

## Targeted demand response for flexible energy communities using clustering techniques



Sotiris Pelekis<sup>a,\*</sup>, Angelos Pipergias<sup>a</sup>, Evangelos Karakolis<sup>a</sup>, Spiros Mouzakitis<sup>a</sup>,  
Francesca Santori<sup>b</sup>, Mohammad Ghoreishi<sup>b</sup>, Dimitris Askounis<sup>a</sup>

<sup>a</sup> Decision Support Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Greece

<sup>b</sup> ASM Terni S.p.A., Italy

### ARTICLE INFO

#### Article history:

Received 30 April 2023

Received in revised form 24 July 2023

Accepted 29 July 2023

Available online 12 August 2023

#### Keywords:

Clustering

Demand response

Energy community

Flexibility

k-means

Load profile

Machine learning

Peak performance score

Reverse power flow

### ABSTRACT

The present study proposes clustering techniques for designing demand response (DR) programs targeting commercial and residential prosumers. The goal is to alter the consumption behavior of the prosumers within a distributed energy community in Italy. This aggregation aims to: (a) minimize the reverse power flow at the primary substation, occurring when generation from solar panels in the local grid exceeds consumption, and (b) shift the system wide peak demand, that typically occurs during late afternoon. Regarding the clustering stage, we consider daily prosumer load profiles and divide them across the extracted clusters. Three popular machine learning algorithms are employed, namely k-means, k-medoids and agglomerative clustering. We evaluate the methods using multiple metrics including a novel metric proposed within this study, namely peak performance score (PPS). The k-means algorithm with dynamic time warping distance considering 14 clusters exhibits the highest performance with a PPS of 0.689. Subsequently, we analyze each extracted cluster with respect to load shape, entropy, and load types. These characteristics are used to distinguish the clusters that have the potential to serve the optimization objectives by matching them to proper DR schemes including time of use, critical peak pricing, and real-time pricing. Our results confirm the effectiveness of the proposed clustering algorithm in generating meaningful flexibility clusters, while the derived DR pricing policy encourages consumption during off-peak hours. The developed methodology is robust to the low availability and quality of training datasets and can be used by aggregator companies for segmenting energy communities and developing personalized DR policies.

© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Climate change alongside the growing population, the increase in electrical energy consumption and the depletion of resources have led to a large-scale deployment of renewable energy sources (RES), increasing their energy share worldwide [1]. Numerous RES technologies have significantly progressed in technical and economic maturity over the past few decades. Yet, their fluctuating and intermittent nature raised concerns related to the balancing and capacity adequacy of an energy supply configuration relying mostly on RES [2] as it usually leads to a gap between the energy produced and the energy demanded during specific hours of the day (peak hours). To this end, battery energy storage systems

(BESS) can be utilized in order to store the surplus of produced energy to make it available during the peak hours [3]. However, it is worth noting that this approach may actually increase energy usage [4]. Furthermore, serving the demand during peak hours may lead to blackouts.

The necessity of increasing the flexibility of the existing bulk system has led researchers to investigate new methodologies for demand side management (DSM) to fully exploit the production of RES. In this direction, demand response (DR), a tool for the management of peak demand, and the balance of generation and consumption in the electrical grid, has emerged recently. However, the implementation of DR programs is still quite limited, especially when it comes to residential consumers. As the cost of equipment that can help homes and businesses participate in DR (i.e. smart meters, controllers and devices) decreases, there is a growing need for the design and implementation of programs to effectively motivate the consumers that are involved in DR. To this end, the I-ENERGY project [5] aims at developing innovative artificial intelligence (AI) services [6,7] for the energy sector by providing better energy forecasts [8,9], flexibility and anomaly

\* Corresponding author.

