Utilities Policy 70 (2021) 101213

Contents lists available at ScienceDirect

Utilities Policy

journal homepage: http://www.elsevier.com/locate/jup

Improvement of customer baselines for the evaluation of demand response through the use of physically-based load models

A. Gabaldón [a,*], A. García-Garre [a], M.C. Ruiz-Abellón [b], A. Guillamón [b], C. Álvarez-Bel [c], L. A. Fernandez-Jimenez [d]

[a] Power Systems Group, Technical University of Cartagena, Dr Fleming S/n, 30202, Cartagena, Spain
[b] Department of Applied Mathematics and Statistics, Technical Un. of Cartagena, 30202, Cartagena, Spain
[c] Institute for Energy Engineering, Universitat Politècnica de Valencia, 46022, Valencia, Spain
[d] Department of Electrical Engineering, Universidad de La Rioja, 26006, Logroño, Spain

ARTICLE INFO

Keywords:
Customer baseline load
Physically-based load modelling for demand response
Measurement and verification

ABSTRACT

Demand Response (DR) is an opportunity and a concern for markets as well as power system flexibility. The deployment of DR depends on both knowledge on its performance and how to measure it effectively to provide adequate economic feedback. DR verification requires a baseline reference. This paper introduces a new baseline that provides an evaluation of response based on simple adjustment factors through physically-based models, tools which are also used in DR. The approach includes the detection of licit and gaming responses before and after DR. Results show that errors decrease by 10–15% with respect to conventional approaches.

1. Introduction

Electricity Networks will need to be much more flexible and innovative than in the past, both from economic and technical points of view, since energy policies in the future will involve a more significant participation of renewable resources in the generation mix. This new mix exhibits more volatility than our conventional power systems. In this scenario, and without the participation of demand-side resources, the objectives for renewable share will not be credible. For this reason, the design of new markets is "customer-centered" (European Commission, 2019/944) and encourages the participation of demand-side through distributed energy resources (DER) to increase the flexibility of power systems. A basic DER option is to develop the portfolio of demand response (DR) on an equal footing with respect to conventional supply-side resources. This includes the payment for the resource's performance (FERC, 2011), (FERC, 2020), a potential that must be measured and verified.

The growth of the DR alternatives needs the engagement of customers, but feedback is necessary to guarantee fair remuneration of resources. For this purpose, the US regulator (FERC) issued Order 745 in 2011 (FERC, 2011). This states that DR providers must be remunerated at the same price paid to generators. That is to say, DR providers will be compensated for the amount of demand that was reduced by DR policies at the full locational marginal pricing (LMP) rate at the time of load flexibility. This proposal generated considerable controversy and some stakeholders raised the issue that paying LMP in all hours presents a significant challenge for the accurate measurement and verification of DR. In 2020, through order 2222 (FERC, 2020) the scenario has changed again. In the European Union (EU) the scenario is quite similar. Article 17 of the European 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" (European Commission, 2019/944).

The achievement of this objective requires right 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 while DR performs. A forecast of demand considering loads and customers is needed. The physical behavior of loads and customers can change due to several parameters: weather, type of day, end-use shares or the frequency in DR calls. It is important to state that both in "real-time" and after a DR event happens, aggregators and System Operators (SOs) should estimate the "steady-state" load of their customers without DR (that is, a Customer Baseline Load, CBL (EnerNOC, 2009)) with respect to smart meter measurement. In "real--time", the baseline shows whether customer and aggregator are meeting DR targets. In the medium-term (monthly charges/revenues), the baseline allows the evaluation of credits to customers, and validates the

* Corresponding author.
E-mail address: antonio.gabaldon@upct.es (A. Gabaldón).

https://doi.org/10.1016/j.jup.2021.101213
Received 5 February 2020; Received in revised form 21 March 2021; Accepted 23 March 2021
Available online 28 April 2021

Table of abbreviations

ANN	Artificial Neural Network
BAWG	Baseline Accuracy Work Group
BRP	Balance Responsible Parties
CAISO	California Independent System Operator
CBL	Customer Baseline Load
DAP	Day-Ahead Prices
DER	Distributed Energy Resources
DR	Demand Response
HVAC	Heat and Ventilation Air Conditioning
ICT	Information and Communication Technology
IESO	Independent Electricity System Operator (Canada)
ISO	Independent System Operator
LMP	Locational Marginal Pricing
MPE	Mean Percentage Error
NAESB	North American Energy Standard Board
NE	New England
NIALM	Non-Intrusive Load Monitoring
nRMSE	Normalized Root Mean Square Error
NYISO	New York System Operator
PBLM	Physically-Based Load Models
RTO	Regional Transmission Operator
RTP	Real-Time Prices
SO	System Operator
STLF	Short-Term Load Forecasting
WH	Water Heater
WS	Weather Sensitive

performance and the potential of DR resources by SOs. The goodness of the CBL can determine the success of the DR programs, because customers need to receive the correct incentive for the effort of meeting DR targets while SOs need to obtain some benefits for developing the DR program. Thus, underpaying customers could lead to customer complaints (conEdison, 2015) and they may be less likely to engage DR or, if they remain in some initiative, they try to reduce their usage in the future assuming they will get underpaid anyway.

Three factors are critical in the development of baselines (EnerNOC, 2009): accuracy, simplicity and integrity. The accuracy of both the baseline estimation and the achieved flexibility is important to avoid paying too high an incentive for DR while still encouraging participation. It is also crucial, for the customers, to recognize their value in participating in DR and also to avoid non-performance penalties or the underestimation of demand reductions. For instance, ISO-NE (an American SO) argues that when a market participant schedules demand reductions for many consecutive days, baselines may no longer reflect a customer's "normal" electricity usage. These remarks are a concern because the interest in DR may be reduced on the premise that its costs are usually high (Smart Energy Demand Coali, 2015) and this casts a doubt on the recovery of the investment made in enabling technologies.

Finally, a baseline methodology must be robust to face to manipulation attempts of some customers or entities. SOs think that customers could simply shift load from day-to-day in different hours to affect the calculation of actual curtailment. These changes in patterns should be detected by baselines to ensure DR integrity. Moreover, CBL methodology should be as simple as possible and should consider the characteristics of customers and markets where DR resources are deployed. If CBL is too complex when it comes to providing a more accurate estimate of "normal" consumption, it could lead to a lack of interest on the part of aggregators and customers. Note that there must be multiple baselines to cover different types of DR activations on a range of different sites (Smart Energy Demand Coali, 2015). For example, a methodology for the evaluation of CBL may be adequate for verifying the provision of ancillary services but it would not be well suited for evaluating the response in energy markets. The same problem also arises in the evaluation of DR, and can be considered as a related problem. The approach presented in this paper covers the improvement of CBL evaluated with similar tools to those used in the planification and operation of DR: the use of aggregated load models in small/medium customers (Gabaldon, 2020).

The main contributions of this paper can be summarized as follows:

1) The paper presents a method to improve the performance of CBLs considering Physically-Based Load Models (PBLM) that are used by aggregators to evaluate the potential and response of main end-uses at residential and commercial levels: electric heaters (EH), heat pumps and water heaters (WH). This approach outperforms conventional CBLs specially when sudden variations of demand take place due to weather changes.
2) The proposed CBL has a double adjustment: forward and backward. The backward adjustment limits the errors due to payback (energy recovery periods) before the control.
3) The proposed alternative can distinguish changes in baselines due to temperature, preheating or precooling from the possibility of gaming, which is an issue reported in the bibliography that affects the fairness of costs and revenues.
4) The methodology can take advantage of other potential tools being used by aggregators or utilities, for example: Non-Intrusive Load Monitoring (NIALM), Short Term Load Forecasting (STLF), renewable forecasting …
5) The proposed CBL can be useful in new services related to the energy recovery of loads after DR and determine the energy balance between aggregators (and customers), Load Serving Entities (LSE) and Balance Responsible Parties (BRP).

The rest of the paper is organized as follows. Section 2 deals with the literature review of CBLs, focusing on their importance in DR programs, the wide range of existing approaches and their relationships with STLF methods. In Section 3, a revision of traditional CBLs and their adjustments are introduced; the PBLM is explained and the procedure to evaluate the DR performance from an economic point of view is presented. Section 4 outlines the case study and illustrates the necessity of new adjustment factors to improve CBLs, whereas Section 5 shows the results obtained for the case study when the proposed method that considers PBLM is applied. The conclusions are presented in Section 6.

2. Literature review

Baseline methods have grown in interest since the last decade due to the forthcoming role of the DR policies in wholesale and retail markets. Moreover, aggregators have gained momentum during this period, and opportunities for them will emerge in future scenarios (EURELECTRIC, 2015). The success of new DR policies, with new customers, markets and services, has led to more complex baselines and to an interest in improving the performance of the methods because payment and revenues are rising (PJM,). For this reason, different research laboratories (Coughlin et al., 2008), (Bertoldi et al., 2016), SO (Lake, 2011), (California, 2017), aggregators (EnerNOC, 2009), utilities (conEdison, 2015), (Willoughby et al., 2013) and energy and environmental agencies (Australia Renewable Energ, 2019), have analyzed different types of baselines, their metrics, and have summarized methods to improve their accuracy. Examples of CBL methods specifying data windows, exclusion rules, and adjustments are reported in (Goldberg and Kennedy Agnew, 2013). It is interesting to note that the US ISO/RTO council periodically summarizes a table (ISO/RTO Council, 2018) that lists the description, measurement and verification parameters for DR programs across different SOs. This proliferation of methodologies makes the management of DR more difficult (Rossetto, 2018). In 2009, the North American Energy Standard Board (NAESB) acknowledged the lack of harmonization as a possible barrier to the development of DR (NAESB, 2009). Consequently, NAESB developed a series of definitions that were later recognized by US authorities (FERC, 2011), (Goldberg and Kennedy Agnew, 2013). European task forces on Smart Grids have also reported the same problems with baselines amongst the different countries and power systems in Europe (European Smart Grids Task, 2019).

