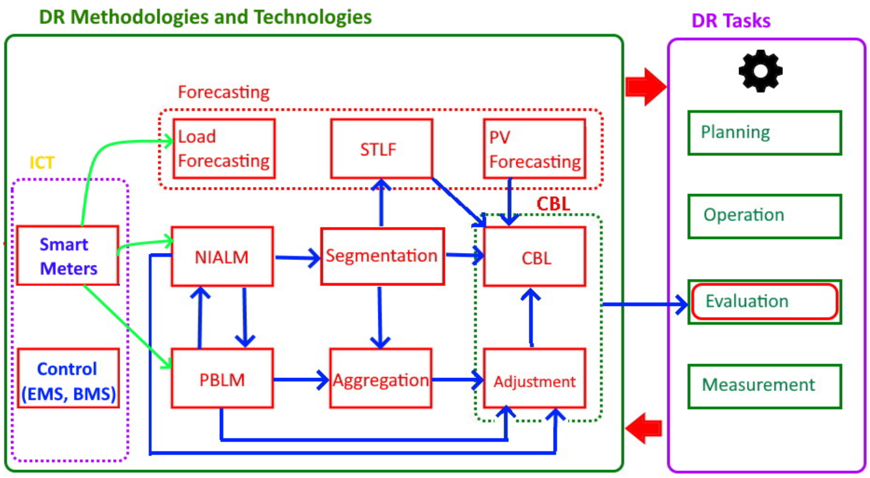
The increase of power systems flexibility in the scenario 2030-50 is based in the flexibility of Demand-Side resources. Four tasks are critical for these resources (specifically for DR): planning, operation, measurement and verification (figure 1). The proposal of this paper is that there are different methodologies that can be used for each task, and, at the same time, some methodologies can be used for more than one task. These synergies among DR tasks can make their management easier. Figure 1 shows a layout of this idea and the interconnections between DR methodologies, developed by the Spanish Research Network REDYD2050.

The evaluation and deployment of DR potential need the use of load or end-use models, such as Physical-Based Load Modeling (PBLM, considered as "grey" or "white" models). Another important issue to determine the potential of flexibility (demand reduction levels, loss of load service, rate of change of demand, energy recovery/snapback…) is the determination of the aggregated response [36] at several levels (from homogeneous to heterogeneous groups of loads, or between loads and other DER resources). Aggregators need some tools to perform DR simulations before the DR event for planning the operation of the responsive loads (usually loads with some kind of energy or product storage [37] such as HVAC, WH, ice or heat storage tanks). In this context, Non-Intrusive Load Monitoring, NIALM, plays an important role. NIALM allows to obtain the end-use demand to tune and validate the parameters of PBLM models, as well as the definition of average end-use patterns (i.e., the calculation of elemental load baselines [38] in customers where sub-metering is not available or not affordable from the point of view of costs), all of them from SM measurements (figure 1).

Regarding the participation in markets (wholesale or local energy markets [39]): the aggregator needs load forecasts to define the energy requirements in day-ahead markets and avoid penalties in balance markets. This can be done through specific forecasts [31], but CBLs can also be used. Weather or gains in efficiency can be evaluated from modelling if these models are physical based. Finally, NIALM should contribute to the verification of the performance of responsive loads before DR (to detect potential gaming) or during DR (load flexibility).

As it has been discussed in section 2, segmentation methods can be crucial for achieving a refinement and a gain in CBL performance, but it is also important to perform load aggregation and evaluate their DR potential [40]. STLF and CBL can also provide feedback for PBLM toolboxes (e.g., the change in customer behaviour due to market prices or due to frequency in event calls) which are common tools for day-to-day operation of aggregators. It is also worth mentioning the linkage in the opposite sense, which is further considered in the adjustment of CBL as it has been proposed in [40].

Figure 1. Interaction among PBLM, CBL, NIALM and STLF tools according to [25] for evaluation purposes in DR.



3.2. Characteristics and methodologies for baselines

As stated in section 2, a main problem with baselines is the lack of standardization. For this reason, the literature describes different CBL methodologies. Some of these, developed "ad hoc", can have excellent results for its specific segments, DR products, markets and situations, but they can fail in different scenarios [18], such as prosumers, small/medium customers or customers with a high share of weather-sensitive loads (for example, commercial customers of buildings). Literature states that "base-profile" baseline methods based on the use of historical data are a sufficient approach for obtaining a good and simple basis to further develop CBLs with high accuracy through adjustment.

US Power Systems have two decades of experience with DR in different markets and consequently with CBL methodologies and their problems and implementation barriers. To overcome these problems, NAESB defined in 2009 five types of methodologies [24]: Maximum Base-Load; Baseline Type-I; Baseline Type-II; Meter Before/Meter After and Metering Generation Output.

Baselines Type-I&II have been adopted as default methodologies by several ISOs, [41]. The Type-I is based on historical demand data, which may also include other variables such as weather and calendar. The Type-II assumes the same idea, but it uses statistical sampling to estimate the aggregated consumption. With the deployment of Smart Meters in last decade [6], [25], this methodology lacks a practical interest. The most common CBLs in the literature are briefly described.

1. <u>Maximum-Base Load (Firm Service Level):</u> is based on the ability of a resource (DER) to reduce its consumption to a specified level: the so-called in some systems as Firm Service Level. The customer should keep is demand below this level to avoid some penalties. Sometimes it is also known as the "non-baseline".

2. <u>*Y-day Simple Average Method:*</u> to predict the CBL, the method uses the average demand over the Y most recent non-DR days immediately before the DR event being considered. Usually from 3-day to 10-day-basis are used for this estimation [42].

3. <u>*Comparable Day Method:*</u> this also considers historical demand data to compute the CBL. In this case, the method only takes one day that is selected for its similar conditions with the event day (temperature, humidity, day of the week…). If sufficient relevant factors are not considered, the forecast trend to be erroneous.

