Online Demand Response Characterization Based on Variability in Customer Behavior

Lester Marrero, Daniel Sbárbaro, *Senior Member, IEEE*, and Luis García-Santander

Abstract—This paper proposes an online framework to characterize demand response (DR) over time. The proposed framework facilitates obtaining and updating the daily consumption patterns of customers. The essential concept of response profile class (RPC) is introduced for characterization and complemented by the measure of the variability in customer behavior. This paper uses a modified version of the incremental clustering by fast search and find of density peaks (CFSFDP) algorithm for daily profiles, considering the multivariate normal kernel density estimator and incremental forms of the Davies-Bouldin (iDB) and Xie-Beni (iXB) validity indices. Case studies conducted using real-world and simulated daily profiles of residential and commercial Chilean end-users have demonstrated how the proposed framework can continuously characterize DR. The proposed framework is proven to achieve realistic customer models for effective energy management by estimating the customer response to price signals at the distribution system operator (DSO) level.

Index Terms—Demand response (DR), incremental indice, online characterization, online clustering, response profile class (RPC).

NOMENCLATURE

α	Shrink factor
γ_i	Product of ρ_i and δ_i
Δt	Interval between two consecutive time points
δ_i	The minimum distance
λ_t	Electricity price at time point t
ρ_i	Local density
A	Auxiliary matrix $\left[\boldsymbol{p}_{k,l}, \boldsymbol{r}_{k,v}^0 \right]$
$\boldsymbol{c}_k^0, \boldsymbol{c}_k$	Initial and non-initial cluster centers of cluster k ($k = 1, 2, …, K$)
e_l	The minimum daily energy of customer l
$\mathbb{E}[\cdot]$	Expectation operator
$\boldsymbol{G}_k^0, \boldsymbol{G}_k$	Clusters k of daily profiles, $\left[\boldsymbol{c}_k^0, \boldsymbol{p}_{k,2}^0, …, \boldsymbol{p}_{k,N_k^0}^0 \right]$ and $\left[\boldsymbol{c}_k, \boldsymbol{p}_{k,2}, …, \boldsymbol{p}_{k,N_k} \right]$
H_l	Variability of customer l
h_t	Smoothing parameter of time point t
L	Set of customers
N_0	Total number of vectors of T-tuples
N_k	Number of daily profiles
N_r	Number of representative points
$\mathbb{P}(\cdot)$	Probability of cluster center followed up to current day
$\boldsymbol{P}_0$	Initial set of daily profiles $\left\{ \boldsymbol{p}_i^0, i = 1, 2, …, N_0 \right\}$
$\boldsymbol{P}$	Set of daily profiles $\left\{ \boldsymbol{p}_l \right\}$
$\boldsymbol{p}_i^0$	The ith initial active power profile $\left[p_{i,1}^0, p_{i,2}^0, …, p_{i,T}^0 \right]^{\text{T}}$
$\bar{\boldsymbol{p}}_l$	Expected active power profile of customer l $\left[\bar{p}_{l,1}, \bar{p}_{l,2}, …, \bar{p}_{l,T} \right]^{\text{T}}$
$\boldsymbol{p}_l$	Active power profile of customer l $\left[p_{l,1}, p_{l,2}, …, p_{l,T} \right]^{\text{T}}$
$p_{l,t}^{\min}, p_{l,t}^{\max}$	The minimum and maximum bounds for $\bar{p}_{l,t}$
$\bar{p}_{l,t}, \bar{q}_{l,t}$	Expected active and reactive power profiles of customer l at time point t
$\boldsymbol{R}_k^0, \boldsymbol{R}_k$	Sets of representative points of cluster k, $\left\{ \boldsymbol{r}_{k,u}^0, u = 1, 2, …, N_r \right\}$ and $\left\{ \boldsymbol{r}_{k,u}, u = 1, 2, …, N_r \right\}$
$r_{l,t}^{\text{d}}, r_{l,t}^{\text{u}}$	The maximum ramp-down and ramp-up rates for $\bar{p}_{l,t}$
$\boldsymbol{r}_{k,u}^0, \boldsymbol{r}_{k,u}$	Representative vector in $\boldsymbol{R}_k^0$ and $\boldsymbol{R}_k$
$s(\boldsymbol{G}_k^0), s(\boldsymbol{G}_k)$	Structure distances of $\boldsymbol{G}_k^0$ and $\boldsymbol{G}_k$
$\mathcal{S}_{l,t}$	Consumption region that contains active and reactive power values of customers at time point t
T	Set of time points

L. Marrero (corresponding author), D. Sbárbaro, and L. García-Santander are with the Department of Electrical Engineering, University of Concepción, Concepción 4070409, Chile (e-mail: lmarrero@udec. cl; dsbarbar@udec. cl; luigarci@udec.cl).

I. INTRODUCTION

THROUGH deep coordination among grid operators and active customers, the capability of facilities for demand

response (DR) and distributed energy resource management will be valuable for ancillary services [1]. In particular, the implementation of DR targets the control of the power-consuming behavior of customers to meet the following objectives: ① reduction of the peak power consumption; ② reduction of the total required power generation, as the main result of the prior objective; ③ change of the demand to follow the available supply, especially with high penetration of renewable energy sources (RESs); and ④ elimination of overloads in the distribution system [2].

The intrinsic socio-demographic characteristics (individual preferences) of customers and real-time externalities (environmental factors) can influence their response [3] in price-based DR programs. Typically, this information is private and unknown. However, understanding how customers respond conditioned by these influences is essential for estimating their potential for flexibility and designing correct pricing schemes to match the operation needs of the distribution system. To this end, a processing and subsequent characterization of daily load profiles are required first. These tasks represent a significant challenge since electricity data are data streams; thus, online clustering is necessary for handling this problem.

Although, in general, there are many studies in the literature about online (or stream) clustering foundations and algorithms (e.g., recent surveys [4]-[6]), few have been reported directly addressing this topic with application to electricity data, which is the focus of this investigation. The same observation is found in related studies [7] and [8].

However, several studies have recently analyzed the electricity consumption of customers by exploiting important offline clustering methods. For example, [9] classifies 2613 households under diverse load conditions such as calendar seasons and days. A clustering is applied in [10] to load patterns represented as images, and then periods with similar consumption levels are identified by considering load variation and uncertainty. A bilevel load shape dictionary is developed in [11], where extracted features such as weekly and seasonal patterns and segment entropy characterize the energy usage of customers. Consumption dynamics for each end-user are formulated first in [12]; then, from a clustering process, an evaluation of variability in the resulting clusters is performed by an entropy analysis. Reference [13] extracts, classifies, and verifies the reliability of the clustering and discovers clusters that describe end-users according to their demand and variability. Reference [14] uses an encoding system with a load-shape dictionary in 44 million daily profiles, focusing on segmenting customers' lifestyles. Finally, a distributed-centralized identification method is proposed in [15] to extract and characterize typical daily patterns for industrial customers.

Despite the valuable contributions of these studies, they lack developing an online characterization. Meanwhile, [16] presents an interim summarization function for the load data streams and an algorithm that incrementally learns and accumulates characterized patterns in six smart meters (SMs). Reference [17] implements the division of the load data streams in time windows, where objects in each window are assigned to different clusters (or concepts) whose structural change over time is analyzed, although they are not updated recursively. Similarly, an approach is presented in [8], which performs change detection and improves forecasting on aggregated time series within cluster by extracting interpretable features from the consumer data. An incremental algorithm is devised in [7] to detect pattern drifts through load pattern extraction, intergradation, and modification, but it works with only one customer simultaneously. Finally, [18] introduces an online adaptive clustering for load profiling. However, these studies mainly focus on analyzing the changes in consumption patterns rather than characterizing the customer behavior. The main advantage of this characterization for the distribution system operator (DSO) is obtaining online mathematical models to estimate the customer response according to consumption preferences and environmental factors. The DSO can then procure electrical energy with more certainty from the electricity pool.