The literature also outlines methods for different customer segments and regions, and some of them establish comparisons between these baselines. Lawrence Berkeley reported in (Coughlin et al., 2008) some methods for non-residential buildings. This research confirmed that morning adjustments improve the performance, but they recommend the use of a different model for different groups of loads according to their weather sensitivity. In (conEdison, 2015) Consolidated Edison presents similar results, but the paper states that simple baselines, such as High3of5 or Mid5of10, perform well. However, when more sophisticated methods such as regression analysis are used, they are inherently inaccurate for other individual customers and days, mainly for residential consumers. Similar results are reported in (Willoughby et al., 2013) by San Diego Gas & Electric through an analysis that covers 21 different methods. Authors conclude that traditional baselines estimate reasonably across all customers and all event days. For instance, High3of5 generally ranks in the first quartile in all accuracy metrics, but any method is accurate for individual customers on individual event days. In conclusion, more complex baselines (e.g. regression methods) only provide marginal improvements in accuracy at high levels of complexity. In (Mohajeryami et al., 2016) authors propose the clustering of customers into different groups to reduce the randomness of each individual demand. Although the idea improves the performance of CBL, the authors conclude that, unlike industrial and commercial customers, the morning adjustment of CBLs produces an adverse impact in their “overall performance index” for residential customers.

The classification of demand into homogeneous or heterogeneous groups has also been applied in the definition of baselines, and it is a pretreatment method that was previously established for DR planning and management (Gabaldon, 2020), (Alvarezet al., 2017), which demonstrates that the operation and verification of DR should share common procedures with CBLs. In (Wijaya et al., 2014), three different baselines are analyzed (exponential moving average, regression and HighXofY) for residential customers and their metrics. Authors conclude that CBL will affect customer decision and participation in future DR events, and that bias is more relevant than accuracy in determining which CBL provides more profit for the stakeholders.

Another possibility stated in the literature is the use of STLF methods to define baselines. Despite being different tools, STLF and CBL share common methodologies. STLF is a research field with a growing interest in the literature. CBL and STLF basically provide demand forecasts in the short-term (24–48 h). The STLF portfolio comprises multiple methodologies. Support Vector Regression (SVR) and Machine (SVM) have been employed to forecast demand (Jiang et al., 2018)– (Pai and Hong, 2005). Hybrid parameter optimization (Jiang et al., 2018) and ant colony optimization (Niu et al., 2010) have been reported to find the optimal parameters for SVR, whereas simulated annealing algorithms were employed to choose the parameters of a SVM model in (Pai and Hong, 2005). Nevertheless, classical statistical methods like ARIMA models still perform well for medium and small customers forecasts and achieve an interesting performance (Ruiz-Abellón et al., 2018). Furthermore, hybrid models that combine two or more different models have showed good results in short-term forecasting, such as the hybrid models of ARIMA and SVM developed in (Nie et al., 2012), (Karthika et al., 2017) and the hybrid of SVM and ANN (Artificial Neural Networks) developed in (Ray et al., 2014). In (Li et al., 2019), a machine learning approach to disaggregate load and PV generation from net load data to obtain CBLs in prosumers is proposed. Authors conclude that reducing errors in the PV output power estimation can improve the CBL performance. Other methods are self-organizing maps and K-means (Bertoldi et al., 2016). In (Wei et al., 2017), a back propagation neural network is adapted to establish baselines in public buildings (Korea and China), taking into account meteorological indices. Finally, large industrial customers are considered in (Zarnikau and Thal, 2013) using historical data to define an appropriate baseline.

Finally, other approaches to obtain CBLs such as control groups are considered in the literature. Control groups include customers that do not participate in DR policies so that their behavior can be compared to those who respond to DR. Because this approach intends to reflect the response of small weather-sensitive loads (e.g. AC cycling and smart thermostats), the customer in the control group should experience the same weather conditions as responsive users. According to (California, 2017), two main alternatives are used: randomized control trial and matched control groups (California, 2017). The randomized trial involves customers who can participate in DR actions but that are randomly selected in advance, and their flexibility is withdrawn during the target period. In the second case, the control group consists of customers which do not participate in DR but have similar characteristics to participants. According to (California, 2017), this specific option outperforms traditional baselines.

CBL for residential customer has also been considered in the literature. In (Lee et al., 2018), CBL estimation is performed through linear regression using historical demand and cooling degree-days as independent variables. In (Wang et al., 2018), a synchronous pattern matching principle-based residential CBL estimation approach without historical data requirement is proposed. Customers are split into 2 groups (DR and control groups). The control group is clustered and then, each DR participant is matched to the most similar cluster to predict their CBLs. The main problem of this approach is that it works better with large enough sample sizes (between 200 and 400 participants) and perhaps aggregators may lack a sufficient customer portfolio for its implementation. Another problem is that these customers can not receive any DR income and the aggregator should rotate the group periodically to achieve equity in the share of customer participation and revenue.

3. Methodology: DR verification through CBL and PBLM

3.1. Overall methodology

Planning, operation, and measurement and verification methodologies are critical in the success of DR (European Smart Grids Task, 2019). Fig. 1 represents the interaction between some of the tools being used by aggregators. First, it can be understood that the evaluation and deployment of DR potential need some models (PBLM) and further aggregation to simulate the aggregated response (Gabaldónet al., 2018) and determine the potential of resources (e.g. minimum reduction levels, loss of load service, energy recovery …). As DR is basically achieved by responsive loads with some kind of storage, for maintaining the customer service (e.g. temperature or hot water in HVAC and WH respectively), aggregators need some tools to perform DR simulations before the event or response is due. For example, the end-use demand can be obtained through NIALM to tune the parameters of each PBLM model. Then, it is necessary to perform a segmentation and define homogeneous and heterogeneous load groups and customers according to the value of those load model parameters and the overall demand. Finally, NIALM can be applied again to achieve average end-use patterns (i.e. end-use load baselines) from smart meter measurements.

Another example of synergy between methodologies is the participation in energy markets: 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 (Ruiz-Abellón et al., 2018) or CBLs that can also be used as a reference

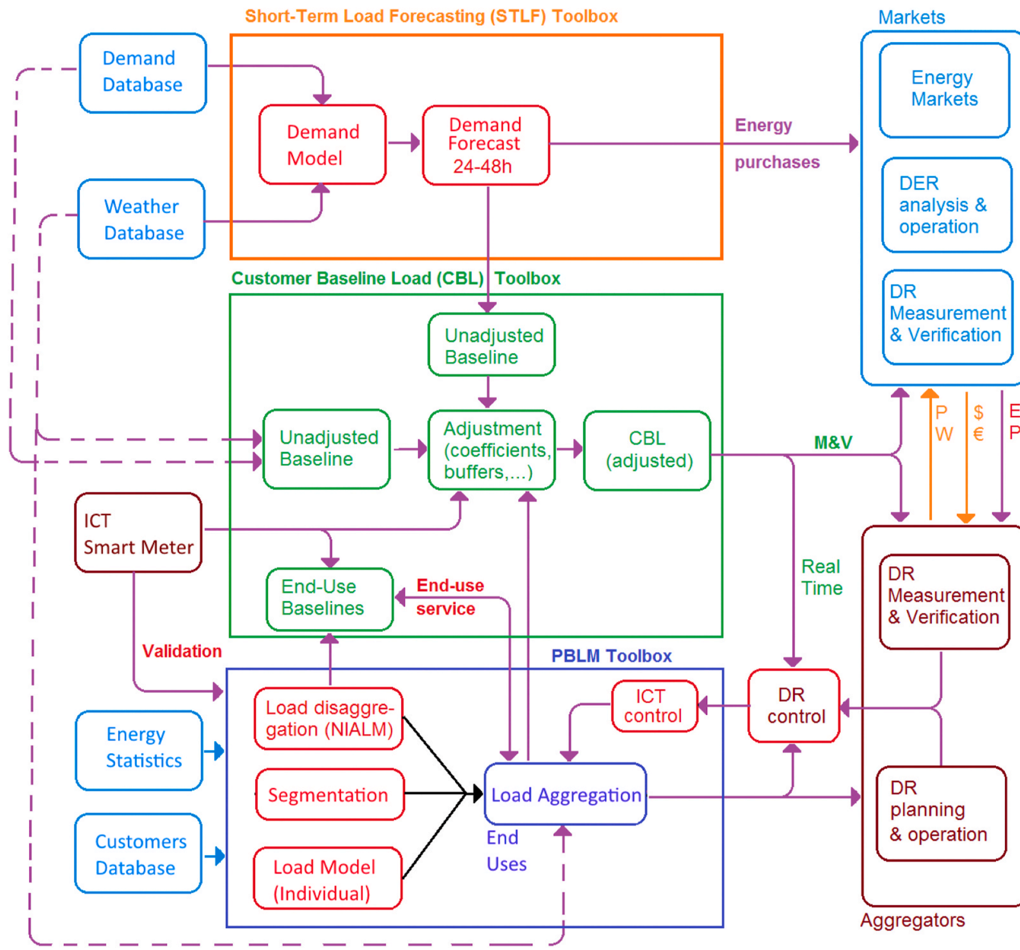

Fig. 1. Interaction between PBLM, CBL and STLF toolboxes.

(orange box in Fig. 1). Finally, NIALM allows a verification of the performance of responsive loads through a representative sample to verify the customer disabling of some devices.

Moreover, the aggregator sometimes requires a simple CBL to perform short-term evaluation of demand changes in customers or net-demand changes due to renewable generation in the case of prosumers (Ruiz-Abellón et al., 2019). The objective is to refine or redefine DR targets that will be sent to control devices (Information and Communication Technology, ICT) and fix the deviations with respect to preliminary targets or SO requests (bottom-right in Fig. 1).

In this way, STLF and CBL provide inputs for PBLM toolboxes (Ruiz-Abellón et al., 2019), and are common tools for the day-to-day operation of aggregators or customers. For example, STLF for the participation in energy markets, see Fig. 1; but it is also necessary to consider the linkage in the opposite sense (from PBLM to CBL toolboxes). This topic is intended to be justified through the results achieved in the methodology proposed in this paper.

3.2. PBLM models

PBLM are considered as grey-box models for end-use loads (HVAC, WH …). These models apply physical laws between loads and their environment (e.g. an energy balance between heat gains, losses and generation; heat storage; energy conversion; water or process flows … that establish their service and state), to determine the load behavior according to the change of inputs in the system (in our case, control of electricity supply or changes in thermostat setpoints). The model is tuned through real data measurements or NIALM, including control response (Fig. 1), to cover actual or future energy responses through DR policies. The advantage of these models is that they can evaluate the effect of a non-electrical input/parameter/variable change.