4. <u>*High/Middle/Low XofY baseline:*</u> the baseline is obtained again by averaging recent historical data. It considers the demand of Y non-DR days preceding the DR event and it uses the average of the X days with the highest (or middle, or lowest) demand within those Y previous days. These baselines apply the so-called exclusion rules [19] (i.e., some days are not considered for the evaluation),

because operators assume that some variables can modify the pattern of demand. Up to 30 or 60 days can be used to define Y [13], but usually shorter periods such as 3, 5 or 10 days are used because long periods could include too much changes in the demand pattern (especially for sensitive loads). The use of High, Low or Middle depends on their use. HighXofY is the most common if DR events are due to peak load periods. Some practical examples of these baselines are High5of10 in California SO [7] or High15of20 in IESO, Canada [20]. These CBLs are calculated as follows:

$$CBL_{XofY}(d,h) = \frac{1}{X}\sum_{i=1}^{X} A(i,h) \tag{1}$$

where $CBL_{XofY}(d,h)$ is the baseline at time h of day d; A(i,h) is the actual load for the i-th highest (middle/lowest) energy day, at time h, among the previous Y non-event days, and X the number of the highest (middle/lowest) days to be averaged in Y after exclusions.

5. *Nearest XofY baseline:* it focuses on the total consumption outside the DR event window to determine which of the Y previous days are more like the event day. The X days (inside the Y interval) having the closest demand to the DR day are selected. Then, the baseline is computed as the average demand of these X days. Exclusion days are also applied.

6. *Weighted Average method:* this is based on a weighted average of the previous day's CBL (some of the above CBL_{XofY}). For instance, this method is used in Korea by KPX [43]. In this case, among the 10 reference days, the two highest and lowest days are excluded, and the remainder is used with different weights. For example, KPX define weights such as 0.10, 0.15, 0.15, 0.20, and 0.25 ranging from the older demand to the most recent demand. The main problem is the estimation of weights, depending on the country and customer segment.

7. *Exponential Moving Average:* a weighted average of customer historical demand is considered again in this option, but the weight decreases exponentially with time. The main difference with the previous CBL is that it considers a broader spectrum for "X" days.

8. *Control Group Methods:* the method considers the possibility that the aggregator (or other power agent) has a demand database with other non-responsive customers (in the same segment) from the DR event day. The customers are clustered in similar groups and the DR customer's load curve is matched to one of these groups. Then, the CBL is calculated by averaging the load curves in the selected cluster in per unit. More complex methods can use a weighted combination of load curves in the cluster, or the load curves of the same customer and other customers on non-DR event days [44].

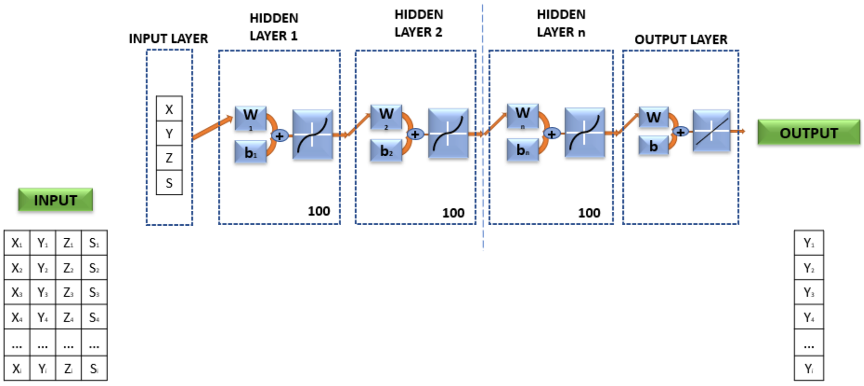
9. *Short-term load forecasting methods:* this cluster of methods comprises a wide range of alternatives. For example, the CBL can be built using a customer-specific regression model, that besides historical loads also weather conditions and calendar features (holidays, season, day of the week...) are considered. In [42], CBL is estimated for campus buildings and with a high penetration of weather sensitive loads (e.g., HVAC). For that, a linear model is fitted over the 5-minute period just before the DR event and the 5-minute period immediately following the set time. Another example, is the use of Neural Networks (NN) to estimate the CBL (a back propagation NN [27], [45]) where NN is adapted to establish baselines in public buildings (South Korea and China, respectively), considering meteorological indices. The close relationship between CBL estimation and STLF models allows that many other methods (e.g., random forest) could be considered as an approach to compute the CBL.

3.3. Neural Network method proposed

Neural networks can be classified as a series of algorithms that endeavour to recognize underlying relationships in a set of data. To compare traditional methods with more advanced techniques for obtaining CBLs, different tests with NN of different complexity have been developed. One-layer, two-layer and three-layer networks will be tested, with different number of neurons, a fixed number of samples for training, validation and testing, and three of the most common used training algorithms: Levenberg-Marquardt (LM) [46], Bayesian Regularization (BR) [47] and Scaled Conjugate Gradient (SCG) [48].The input and output layers just pass on/out the information to/from the hidden layers, whereas these apply the sigmoid activation function.

Data from a commercial premise in the service sector in the south-east of Spain will be used, 2017 and 2018 data are used as the training validation and verification set for the NN (60/20/20 ratio used), and 2019 data for testing the performance of the NN. In order to have a higher degree of correlation with traditional methods, we will use calendar data as input to the neural network. Four input variables repeated for each hour are established. Namely, the hour number from 0 to 23, the day of the month from 1 to 31, the month of the year from 1 to 12, and the type of day of the week from 1 to 7. To reduce the input noise, type 1 days (Sundays) are not considered. This is because, on this day, most of the commercial premises are closed. Therefore, DR measures cannot be applied. As output, we have the value of the hourly power demand in kW. Consequently, there are five types of variables counting the inputs and the output for each hour of the year.

Figure 2. Structure of a Neural Network with four input parameters and one output.



Before obtaining the final solution, the use of other variables as an input was investigated. Variables such as hourly temperature, or whether the type of day is a holiday or not, have been tested to develop the NN. However, the inclusion of these variables was returning similar results in the prediction, so finally they are not considered in the current research. It is considered that the 4 inputs variables that are explained before (hour number, day number, month number and type of day) are enough to develop the NN as including the weather conditions do not improve the performance of the NN.