This paper proposes an online framework to characterize the DR from the DSO perspective. The continuous processing of daily profiles makes it possible to know the response profile classes (RPCs) of customers and compute their variability. The framework comprises a modified version of the incremental clustering by fast search and find of density peaks (CFSFDP) algorithm [19] based on the seminal work in [20]. Previous experiences using the CFSFDP algorithm for load profiling in related studies [12], [15] support this selection. The incremental formulation adapts this original offline algorithm to work in an online setup. The proposed framework considers the multivariate normal kernel density estimator, which is robust to the number of objects in clusters during online processing, and incremental forms of the Davies-Bouldin (iDB) and Xie-Beni (iXB) validity indices [21] to provide the information about the algorithmic performance. The DSO can obtain the expected consumption of customers from their corresponding RPCs.

The customer response to the price signal has also been studied recently in the literature in the context of DR pricing. For example, in [22], the strategies for setting real-time prices are developed by implicitly learning the price elasticity of consumers online, although only own-price elasticity is in load changes. In [23], a model from the aggregator perspective runs a pricing program with distribution system constraints and learns the price sensitivities of customers. Reference [24] adopts a non-intrusive load monitoring-based pricing approach that estimates the DR potential of thermostatically controlled loads, and then models the price responsiveness of customers. While how to generate time-varying price signals is beyond the scope of this paper, the proposed framework provides the DSO with two powerful instruments to achieve the following goals: ① the updated customer models for the estimation of responses to different price signals; and ② the modeling of the underlying probability distribution of random deviations in demand from expected values, inherent to the stochastic behavior of customers. According to [25], demand uncertainty generally has a normal distribution; therefore, the proposed framework is suitable for finding the information about the statistical moments.

The main contributions of the paper are summarized below.

1) An innovative online framework is proposed to characterize the DR. The paper presents the modified incremental CFSFDP algorithm, defined to work in a Hilbert space. Online clustering introduces the multivariate normal kernel density estimator for robustness to the number of objects in clusters and the monitoring of algorithmic performance through the iDB and iXB validity indices.

2) Application of the proposed framework allows the DSO to perform two essential activities: updating the RPC and variability when customer response materializes (at the end of the previous day) and estimating the customer behavior to price signals based on a known RPC (within the current day).

3) The proposed framework is tested with real-world and simulated daily profiles. Results show the online process for obtaining RPCs of residential and commercial Chilean end-users. This paper also provides a comparison analysis with the online algorithm in [18] and a sensitivity analysis employing different combinations for the number of representatives and shrink factor.

The organization of this paper is as follows. Section II describes the proposed framework and provides theoretical foundations and mathematical models. Section III describes the solution methodology. Section IV presents two case studies with real-world and simulated daily profiles to verify the proposed framework. The performance monitoring and the comparison and sensitivity analysis are discussed in Section V. Finally, Section VI concludes this paper.

II. Proposed Framework

The DSO needs to ensure the reliability of distribution system, which may include small distributed solar generation units and generally has tight capacity constraints. By appropriately choosing dynamic price signals to be broadcasted to consumers enrolled in a price-based DR program, the DSO can reduce the distribution system costs and increase its reliability, for example, by shifting flexible consumption to the periods with high stochastic production [26].

This paper considers the price-setting DSO that aims at managing the demand flexibilities and pursues the estimation of customers' behavior to price signals. To this end, the DSO performs the daily processing of load profiles through the modified incremental CFSFDP algorithm to update the RPCs and variability of customers. Using the corresponding RPCs, the DSO can generate the expected consumption profiles in response to control price signals. Therefore, it can decide with high certainty the amount of electricity to trade, for example, in the balancing market within the day. The proposed framework, focused on residential and commercial customers, comprises both the previous day, at the end of which the DSO knows the power responses of customers and updates their RPCs and individual variability, and the current day, during which the DSO estimates the customers' response to a price signal based on known RPCs.

Let $\bar{s}_{l,t} = \bar{p}_{l,t} + j\bar{q}_{l,t}$ designate an expected complex power

value to be consumed at time point t by a customer l under a contract. $\bar{s}_{l,t} \in \mathcal{S}_{l,t}$, where $\mathcal{S}_{l,t}$ is positioned in the complex plane. It is possible to obtain a bounded and convex approximation $\mathcal{S}'_{l,t}$ of this region given practical bounds, both for active and reactive power. Although the estimation of $\bar{q}_{l,t}$ is fundamental for the analysis at the distribution system level, this paper focuses specifically on the active power responses of end-users. The following linear model for setting a flexible active power profile is defined [26]:

$$p_{l,t}^{\min} \leq \bar{p}_{l,t} \leq p_{l,t}^{\max} \quad l \in L, t \in T \tag{1}$$

$$-r_{l,t}^{d} \Delta t \leq \bar{p}_{l,t} - \bar{p}_{l,t-1} \leq r_{l,t}^{u} \Delta t \quad l \in L, t = 2, 3, ..., T \tag{2}$$

$$\sum_{t=1}^{T} \bar{p}_{l,t} \Delta t \geq e_l \quad l \in L \tag{3}$$

Formula (1) provides the expected response $\bar{p}_{l,t}$ between a minimum $p_{l,t}^{\min}$ and maximum $p_{l,t}^{\max}$ bound for customer l at time point t. Also, $\bar{p}_{l,t}$ can increase or decrease depending on the market price due to the combined use of shifting and shedding loads. Formula (2) forces ramp limits on the decrease and increase of active power in two successive time points. Finally, a minimum daily energy e_l is specified by (3) to account for basic activities.

An important observation to consider actual consumption features of consumers is that each region $\mathcal{S}'_{l,t}$ is time-varying since the bounds vary over time based on their preferences and environmental factors. This is exploited in this paper by developing an online processing and subsequent characterization of daily load profiles. From the corresponding outcome, it is attainable to differentiate the behaviors of consumers through RPCs, where each RPC represents a portion (of similar daily profiles or vectors) of the polytope that entirely contains the customer's load scenarios in the vector space. Therefore, based on these classes or portions, a more refined estimation of the consumption activity is feasible.

From a set of daily profiles associated with a RPC, each pair of parameters $p_{l,t}^{\min}$ and $p_{l,t}^{\max}$ of the model can be obtained as the corresponding extreme values, providing the convex (inner) approximation of the active power.

Concerning the values of the maximum ramp rates in (2), as they are related to the speed at which the consumer can decrease or increase its demand, they differ for each time point of the day and between RPCs. The strategy for its online determination is to consider the load changes from time point $t-1$ to t ($t = 2, 3, ..., T$) within the set of daily profiles of the RPC. Then, the following expressions are obtained:

$$r_{l,t}^{d} \Delta t = \max_{\{p_{l,t-1} - p_{l,t}\}} \left(p_{l,t-1} - p_{l,t} \right) \quad l \in L, t = 2, 3, ..., T \tag{4}$$

$$r_{l,t}^{u} \Delta t = \max_{\{p_{l,t} - p_{l,t-1}\}} \left(p_{l,t} - p_{l,t-1} \right) \quad l \in L, t = 2, 3, ..., T \tag{5}$$

Formulas (4) and (5) indicate a decrease and an increase in demand concerning the previous time point, respectively.

Lastly, the minimum daily energy can be the lowest total consumption among all the profiles within the RPC.

III. Solution Methodology

Daily processing of load profiles favors the appropriate

characterization of customer behavior. As consumers can be associated based on the similarity of their consumption patterns, to make this analysis scalable, this paper presents an incremental clustering, which is described in this section. Then, the section considers the performance monitoring and the estimation of the variability of customer response.

Algorithm 1 outlines the online workflow that runs at the end of each day when the customer response materializes.