E-mail addresses: [spelekis@epu.ntua.gr](mailto:spelekis@epu.ntua.gr) (S. Pelekis), [aggelospip@gmail.com](mailto:aggelospip@gmail.com) (A. Pipergias), [vkarakolis@epu.ntua.gr](mailto:vkarakolis@epu.ntua.gr) (E. Karakolis), [smouzakitidis@epu.ntua.gr](mailto:smouzakitidis@epu.ntua.gr) (S. Mouzakitis), [francesca.santori@asmterni.it](mailto:francesca.santori@asmterni.it) (F. Santori), [mohammad.ghoreishi@asmterni.it](mailto:mohammad.ghoreishi@asmterni.it) (M. Ghoreishi), [askous@epu.ntua.gr](mailto:askous@epu.ntua.gr) (D. Askounis).

## Nomenclature

BESS	Battery energy storage system
CPP	Critical peak pricing
DBI	Davies–Bouldin validity index
DLC	Direct load control
DR	Demand response
DSM	Demand side management
DTW	Dynamic time warping
EMC	Energy management controllers
EPES	Electrical power and energy systems
PMS	Peak match score
PPS	Peak performance score
PV	Photovoltaic
RES	Renewable energy sources
RTP	Real-time pricing
SOM	Self-organizing maps
TOU	Time of use
TSO	Transmission system operator

detection services [10], amongst others. These services cover the entire energy value chain.

In this study, we aim to segment a flexible energy community of heterogeneous prosumers pertaining to a power distribution grid fraction in central Italy [11,12]. The municipal enterprise, that owns and manages the electricity distribution network, aspires to act as an aggregator and identify prosumers suitable for DR, therefore aiming to mitigate two shortcomings related to the active power load profiles within the energy community: (i) the problem of reverse power flow, where the total energy production in the community is greater than the total consumption; this occurs in the early afternoon when production from photovoltaic (PV) installations peaks and leads to an injection of load from the network managed by the aggregator to the network that supplies it, which is undesirable for the aggregator; (ii) the reduction in electricity demand during peak hours in the late afternoon, through peak shaving and/or peak shifting techniques. To this end, we perform a cluster analysis, following proper data pre-processing, on the daily load/generation profiles as recorded by smart meters measuring the active power of participant prosumers. In this direction, different clustering algorithms (k-means, k-medoids, agglomerative clustering) are compared. Subsequently, we analyze the extracted clusters aiming to distinguish those that can contribute to the above mentioned optimization goals through targeted, general-purpose DR pricing policies proposed to their pertaining prosumers. Note here that the datasets have been collected following a centralized approach by the utility. In this context, to address any data privacy issues, the informed consent of the participants has been acquired and, subsequently, the datasets have been anonymized by removing any personally identifiable information (PII), as suggested by the General Data Protection Regulation [13]. Therefore, only smart meter IDs have been kept as the main identifier of each prosumer. Additionally, to ensure the confidentiality and integrity of our smart meter measurements, the database for storing the relevant datasets has been secured by an identity and access management mechanism based on Keycloak [14]. The present study has also been conducted under the context of I-ENERGY project and is targeted to the electrical power and energy systems (EPES) domain and specifically transmission system operators (TSO), suppliers or aggregators that aim to organize, and segment flexible energy communities within DR settings.

The rest of the paper is organized as follows. The rest of Section 1 presents the current status quo in energy flexibility, the electricity market, and existing demand response schemes, alongside several machine learning-driven clustering techniques applied for flexibility assessment and demand response. Finally, the main contributions of our study are summarized. Section 2 describes our methodological approach, including the data-preprocessing, the daily load profile clustering methods, and the formalization of the peak performance score metric (PPS). Section 3 presents the results of the clustering process while also discussing the personalized DR schemes based on the derived prosumer clusters. Lastly, Sections 4 and 5 wrap up the paper, presenting concluding remarks and future perspectives respectively.