3.4. The adjustment of baselines

The base profile defined by the baseline can be improved using adjustment methods that consider the possibility that demand could change in the short-term with respect the first estimation done by the methods described in paragraph 3.2. There are two main methods: multiplicative and additive adjustment that basically represent the same idea. The objective of these adjustments is to modify the preliminary CBL to adapt it to weather and demand conditions on the DR event day. The easiest way to evaluate these factors is the use of pre-event DR data, and then calibrate the baseline using the observed non-event hours prior to DR periods. In [9] the adjustment factor is defined by:

$$amf(d) = \frac{\sum_{k=1}^{a1} A(d, h_0 - (b1 + k))}{\sum_{k=1}^{a1} P(d, h_0 - (b1 + k))} \tag{2}$$

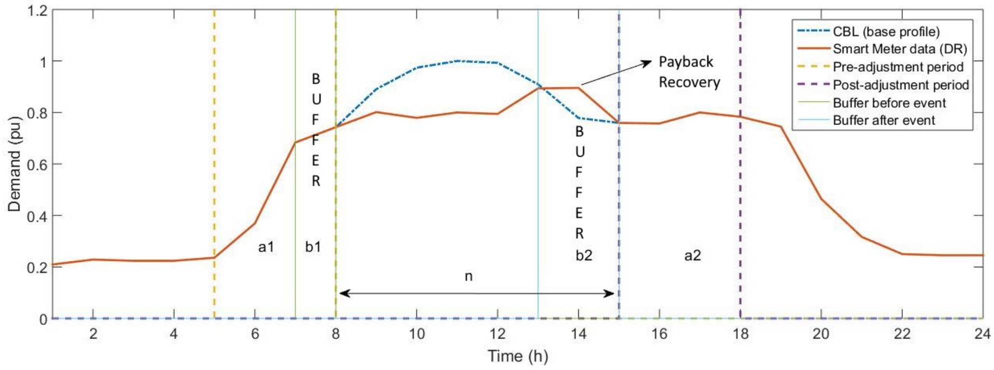
where *amf(d)* denotes the adjusted multiplicative factor for the day *d*; *A(d, h)* is again the actual load of day *d* at time *h*; *P(d,h)* is the predicted load (from unadjusted baseline or short-term load forecasting methods [31]) of day *d* at time *h*; *h0* is the start time of the DR event; *b1* is the buffer-time and *a1* is the length of the pre-adjustment band (Figure 3). Then, the new CBL is evaluated by:

$$CBL_{adj}(d,h) = amf(d) * CBL(d,h) \tag{3}$$

Some SOs uses pre and post DR adjustment factors combined in the same baseline [7]. The idea is that the post-event factor gives additional information about the boundary conditions throughout the DR-event day (e.g., weather changes that modify demand). CAISO Baseline Accuracy Work Group justifies this approach to avoid contamination of baseline both for pre-cooling and snapback periods to occur in the hours directly before and after the DR event [7].

Figure 3 depicts this idea. In this case, the DR event period ranges from 8:00 to 13:00 and the pre-adjustment period uses data from 5:00 to 7:00 (*a1=2h*) while the post-adjustment period uses data from 15:00 to 18:00 (*a2=3h*). Obviously, these periods do not overlap. The consideration of two periods (pre buffer *b1=1h* and post buffer *b2=2h*), limits the possibility of perturbations like gaming just before the baseline method takes demand data from SM to adjust its forecast. Pre-adjustment buffers are applied in several systems in the USA (e.g., NYISO [49] which uses a two hours buffer *b1*). It seems necessary that the definition and the duration of both "buffers" should be justified by load mix and behaviour (e.g., through load modelling, the approach proposed in [40]). In Figure 3, *n* represents the period in which the consumption is affected by the DR event, that is, the sum of de DR period and the post-buffer period.

Figure 3. Example of baseline adjustment and periods being used in equation (2)



3.5. Evaluation of CBLs performance

Regarding the error metrics, the Mean Percent Error (MPE) has been selected to describe the magnitude and direction of the estimation bias. MPE reflects the percentage by which the baseline, on average, over or underestimates the "true demand" in absence of a DR event. To evaluate the precision, both the Mean Percent Average Error (MAPE) and the normalized Root Mean Squared Error (nRMSE) have been selected. The lower MAPE and nRMSE are, the more precise the baseline is. Note that metrics are defined through relative errors, so they can be used to compare accuracy and precision of CBLs measured in different scales. Mathematically, these metrics are defined as follows, [7]:

$$MPE = \frac{100}{n}\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{\bar{y}} \quad \textbf{(4)}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{\bar{y}}\right| ; \; nRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\bar{y}} \quad (5)$$

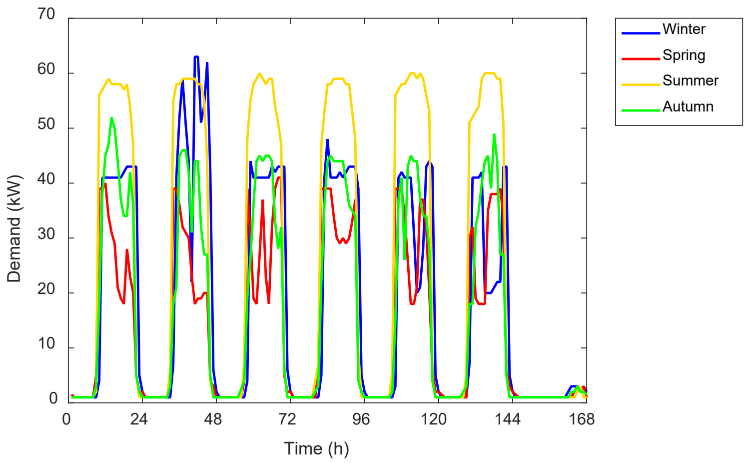
where y_i is the real demand at time i, $\hat{y}_i$ is the CBL (forecasted demand) at time *i*, and $\bar{y}$ is the mean of the real demand for the *n* values. Remark that *n* refers to the length of the DR evaluation period.

4. Results and Discussion

4.1. Case of study

To test and validate the methodologies for the calculation of CBLs described above, a representative commercial customer from the south-east of Spain has been selected. The reason for choosing this customer is that its high consumption and high share of weather-sensitive loads (mainly HVAC), compared to individual residential customers, makes them a target for their engagement in DR policies. At the same time, their management and aggregation are easy for SOs, as their consumption is "scheduled" and less variable than residential demand. Figure 4 presents the weekly demand for the different seasons of a year. As it can be seen, the hours of "start" and "end" consumption during a working day are always the same, but there is some variability during the day, mainly because of the weather. It is remarkable that in summer, the consumption is the highest as the HVAC is working all day because of the high ambient temperatures (average temp. max: 34ºC, min: 21ºC).

Figure 4. Weekly demand of a commercial customer on the four seasons of a year



4.2. Neural Network CBL developed

For the development of the NN and the performance of different tests, the program "MATLAB" is used. The input data is loaded with the format described for the years 2017 and 2018, performing the training, validation and verification of the NN with the output of the same years. Then, the NN is applied to the 2019 input data, evaluating the prediction output with real data for testing purposes.