Algorithm 1: online DR characterization

Input: set P, G_k^0, R_k^0, $k = 1, 2, ..., K$

1: **for** $t \leftarrow T$ **do**

2:    Assign each object p_l to the corresponding cluster by (14) and solve (15) to update the representative

3:    **for** each cluster G_k, **do**

4:        Compute the local density by (7) and the minimum distance by (8) and (9) for each object, and select new cluster centers through their products if they exist

5:        Assign each remaining object to the corresponding cluster and select the representatives by (10)-(13)

6:    **end for**

7:    **if** new clusters arise **then**

8:        Compute the minimum distance between all pairs of clusters by (16) and merge accordingly

9:        In the merged clusters, compute the local density by (7) and the minimum distance by (8) and (9) for each object, and select as center the object with the highest product value and the representatives by (10)-(13)

10:    **end if**

11:    Compute the iDB and iXB validity indices

12:    Update the RPCs and variability of customers by (17)

13: **end for**

A. Modified Incremental CFSFDP Algorithm

Let $P_0 = \{p_l^0\}$ denote an initial set of power profiles collected during τ days from L customers equipped with SMs, with each vector $p_l^0 = \left[p_{l,1}^0, p_{l,2}^0, ..., p_{l,\tau T}^0 \right]^T$. Let the initial power profile of each customer recast into τ daily load profiles. By gathering all these new profiles, the initial set is reformulated then with a total of $N_0 = \tau L$ vectors of T-tuples, i.e., $P_0 = \left\{ p_i^0, i = 1, 2, ..., N_0 \right\}$, with each vector $p_i^0 = \left[p_{i,1}^0, p_{i,2}^0, ..., p_{i,T}^0 \right]^T$. Furthermore, a set $P = \{p_l\}$ of load profiles is processed daily after this historical collection.

Any structure in the vector space depends on the similarity metric and the clustering criterion, which expresses how to use the metric. In the paper, the metric $d_2 : \mathbb{R}^T \times \mathbb{R}^T \rightarrow \mathbb{R}$, defined in (6) in terms of the l_2 norm for any pair of vectors p_i^0 and p_j^0, is employed. Hence, $\mathbb{R}^T$ is a normed linear vector space, particularly a Hilbert space, due to the induced norm [27].

$$d_2\left(p_i^0, p_j^0 \right) = \left\| p_i^0 - p_j^0 \right\|_2 \quad \forall p_i^0, p_j^0 \in P_0 \tag{6}$$

The implementation of the modified incremental CFSFDP algorithm involves the following stages.

1) Application of CFSFDP Algorithm

The CFSFDP algorithm uses the initial set P_0, where the clustering criterion relies on the computation of two magnitudes for each object p_i^0, i.e., the local density ρ_i and the mini-

mum distance δ_i concerning the vectors of higher density:

$$\rho_i = \frac{1}{N_0 h_1 h_2 ... h_T} \sum_{j=1}^{N_0} \left(\prod_{t=1}^{T} k \left(\frac{p_{i,t}^0 - p_{j,t}^0}{h_t} \right) \right) \quad i = 1, 2, ..., N_0 \tag{7}$$

$$\begin{cases} \delta_i = \min\limits_{p_j^0 \in P_0} d_2\left(p_i^0, p_j^0 \right) \\ \text{s.t. } \rho_j \geq \rho_i \quad i, j = 1, 2, ..., N_0 \end{cases} \tag{8}$$

Unlike [19] and [20] that use the cutoff distance of high sensitivity to a small number of objects, (7) denotes the general form of a multivariate product kernel estimator at point p_i^0 of the probability density function, in which the same (univariate) kernel function $k(\cdot)$ is used at each time point t with a different smoothing parameter h_t. Without loss of generality, let the normal kernel be selected. Then, according to Scott's rule [28] in $\mathbb{R}^T$, each smoothing parameter can be approximated as $h_t = \sigma_t N_0^{-1/(T+4)}$, with σ_t as the standard deviation for time point t.

For the object p_i^0 with the highest density, the distance is:

$$\delta_i = \max\limits_{p_j^0 \in P_0} d_2\left(p_i^0, p_j^0 \right) \tag{9}$$

Each cluster center c_k^0 ($k = 1, 2, ..., K$) describes a dominant consumption pattern and is expected to be surrounded by a neighborhood with lower local density and at a relatively large distance from any object with a higher local density. To identify the centers, both the plot of δ_i as a function of ρ_i for each object (the decision graph) and the plot of quantities $\gamma_1, \gamma_2, ..., \gamma_{N_0}$ in decreasing order (each $\gamma_i = \rho_i \delta_i$) can be used [20]. After detecting the centers, each remaining object is assigned to the same cluster as its nearest neighbor with higher density. Each cluster is denoted as $G_k^0 = \left[c_k^0, p_{k,2}^0, p_{k,3}^0, ..., p_{k,N_k^0}^0 \right]$, where N_k^0 is the number of daily profiles.

2) Determination of Representatives

Each cluster can be represented by a fixed number of representative points generated by selecting well-scattered points and then shrinking them toward the center by a specified fraction. This approach helps to identify the clusters with non-spherical shapes and wide variances in size [29].

Let N_r be the number of representative points of clusters and $R_k^0 = \left\{ r_{k,u}^0, u = 1, 2, ..., N_r \right\}$ be the set of these points to select for any cluster k. Their determination follows a sequential order. A selected vector $p_{k,i}^0$ becomes a representative vector $r_{k,u}^0$. The first representative is:

$$r_{k,1}^0 = \arg\max\limits_{p_{k,i}^0 \in G_k^0} d_2\left(p_{k,i}^0, c_k^0 \right) \tag{10}$$

And the rest of them are selected one by one as:

$$\begin{cases} r_{k,u}^0 = \arg\max\limits_{p_{k,i}^0 \in G_k^0 \setminus R_k^0} \left(\min\limits_{r_{k,u}^0 \in R_k^0} d_2\left(p_{k,i}^0, r_{k,u}^0 \right) \right) \\ \text{s.t. } d_2\left(p_{k,i}^0, r_{k,u}^0 \right) \geq e^{-s\left(G_k^0 \right)} \quad k = 1, 2, ..., K, u = 2, 3, ..., N_r \end{cases} \tag{11}$$

To select the second representative, $r_{k,u}^0$ is $r_{k,1}^0$; to select the third representative, $r_{k,u}^0$ could be $r_{k,1}^0$ or $r_{k,2}^0$, and so on. The structure distance $s\left(G_k^0 \right)$ is expressed in (12) based on

the mean $\mu_{k,t}$ and standard deviation $\sigma_{k,t}$ of time point t [19].

$$s\left(G_k^0\right) = \sqrt{\sum_{t=1}^{T}\left(\frac{\sigma_{k,t}}{\mu_{k,t}}\right)^2} \frac{1}{T}\sum_{t=1}^{T}\mu_{k,t} \quad k=1,2,\dots,K \quad (12)$$

The shrinking process of the representative points depends on the (user-defined) shrink factor $\alpha \in [0,1]$:

$$\boldsymbol{r}_{k,u}^0 = \boldsymbol{r}_{k,u}^0 + \alpha\left(\boldsymbol{c}_k^0 - \boldsymbol{r}_{k,u}^0\right) \quad k=1,2,\dots,K, u=1,2,\dots,N_r \quad (13)$$

Shrinking the scattered points toward the center undoes surface abnormalities and mitigates the effect of outliers since these are typically further away from the cluster center [29].