This subsection presents an example of elemental PBLM for heating (and cooling) devices in the proposed example (a university building). The model involves the development of a thermal-electrical equivalent model, which is a lumped RC network, usually called 3R2C, 2R2C, 2R1C or 1R1C, depending on the number of lumped RC parameters that have been chosen to reproduce the admittance or transmittance of each wall for the overall model: 3, 2 or 1 (Li and Wen, 2014), (Hu et al., 2019). Fig. 2 represents the energy balance between an appliance/load, the dwelling where the load renders the service (indoor temperature X in this case) and the environment. All these PBLM consist of several sub-models to allow optimal flexibility and low computer costs in modelling processes.

Main features included in the model are:

- Heat gains: solar radiation (H_{sw}, H_w) or internal gains due to inhabitants (H_r) or internal appliances (H_a), for example, lighting or air renovation, especially important during the COVID pandemic.
- Heat storage: from the specific heat of external walls (C_w), indoor mass (C_a) or roof/ground (C_{rg}).
- Control mechanisms: which drive DR policies. F for instance, smart thermostats (m(t) in TCL), or direct ON/OFF control of supply in other loads.
- State variables are temperatures: indoor (X) that conditions the load service (and the minimum level of service admissible by customers during control); walls (X_w) and roof/ground (X_{rg}), a characteristic which allows an easy evaluation of energy stored in walls (in analogy with the energy stored by capacitors or batteries, i.e. the indirect

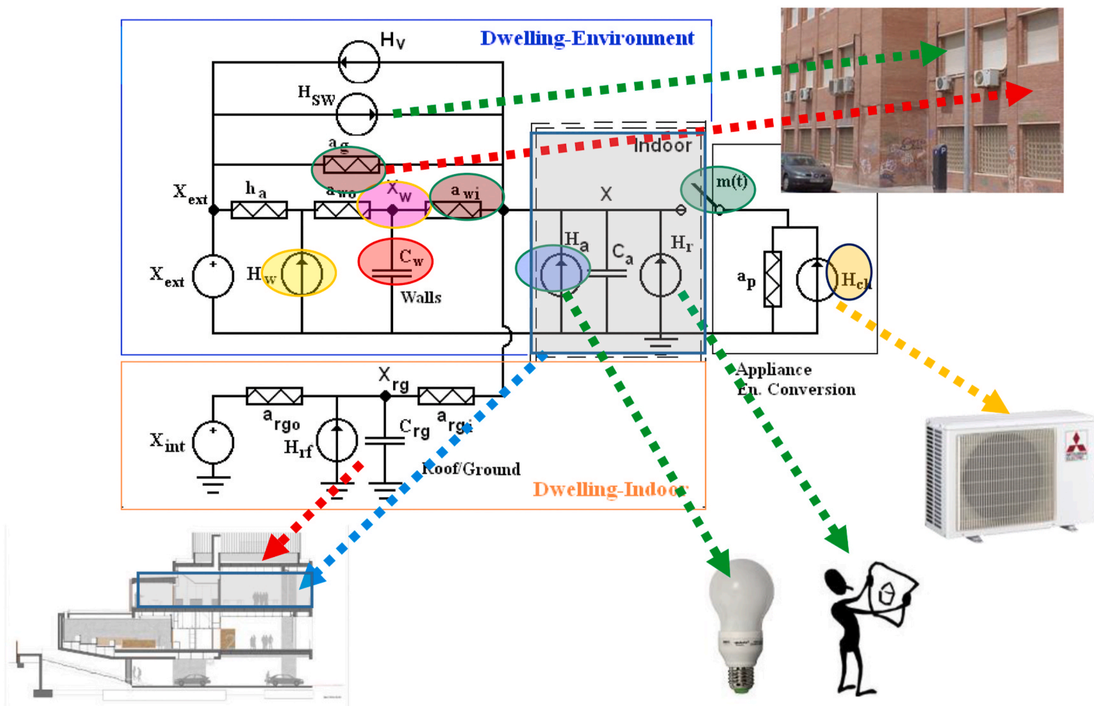

Fig. 2. Load (HVAC) and an example of PBLM equivalent (a desk) in a university building.

capacity of storage in buildings). These last variables are very important during preheating and precooling actions, to help in their evaluation.

More detailed information about formulation and the physical meaning, evaluation and identification of the parameters of this kind of model with additional research papers, scripts, simulation examples, coefficient values and data for electric heaters, HVAC and WH loads (demonstrators) can be found in (Gabaldon, 2020).

3.3. Traditional (unadjusted) baselines

Many new complex methodologies to obtain CBLs are described in the literature. Specific methodologies can be defined with excellent results for specific segments, DR products, markets and situations. However, they can fail in different scenarios (California, 2017): small/medium customers, prosumers, customers with a high share of weather-sensitive loads, loads that vary from day to day in a consistent pattern (some industrial facilities), or customers that usually are responding to DR calls. The literature agrees that "traditional" baselines are sufficient for obtaining a good and simpler basis for developing CBLs.

US power systems have significant experience with DR and consequently with CBL methodologies. For this reason, NAESB has defined five types of methodologies: a) Maximum base load; b) Meter before/meter after; c) Metering generation output; d) Baseline Type-I and e) Baseline Type-II (NAESB, 2009). They can be applied in several markets except for types b&c. Nevertheless, for the evaluation of DR in some markets, several SOs adopt a default methodology: baselines Type-I&II (Gabaldon, 2020), (Willoughby et al., 2013).

Baseline Type-I is based on the demand resource's historical meter data which can also include other variables such as weather and calendar data. Baseline Type-II assumes the same idea, but it uses statistical sampling to estimate the aggregated consumption. With the increase in the deployment rate of smart meters, reaching 100% in some European countries (FERC, 2020), this methodology lacks an important part of its practical interest. Other methodologies, such as Maximum Base Load, are much simpler because they are aimed at reducing the consumption of a demand-side resource to a specific level, regardless of its demand before deployment.

Several approaches have been proposed for baselines Type-I and Type-II. These include High (Low/Mid) XofY (Wijaya et al., 2014), exponential moving average and load forecasting baselines (EnerNOC, 2009), (Goldberg and Kennedy Agnew, 2013):

a) High(Low/Mid) XofY baseline: it considers the consumption of *Y* non-DR days before the day DR is deployed. The baseline is the average demand of the *X* highest (lowest/medium) demand days within those *Y* days. Some *Y* days are excluded, by the so called exclusion rules (Goldberg and Kennedy Agnew, 2013), because operators assume that some variables can modify the pattern of demand (e.g. a broader lock-back window from 30 to 50 days is used to define *Y* (EnerNOC, 2009)). By far, HighXofY is the most common baseline, because DR events are usually correlated with load peaks, but it is also possible to have some events in spring or fall seasons when the load does not peak (this is the case of Mid/LowXofY baselines). Some practical examples of these baselines are the use of High10of10 in CAISO (California, 2017), High5of10 in NYISO (Ruiz-Abellón et al., 2018) or High15of20 in IESO (ISO/RTO Council, 2018). These unadjusted CBLs are calculated by:

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

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

b) Exponential Moving Average: is a weighted average of the customer historical database, where the weight decreases exponentially over time. It is a similar procedure to CBL_{XofY} but the days before DR deployment have different weights, and it considers a broader spectrum for "*X*" days.

c) Short-term load forecasting (STLF) methods: the baseline is computed using a regression whose parameters are evaluated based on historical demand data. These techniques are used to find the days that have the most similar load patterns to the day of the DR event.

The proposed method sets out to implement some of these methodologies and then define new and improved adjustment coefficients through PBLM.

3.4. Adjustment coefficients for baselines

Unadjusted baselines can be tuned and improved by the use of adjustment methods. The conventional CBL_{XofY} is modified to adapt it to actual weather and demand conditions. In (Coughlin et al., 2008) the adjusted factor is obtained by:

$$af(d) = \frac{\sum_{k=1}^{pa1} A(d, h_0 - (b1 + k))}{\sum_{k=1}^{pa1} P(d, h_0 - (b1 + k))} \tag{2}$$

where *af(d)* denotes the adjusted factor for the day *d*, *A(d, h)* is the actual load of day *d* at time *h*, *P(d, h)* is the predicted load (from unadjusted baseline or STLF methods, e.g. (Ruiz-Abellón et al., 2018)) of day *d* at time *h*, h_0 is the start time of the DR event, *b1* is the buffer-time and *pa1* is the length of the pre-adjustment band (Fig. 3). Then, the adjusted CBL is obtained by:

$$CBL_{adj}(d, h) = af(d)^{*}CBL_{XofY}(d, h) \tag{3}$$

Sometimes, for example in (Bertoldi et al., 2016), an additive adjusted factor is used instead of a multiplicative one. The most usual way to evaluate these factors is the use of pre-event DR data, and calibrate the baseline using the observed non-event hours prior to DR periods. In addition, other ISOs use pre and post DR adjustment factors combined in some baseline (California, 2017). The idea is that post-event factor gives additional information about the boundary conditions throughout the DR day. CAISO Baseline Accuracy Work Group (BAWG) justifies this approach to avoid contamination of baseline, both for pre-cooling and snapback (payback of loads), occurring in the hours directly before and after the DR event (Fig. 3, pre and post-DR buffer period). BAWG recommends a 2-h buffer before and after DR. The problem is that the duration of this buffer is not justified from the point of view of end-uses and customer demand. BAWG reports changes in the adjustment coefficient at around 3–4% using both pre and post buffer periods (California, 2017).

Fig. 3 illustrates the use of pre and post adjustment periods (parameters a1 and a2 of this figure): the DR period ranges from 8:00 to 13:00 and the pre-adjustment period uses data from 5:00 to 7:00 while the post-adjustment period uses data from 15:00 to 18:00. Notice that both periods do not overlap. This is due to the inclusion of two "buffers" between the adjustment and the DR periods; b1 is the buffer period before the DR event and b2 is the buffer period after the DR event. The buffer periods try to avoid the possibility of perturbations, for example, an increase in customer demand due to preheating/precooling or "gaming". These kind of periods are implemented in several systems in the USA, for instance, NYISO (NYISO, 2019) uses a 2-h buffer. In our proposal, the duration of both "buffers" is determined and justified by PBLM. In the figure, *n* represents the period in which the consumption is affected by the DR event, that is, the sum of the DR period and the post-buffer period.