The three methods discussed in Section 3.3 are used to train the NN. In these tests, a criterion is followed to consider that the training is finished: that a thousand epochs of calculation have been performed, that the mean squared error is too high (5.57e3) or 0, that the gradient between points is too high (1.47e4) or low (1e-7) or that the training time reaches twenty minutes. Some of the criteria are based on early stopping criteria to avoid excessive adjustment, stopping the execution in case of instability, or that the improvement with each epoch is not significant. The last condition is imposed in order to be suitable for use in one hour ahead forecasts for the electricity market.

The tables below (1-2) present the main results of the most significant tests. Although other networks with a higher number of layers and neurons have been tested, they do not yield significant results in an allowable computation time.

Table 1. Results of the NN training performance with 10 neurons/layer

		Train		Validation
Neural Network	**Algorithm**	**RMSE (kW)**	**Time (sec)**	**RMSE (kW)**
One-Layer NN	LM	10,05	2	10,21
	BR	10,00	10	10,08
	SCG	11,09	2	10,62
Two-Layers NN	LM	9,36	2	10,17
	BR	9,03	35	9,85
	SCG	**9,98**	**3**	**9,80**

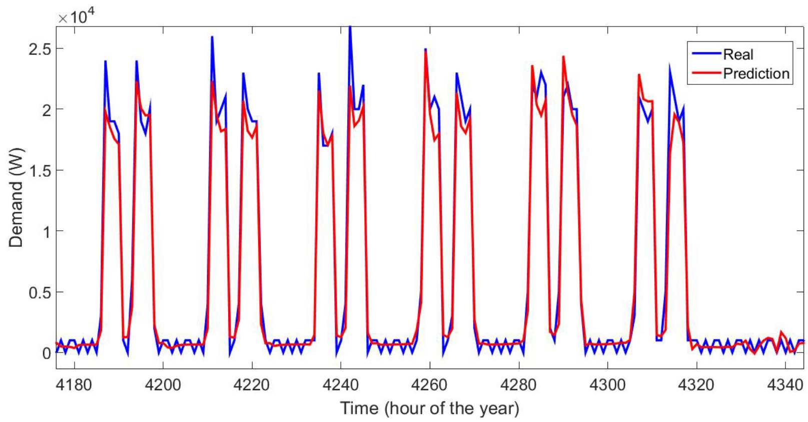
Three-Layers NN	LM	7,89	4	13,97
	BR	7,18	83	15,49
	SCG	9,88	3	13,39
Five-Layers NN	LM	8,58	4	13,35
	BR	5,61	145	16,19
	SCG	10,25	4	13,59

Table 2. Results of the NN training performance with 50 neurons/layer

			Train	**Validation**
Neural Network	**Algorithm**	**RMSE (kW)**	**Time (sec)**	**RMSE (kW)**
One-Layer NN	LM	8,95	4	10,22
	BR	8,66	71	10,54
	SCG	11,62	4	10,93
Two-Layers NN	LM	7,84	175	10,68
	BR	4,43	1200	14,16
	SCG	9,18	18	10,49
Three-Layers NN	LM	5,59	558	14,73
	BR	9,06	1200	13,17
	SCG	8,93	22	13,65
Five-Layers NN	LM	3,02	1200	17,04
	BR	15,72	1200	15,80
	SCG	8,8	45	14,03

After evaluating all the tests, Scaled Conjugate Gradient (SCG) is chosen as the most reliable method for training the network. Due to its reduced training time and its good performance compared to the other methods. Among the tested networks, the one formed by two layers with ten neurons each, is the one that returns the best results as shown in Table 1. Figure 5 shows the comparison of the CBL predicted with the NN selected and the real data.

Figure 5. Neural network prediction against real data



4.3. Comparison of unadjusted and neural network baselines

In this study, we have obtained CBLs for a commercial customer with six different methods, to compare the performance of each method. The methodologies analyzed are the High3of5, Low3of5, Mid4of6, Nearest5of10, Mid6of10-weigthed and the NN baseline selected in section 4.2.

Data from January 2017 to December 2019 was available for obtaining the CBLs. As the NN method needs a great database to train the net, data from 2017 and 2018 was used with this objective, and then, we have tested the methodology with the 2019 data. For comparison purposes, the other traditional methods were only evaluated in this period (year 2019). The CBLs have been calculated only for workdays and work hours, that is, the days and hours in which the commerce is opened (from Monday to Saturday and from 10 to 23h). That is because there is no possibility of applying DR policies when the commerce is closed, as the consumption is minimum. This can be appreciated in Figure 4 (hours 144 to 168, Sunday).

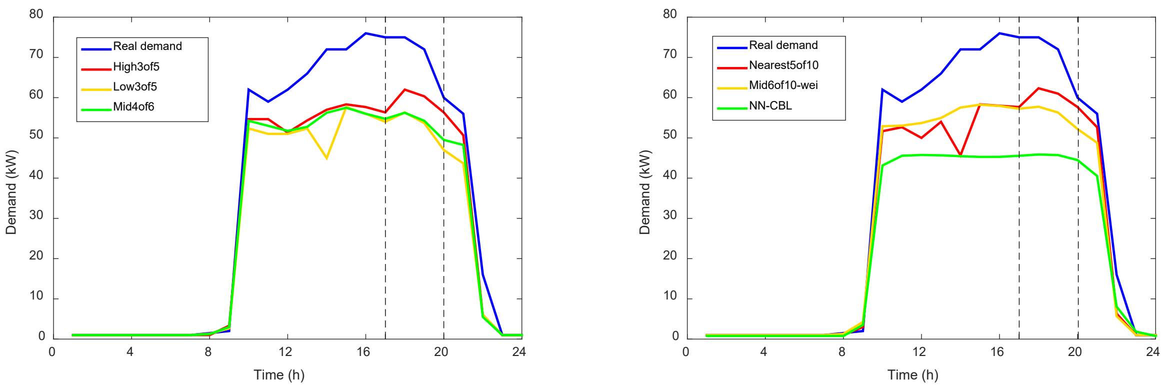
Table 3 presents the metrics for the different CBLs analysed. As it can be seen, the method with the lowest index of error is Nearest5of10, with a 12.38% of MAPE and a 17.78% of nRMSE. It is also remarkable that all the CBLs overestimate the demand, as in all methods, MPE has positive values.