3) Assignment of New Load Profiles

With a new daily set $\boldsymbol{P}$, each object $\boldsymbol{p}_l$ is assigned to the cluster G_k^0 with the nearest representative. To this end, the local density ρ_l is set initially to be zero, and the minimum distance δ_l is given as:

$$\delta_l = \min_{\boldsymbol{r}_{k,u}^0 \in \boldsymbol{R}_k^0} d_2\left(\boldsymbol{p}_l, \boldsymbol{r}_{k,u}^0\right) \quad (14)$$

The assignment causes a change in the cluster structure, which requires updating the representative points. Assuming $\boldsymbol{r}_{k,u}^0$ as the nearest representative to the assigned vector $\boldsymbol{p}_{k,l}$, the solution of the problem below is found:

$$\begin{cases} \arg\max_{\boldsymbol{p}_{k,l},\boldsymbol{r}_{k,u}^0 \in \boldsymbol{A}} \left(\min_{\boldsymbol{r}_{k,u}^0 \in \boldsymbol{R}_k^0 \setminus \boldsymbol{r}_{k,u}^0} d_2\left(\boldsymbol{A}(1:T,c), \boldsymbol{r}_{k,u}^0\right)\right) \\ \text{s.t. } d_2\left(\boldsymbol{A}(1:T,c), \boldsymbol{r}_{k,u}^0\right) \geq e^{-s\left(G_k^0\right)} \quad k=1,2,\dots,K, l\in L, c=1,2 \end{cases} \quad (15)$$

where the auxiliary matrix is $\boldsymbol{A} = \left[\boldsymbol{p}_{k,l}, \boldsymbol{r}_{k,u}^0\right]$ with the column c.

From this result, if $\boldsymbol{p}_{k,l}$ produces the maximum value, $\boldsymbol{p}_{k,l}$ replaces $\boldsymbol{r}_{k,u}^0$ as the new representative [19]. Each cluster can be denoted as $G_k = \left[\boldsymbol{c}_k^0, \boldsymbol{p}_{k,2}, \boldsymbol{p}_{k,3}, \dots, \boldsymbol{p}_{k,N_k}\right]$.

4) Splitting Procedure

In this stage, the algorithm looks for the cluster with more than one dominant pattern. Specifically, in each new cluster G_k, parameters ρ_i and δ_i, and the quantity γ_i, are first obtained for each object $\boldsymbol{p}_{k,i}$. Based on these product values, new cluster centers can be identified, and the remaining objects can be assigned as in the CFSFDP algorithm [19]. Since only very few clusters arise in practice, an empirical criterion by computing the mean of the ten highest quantities can deliver new centers if these objects have a value higher than this.

5) Merging Procedure

This stage happens if new clusters arise, looking for those containing a similar dominant pattern. Specifically, the connected graph is constructed to find the components connected between clusters. The minimum distance between any two clusters G_k and G_m $(k,m=1,2,\dots,K)$ is computed as:

$$D\left(G_k, G_m\right) = \min_{\boldsymbol{r}_{k,u} \in \boldsymbol{R}_k, \boldsymbol{r}_{m,u} \in \boldsymbol{R}_m} d_2\left(\boldsymbol{r}_{k,u}, \boldsymbol{r}_{m,u}\right) \quad (16)$$

If $D\left(G_k, G_m\right) \leq s\left(G_k\right)$ and $D\left(G_k, G_m\right) \leq s\left(G_m\right)$, an edge is added between them. After adding all the edges, the graph with multiple components results, and the clusters with the same component can be merged [19].

B. Incremental Indices for Performance Monitoring

Lately, [21], [30], and [31] have extended several of the offline cluster validity indices to deal with the clustering over data streams. This paper implements the iDB and iXB validity indices to monitor the performance of the online algorithm.

For the computation of validity indices, the compactness term is essential. This paper considers the incremental formulation proposed in [21], which represents a counterpart hard for the calculation of the compactness, as considered in [30] (in the context of fuzzy clustering). This formulation is applied to clusters the that do not undergo splitting or merging, and for calculating both indices, the corresponding equations in [21] are employed. Instead, this paper applies the offline forms of both indices to the clusters that arise due to a split or merger. However, the compactness formulation in this paper is more straightforward than that described in [21] due to the following differences: ① the calculation is executed after the assignment of objects in $\boldsymbol{P}$ to clusters and not after the assignment of each; and ② the use of a cluster center $\boldsymbol{c}_k$ instead of the centroid, which changes with the added objects. In both indices, smaller values mean better solutions, whereas sudden changes indicate the changes in the cohesion and separation of clusters produced by the online algorithm [6].

C. Variability of Customer Response

To measure the uncertainty about the daily RPC for each customer, the entropy value [27] is used:

$$H_l = -\mathbb{E}\left[\lg \mathbb{P}\left(\boldsymbol{c}_k\right)\right] \quad l\in L, k=1,2,\dots,K \quad (17)$$

Equation (17) is solved after each daily observation of RPCs and gives the amount of information inherent to each, i.e., the variability in customer behavior.

IV. CASE STUDIES

This section presents two case studies to demonstrate the benefits of the proposed framework. The first case study involves residential and commercial electricity data recorded over six weeks (February 1 to March 13, 2020), with 15 min intervals. The Hilbert space is then $\mathbb{R}^{96}$. The number of customers is 925, charged with a regulated tariff; however, this paper assumes that they practice an optimizing behavior, using electricity in known off-peak periods. Therefore, Algorithm 1 is applied continuously when each day ends, which simulates the succession of days during this period. As a result, the DSO can update the RPCs and variability of customers at the end of the day. The second case study considers numerical simulations for an additional week, where the DSO broadcasts price signals daily. Based on each daily update of RPCs from Algorithm 1, this case study allows obtaining the expected load profiles of customers for the next day.

A. Case Study 1: Using Real-world Data Set

Incompleteness in electricity data is a common trend. Then, cleaning is executed by identifying and discarding daily profiles with missing and inconsistent values. The initial set $\boldsymbol{P}_0$ considers the first week of measurement data. Thus,

each set $\boldsymbol{P}$ of profiles corresponding to the rest of the days is processed incrementally. Likewise, the normalization of each daily profile concerning its maximum value is implemented to facilitate the clustering process.

The application of the CFSFDP algorithm to $\boldsymbol{P}_0$ allows the identification of four initial clusters through the decision graph and the plot of quantities $\gamma_1, \gamma_2, \ldots, \gamma_{N_0}$ in decreasing order. The results are depicted in Fig. 1 and the selected centers are highlighted with bigger blue dots. One of them is remarkably different, whereas the remaining three differ from the rest mainly due to the distance parameter, which means that these points do not have a neighborhood as high as the first.
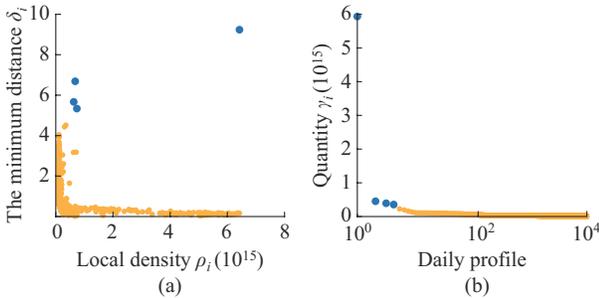
Fig. 1. Plots for identification of initial clusters. (a) Decision graph. (b) Quantities $\gamma_1, \gamma_2, \ldots, \gamma_{N_0}$ in decreasing order.

Representative points in clusters are the basis for assigning new profiles. The incremental clustering uses the following parameter values: $N_r = 8$ and $\alpha = 0.4$. To better observe the shrinking of representatives, Fig. 2 shows the shrinking process for the points that belong to the initial cluster 3, with the lowest cardinality, into a two-dimensional space using the principal component analysis (PCA) [27]. Representative points and the cluster center are highlighted in bigger black dots and a blue diamond, respectively. In $\mathbb{R}^{96}$, this comprises the translation of the points around the center.
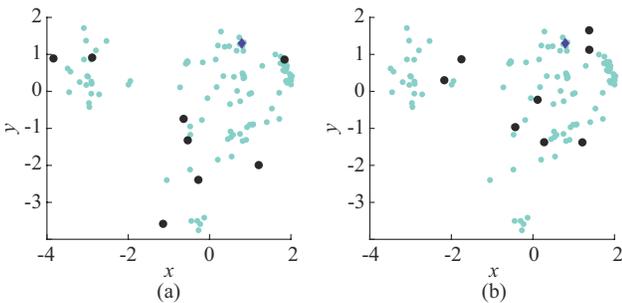

Fig. 2. Scatter data of cluster 3 for PCA. (a) Before shrinking process. (b) After shrinking process.