### 1.1. Energy flexibility

Energy flexibility describes the amount of energy consumption that can be shifted or changed in time. Most often, assessing and forecasting grid flexibility aims to reduce fluctuations in demand (and generation) of electricity, i.e. the smoothing of the demand curve and shaving load peaks. Flexibility is also the key to achieve maximum RES penetration in the power grid. However, in terms of formalization, the landscape of energy flexibility is quite fuzzy as different studies come up with varying definitions. In practice, energy flexibility is often defined in conjunction with other quantities that denote the motive for providing it, such as energy prices [15]. In this context, for the needs of our study, energy flexibility can be intuitively perceived as  $-\frac{\% \Delta E}{\% \Delta p}$ , where E the energy demand, p the energy price and  $\Delta$  the delta operator. This formalization implies that the larger the motive (price increase/decrease) required for a certain variation (decrease/increase) of the energy consumption, the lower the flexibility of the specific consumer and vice-versa. Note here that the minus sign denotes that an increase in price leads to a decrease in consumption and vice-versa.

From a demand side perspective, the use of flexibility within a power grid is achieved through DR. As the term refers specifically to adjustments in the electricity demand by end-users (industrial, commercial or domestic), DR can be considered as one of the main DSM mechanisms. In this context, prosumers commit to change their conventional consumption patterns by using temporary on-site power generation or reducing/shifting electricity consumption away from periods of low RES generation and/or high demand. This can be achieved by responding to signals from the network operator or electricity provider [16].

### 1.2. Market and stakeholders overview

The electricity market is divided into the energy market, the capacity market, and the ancillary services market, which are designed to provide financial incentives to the various stakeholders to contribute to energy supply and normal grid operation. DR is mainly related to energy and ancillary services trading [17–19]. Depending on the country, contracts between interested parties in the market can take place through bilateral (off market) transactions or through an established market (exchanges, auction with clearing prices). An overview of the market and its stakeholders will contribute to the familiarization of the reader with the underlying concepts of flexibility market and DR services:

- TSOs are market facilitators who ensure that every transaction meets the network constraints. TSOs can usually buy or sell products in all electricity markets.

- Suppliers participate in the market and ensure that the amount of energy committed to the market will balance the consumption of their end users in their portfolio. They can offer their customers specific contracts as flat rates or DR schemes. When proposing DR schemes, the challenge for suppliers is to evaluate how these programs will affect the consumption of their portfolio.
- End-users buy electricity from suppliers. When signing up for a DR program, they can either respond to a request or price manually, or through an in-house power management system. In this context, AI methods are significantly useful for facing the challenges of automatic consumer response.
- Producers generate electricity and offer their production at a certain price in the markets. Their products can be either energy services and/or frequency response (ancillary) services.
- Aggregators bring together end customers or small producers in order to achieve the minimum capacity that allows to provide energy flexibility products and ancillary services. Therefore, they have direct contracts with end-users offering their flexibility to suppliers or grid operators. With regard to suppliers, it must be ensured that end customers are committed to the flexibility they negotiate in the wholesale market.

Once a supplier commits to deliver a certain amount of power to the grid, compliance is expected, otherwise, there is a penalty. Thus, it is very important for aggregators to ensure that end-users can provide the flexibility they have been engaged to [20]. In this context, the criteria, and therefore the implementation of DR schemes, may be aimed at smoothing the electricity demand curve, maximizing RES penetration, minimizing greenhouse gas emissions, maximizing electricity exports, minimizing electricity imports etc. All the above should always be in compliance with the requirements of network stability and the respective regulations for reserves, taking into account the multitude of network stability levels (frequency, voltage, spinning and not spinning reserve, post-black out reserve etc.). In general, DR schemes can range from hour to seconds scale in terms of time resolution.

In markets and stock exchanges, where the equilibrium price results from tenders and auctions (bidding), the place of a DR provider is highly sensitive as an overestimation of power can lead to inability to cover and possible fine while an underestimation can lead to profit losses [21]. Indicatively, it has been estimated that the proper implementation of DR schemes can reduce the load peak in Europe by 10% [22].

### 1.3. Demand response programs

There are two main types of demand response programs: price or time-based programs and incentive-based programs [23] as illustrated in Fig. 1. In price-based DR, the price of electricity changes throughout the day and consumers benefit from consuming more electricity during low-price periods and less during high-price periods. In incentive-based programs, the price of electricity is constant, but the program administrators reward consumers who manage to consume less than their forecasted consumption during periods when the system is congested (e.g., when demand is very high and production is expensive). As our study predominantly focuses on price-based DR, from now on we limit our analysis to such programs. However, the core of the insights and conclusions drawn from this study can partially apply to incentive-based DR as well. Popular incentive-based DR schemes have also been listed in Fig. 1, however, the reader is referred to Vardakas et al. [23] for more details.