Table 3. Error metrics of the unadjusted CBLs

CBL	MAPE (%)	MPE (%)	RMSE (kW)	nRMSE (%)
High3of5	15.55	8.79	6.04	21.00
Low3of5	12.78	0.61	5.96	19.86
Mid4of6	12.91	4.65	5.74	18.46
Nearest5of10	12.38	4.00	5.19	17.78
Mid6of10-weighted	14.49	5.35	5.81	19.93
NN-CBL	17.60	2.52	6.75	21.55

Figure 6 depicts the comparison of the different methodologies for one day. There is marked with dashed lines the period that is going to be used as DR event for the adjustment of the CBLs, i.e., from 17:00 to 20:00 (see next Section).

Figure 6. Comparison of unadjusted CBLs to real demand on a specific day



4.4 Adjusted baselines: WS, PBLM and backward coefficients

Sometimes, SOs use different adjustment coefficients to improve the performance of the CBLs. These coefficients are calculated with data from the consumption of the previous hours to the DR event. However, as the DR revenues are normally obtained after the event, there is also the possibility to use adjustment coefficients calculated with data from the hours after the event. In this paper, we have calculated two multiplicative pre-event (Weather Sensitive, WS and Physically Based, PBLM) and one post-event (backward, BW) adjustment coefficients. As it was explained in section 3.4, it can be necessary to consider two buffer periods for DR events to reduce the perturbations in the calculations of CBLs that could be caused by gaming, preheating or precooling strategies or also the normal increase of consumption after the event to compensate the demand reduction during the application of DR policies. This rise in consumption after a DR event is called "energy recovery period". In table 4 there are shown the definition of the adjustment coefficient and buffer periods applied.

Table 4. Definition of the adjustment coefficients for the CBLs analysed

Adjustment coefficient	Data	Period	Buffer
Weather sensitive (WS)	Pre-DR	First 2 hours of the 4-hour period pre-DR event	2 hours before event
Physically based (PBLM)	Pre-DR	Last 2 hours of the 4-hour period pre-DR event	No buffer
Backward (BW)	Post-DR	Last 2 hours of the 3-hour period post-DR event	1 hour post event

A DR event is defined during the working hours, in this case, an event that starts at 17:00 and ends at 20:00. According to the definitions of the table 4, the adjustment coefficients are calculated as follows:

- The WS coefficient, commonly used by NYISO, is calculated with the data from the two first hours of the four-hour period before the start of the DR event, that is, from 13:00 to 15:00. In Table 5, there are shown the error metrics for the WS adjusted CBLs. As it can be appreciated, in all case, the error is reduced, e.g., MAPE is reduced from 17-12% of the unadjusted baselines to around 7% for the WS adjustment. Figure 7 shows the WS-adjusted CBLs for a specific day of 2019 compared to the real demand.

Table 5. Error metrics of adjusted WS-CBLs in DR period

CBL	MAPE (%)	MPE (%)	RMSE(kW)	nRMSE (%)
WS-High3of5	7.03	2.19	4.17	8.53
WS-Low3of5	7.36	0.85	4.44	8.94
WS-Mid4of6	6.99	1.94	4.18	8.48
WS-Nearest5of10	6.99	2.15	4.21	8.48
WS-Mid6of10-weighted	6.78	1.78	4.02	8.17
WS-NN-CBL	7.49	-0.10	4.51	9.00

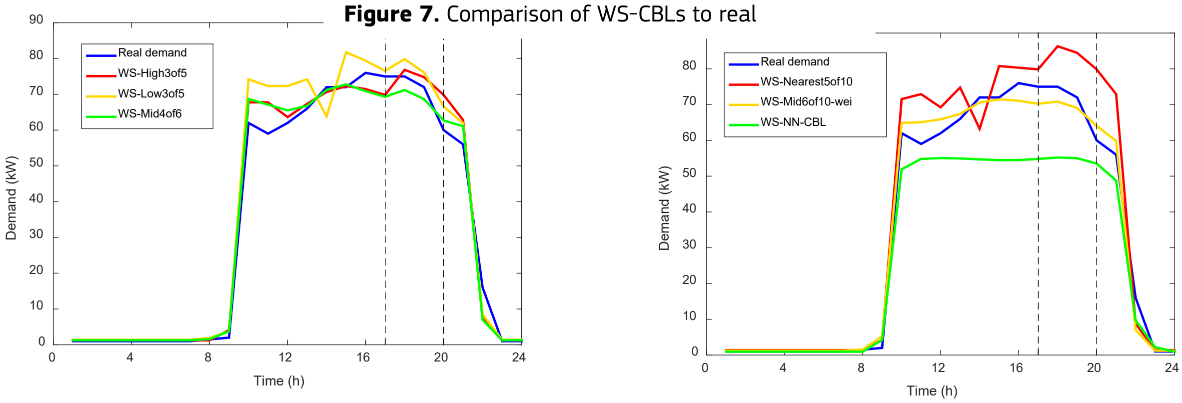


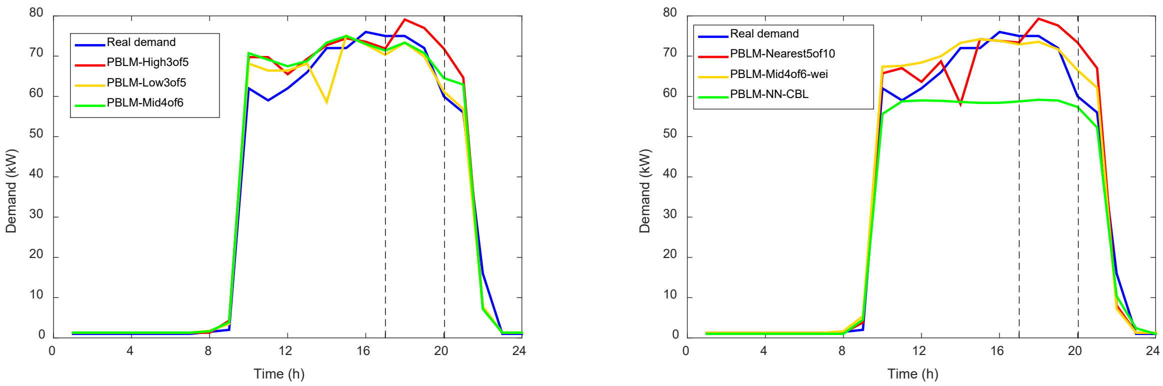
Figure 7. Comparison of WS-CBLs to real

- The PBLM coefficient, defined by Physically Based Load Models [40], is calculated with data from the two hours before the start of the DR event, without any buffer, so the period used for the calculation is from 15:00 to 17:00. Table 6 present the performance of the PBLM adjusted CBLs. As in the WS adjustment, PBLM adjusted baselines reduce the error from their unadjusted versions. In addition, the PBLM adjustment coefficient slightly improves the performance compared to the WS coefficient (MAPE is reduced around 1.5-2%). Figure 8 present the different PBLM-adjusted CBLs compared to the real demand for a specific day.