The online algorithm produces five final clusters considering the real-world data set. Figure 3 illustrates their evolution starting from the initial four. Following the trajectories, it is possible to identify the origin for each cluster, i.e., the one from which they arise, and the corresponding date. In the same way, for those clusters that fade, it is noticeable the one where they merge and the corresponding date. The colored dots in Fig. 3 indicate the sequence of days, which

helps to identify the dates of clusters that arise and fade. Black lines are used to represent the splits and mergers of the different-colored clusters. Cluster 1 is the only one that maintains from the beginning. Also, several clusters arise and fade in the last days of the period, which is attributed to a change in the consumption behavior of most customers since February is a vacation month and March is a month of more work activities. Finally, some isolated objects (with small density and great distance) are classified as outliers and discarded during online processing.
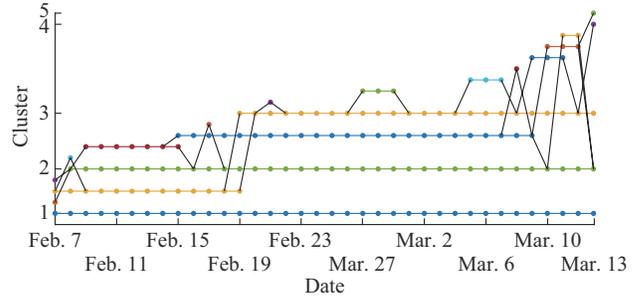

Fig. 3. Evolution of clusters for measurement period.

Since customers generally have well-defined behaviors (at least for a specific period) and these behaviors repeat between them, continuous splitting of clusters is uncommon. Also, a splitting rarely generates more than two clusters, which also happens with merging. Figure 3 confirms these observations, with some exceptions in the last days when splits and mergers have increased. Furthermore, since a cluster might contain at most two dominant patterns, the empirical criterion presented in the splitting procedure to split it directly (without analysis as in Fig. 1) is simple but effective. This includes the isolated objects which are treated as outliers.

Figure 4 depicts the response profiles for the final clusters with cluster centers in black in case study 1, and Table I gives their cardinality and daily average consumption.

The most increased percentage of cluster 2 (over 70%) concludes that this pattern is present in most residential end-users; however, it also includes commercial establishments and small businesses. Cluster 1 presents the highest consumption and more stable behavior. Cluster 3 shows a slightly lower consumption in the morning. Daily profiles with irregular behavior are much more noticeable in cluster 4, and cluster 5 has a typical residential pattern with low consumption.

The appropriate characterization of customer behavior is essential in a control-by-price strategy. Table II summarizes the RPCs of customers and their combinations, which is the main benefit of the investigation. Most consumers use four RPCs within the measurement period.

To complement this, the variability between the patterns in the combination is fundamental. For example, customers 99 and 630 (which are internal identifications for privacy-preserving) equally have the most common combination: 1-2-3-4; however, the probabilities on energy usage are very different, which are 0.047-0.905-0.024-0.024 and 0.286-0.238-0.286-0.19, respectively, and their corresponding entropy values are $H_{99} = 0.1796$ and $H_{630} = 0.5965$ (with the general mean of 0.2436), respectively.
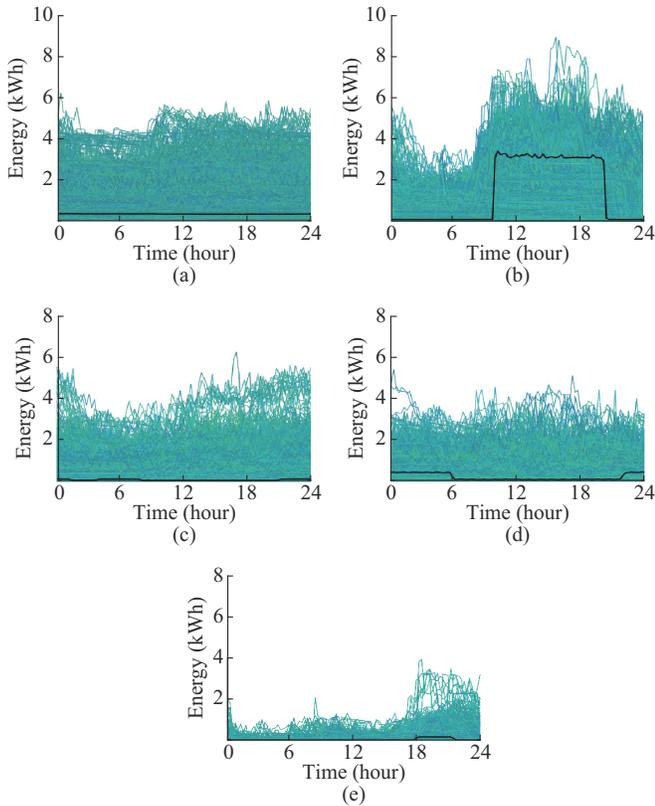
Fig. 4.   Daily consumption profiles of final clusters in case study 1. (a) Cluster 1. (b) Cluster 2. (c) Cluster 3. (d) Cluster 4. (e) Cluster 5.

TABLE I
CARDINALITY AND DAILY AVERAGE CONSUMPTION IN FINAL CLUSTERS IN CASE STUDY 1

Final cluster	Cardinality	Average consumption (kWh)
1	4153	50.5
2	26684	30.3
3	2643	37.4
4	3629	27.8
5	349	19.9

TABLE II
RPCs OF CUSTOMERS AND THEIR COMBINATIONS IN CASE STUDY 1

RPC	Number of customers	Combination of RPCs (number of customers of combination)
1	130	2 (122), 4 (4), 1 (4)
2	215	1-2 (85), 2-4 (100), 1-4 (2), 2-3 (8), 1-3 (6), 2-5 (12), 3-4 (2)
3	264	1-2-3 (58), 1-2-4 (69), 2-4-5 (32), 2-3-4 (91), 1-3-4 (6), 2-3-5 (4), 1-2-5 (4)
4	294	1-2-4-5 (14), 1-2-3-4 (248), 2-3-4-5 (29), 1-2-3-5 (3)
5	22	1-2-3-4-5 (22)

The lower value is because of the predominance of the second pattern and the very low probability of the rest, which implies less uncertainty about the daily RPC for customer 99. Figure 5 confirms this information based on the daily consumption profiles of both consumers for the mea-

surement period, where a more uncertain daily behavior characterizes customer 630.
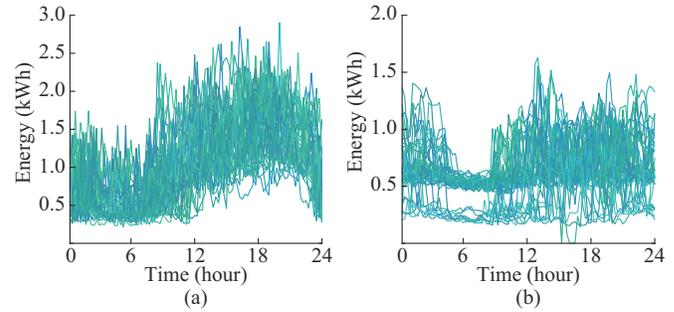

Fig. 5.   Daily consumption profiles for measurement period. (a) Customer 99. (b) Customer 630.

B. Case Study 2: Using Simulated Data Set

This case explores the effect of relaxing the regulatory condition to allow the price to fall in off-peak periods and to increase in peak periods.