3.4.1. The risk of gaming

Many approaches include a cap in the adjustment factor from ±20 to ±40%. This involves a limit to how much the CBL can be adjusted to account for differences in patterns of consumption in the very short-term. Some reports argue (Australia Renewable Energ, 2019) that the use of a larger factor can create incentives and opportunities for gaming. For instance, supposing that an aggregator or a large customer knows that a DR event will happen, they could increase the consumption in the pre-DR period thereby increasing their CBL. So, the larger the cap, the greater the opportunity for gaming arises. However, this behaviour can also appear as a logic response of the customers to face DR events, for example, pre-heating and pre-cooling policies. The price-elasticity and price-response of demand should be improved and considered for CBLs too (Gabaldónet al., 2018). For these reasons, the amount of notice given for an event (Australia Renewable Energ, 2019) ought to be reconsidered because it can decrease DR response. Nevertheless, the underestimation of DR may deter customers from their engagement in DR policies (Rossetto, 2018). It may be more interesting to consider the length and proximity of the adjustment window period to the DR period, but based on the physical response of loads. PBLM can help to distinguish between gaming and pre-heating or pre-cooling policies. A similar role can be played by NIALM, even more precisely, because it can extract elemental load patterns in the aggregated demand and define the change of patterns of responsive loads prior to the DR event (Fig. 1).

3.5. Assessment of baseline characteristics

An ideal baseline should be both accurate and precise. Accuracy refers to the lack of bias. An estimator is said to be biased if it overestimates (negative bias) or underestimates (positive bias) the target in the average. In our context, errors (the difference between the real consumption series and the baseline) should be unbiased with respect to zero. On the other hand, precision refers to the dispersion or variability

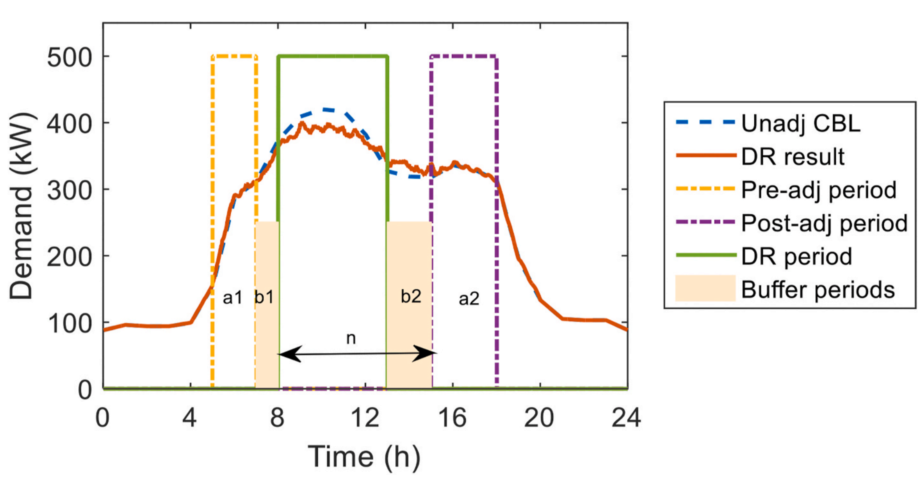

Fig. 3. Example of baseline adjustment and periods being used in (2).

of the estimations. An estimator is said to be precise if it has low variability. In our context, errors should have low dispersion around zero. Both aspects (accuracy and precision) are relevant for assessing the performance of baselines (see Fig. 4). Some baselines have a good accuracy but can be imprecise, whereas other baselines exhibit high precision but can be biased.

Two common metrics included in (California, 2017) have been used to assess baselines. The Mean Percent Error (MPE) has been selected to measure accuracy because it can describe the magnitude and direction of the bias. MPE indicates the percentage by which the baseline, on average, over or underestimates the true demand. The closer to zero the MPE is, the more accurate the baseline is. To evaluate the precision, the normalized root mean squared error (nRMSE) has been selected. This metric normalizes the RMSE by dividing it by the average of the actual demand. The lower the nRMSE is, the more precise the baseline is. Note that MPE and nRMSE are relative measurements, so they can be used to compare the accuracy and precision of several baselines measured in different units or scales. Mathematically:

$$MPE = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)}{\overline{y}} \tag{4}$$

$$nRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}}{\overline{y}} \tag{5}$$

where y_i is the real demand at time i, $\widehat{y}_i$ is the CBL (predicted demand) at time i, and $\overline{y}$ is the mean of the real demand for the n values. Note that the accuracy and precision metrics are calculated just from the beginning of the control period to the end of the recovery period, therefore n denotes the length of the evaluation period.

3.6. Economic evaluation of DR policies

The calculation of CBLs is crucial when determining revenues. The success of DR depends on the incentives that customers receive. If customers didn't perceive a significant benefit from their effort to reduce or shift their consumption, they would finally end up leaving the program. Moreover, utilities should strike a balance between the benefits they obtain with DR and the revenues they are willing to share with their customers. Consequently, it seems necessary to measure the impact of CBLs in net revenues.

As mentioned in Section 1, US FERC Order 745 (FERC, 2011) and Article 17 of the European Directive 2019/944 (European Commission, 2019/944) state that markets should treat demand reduction as if it were a supply source, so DR providers must be remunerated at the same price paid to energy producers. In this paper, it is proposed that the amount of demand reduced through DR is paid to participants at the full LMP at the time of response. This compensation model has been a subject of controversy. For instance, the Electric Power Supply Association (EPSA) with the support of a group of economists argue that paying demand response resources full LMP overcompensates those resources, because they are receiving not only the incentive payments but also the benefit of not paying the cost of the energy that was not consumed. This issue has been discussed in detail in (Spees et al., 2020).

In our study, the main goals are to properly evaluate the amount of energy reduced during DR events and demonstrate that the accuracy of this evaluation could affect the economic revenues of DR. LMP prices of PJM energy markets have been applied (Fig. 5, accessible through Data Miner 2 (PJM,)) for comparison purposes. Specifically, savings and penalties for DR programs have been obtained according to Real-Time Energy Market prices (hourly LMPs). The total cost of the energy consumption has been calculated using Day-Ahead Hourly LMPs and it is used as a reference to obtain a reduction in the consumer's electricity bill (the so-called DR perceived savings). Energy costs, DR savings and penalties, have been evaluated by:

$$DR\ \ Savings(\$) = \sum_{i=hini}^{hfin}(CBL(d,i) - DR(d,i))^{*}RTP(d,i) \tag{6}$$

$$DR\ \ Penalties(\$) = \sum_{i=hfin}^{hfin+b2}(DR(d,i) - CBL(d,i))^{*}RTP(d,i) \tag{7}$$

$$Daily\ \ total\ \ cost(\$) = \sum_{i=1}^{24} Real\ \ Demand(d,i)^{*}DAP(d,i) \tag{8}$$

where *hini* and *hfin* represent the start and end hours of the declared DR event on day *d*, *b2* is the post-buffer time, *CBL(d, i)* correspond to the energy consumption forecasted for the baseline analyzed at hour *i*, *DR(d, i)* is the energy consumed when applying DR strategies and *Real Demand (d, i)* is the energy consumption measured at the smart meter. *RTP(d,i)* and *DAP(d, i)* are, respectively, the Real-Time and Day-Ahead Price for PJM markets at day *d* and hour *i*.

Finally, the rate of perceived savings obtained with DR policies are calculated by:

$$Perceived\ \ savings\ (\%) = \frac{DR\ \ savings(\$)}{Daily\ \ total\ \ cost(\$)}{*}100 \tag{9}$$

Notice that for calculating the "real DR savings" (i.e. fair and right revenue that customers should receive for applying the DR policies), it is only necessary to replace *CBL(d, i)* with the *Real Demand(d, i)* in equation (6).

4. Case of study

To explain the proposed methodology, the demand of a university building in Cartagena (Spain) has been selected as an example. End-uses include around 50 HVAC that explain around 40–50% of the overall demand of the building. Note that they are weather-sensitive (WS) loads often used in DR programs. A simple and conventional CBL such as the HighXofY provides a good basis that can be used as an input to be tuned through adjustment coefficients (Fig. 1).

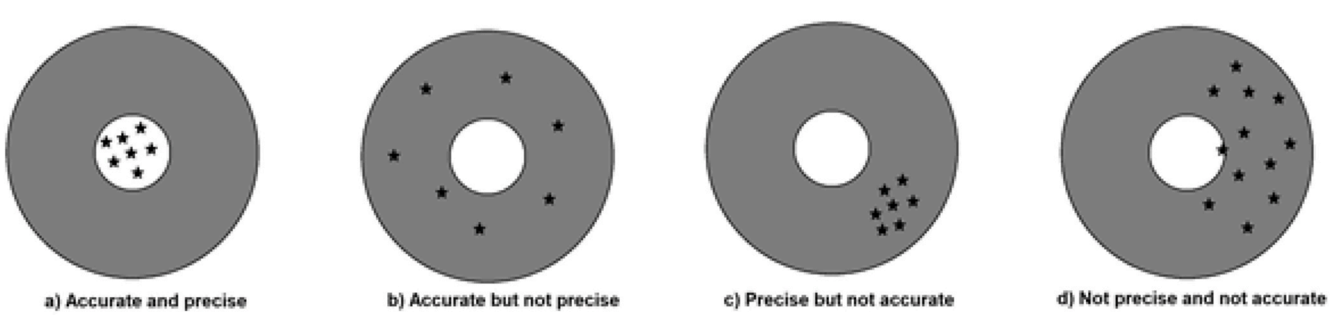

Fig. 4. Accuracy and precision.