Table 6. Error metrics of adjusted PBLM-CBLs in DR period

CBL	MAPE (%)	MPE (%)	RMSE(kW)	nRMSE (%)
PBLM-High3of5	5.09	2.20	3.16	6.32
PBLM-Low3of5	5.06	0.46	3.17	6.32
PBLM-Mid4of6	4.75	1.86	2.98	5.98
PBLM-Nearest5of10	4.72	1.53	2.94	5.90
PBLM-Mid6of10-weighted	4.78	1.78	4.03	5.93
PBLM-NN-CBL	6.08	1.25	3.68	7.50

Figure 8. Comparison of PBLM-CBLs to real demand on a specific day



- Finally, the BW coefficient, defined also by PBLM [40], uses data from the last two hours of the three-hour period after the end of the DR event, that is, use a one-hour buffer period post-event. Consequently, the BW adjustment coefficient is calculated from 21:00 to 23:00. In this case (Table 7), the BW adjusted CLBs reduce the error compared to the unadjusted ones, but their performance is worse than the other adjustment coefficient, except from the case of the NN-CBL, in which the BW-CBL is the most accurate.

Table 7. Error metrics of adjusted BW-CBLs in DR period

CBL	MAPE (%)	MPE (%)	RMSE(kW)	nRMSE (%)
BW-High3of5	11.23	2.27	3.11	10.11
BW-Low3of5	9.72	1.71	2.97	9.60
BW-Mid4of6	9.95	1.63	2.89	9.34
BW-Nearest5of10	10.06	1.74	2.90	9.30

BW-Mid6of10-weighted	9.77	2.07	2.84	9.18
BW-NN-CBL	4.24	-0.90	2.50	5.85

The reason for the rise of the error in this case can be the hours in which the BW coefficient is calculated. From 21:00 to 23:00 the commerce is closing, and the consumption is reduced, so the calculation of the adjustment could be more imprecise than the other ones. For this reason, it is necessary to analyze the convenience of each adjustment coefficient to the type of customer we are working on, as each coefficient can be more adequate for one customer but not for others, depending on its consumption routines and behaviours. Figure 9 shows the six different BW adjusted CBLs studied, compared to the real demand on a specific day.

Figure 9. Comparison of BW-CBLs to real demand on a specific day

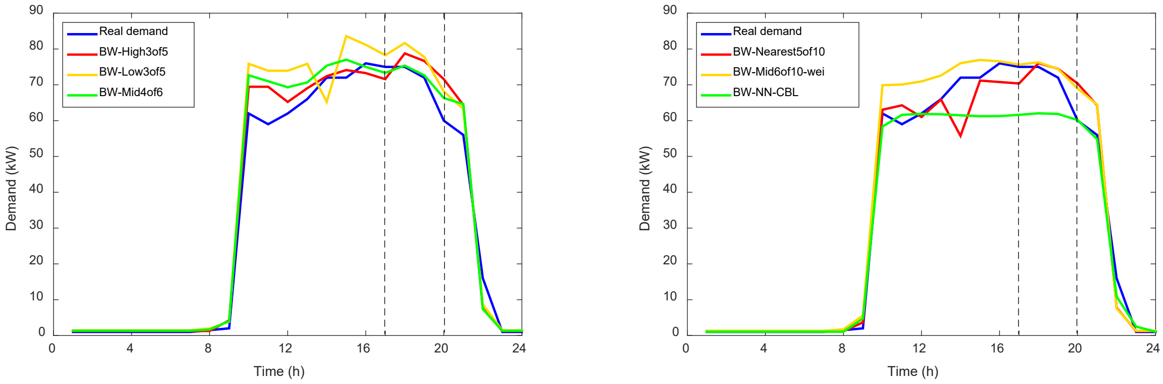
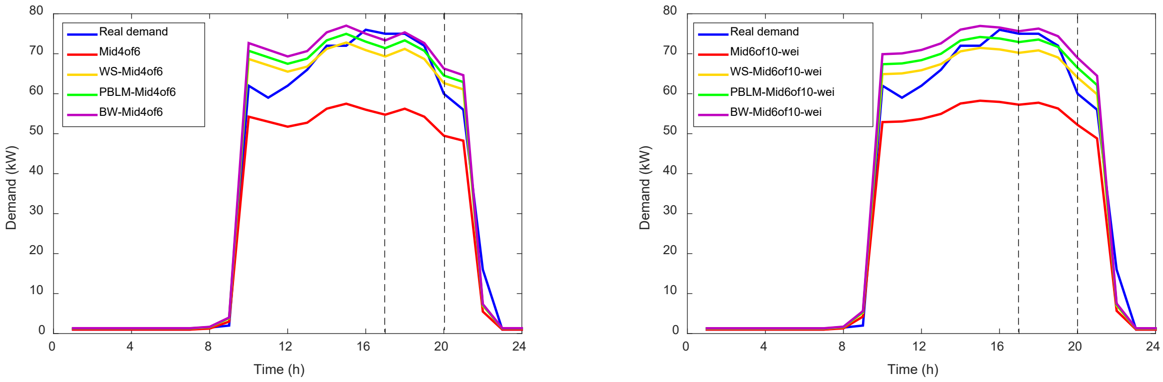


Figure 10 shows the Mid4of6 (8a) and the Mid6of10-weighted (8b) CBLs with all the adjustment coefficients presented in the paper, compared to the real demand (i.e., SM data), on a specific day. We can conclude that adjustment coefficients improve the performance of the unadjusted CBL, , as it can be appreciated in the Figure 10.

Figure 10. Unadjusted and adjusted CBLs compared to real demand on a specific day (a) Mid4of6 (b) Mid6of10-weighted



(a) (b)

5. Conclusions

The verification of demand flexibility becomes a key issue for the development of DR. Baselines are the basis to provide this verification, which defines the subsequent revenue and payment. An accurate CBL also emerges as catalyser for engaging new customers in markets recognizing and given credit for their real flexibility.