For simplicity, the DSO broadcasts a single price signal to all consumers each day of the specified week (from March 14 to March 20, 2020); however, customized price signals can be designed according to customer behavior and broadcasted, e.g., each hour of the day to exploit newly available information of system states. Figure 6 depicts these price signals generated according to the demand of Chilean power system.
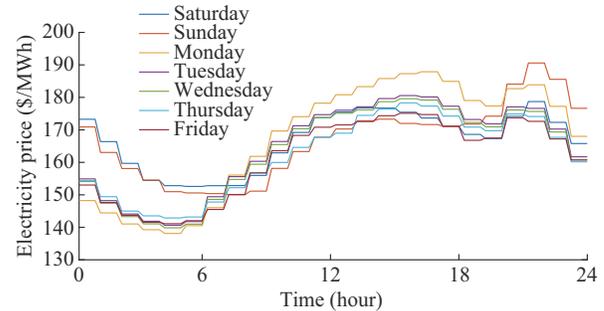

Fig. 6.   Price signals generated according to demand of Chilean power system.

The execution of numerical simulations considers the following ideas: ① a realistic situation is applied where consumers use the daily RPC with the highest probability for that day; ② to estimate active power responses to the price signal, customers practice a daily cost minimization; hence, this paper employs the linear programming problem in (18) [26]; and ③ to account for the presence of stochasticity in the responses, each random deviation from the expected active power value at time point t follows a normal distribution with mean zero and variance of 0.01 kWh^2.

$$\begin{cases} \min_{\bar{p}_l} \left(\sum_{t=1}^{T} \lambda_t \bar{p}_{l,t} \Delta t \right) \\ \text{s.t. (1)-(3)} \end{cases} \qquad (18)$$

For example, considering the same two customers on March 14, the RPC with the highest probability for this day results in RPC 2 for customer 99 and RPC 3 for customer

630. Figure 7 represents their corresponding mathematical models according to the RPCs, from which the optimizing behavior can be found. Each model involves the minimum and maximum bounds (yellow curves), which in general follow the shape of clusters 2 and 3 in Fig. 4, respectively, and the maximum ramp values given by products $r_{l,t}^{d}\Delta t$ and $r_{l,t}^{u}\Delta t$. Both consumers present the time points with a high predominance of one of the ramp rates, which implies a higher flexibility of that ramp rate. For customer 630, lower values around the 6th hour indicate less flexibility for this time. Finally, the minimum daily energy is 73.73 kWh for customer 99 and 34.45 kWh for customer 630.
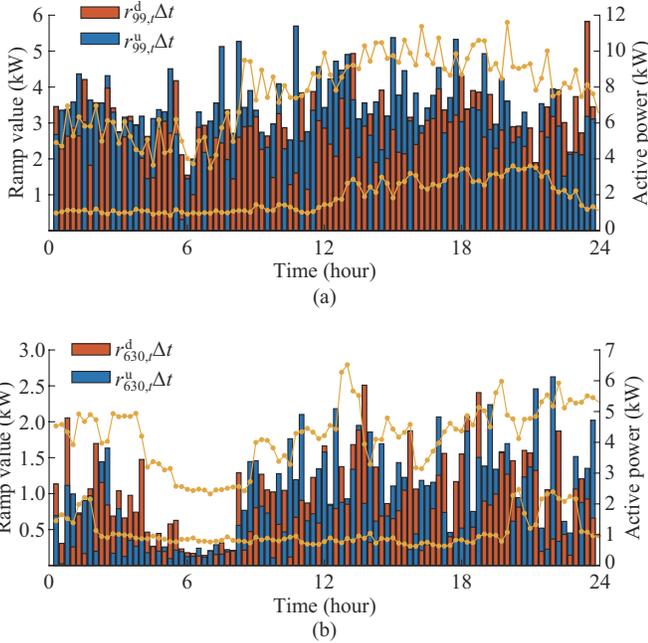

Fig. 7. Ramp values $r_{l,t}^{d}\Delta t$ and $r_{l,t}^{u}\Delta t$, and the minimum and maximum bounds. (a) Customer 99. (b) Customer 630.

After the simulation period, the final clusters increase to six. Figure 8 illustrates their response profiles, and Table III gives their cardinality and daily average consumption.

The main difference concerning the case with the real-world data set is the new arising cluster 5, related in shape to the previous cluster 2. Most of the new daily profiles under the effect of the price signals are added to this new cluster, which is also indicated in Table III. Considering the shape of the control signals with the lower prices in the early morning and the corresponding constraints on bounds and the maximum ramp rates during these hours, most of the new daily profiles present lower consumption to reduce the final cost. In addition, the algorithm assigns the lower consumption profiles of cluster 2 to this new cluster. Based on these factors, cluster 5 shows the lowest energy value, whereas for cluster 2, the cardinality decreases notably, and its average consumption increases. Finally, the same cluster centers hold from the case using the real-world data set.

Lastly, Table IV summarizes the RPCs and their combinations including the simulation period. Most consumers continue to use four RPCs, but the most common combination is now 1-2-3-4-5.
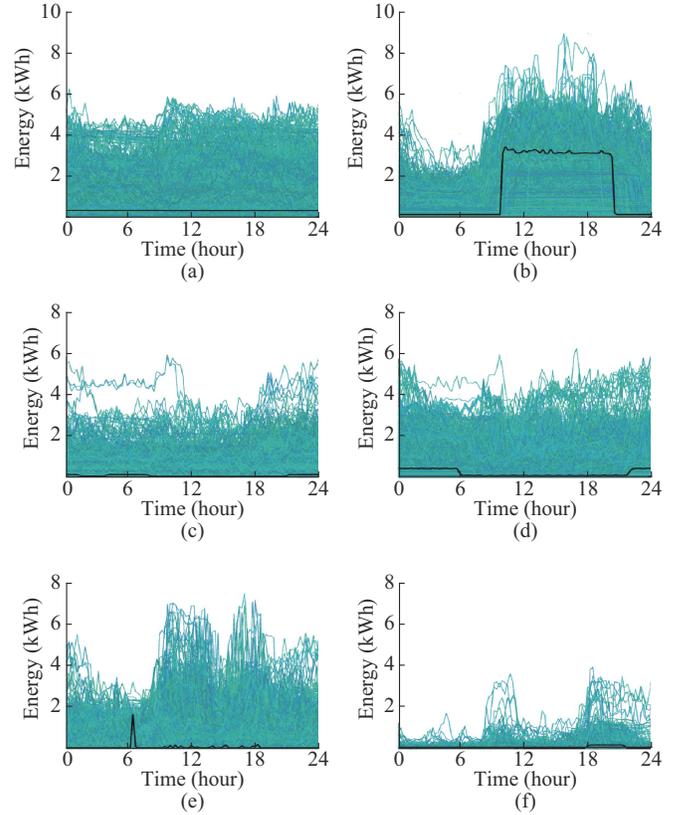

Fig. 8. Daily consumption profiles of final clusters in case study 2. (a) Cluster 1. (b) Cluster 2. (c) Cluster 3. (d) Cluster 4. (e) Cluster 5. (f) Cluster 6.

TABLE III
CARDINALITY AND DAILY AVERAGE CONSUMPTION IN FINAL CLUSTERS IN
CASE STUDY 2

Final cluster	Cardinality (ratio of new daily profiles (p.u.))	Average consumption (kWh)
1	4227 (0.01)	51.7
2	12003 (0.05)	45.4
3	2467 (0.03)	34.2
4	6041 (0.32)	31.1
5	18960 (0.58)	17.5
6	190 (0.01)	21.0

TABLE IV
RPCS OF CUSTOMERS AND THEIR COMBINATIONS IN CASE STUDY 2

RPC	Number of customers	Combination of RPCs (number of customers of combination)
1	9	2 (5), 5 (3), 4 (1)
2	67	4-5 (34), 2-5 (27), 3-4 (2), 1-2 (1), 1-4 (1), 1-5 (1), 3-5 (1)
3	242	1-4-5 (6), 2-4-5 (172), 3-4-5 (29), 1-2-4 (7), 1-3-4 (11), 2-5-6 (7), 1-2-5 (10)
4	305	1-2-4-5 (135), 1-3-4-5 (40), 2-3-4-5 (85), 2-4-5-6 (24), 1-2-3-5 (5), 1-2-3-4 (12), 1-4-5-6 (3), 1-2-5-6 (1)
5	282	1-2-4-5-6 (17), 1-2-3-4-5 (252), 1-3-4-5-6 (3), 2-3-4-5-6 (10)
6	20	1-2-3-4-5-6 (20)

V. PERFORMANCE MONITORING AND COMPARISON AND SENSITIVITY ANALYSIS

Considering the real-world data set, this section first analyzes the monitoring of the clustering algorithm based on the iDB and iXB validity indices. The comparison with the online algorithm in [18] and the sensitivity analysis employing the number of representatives and the shrinking factor are also investigated.