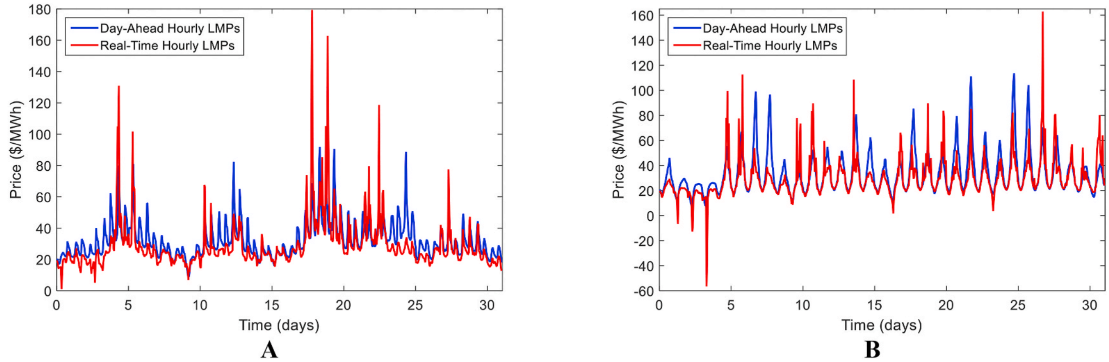

Fig. 5. Energy prices profiles of the PJM Day-Ahead (blue) and Real-Time (red) Energy Markets: a) January b) July.

As stated before, some literature approaches include the temperature of several days prior to DR event, and its effect on the consumption to improve CBLs (Sharifi et al., 2016). Two days have been selected for the analysis: one day in January and another one in July 2016. Fig. 6 compares the actual demand curve with a traditional baseline (High5of10) and its adjusted version (WS-CBL, adjusted through conventional methods (Lake, 2011)).

A "placebo/qualifying curve", that is, a measured curve without DR control, has been used to test CBLs in the case that a fictitious DR event starts around 9:00 h. Fig. 6a&b depict that the unadjusted baseline exhibits greater errors than its adjusted version, but the adjusted versions still fail. The real load would be underestimated in January (Fig. 6a) and overestimated in July (Fig. 6b). Moreover, errors are more significant in peak periods (around 10–15%) than in daily load profile. Thus, DR evaluation through these CBLs needs an improvement. The proposal is to study the weather impact in demand through PBLM simulations and a further analysis of the periods in which changes of demand patterns clearly arise. The right identification of these periods results in the improvement of the CBL metrics and this is possible when the user (or a third party, such as the aggregator, LSE, BRP) considers the transient behavior of the load by means of a PBLM.

Summarizing, the previous example justifies the necessity of considering a new pre-adjustment of the CBL in equation (2) and a refinement through post-adjustment feedback. In our case, the selection of the hours for the adjustment factors (parameters *b1-b2* and *a1-a2* in Fig. 3) is done through PBLM. On the other hand, PBLM allow us to filter disturbances in the evaluation of CBL in order to distinguish a normal pattern (to maintain load service during DR) from "gaming" (to alter CBL evaluation). Next section touches upon this concern.

4.1. Detecting gaming and pre-heating through PBLM

Some simulations have been performed to exemplify the proposed method and results are shown in Fig. 7. Fig. 7 depicts the simulation of a group of HVAC loads (around 50 loads distributed in classroom and desks of the building, Fig. 2) with different weather conditions. The load group performs with: the outdoor average temperature; the average minus 3 °C and the average minus 6 °C. A "service function" of HVACs starting at 6:00 and finishing at 13:00, and restarting from 15:00 to 20:00, has been chosen. Fig. 7a represents the results for a homogeneous group (Gabaldon, 2020) of loads.

It can be observed that the demand is basically the same at the beginning of their service period (from 6:00 to 9:00, all loads are in ON state, irrespective of the weather because all spaces remained without service during the night). However, they are different in a second period (from 9:00 to 13:00 and 15:00 to 20:00 h), when some loads in the group reach to its "steady-state" service depending on weather or occupancy. These variations in demand due to weather justify the definition of periods taken in equation (2) for the evaluation of adjustment factors. It also justifies the failure of the conventional WS-CBL because the traditional adjustment procedure can consider a wrong period irrespective of weather. Fig. 7b shows the demand for a heterogeneous group of HVAC loads (with different thermal and electrical characteristics). The load remains the same for different temperatures until 8:00 but increasingly changes from 8:00 to 10:00 depending on the weather. If the DR event starts at 9:00 and the adjustment factor only considers the first 2 h (5:00 and 6:00) of the 4-h period prior to the start of the DR event (NYISO, 2019), this factor fails to detect changes in load behavior. In this case, PBLM provides an adjustment period which only considers the last 2 h without any buffer period. The necessity of a buffer is justified by some utilities because this buffer-period is a potential "gaming period" and distorts DR potential, but sometimes the buffer hides actual load changes and the service provided to the customer.

PBLM also helps in differentiating gaming from pre-heating or pre-cooling policies, which are policies to maintain customers' service during DR periods through the possibility of indirect thermal energy storage in walls or inside the dwelling ("virtual storage" on C_a and C_w capacitors of PBLM, see Fig. 2). Gaming strategies have been discussed

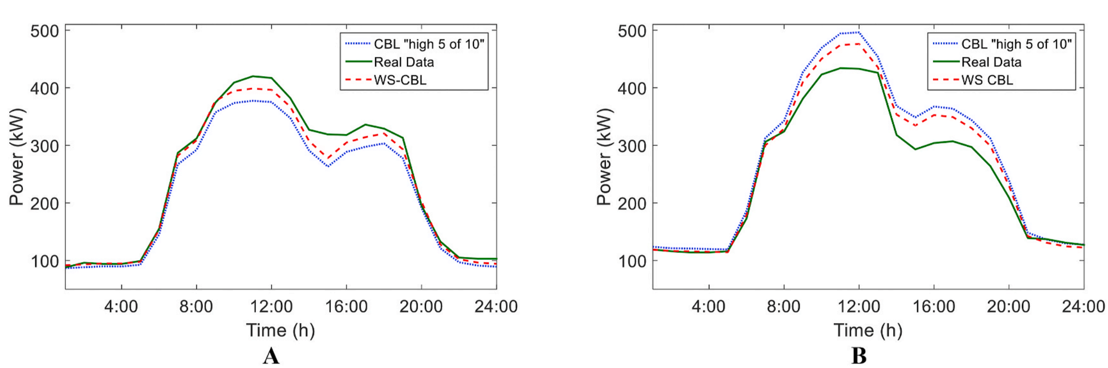

Fig. 6. Comparison between "High5of10" CBL and its corresponding adjusted CBL (WS-CBL) for morning DR events: (a) January 18th, (b) July 18th.

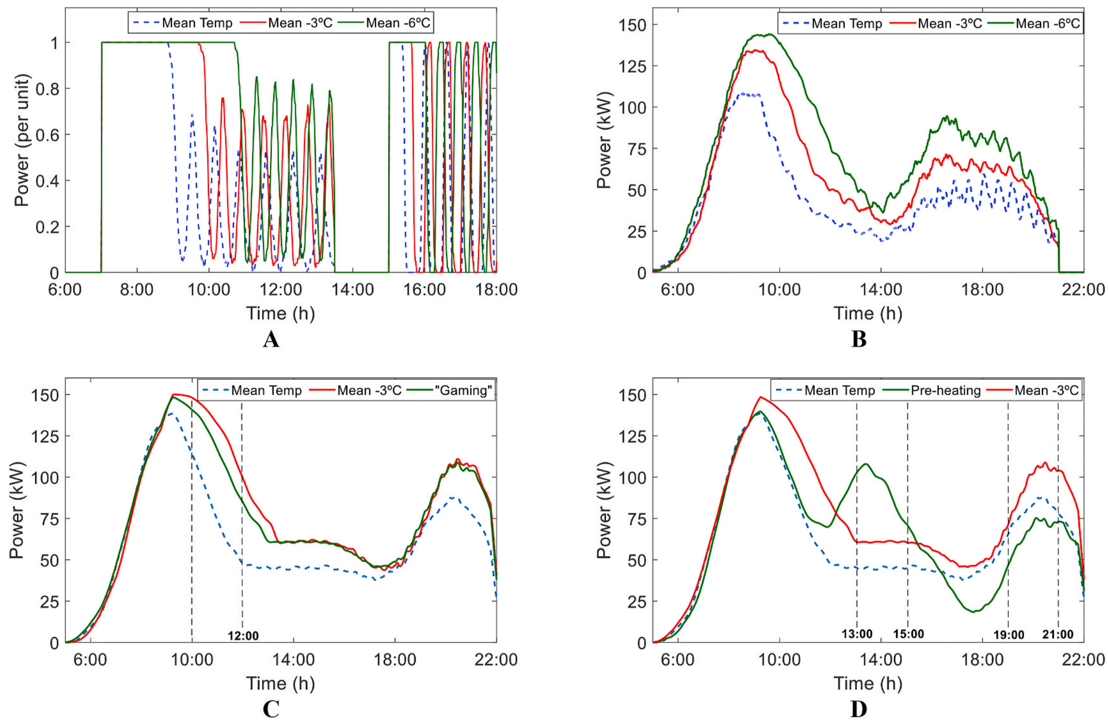

Fig. 7. Simulation of HVAC demand: (a) Homogeneous Group of HVAC; (b) Heterogeneous Group of HVAC; **(c)** Customer "gaming" simulating adverse weather; **(d)** Pre-heating policy and payback reduction (red arrows).

extensively in (Chen and Kleit, 2016). Fig. 7c&d presents the response of HVAC loads when an event is called from 16:00 to 19:00 in three scenarios: the average temperature (unadjusted CBL), the average -3 °C in which an adjustment is required to fit CBL to new patterns and for an increase of +3 °C in the thermostat setpoint to feign a higher demand due to weather that affects the adjustment of baseline(gaming). Mathematically, the use of the two first hours of the four ones before the event enables the manipulation of CBL (from 10:00 to 12:00).

These attempts of gaming can be detected through PBLM during pre-adjustment period a1 (Fig. 7c) but the identification of gaming is much more accurate after the DR period during the energy payback. Fig. 7d presents the case of preheating before the DR period. Considering the post-adj period a2 (19:00–21:00), this energy recovery period after a preheating is quite different from a decrease of the external temperature or from gaming (e.g. switching ON of non-controlled load, such as electric heaters in winter or dishwashers before DR). These results justify the need for two adjustment factors and the usefulness of PBLM feedback. These factors can be enhanced through the use of NIALM tools (Fig. 1) if granularity of smart meters is available (Gabaldon, 2020), because elemental end-use patterns can be recognized out of their normal pattern in the overall demand (Gabaldón et al., 2017).