Moreover, demand resources from small and medium segment (an SME, in the case presented in this paper) need an accurate and fair verification of their response. In this proposal, the accuracy of demand forecasts with DR (using PBLM modelling to tune adjustment factors) is improved with respect to a simple base profile. Some other approaches involve specific and complex methodologies that have proven to define accurate CBLs. The main drawback of more complex proposals is that these options usually increase the complexity of DR and they sometimes require different models for different customers and segments. Moreover, some paper in the literature report that these kind models can present problems if DR performs periodically when the customer changes its behaviour, or simply if the aggregator develops more complex products (e.g., the participation of demand in several markets and services). Literature shows that base profile (unadjusted) CBLs are not the best option, but they can improve their performance through adjustment factors. Until now, these factors have been proposed based on experience. This paper highlights the convenience of using adjustment factors explained by PBLM in other segments, such as commercial SMEs, and that a double-adjusted CBL also displays an improvement in performance. Thus, the proposed methodology arises as an adequate, accurate, simple and understandable estimator, i.e., the main characteristics of a good CBL.

This paper also presents the synergies associated with the use of other tools used by aggregators, such as NIALM, customer segmentation and enabling technologies to verify load flexibility. In this case, the adjustment period can be justified and improved both before and after the period of DR events. In this way, different DR actors can obtain necessary feedback to perform a better evaluation of the DR potential, necessary for the customer-centred markets to be envisaged in the 2050 horizon.

Acknowledgements

This work has been supported by the Agencia Estatal de Investigación, Ministerio de Ciencia e Innovación (Project RED2018-102618-T funded by MCIN/AEI/10.13039/501100011033); and the Ministerio de Educación (Spanish Government) under grant FPU17/02753.

References

[1] European Commission, “Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU.” [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L0944. [Accessed: 13-Jan-2020].

[2] N. Good, K. A. Ellis, and P. Mancarella, “Review and classification of barriers and enablers of demand response in the smart grid,” *Renew. Sustain. Energy Rev.*, vol. 72, May 2017.

[3] NYISO, “Distributed Energy Resources Market Design Concept Proposal,” 2017. [Online]. Available: https://www.nyiso.com/documents/20142/1391862/Distributed-Energy-Resources-2017-Market-Design-Concept-Proposal.pdf/122a815f-b767-e67f-0a8f-323e5489c2b1. [Accessed: 28-Dec-2021].

[4] J. Mcanany, “2021 Demand Response Operations Markets Activity Report: December 2021,” 2021.

[5] RTE (France), “Règles pour la valorisation des effacements de consommation sur les marchés de l’énergie NEBEF 3.1,” 2018.

[6] C. Lake, “PJM Empirical Analysis of Demand Response Baseline Methods.,” 2011. [Online]. Available: https://www.pjm.com/-/media/markets-ops/demand-response/pjm-analysis-of-dr-baseline-methods-full-report.ashx?la=en. [Accessed: 15-Jan-2020].

[7] California ISO, “Baseline Accuracy Work Group Proposal,” 2017. [Online]. Available: http://www.caiso.com/Documents/2017BaselineAccuracyWorkGroupFinalProposalNexant.pdf#search=customer baseline. [Accessed: 13-Jan-2020].

[8] PJM, “Demand Response in PJM Markets and Operations.” [Online]. Available:

https://www.pjm.com/markets-and-operations/demand-response.aspx. [Accessed: 13-Jan-2020].

[9] K. Coughlin, M. A. Piette, C. Goldman, and S. Kiliccote, “Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California Environmental Energy Technologies Division,” Berkeley, 2008.

[10] P. Bertoldi, P. Zancanella, and B. Boza-Kiss, “Demand Response status in EU Member States,” 2016.

[11] N. Dawood, “Short-Term Prediction of Energy Consumption in Demand Response for Blocks of Buildings: DR-BoB Approach,” *Build. 2019, Vol. 9, Page 221*, vol. 9, no. 10, Oct. 2019.

[12] “DRIMPAC H2020 project – Unified DR interoperability framework enabling market participation of active energy consumers.” [Online]. Available: https://www.drimpac-h2020.eu/. [Accessed: 28-Dec-2021].

[13] EnerNOC, “The Demand Response Baseline. White Paper,” 2009.

[14] L. Willoughby, J. Bode, M. St, and S. Francisco, “2012 San Diego Gas & Electric Peak Time Rebate Baseline Evaluation Prepared for : San Diego Gas & Electric Prepared by : The FSC Group Table of Contents,” 2013.

[15] conEdison, “Advanced Metering Infrastructure Business Plan,” 2015. [Online]. Available: http://documents.dps.ny.gov/public/Common/ViewDoc.aspx?DocRefId=%7B0864BF10-42C8-40AF-8E57-833CD4FE6B07%7D. [Accessed: 20-Jan-2020].

[16] Australia Renewable Energy Agency (ARENA), “Baselining the ARENA-AEMO Demand Response RERT Trial (September 2019),” 2019.

[17] N. Rossetto, “Measuring the Intangible: An Overview of the Methodologies for Calculating Customer Baseline Load in PJM,” *Florence School of Regulation*, no. 2018/5, pp. 1–10, 2018.

[18] FERC, “Demand Response Compensation in Organized Wholesale Energy Markets (Final Rule), Order 745, June 2011.” [Online]. Available: https://www.ferc.gov/legal/maj-ord-reg.asp. [Accessed: 13-Jan-2020].

[19] M. L. Goldberg and G. Kennedy Agnew, “Measurement and Verification for Demand Response Prepared for the National Forum on the National Action Plan on Demand Response: Measurement and Verification Working Group,” 2013.

[20] ISO/RTO Council, “North American Wholesale Electricity Demand Response Program Comparison,” 2018. [Online]. Available: https://isorto.org/reports-and-filings/. [Accessed: 13-Jan-2020].

[21] F. Pallonetto, M. De Rosa, F. D’Ettorre, and D. P. Finn, “On the assessment and control optimisation of demand response programs in residential buildings,” *Renew. Sustain. Energy Rev.*, vol. 127, p. 109861, Jul. 2020.

[22] S. Mohajeryami, M. Doostan, and P. Schwarz, “The impact of Customer Baseline Load (CBL) calculation methods on Peak Time Rebate program offered to residential customers,” *Electr. Power Syst. Res.*, vol. 137, pp. 59–65, Aug. 2016.