A. Performance Monitoring of Online Clustering

To assess the cohesion and separation of clusters produced in the online processing, Fig. 9 depicts the iDB and iXB validity indices. A high correlation characterizes the measurement period, and the resulting trend of the values exhibits good performance for the algorithm based on an adequate assignment of objects.
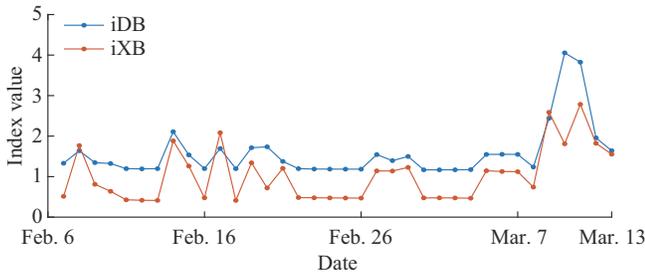

Fig. 9.   Evolution of iDB and iXB validity indices.

On the days when splitting or merging events happen (as shown in Fig. 3), the indices almost always show a slight increase or decrease, respectively. Conversely, both indices remain constant when clusters hold from one day to the next. Then, the usefulness of these incremental forms for monitoring the performance of the online algorithm is proved, as in other correctly and poorly partitioned data sets [21]. The higher values in the last few days are due to the change in daily patterns attributed to most consumers.

B. Comparison Analysis

This paper uses the algorithm in [18] for comparison purposes since it also processes daily profiles. Concerning the proposed framework, this algorithm has two main differences: ① it uses centroids, which is not always valid in practice since clusters ideally need to have a spherical structure in the vector space; and ② it discards past daily profiles and preserves this information only through a distance matrix and the centroids themselves, which are updated daily.

The algorithm is applied using the l_2 norm and the first week for the consensus clustering [18], in which the k-means method uses the set $\{2, 3, …, 8\}$ for the number of clusters per instance. Table V gives the result of two cases with different parameter values for the facility cost C^f that decides the split and the minimum distance between centroids d_{min} that favors the merger, where D_p is the mean of the probability distances. The rest of the parameter values comprise the initial number of clusters K^0 obtained from the corresponding dendrogram, the exponential forgetting v, and the number of disruptive loads γ_{min}, which remain as in the application with electricity data in [18]. Specifically, different parameter values in the second case produce many more splits and mergers.

TABLE V
PARAMETER VALUES AND FINAL CLUSTERS

Case	Parameter value					Final cluster
	K^0	v	γ_{min}	C^f	d_{min}	
1	7	0.85	5	$6D_p$	$0.5D_p$	6
2	7	0.85	5	$5D_p$	$0.8D_p$	8

Figure 10 illustrates the evolution of the iDB and iXB validity indices in these two case studies with different parameter values using the algorithm in [18], where high-magnitude spikes that suggest a poorer assignment of daily profiles are present in both.
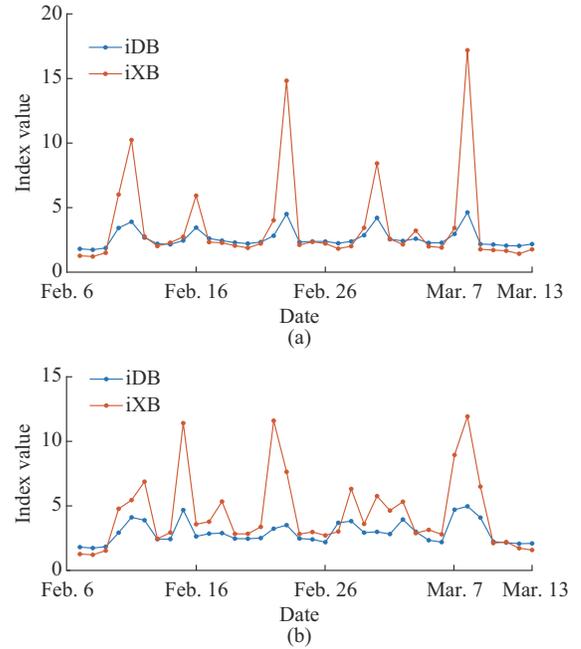

Fig. 10.   Evolution of iDB and iXB validity indices in two case studies. (a) Case study 1. (b) Case study 2.

C. Sensitivity Analysis

The sensitivity analysis evaluates the influence of changes on the number of representatives N_r and the shrinking factor α. The solutions for a group of selected values of these parameters are given, as shown in Table VI.

Considering N_r, the trend is to obtain fewer clusters as this number increases. With one representative (the cluster center itself), the result of nine final clusters is obtained, which is close to the eight final clusters of case study 2 using the algorithm in [18] (which employs the centroid). With $N_r = 4$, a high difference between the maximum and minimum values is obtained for the iDB index. Generally, it is hard to establish the final number of these scattered points in the vector space. With a relatively small number, a large cluster can split incorrectly, but with a large number, two close clusters are more likely to merge. According to the values of the indices in Table VI with the real-world data set, using $N_r \geq 6$ produces adequate results.

Considering α, it is not practical to take very low or very high values [29]. In addition to $\alpha = 1$, Table VI presents the results with $\alpha = 0.3$ and $\alpha = 0.5$. Similarly, with a smaller val-

ue of α, the scattered points shrink little, and merging between two clusters is more likely to occur, whereas larger values cause these points to move more toward the cluster center and favor the splitting.

TABLE VI
FINAL CLUSTERS, MEAN, STANDARD DEVIATION, AND DIFFERENCE BETWEEN THE MAXIMUM AND MINIMUM iDB AND iXB VALIDITY INDICES FOR DIFFERENT VALUES OF N_r AND α

N_r	α	Final cluster	Mean, standard deviation, and difference between the maximum and minimum iDB validity indices	Mean, standard deviation, and difference between the maximum and minimum iXB validity indices
8	0.4	5	1.57, 0.65, 2.89	1.02, 0.65, 2.37
4	0.4	6	2.92, 5.81, 29.68	1.52, 0.83, 3.27
6	0.4	5	1.47, 0.30, 1.04	1.20, 0.67, 2.23
10	0.4	3	1.27, 0.20, 0.88	0.61, 0.31, 1.36
12	0.4	3	1.19, 0.16, 0.76	0.60, 0.35, 1.34
8	0.3	3	1.41, 0.52, 3.23	1.14, 0.78, 3.50
8	0.5	5	1.28, 0.21, 0.96	0.75, 0.42, 1.52
1	1.0	9	1.86, 0.19, 0.76	1.34, 0.25, 1.22

VI. CONCLUSION

The main idea of this paper is to propose an online framework for DR characterization. In particular, the DSO can use the proposed framework to obtain and update daily the RPCs and variability of customers, and estimate the customer response to a price signal based on a known RPC, which is suitable for effective energy management on the demand side.

Furthermore, the underlying probability distribution of random deviations in demand from expected values is generally unknown. However, the proposed framework contributes to modeling it since each daily deviation can be considered a realization of the corresponding random variable. The use of parameters of this empirical distribution overcomes the limitation of making distributional assumptions that can result in risky or more conservative costly solutions.

Some technical and practical complications arise with each set of load profiles and the higher amount that needs to be processed daily. In particular: ① more computational effort is demanded; ② as clusters extend in the vector space, a larger N_r might eventually be more appropriate for capturing their geometry; and ③ the time to find the information about the RPCs of customers and estimate their responses could be longer than that available by the DSO. Thus, a suitable strategy to be included in the online algorithm is the well-known sliding window [4], according to which the number of daily profiles remains practically constant, and the information of interest is present. Algorithm 1 favors this inclusion.