#### 1.3.1. Time of use

Time of use (TOU) programs divide time into discrete periods and offer a predefined current price for each period [24]. A period can last from a few hours to a few days. When designing the scheme, aggregators set high prices during the time intervals that they intend to reduce consumption (usually peak demand hours) and lower prices otherwise (off peak hours).

TOU programs have the advantage that they allow consumers to plan the consumption patterns of different appliances during the day at the beginning of the program and then follow this planning [24]. They do not need to be flexible to the extent of responding to real-time signals nor do they need to have intelligent controllers for this purpose. They can also a priori calculate the savings they will achieve on their electricity bill if they participate in the program, which is more difficult in programs that have a more complex structure and the value for each period of the day is not predetermined.

From the aggregator's perspective, TOU programs are simpler to design and manage (since they do not require to monitor and change the price or communicate with consumers on a daily basis) and seem to be more popular for consumers than dynamic price adjustment programs, such as real time pricing (RTP) [25]. However, they may create new sharp spikes in demand when the price drops from a high to a lower level [26,27]. This phenomenon limits the number of consumers that can join the same TOU plan (or plans with similar prices and periods), since a large number of consumers in the same plan would create a new peak for the distribution system.

With respect to literature relating to TOU schemes, Yan et al. [24] report the current results from the implementation of TOU programs in different parts of the world. These programs have achieved on average a peak demand reduction of about 10%. Other sources estimate it to 5% [28] for simple TOU programs. This percentage may seem small but it is not negligible. For example, Rosenzweig et al. [29] estimate that a reduction in consumption of only 2%–5% during peak hours would reduce the spot price of electricity by 50% or more. However, there are also TOU programs that do not reach their target. Yan et al. [24] report programs that did not achieve significant results.

#### 1.3.2. Critical peak pricing

Unlike TOU schemes, in which the energy price for each hour of the day or week is fixed, the price in critical peak pricing (CPP) schemes can change for certain hours when it is predicted that there will be heavy congestion on the grid [24]. The periods when this price change occurs are called events. Consumers are usually informed about events the day before (from newspapers, social media, mobile messages, etc.) and are motivated to reduce their consumption during an event by very high prices, which are usually higher compared to off-peak prices than they are in a TOU program [30].

The start and end of an event are determined by the aggregator and are usually constrained by the program terms and conditions together with the total number of events that can occur within a predefined period of time [23]. Therefore, we can observe that a CPP program is not a solution for DR on a daily basis [24], but rather at exceptional cases where consumption is too high, or conditions are such that system reliability is at risk. To address this weakness, aggregators often combine CPP events with different price levels during normal conditions. Practically, such an approach essentially means that we have a CPP program running "on top" of a TOU one.

From the prosumer's perspective, CPP programs are generally easy to implement, since they do not require effort on a daily basis, except when there is a DR event. In fact, it has been observed that households are more willing to significantly reduce