The "payback" period, i.e. the energy recovery after DR (19:00–21:00) raises interest because also involves changes in the balance of energy which affects the aggregators and third parties (e.g. BRP) and their economic flows. The assumption of several roles by aggregators can be problematic with regard to market competition legislation (for example, article 102 of Treaty on the Functioning of the EU on dominant positions in internal markets (Park et al., 2015),). Fig. 7d shows that both during the DR events and the payback periods, the customer change their consumption patterns. This might be significant in balance markets and specially with the increasing need to balance renewables in the near future, for example through the use of the DR potential (Ruiz-Abellón et al., 2019) which must be analyzed and adequately rewarded. Some EU states, for instance France, have decided that the aggregator should pay BRP/suppliers for these energy changes (Park et al., 2015), and this has an impact on DR cost-effectiveness that must be evaluated: for instance a mechanism through specific CBL designs that are able to take balancing into account. Changes in demand during those periods are also a concern or utilities. For example, EURELECTRICS (EURELECTRIC, 2015), recommends to improve DR and aggregation development, "… ensuring that BRP/suppliers are able to renegotiate supply contracts to take into account the indirect effects of demand response (i.e. rebound effects) and consequents impacts on sourcing costs". PBLM backward adjustment also supports measurement and verification of these payback periods.

5. Results and discussion

5.1. CBL with PBLM adjustment

To illustrate the practical use of PBLM in the adjustment of CBLs, the response of a heterogeneous control group of HVAC has been simulated through PBLM. Specifically, in January, the external temperature has been chosen lower than the average (mean -3 °C) while in July it has been chosen greater than the average (mean+3 °C) according to the assumption of higher levels of demand arise during emergency events or taking into account weather prediction (Lazos et al., 2014). Moreover, a load control policy to reduce around 20% of heat pumps demand from 9:00 to 13:00 has been considered (a value that seems reasonable to avoid an excessive lack of comfort during the control period. Notice that PBLM supply feedback about the lack of comfort, in this case the internal temperature X in Fig. 2). Fig. 8a&b present some results in January. Fig. 8a depicts some simulation results for the control period in the morning: the demand curve (without DR); the demand curve with DR control of HVAC demand and the curve representing the load change (i.e. demand flexibility). To take into account other periods for power system events, an alternative DR control from 15:00 to 19:00 has been considered too (Fig. 8b). Fig. 8c&d shows some DR simulation results in July. Fig. 8c depicts some results for a control period in the morning (9:00 to 13:00) and Fig. 8d shows the simulation curves during a control period in the afternoon (15:00 to 19:00).

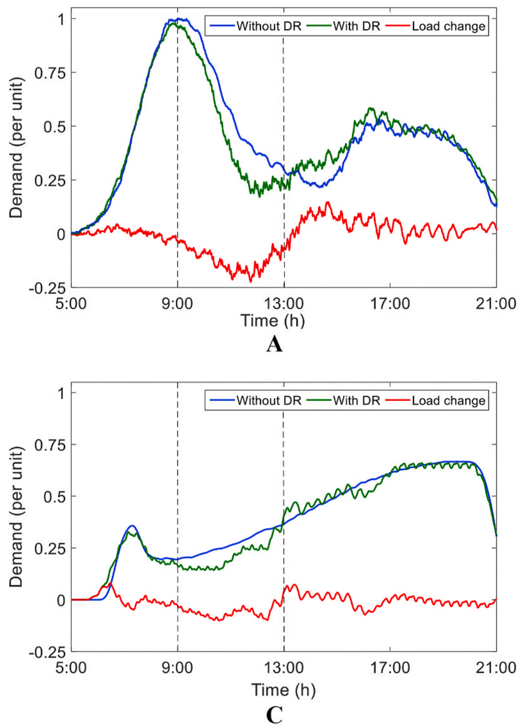

Fig. 8. Simulation of the response of a heterogeneous control group of heat pumps during DR events. Control period: a) In the morning (January); b) In the afternoon (January); c) In the morning (July); d) In the afternoon (July).

5.2. Improvements for CBL: the backward adjustment

It is important to consider that energy savings reported during DR control periods are not fully recovered after control as it can be observed in the payback time from 13:00 to 15:00 (Fig. 8a). This can be explained because external temperature (in general, the weather, including solar radiation) contributes to reduce demand in the early afternoon in winter and, in this way, the load service changes along the day and specially after control periods. Another concern is the change of control performance due to changes in the customer behavior (Yang et al., 2017). Thus, it is also important to analyze the "energy recovery period", i.e., the period where a percentage of the energy reduced during control is

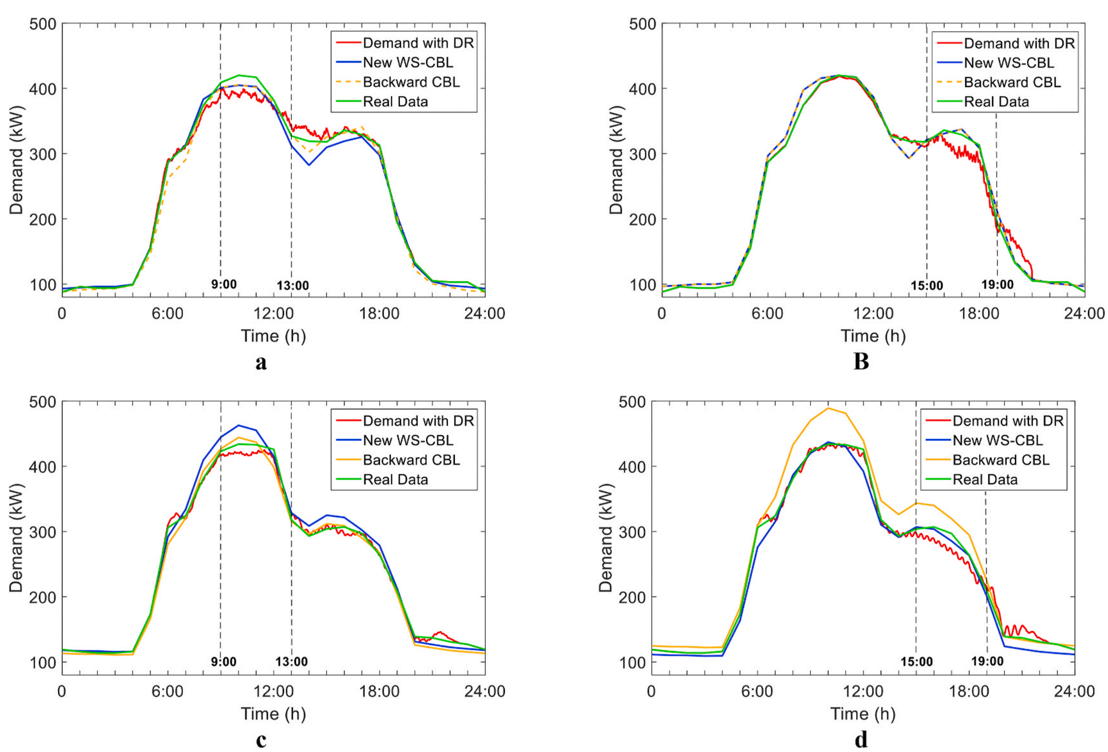

Fig. 9. CBLs' comparison for DR policies: Control period: a) In the morning (January); b) In the afternoon (January); c) In the morning (July); d) In the afternoon (July).

recovered to restore "normal" load service, in this case, the internal temperature of dwellings (from 13:00 to 15:00 in Fig. 8a; or from 19:00 to 21:00 in Fig. 8b).

PBLM provides another important insight to define an additional factor for the adjustment (section 3): the load does not recover "business as usual" or "steady-state" values of demand until some hours after the control (Fig. 8a, curve labelled as "load change"). For this reason, adjustment values are not taken at the end of control (13:00, when the transient remains) but in the period from 15:00 to 19:00. So, the adjustment period is the 4-h period beginning the start of the hour that is 2 h after the end of the reliability event, that is, the end of the recovery period defined through the simulation of PBLM. Notice that this new CBL (Fig. 9), called "Backward-CBL", is only valid from 13:00 to 00:00 (outside the control period), and therefore the energy saving of that CBL does not make sense (see Table 1) and is not considered in practice. In the case of an afternoon control period (from 15:00 to 19:00), the backward factor is calculated by taking adjustment values from 21:00 to 00:00. In this case, the Backward-CBL is only valid from 19:00 to 00:00.

The load response to morning and afternoon events, through the use of four different CBLs, has been analyzed (Table 1):

Fig. 6 compares the conventional and the Standard WS-CBL whereas Fig. 9 shows the results obtained for the New-WS-CBL and the Backward-CBL proposed in the paper.

From Tables 2–5, it can be concluded that in the simulation example, consisted on a small reduction of demand with a medium size customer as DR target, a simple forecasting method (i.e. "High5of10") modulated through new and specific adjustment factors and periods justified by PBLM simulation exhibits a better performance than using the traditional CBL with a conventional adjustment method, and take profit from other methodologies (Fig. 1) needed for the management of DR. For example, in Table 2, the error is reduced from 230.9% to 63.6% and in Table 5 is reduced from 167.8% to 27.1%. This method, that still have significant errors, can be also refined with the consideration of demand forecasting values (Ruiz-Abellón et al., 2018), (Sharifi et al., 2016), as CBL basis in equation (2) if the aggregator uses STLF methods to accomplish other objectives but with the increase of the complexity of the process. In Tables 2–5, positive values mean that demand when applying DR strategies is greater than the baseline analyzed. MPE and nRMSE are calculated only in the control and recovery periods, that is from 9:00 to 15:00 in morning events and from 15:00 to 21:00 in afternoon events. MPE and nRMSE of Backward-CBL are only calculated in the recovery periods (from 13:00 to 15:00 in morning events and from 19:00 to 21:00 in afternoon events) as before 13:00 (morning) and 19:00 (afternoon) this baseline has not sense. In the case of the MPE, positive values mean that the baseline is overestimating the real demand whereas negative values mean that they are underestimating the real demand.