[23] Y. Weng, J. Yu, and R. Rajagopal, “Probabilistic baseline estimation based on load patterns for better residential customer rewards,” *Int. J. Electr. Power Energy Syst.*, vol. 100, pp. 508–516, Sep. 2018.

[24] C. Alvarez *et al.*, “Methodologies and proposals to facilitate the integration of small and medium consumers in smart grids,” in *CIRED - Open Access Proceedings Journal*, 2017, vol. 2017, no. 1, pp. 1895–1898.

[25] A. Gabaldon, “REDYD 2050 Research Network on Distributed Energy Resources web page.” [Online]. Available: http://www.demandresponse.eu/. [Accessed: 25-Jan-2020].

[26] E. Lee, D. Jang, and J. Kim, “A Two-Step Methodology for Free Rider Mitigation with an Improved Settlement Algorithm: Regression in CBL Estimation and New Incentive Payment Rule in Residential Demand Response,” *Energies*, vol. 11, no. 12, p. 3417, Dec. 2018.

[27] S. Pati, S. J. Ranade, and O. Lavrova, “Methodologies for customer baseline load estimation and their implications,” *2020 IEEE Texas Power Energy Conf. TPEC 2020*, Feb. 2020.

[28] H. Jiang, Y. Zhang, E. Muljadi, J. J. Zhang, and D. W. Gao, “A Short-Term and High-Resolution Distribution System Load Forecasting Approach Using Support Vector Regression with Hybrid

Parameters Optimization," *IEEE Trans. Smart Grid*, vol. 9, no. 4, Jul. 2018.

[29] D. Niu, Y. Wang, and D. D. Wu, "Power load forecasting using support vector machine and ant colony optimization," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2531–2539, Mar. 2010.

[30] P. F. Pai and W. C. Hong, "Support vector machines with simulated annealing algorithms in electricity load forecasting," *Energy Convers. Manag.*, vol. 46, no. 17, pp. 2669–88, Oct. 2005.

[31] M. del C. Ruiz-Abellón, A. Gabaldón, and A. Guillamón, "Load forecasting for a campus university using ensemble methods based on regression trees," *Energies*, vol. 11, no. 8, p. 2038, Aug. 2018.

[32] H. Nie, G. Liu, X. Liu, and Y. Wang, "Hybrid of ARIMA and SVMs for short-term load forecasting," in *Energy Procedia*, 2012, vol. 16, no. PART C, pp. 1455–1460.

[33] S. Karthika, V. Margaret, and K. Balaraman, "Hybrid short term load forecasting using ARIMA-SVM," in *2017 Innovations in Power and Advanced Computing Technologies, i-PACT 2017*, 2017, vol. 2017-January, pp. 1–7.

[34] P. Ray, S. Sen, and A. K. Barisal, "Hybrid methodology for short-Term load forecasting," in *2014 IEEE International Conference on Power Electronics, Drives and Energy Systems, PEDES 2014*, 2014.

[35] K. Li, F. Wang, Z. Mi, M. Fotuhi-Firuzabad, N. Duić, and T. Wang, "Capacity and output power estimation approach of individual behind-the-meter distributed photovoltaic system for demand response baseline estimation," *Appl. Energy*, vol. 253, p. 113595, Nov. 2019.

[36] A. Gabaldón *et al.*, "Integration of Methodologies for the Evaluation of Offer Curves in Energy and Capacity Markets through Energy Efficiency and Demand Response," *Sustainability*, vol. 10, no. 2, p. 483, Feb. 2018.

[37] M. H. Shoreh, P. Siano, M. Shafie-khah, V. Loia, and J. P. S. Catalão, "A survey of industrial applications of Demand Response," *Electr. Power Syst. Res.*, vol. 141, pp. 31–49, Dec. 2016.

[38] C. Dinesh, S. Welikala, Y. Liyanage, M. P. B. Ekanayake, R. I. Godaliyadda, and J. Ekanayake, "Non-intrusive load monitoring under residential solar power influx," *Appl. Energy*, vol. 205, pp. 1068–1080, Nov. 2017.

[39] E. G. Cazalet, M. Kohanim, and O. Hasidim, "Complete and Low-Cost Retail Automated Transactive Energy System (RATES) , California Energy Commission," 2020.

[40] A. Gabaldón, A. García-Garre, M. C. Ruiz-Abellón, A. Guillamón, C. Álvarez-Bel, and L. A. Fernandez-Jimenez, "Improvement of customer baselines for the evaluation of demand response through the use of physically-based load models," *Util. Policy*, vol. 70, p. 101213, Jun. 2021.

[41] NAESB, "Business Practices for Measurement and Verification of Wholesale Electricity Demand Response," 2009.

[42] S. Lei, J. Mathieu, and R. Jain, "Performance of existing Baseline Models in quantifying the effects of Short-Term Load Shifting of campus buildings," *SLAC-R-11*, 2019.

[43] S. Lee, "Comparing Methods for Customer Baseline Load Estimation for Residential Demand Response in South Korea and France: Predictive Power and Policy Implications Chaire European Electricity Markets," *Work. Pap. 39*.

[44] 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, Nov. 2018.

[45] C. Yang, Q. Xu, and X. Wang, "Strategy of constructing virtual peaking unit by public buildings' central air conditioning loads for day-ahead power dispatching," *J. Mod. Power Syst. Clean Energy*, vol. 5, no. 2, pp. 187–201, Mar. 2017.

[46] L. M. Saini and M. K. Soni, "Artificial neural network based peak load forecasting using Levenberg-Marquardt and quasi-Newton methods," *IEE Proc. Gener. Transm. Distrib.*, vol. 149, no. 5, pp. 578–584, Sep. 2002.

[47] L. M. Saini, "Peak load forecasting using Bayesian regularization, Resilient and adaptive backpropagation learning based artificial neural networks," *Electr. Power Syst. Res.*, vol. 78, no. 7, pp. 1302–1310, Jul. 2008.

[48] L. M. Saini and M. K. Soni, “Artificial neural network-based peak load forecasting using conjugate gradient methods,” *IEEE Trans. Power Syst.*, vol. 17, no. 3, Aug. 2002.

[49] NYISO, “Emergency Demand Response Program Manual.,” 2019. [Online]. Available: https://www.nyiso.com/demand-response. [Accessed: 23-Jan-2020].