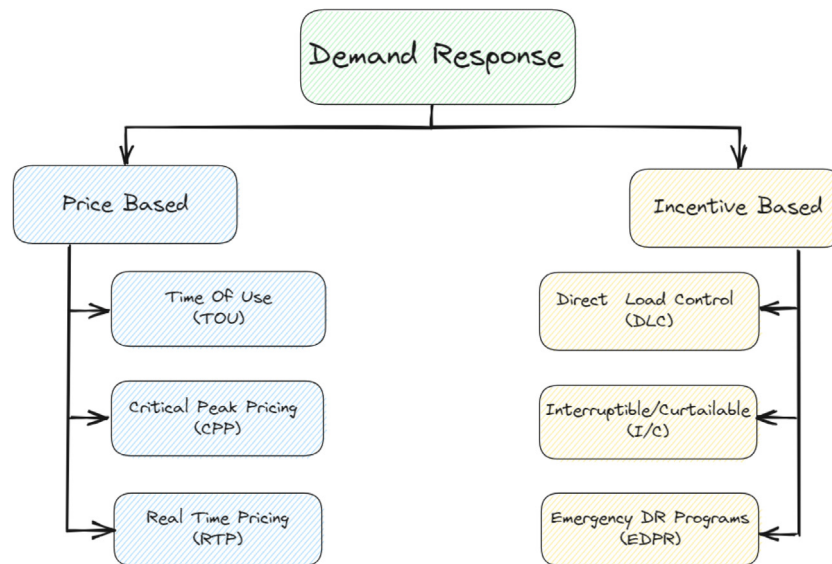


Fig. 1. The categories of demand response programs.

their consumption during a DR event that occurs occasionally than during the peak hours of a TOU program, which requires daily engagement [24]. However, due to the occasional nature of CPP programs, consumers typically do not gain significant cost reductions from participating in them, and if for some reason they are not notified of an event, they are charged at very high rates.

With respect to related literature, several studies discuss the application and effectiveness of CPP schemes. Faruqi and Sergici [31] provide a summary of 15 experimental DR programs. Among the CPP programs, a 13%–20% reduction in consumption was observed during DR events in homes without direct load control (DLC) technologies (systems that allow automatic reduction of consumption during an event, affecting loads such as air conditioners), and the reduction was even greater when DLC technologies were present, from 27 to 44%. Newsham and Bowker [28] confirm the above. They argue that when DLC systems are in place, a CPP program can lead to at least a 30% reduction in consumption during an event, however they believe that similar reductions can be achieved without such systems, provided that consumers participating in the program are carefully selected and offered significant support. These reductions in consumption are considerably larger than those expected in TOU programs. Newsham and Bowker [28] detect two possible reasons for this difference. The first one has to do with price. In general, the energy price during a CPP event is much higher compared to off-peak values than it is in a TOU program [30]. This means that consumers also have a greater incentive to reduce their consumption. The second reason relates to the frequency of CPP events, during participants in a program are requested to reduce their consumption. In this direction, it may be easier for a consumer to abruptly reduce their consumption 25–30 times a year, knowing that this will bring high profit than completely changing their habits, as required by a TOU program.

### 1.3.3. Real time pricing

In real time pricing (RTP) programs, the price of electricity changes during the day depending on the wholesale price of electricity and the conditions prevailing in the grid. The prices within an RTP program are usually announced to consumers one hour in advance, but there are also programs in which they are announced the day before [24]. Especially in the first case, where a consumer does not have the time to plan their daily usage of their appliances, the existence of a mechanism for the

prices to be announced immediately, alongside energy management controllers (EMC) at the prosumer's side are required. RTP programs allow suppliers to vary the selling price of electricity to better reflect the conditions in the network and the electricity production-demand relationship [24]. This is particularly advantageous for both the provider and society as a whole when it leads to a reduction in overall consumption during peak hours, when electricity is generally more expensive and polluting.

Nevertheless, the implementation of RTP schemes for residential consumers is still very limited. There are two main reasons for this. On the one hand, many households do not have the equipment to be able to join an RTP program. The use of smart appliances and controllers is still limited, and in many areas smart meters are not even present in households. At the same time, most residential consumers prefer to avoid risky investments, and perceive the obligation to respond to real-time price signals as a burden [32]. This seems rational given that without automatic controllers and smart appliances, manually adjusting consumption to prices that change so frequently is difficult, complex and time-consuming. The barriers for commercial consumers are similar, apart from the fact that they generally have more resources to invest in relevant technologies and to analyze the economic viability of a related investment. However, as also described in the following section, the low variability of commercial loads suggests that it may be more advantageous to join a program with day-ahead pricing (TOU) than a program with dynamic pricing (RTP).