The estimation of the DR effects requires a precise knowledge of demand by CBLs both during and after DR control. Tables 2–5 show that the conventional CBL and the standard WS-CBL have problems in the recovery period, especially if there is a DR morning event (errors of 409% and 203% in Table 2 or 668% and 314% in Table 4). The PBLM pre-adjustment (New-WS-CBL) could reduce the error in recovery period (from 314% to 13% in Table 4). Notice that revenues are not on real-time (for instance, until 75 days after DR event (NYISO, 2019)) and aggregators and customer have access to real demand measurement after control periods. Using the backward adjustment, the error of DR evaluation in recovery period is reduced significantly (from 203% to 35% in Table 2 or for 126%–11.9% in Table 5). This is a fair value for a "weak" DR action as the one used for this example (8% of peak reduction). It can be highlighted that Backward-CBL gives a worse result in Table 4. This is attributable to the reduced control effort requested to HVACs and consequently the low energy recovery of the load group (Fig. 9c). This is also a criterion to be considered from the aggregators when they apply both adjustments and is an issue for future works.

Table 1
Definition of the CBLs analyzed and their buffer periods.

CBL	Description	Buffer
Real demand	Real consumption (optimal but unknown CBL)	N/A
CBL	Unadjusted High5of10	N/A
Standard WS-CBL	Adjusted High5of10	First 2 h of 4-h period pre-DR event
New-WS-CBL	Adjusted High5of10	Defined by PBLM models
Backward-CBL	Adjusted New-WS-CBL	Last 2 h of 4-h period post-DR event defined through PBLM simulation

5.3. Economic analysis of demand response through CBL and PBLM

As it was explained in Section 3.2, it is necessary to evaluate how the accuracy of CBL calculation methods financially affects the economic flows due to DR. To accomplish this task, four different days in January and July have been selected, in which there were peaks of prices in the PJM Real-Time Energy Market, so there are suitable scenarios for DR events. The prices' profiles for these selected days are shown in Fig. 10.

Tables 6 and 7 show the savings and penalties that could be obtained by applying a DR policy that intends 20% reduction of the HVAC demand, which have been calculated using the methodology explained in Section 3.6.

Regarding savings, notice that the "real demand" results reflect the "correct" incentive that should be compensated to the DR participants. As it was mentioned in Section 4, both WS-CBLs (Standard and New) exhibits greater performances than the conventional CBL (for example, in Table 6: 4.89$ and 5.52$ of revenue vs. 0$ of revenue, being the "correct" incentive 9.9$). At the same time, it can be deduced that both CBLs underestimate the real consumption in January (see Fig. 6), making the revenues for the DR participants lower than it should be. However, in July, the reverse situation occurs (Table 7). As the CBLs overestimate the load consumption, the revenue that the customers could receive is greater than it should be according to the "correct" incentive.

In the case of potential penalties applied to customers for increasing their consumption in recovery periods (e.g. a lack of balance in distribution detected by BRPs), a similar analysis could be done. WS-CBLs works better that the conventional baseline, so the penalties calculated with the pre-adjusted versions are more accurate. However, the use of the backward adjustment could improve this performance significantly (for example in Table 6: 2.25$ with Backward-CBL vs. 5.68 $ with Standard WS-CBL, being the "correct" penalties 1.97$). Otherwise, overestimated CBLs (July) reduce the penalties that could be charged to DR participants, which could be a decisive factor to increase the engagement of consumers in DR and electricity markets.

Considering that LMP prices are applied to customers, Table 8 presents the total costs of the energy consumed in DR event days (PJM Day-Ahead Energy Market, equation (8)) and Table 9 shows the perceived savings on these days calculated thought equation (9). Other prices can be applied for simulation, but this scenario has been applied for comparison purposes (technical and revenue errors through CBLs) and was a real scenario for aggregators in PJM (FERC, 2011).

Table 9 depicts that the perceived savings increase when there are declared DR afternoon events. The heat pumps energy consumption is greater in the afternoon so reducing it is more feasible (Fig. 8). The "correct" incentive varies from 1.2% to 4.1% while the different CBLs obtain savings that range from 0% to 9.3%. As it was mentioned when penalties were discussed, overestimated CBLs stimulate the engagement of customers in DR programs as they increase the perceived savings, however, much overestimated CBLs such as conventional CBL in July could not be sustainable from the point of view of the utility companies and aggregators.

Table 2
Evaluation of DR performance (morning control) with different CBLs (January).

CBL	Energy savings (control period, kWh)	Error (control period, %)	Recovery (after control, kWh)	Error (after control, %)	nRMSE	MPE
Real demand	−78.22	0	30.12	0	–	–
CBL	102.42	230.9	153.40	409.3	0.1096	−0.1085
Standard WS-CBL	−28.50	63.6	91.47	203.7	0.0475	−0.0436
New-WS-CBL	−28.52	63.5	90.93	201.9	0.0475	−0.0436
Backward-CBL	N/A	N/A	40.68	35.1	0.0236	−0.0109

Table 3
Evaluation of DR performance (afternoon control) with different CBLs (January).

CBL	Energy savings (control period, kWh)	Error (control period, %)	Recovery (after control, kWh)	Error (after control, %)	nRMSE	MPE
Real demand	−70.2	0	41.24	0	–	–
CBL	46.42	166.1	68.97	67.25	0.0994	−0.0884
Standard WS-CBL	−79.06	−12.63	28.60	−30.7	0.0253	0.0138
New-WS-CBL	−77.32	−10.15	29.16	−29.3	0.0248	0.0124
Backward-CBL	N/A	N/A	28.83	−30.1	0.0547	0.0419

Table 4
Evaluation of DR performance (morning control) with different CBLs (July).

CBL	Energy savings (control period, kWh)	Error (control period, %)	Recovery (after control, kWh)	Error (after control, %)	nRMSE	MPE
Real demand	−44.04	0	9.95	0	–	–
CBL	−244.72	−455.7	−56.48	−668.4	0.1033	0.0987
Standard WS-CBL	−125.81	−185.7	−21.30	−314.4	0.0492	0.0411
New-WS-CBL	−15.8	64.13	11.24	13.1	0.0292	−0.0122
Backward-CBL	N/A	N/A	4.17	−58.0	0.0150	0.0117

Table 5
Evaluation of DR performance (afternoon control) with different CBLs (July).

CBL	Energy savings (control period, kWh)	Error (control period, %)	Recovery (after control, kWh)	Error (after control, %)	nRMSE	MPE
Real demand	−66.67	0	34.72	0	–	–
CBL	−178.57	−167.8	31.38	−9.6	0.0953	0.0641
Standard WS-CBL	−48.57	27.1	78.67	126.5	0.0550	−0.0477
New-WS-CBL	−65.60	1.61	72.47	108.7	0.0464	−0.0330
Backward-CBL	N/A	N/A	30.60	−11.9	0.0323	0.0090

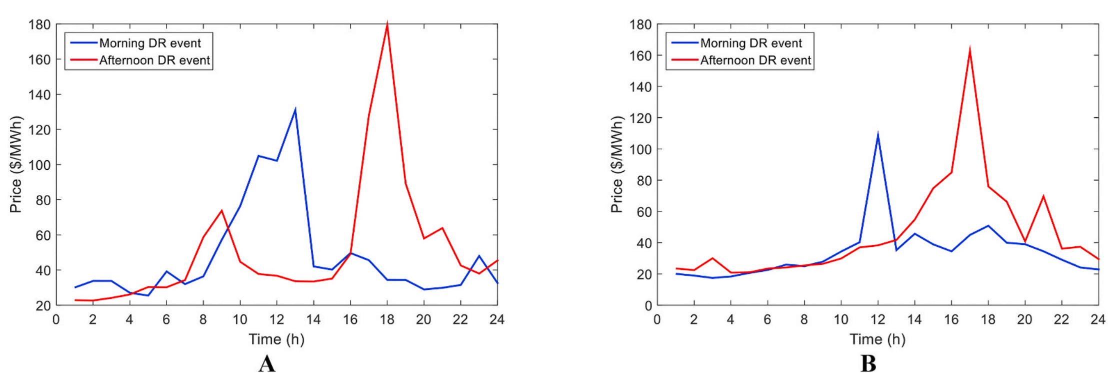

Fig. 10. PJM Real-Time energy market profiles for selected DR event days: a) January b) July.

6. Conclusions and policy implications

Baselines have been an important concern to provide accurate estimates for the operation and planning of power systems, but it can also arise as an important tool to engage and empower customers in markets, for example for decision making in electricity markets. Moreover, DR policies are necessary tools to make credible the rise of renewable share into the power generation mix in the 2030–2050 horizon and need to be also evaluated in balance services and markets.

From the point of view of aggregators and customers, the demand needs an accurate and fair verification of their flexibility. In this way, the accuracy of demand's forecasts with DR (PBLM modelling) and without DR (CBL), with incomes and penalties associated, are crucial factors for the development of DR and DER. In the energy sector, aggregators, suppliers, operators or BRP, need accurate demand models since most of their decisions and economic transactions are based on forecasts of demand and its potential flexibility. Specific and complex methodologies (ANN, machine learning …) are able to define CBLs, but this option increases the complexity of DR and needs a model for each specific customer or segment of demand. These models can present problems if

Table 6
Economic evaluation of DR performance in January.

CBL	Saving in morning control period ($)	Penalties in morning recovery period ($)	Saving in afternoon control period ($)	Penalties in afternoon recovery period ($)
Real demand	9.90	1.97	10.30	2.93
CBL	0	13.70	0.43	5.39
Standard WS-CBL	4.89	5.68	10.88	2.60
New-WS-CBL	5.52	5.39	10.68	2.63
Backward-CBL	N/A	2.25	N/A	2.10

Table 7
Economic evaluation of DR performance in July.

CBL	Saving in morning control period ($)	Penalties in morning recovery period ($)	Saving in afternoon control period ($)	Penalties in afternoon recovery period ($)
Real demand	3.18	1.04	7.57	1.96
CBL	13.97	0	21.97	1.80
Standard WS-CBL	7.91	0.18	6.36	3.98
New-WS-CBL	2.61	1.56	8.13	3.65
Backward-CBL	N/A	0.71	N/A	1.05

Table 8
Daily total energy costs in Day-Ahead Energy Market.