While processing and analyzing load data streams represent a significant challenge, case studies using real-world and simulated daily profiles of Chilean end-users have demonstrated the applicability of the proposed framework. According to the results, most consumers use four RPCs within the measurement period. Furthermore, the behavior of iDB and iXB validity indices and the comparison analysis have verified the adequate assignment of objects of the online

clustering. Future work needs to address the following issues in more depth: ① the impact of dimensionality; ② the definition of the most reliable values for parameters N_r and α; and ③ the correct RPC for estimating the expected response of customers (not necessarily the RPC with the highest probability).

REFERENCES

[1] K. Schmitt, R. Bhatta, M. Chamana *et al.*, "A review on active customers participation in smart grids," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 1, pp. 3-16, Jun. 2023.

[2] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, "A survey on demand response programs in smart grids: pricing methods and optimization algorithms," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 152-178, Jan.-Mar. 2015.

[3] X. Chen, Y. Li, J. Shimada *et al.*, "Online learning and distributed control for residential demand response," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 4843-4853, Nov. 2021.

[4] J. A. Silva, E. R. Faria, R. C. Barros *et al.*, "Data stream clustering: a survey," *ACM Computing Surveys*, vol. 46, no. 1, pp. 1-31, Oct. 2013.

[5] M. Carnein and H. Trautmann, "Optimizing data stream representation: an extensive survey on stream clustering algorithms," *Business & Information Systems Engineering*, vol. 61, pp. 277-297, Jun. 2019.

[6] J. C. Bezdek and J. M. Keller, "Streaming data analysis: clustering or classification?" *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 91-102, Jan. 2021.

[7] Z. Jiang, R. Lin, and F. Yang, "An incremental clustering algorithm with pattern drift detection for IoT-enabled smart grid system," *Sensors*, vol. 21, no. 19, p. 6466, Oct. 2021.

[8] P. Laurinec and M. Lucká, "Interpretable multiple data streams clustering with clipped streams representation for the improvement of electricity consumption forecasting," *Data Mining and Knowledge Discovery*, vol. 33, pp. 413-445, Mar. 2019.

[9] M. Sun, I. Konstantelos, and G. Strbac, "C-vine copula mixture model for clustering of residential electrical load pattern data," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2382-2393, May 2017.

[10] M. Charwand, M. Gitizadeh, P. Siano *et al.*, "Clustering of electrical load patterns and time periods using uncertainty-based multi-level amplitude thresholding," *International Journal of Electrical Power & Energy Systems*, vol. 117, p. 105624, May 2020.

[11] H. Liang and J. Ma, "Develop load shape dictionary through efficient clustering based on elastic dissimilarity measure," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 442-452, Jan. 2021.

[12] Y. Wang, Q. Chen, C. Kang *et al.*, "Clustering of electricity consumption behavior dynamics toward big data applications," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2437-2447, Sept. 2016.

[13] S. Haben, C. Singleton, and P. Grindrod, "Analysis and clustering of residential customers energy behavioral demand using smart meter data," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 136-144, Jan. 2016.

[14] J. Kwac, J. Flora, and R. Rajagopal, "Lifestyle segmentation based on energy consumption data," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2409-2418, Jul. 2018.

[15] F. Lu, X. Cui, J. Xing *et al.*, "Electricity load profile characterisation for industrial users based on normal cloud model and iCFSFDP algorithm," *IEEE Transactions on Power Systems*, vol. 38, no. 4, pp. 3799-3813, Jul. 2023.

[16] D. de Silva, X. Yu, D. Alahakoon *et al.*, "A data mining framework for electricity consumption analysis from meter data," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 399-407, Aug. 2011.

[17] M. A. Masud, J. Huang, M. Zhong *et al.*, "Cluster survival model of concept drift in load profile data," *IEEE Access*, vol. 6, pp. 51269-51285, Sept. 2018.

[18] G. L. Ray and P. Pinson, "Online adaptive clustering algorithm for load profiling," *Sustainable Energy, Grids and Networks*, vol. 17, p. 100181, Mar. 2019.

[19] L. Zhao, Z. Chen, Y. Yang *et al.*, "ICFS clustering with multiple representatives for large data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 728-738, Mar. 2019.

[20] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492-1496, Jun. 2014.

[21] L. E. B. da Silva, N. M. Melton, and D. C. Wunsch, "Incremental cluster validity indices for online learning of hard partitions: extensions and comparative study," *IEEE Access*, vol. 8, pp. 22025-22047,

Jan. 2020.

[22] S.-J. Kim and G. B. Giannakis, "An online convex optimization approach to real-time energy pricing for demand response," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2784-2793, Nov. 2017.

[23] N. Tucker, A. Moradipari, and M. Alizadeh, "Constrained Thompson sampling for real-time electricity pricing with grid reliability constraints," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 4971-4983, Nov. 2020.

[24] Y. Tao, J. Qiu, S. Lai *et al.*, "Customer-centered pricing strategy based on privacy-preserving load disaggregation," *IEEE Transactions on Smart Grid*, vol. 14, no. 5, pp. 3401-3412, Sept. 2023.

[25] L. Roald, F. Oldewurtel, B. Van Parys *et al.* (2015, Aug.). Security constrained optimal power flow with distributionally robust chance constraints. [Online]. Available: https://arxiv.org/abs/1508.06061

[26] J. M. Morales, A. J. Conejo, H. Madsen *et al.*, *Integrating Renewables in Electricity Markets: Operational Problems*. New York: Springer, 2014.

[27] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River: Prentice Hall, 2000.

[28] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, 2nd ed. Hoboken: John Wiley & Sons, 2015.

[29] S. Guha, R. Rastogi, and K. Shim, "Cure: an efficient clustering algorithm for large databases," *Information Systems*, vol. 26, no. 1, pp. 35-58, Mar. 2001.

[30] M. Moshtaghi, J. C. Bezdek, S. M. Erfani *et al.*, "Online cluster validity indices for performance monitoring of streaming data clustering," *International Journal of Intelligent Systems*, vol. 34, no. 4, pp. 541-563, Apr. 2019.

[31] O. A. Ibrahim, J. M. Keller, and J. C. Bezdek, "Evaluating evolving structure in streaming data with modified Dunn's indices," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 2, pp. 262-273, Apr. 2021.

Lester Marrero received the B.Sc. and M.Sc. degrees in electrical engineering from Universidad Central de Las Villas, Santa Clara, Cuba. He is currently pursuing the Ph.D. degree in electrical engineering at Universidad de Concepción, Concepción, Chile. His research interests include modeling and optimization of electric distribution systems, distributed energy resources, and data analytics in smart grids.

Daniel Sbárbaro received the electrical engineering degree from Universidad de Concepción, Concepción, Chile, in 1984, and the Ph.D. degree in electrical engineering from Glasgow University, Glasgow, U.K., in 1993. In 1998, he was an Alexander von Humboldt Fellow with the Control Engineering Laboratory, Ruhr University, Bochum, Germany. He also held Visiting Research positions with the Daimler-Benz Research Institute, Berlin, Germany, and Stuttgart University, Stuttgart, Germany, in 1993 and 2009, respectively. He is currently a Professor with the Department of Electrical Engineering, Universidad de Concepción. His current research interests include modeling and control of energy systems and development of technologies for improving sustainability of main Chilean productive sectors.

Luis García-Santander received the Ph.D. degree in electrical engineering from Supélec, Paris, France, in 2003. Since 1996, he has been with the Department of Electrical Engineering, Universidad de Concepción, Concepción, Chile, where he is currently an Associate Professor. His research interests include renewable generation, planning and operation of distribution systems, energy management and efficiency, smart grid, smart metering, and electromobility.