Day	DR morning event in January	DR afternoon event in January	DR morning event in July	DR afternoon event in July
Total Costs ($)	258.83	253.39	263.44	234.94

Table 9
Perceived savings on DR event days.

CBL	DR morning event in January (%)	DR afternoon event in January (%)	DR morning event in July (%)	DR afternoon event in July (%)
Real demand	3.8	4.1	1.2	3.2
CBL	0	0.17	5.3	9.3
Standard WS-CBL	1.9	4.3	3.0	2.7
New-WS-CBL	2.1	4.2	1.0	3.5
Backward-CBL	N/A	N/A	N/A	N/A

DR response performs periodically (i.e. the change of demand patterns). Literature shows that unadjusted baselines are not the best option but perform well in many cases, with the help of some adjustment factors. The problem is that these factors are based on trial and error methods, basically a black-box approach. This paper introduces the idea that adjustment factors can be physically explained by PBLM, and that a double-adjusted CBL exhibits even a better performance (reduces error of usual methods by 10–15%). The method considers the energy recovery periods that usually appears after control policies which involve weather-sensitive loads, significantly reduces the error. Thus, this methodology of adjustment arises as an adequate baseline estimator.

This work also states the benefits of synergies associated with the use of other aggregator tools, such as NIALM, customer segmentation, and ICT resources, to verify load response. In this case, the adjustment period can be justified and improved both before and after the period of DR events (i.e. adjusted and backward baseline, respectively) to improve the evaluation of DR policies and their changes for energy balance evaluations between the different agents involved. In this way, customers, aggregators and SOs can obtain a necessary feedback to define demand patterns and perform a better evaluation of the DR potential.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Agencia Estatal de Investigación, Ministerio de Ciencia e Innovación (projects ENE-2016-78509-C3-2 P, RED2018-102618-T and ENE-2016-78509-C3-3P/AEI/10.13039/501100011033); the Ministerio de Educación (Spanish Government) under grant FPU17/02753, and EU-ERDF funds.

References

Alvarez, C., et al., 2017. Methodologies and proposals to facilitate the integration of small and medium consumers in smart grids. in CIRED - Open Access Proceed. J. 2017 (1), 1895–1898.
Australia Renewable Energy Agency (ARENA), 2019. Baselining the ARENA-AEMO Demand Response RERT Trial (September 2019).
NAESB, 2009. Business Practices for Measurement and Verification of Wholesale Electricity Demand Response.
Bertoldi, P., Zancanella, P., Boza-Kiss, B., 2016. Demand Response Status in EU Member States.
EURELECTRIC, 2015. Designing Fair and Equitable Market Rules for Demand Response Aggregation [Online]. Available: http://www.eurelectric.org. (Accessed 8 January 2020).
California, I.S.O., 2017. Baseline Accuracy Work Group Proposal [Online]. Available: http://www.caiso.com/Documents/2017BaselineAccuracyWorkGroupFinalProposalNexant.pdf#search=customer baseline. (Accessed 13 January 2020).
Chen, X., Kleit, A.N., 2016. Money for nothing? Why FERC order 745 should have died. Energy J. Volume 37 (Number 2).
conEdison, 2015. Advanced Metering Infrastructure Business Plan [Online]. Available: http://documents.dps.ny.gov/public/Common/ViewDoc.aspx?DocRefId=%7B0864BF10-42C8-40AF-8E57-833CD4FE6B07%7D. (Accessed 20 January 2020).
Coughlin, K., Piette, M.A., Goldman, C., Kiliccote, S., 2008. Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-residential Buildings in California. Environmental Energy Technologies Division, " Berkeley.
EnerNOC, 2009. The Demand Response Baseline. White Paper.
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 January 2020).
European Smart Grids Task Force, 2019. Demand Side Flexibility [Online]. Available: https://ec.europa.eu/energy/.
Gabaldon, A.. Demand Response (DR) web page [Online]. Available: http://www.demandresponse.eu/. (Accessed 25 January 2020).
Gabaldón, A., Molina, R., Marín-Parra, A., Valero-Verdú, S., Álvarez, C., 2017. Residential end-uses disaggregation and demand response evaluation using integral transforms. J. Mod. Power Syst. Clean Energy 5 (1), 91–104. Jan.
Gabaldón, A., et al., 2018. Integration of methodologies for the evaluation of offer curves in energy and capacity markets through energy efficiency and demand response. Sustainability 10 (2), 483. Feb.
Goldberg, M.L., Kennedy Agnew, G., 2013. Measurement and Verification for Demand Response Prepared for the National Forum on the National Action Plan on Demand Response: Measurement and Verification Working Group.
Hu, M., Xiao, F., Wang, L., 2019. Investigation of demand response potentials of residential air conditioners in smart grids using grey-box room thermal model. Appl. Energy 207 (2017), 324–335.
Jiang, H., Zhang, Y., Muljadi, E., Zhang, J.J., Gao, D.W., 2018. A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization. IEEE Trans. Smart Grid 9 (4), 3331–3350. Jul.

Karthika, S., Margaret, V., Balaraman, K., 2017. Hybrid short term load forecasting using ARIMA-SVM. In: 2017 Innovations in Power and Advanced Computing Technologies, I-PACT 2017, 2017-January, pp. 1–7.

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

Lazos, D., Sproul, A.B., Kay, M., 2014. Optimisation of energy management in commercial buildings with weather forecasting inputs: a review. Renew. Sustain. Energy Rev. 39, 587–603. Nov.

Lee, E., Jang, D., Kim, J., 2018. 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 11 (12), 3417. Dec.

Li, X., Wen, J., 2014. Review of building energy modeling for control and operation. Renew. Sustain. Energy Rev. 37, 517–537. Elsevier Ltd.

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

PJM. Demand response in PJM markets and operations [Online]. Available: https://www.pjm.com/markets-and-operations/demand-response.aspx. (Accessed 13 January 2020).

Mohajeryami, S., Doostan, M., Schwarz, P., 2016. The impact of Customer Baseline Load (CBL) calculation methods on Peak Time Rebate program offered to residential customers. Elec. Power Syst. Res. 137, 59–65. Aug.

Nie, H., Liu, G., Liu, X., Wang, Y., 2012. Hybrid of ARIMA and SVMs for short-term load forecasting. Energy Procedia 16 (PART C), 1455–1460.

Niu, D., Wang, Y., Wu, D.D., 2010. Power load forecasting using support vector machine and ant colony optimization. Expert Syst. Appl. 37 (3), 2531–2539. Mar.

NYISO, 2019. Emergency Demand Response Program Manual [Online]. Available: https://www.nyiso.com/demand-response. (Accessed 23 January 2020).

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

Pai, P.F., Hong, W.C., 2005. Support vector machines with simulated annealing algorithms in electricity load forecasting. Energy Convers. Manag. 46 (17), 2669–2688. Oct.

Park, S., Ryu, S., Choi, Y., Kim, J., Kim, H., 2015. Data-driven baseline estimation of residential buildings for demand response. Energies 8 (9), 10239–10259. Sep.

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

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

Ruiz-Abellón, M. del C., Gabaldón, A., Guillamón, A., 2018. Load forecasting for a campus university using ensemble methods based on regression trees. Energies 11 (8), 2038. Aug.

Ruiz-Abellón, M.C., Fernández-Jiménez, L.A., Guillamón, A., Falces, A., García-Garre, A., Gabaldón, A., 2019. "Integration of demand response and short-term forecasting for the management of prosumers' demand and generation. Energies 13 (1), 11. Dec.

Sharifi, R., Fathi, S.H., Vahidinasab, V., 2016. Customer baseline load models for residential sector in a smart-grid environment. Energy Rep. 2, 74–81. Nov.

Smart Energy Demand Coalition (SEDC), 2015. Mapping Demand Response in Europe Today [Online]. Available: http://www.smarten.eu/wp-content/uploads/2015/09/Mapping-Demand-Response-in-Europe-Today-2015.pdf. (Accessed 27 January 2020).

Spees, K., Pfeifenberger, J.P., Lam, L., 2020. Energy-Market Payment Options for Demand Response in Ontario.

Wang, F., Li, K., Liu, C., Mi, Z., Shafie-Khah, M., Catalao, J.P.S., 2018. Synchronous pattern matching principle-based residential demand response baseline estimation: mechanism analysis and approach description. IEEE Trans. Smart Grid 9 (6), 6972–6985. Nov.

Wei, W., Wang, D., Jia, H., Wang, C., Zhang, Y., Fan, M., 2017. Hierarchical and distributed demand response control strategy for thermostatically controlled appliances in smart grid. J. Mod. Power Syst. Clean Energy 5 (1), 30–42. Jan.

Wijaya, T.K., Vasirani, M., Aberer, K., 2014. When bias matters: an economic assessment of demand response baselines for residential customers. IEEE Trans. Smart Grid 5 (4), 1755–1763.

Willoughby, L., Bode, J., St, M., Francisco, S., 2013. 2012 San Diego Gas & Electric Peak Time Rebate Baseline Evaluation Prepared For : San Diego Gas & Electric Prepared By. The FSC Group Table of Contents.

Yang, C., Xu, Q., Wang, X., 2017. "Strategy of constructing virtual peaking unit by public buildings' central air conditioning loads for day-ahead power dispatching. J. Mod. Power Syst. Clean Energy 5 (2), 187–201. Mar.

Zarnikau, J., Thal, D., 2013. The response of large industrial energy consumers to four coincident peak (4CP) transmission charges in the Texas (ERCOT) market. Util. Pol. 26, 1–6. Sep.

FERC, 2011. Demand response compensation in organized wholesale energy markets (final rule), order 745. june-[Online]. Available: https://www.ferc.gov/legal/maj-ord-reg.asp. (Accessed 13 January 2020).

FERC, 2020. Participation of distributed energy resource aggregations in markets operated by regional transmission organizations and independent system operators, order 2222. sept-[Online]. Available: https://www.ferc.gov/sites/default/files/2020-09/E-1_0.pdf.