As dynamic pricing still has little application, there are not many results from actual programs to assess its effect on peak hour consumption, unlike TOU and CPP programs. Two pilot programs applied to residential consumers in Washington [33] and Chicago [34] in 2007 and 2009, respectively, gave promising results. At first, controllers were installed in consumers' houses allowing them to determine automatic heating adjustments based on their electricity price and temperature. These systems, together with some sources of flexibility in the local network (water pumps, generators) managed to yield a 5%–20% reduction in consumption during peak hours (among the project participants). In the second project, consumers were not given automatic controllers but some lamps that changed color depending on the level of the electricity price. It was their responsibility to control the various appliances in the home. The result was again impressive: a reduction in consumption during peak hours ranging from 5 to 14%.

### 1.3.4. The importance of entropy in DR schemes

The term “load variability” refers to whether the consumption behavior of the load is stable over time (low variability) and thus easily predictable or highly variable and therefore hard to predict. Another term often used instead of variability is entropy. There are several ways of measuring entropy in the literature. Smith et al. [35] do it by employing k-means on the daily profiles for each load. A load whose daily profiles belong to several different clusters is considered more unpredictable and thus has high entropy. Kwac et al. [36] and Zhou et al. [37] express this concept mathematically by the sum of Eq. (1).

$$S_n = \sum_{i=1}^K (p(C_i) * \log(p(C_i))) \quad (1)$$

where  $K$  is the number of clusters,  $C_i$  is the center of cluster  $i$  and  $p(C_i)$  is the probability that a load curve from the  $n_{th}$  load belongs to the cluster with center  $C_i$ .

In general, research has demonstrated [38] that consumers with low entropy are suitable for incentive-based schemes (e.g., with direct control) because it is easier to predict and schedule their consumption. On the other hand, consumers with high entropy are suitable for price-based programs because they have more flexibility in their consumption and can make real-time decisions more easily. Based on these results, it seems rational that also among price-based programs, those that follow the same pricing policy every day (TOU programs) will favor prosumers with low variability as such groups are expected to be more capable of scheduling their daily operations according to the price and then follow this schedule every day (more easily than consumers with high variability). On the other hand, programs that are occasional (CPP programs) or where the prices change hourly (RTP programs) are expected to favor consumers with high variability as they are more flexible in their consumption behavior than those with low variability and will be eventually capable of adopting an occasional and well-paid behavior shift request issued by the aggregator.

### 1.4. Machine learning driven clustering for flexibility assessment and demand response

Si et al. [38] review the clustering methods applied to electricity consumption data. The authors conclude that the most common clustering algorithm in smart meter data is k-means, with the euclidean distance to be the most common dissimilarity measure. Although k-means is a simple algorithm and handles large, high-dimensional datasets well, it is very sensitive to outliers and noisy data. Variants of k-means, such as k-medians and k-medoids, can address these problems, in exchange for higher computational costs. There are also other alternatives, such as self-organizing maps (SOM) and various hierarchical algorithms. In addition to these methods, and given that load curves have time series characteristics, various algorithms and proximity measures have been discussed in the literature of time series analysis. An indicative example is dynamic time warping (DTW), an algorithm that aligns time series based on their shape. By applying DTW-based clustering algorithm, the load curves with similar shapes can be grouped more effectively than using the euclidean distance [39]. However, effectively grouping load curves that have similar shapes but peak at different times of the day (e.g. one in the morning and the other in the afternoon) is a problem for traditional DTW methods. In this context, constrained DTW [40,41] –that can prevent time series alignment when the time shift is large – and weighted DTW [39] –that can impose a size-dependent penalty on the similarity measure – can be used as alternatives in this case. Deep learning-based time

series clustering techniques have also been proposed in literature and often lead to more flexible alternatives [42].

A concern that arises here is the evaluation and selection of the appropriate clustering algorithms. Clustering validity indices can solve this problem. Such metrics quantify either how “close” the center of a cluster is to its samples (e.g. mean squared error) or how “far” different clusters are from each other. McLoughlin et al. [43] use the Davies–Bouldin validity index (DBI) to select an appropriate clustering algorithm alongside an optimal proper number of clusters. Cao et al. [44] use two metrics, one to evaluate how well the peaks of the load curve samples match the peaks of the profiles of the clusters to which they belong – namely peak match score (PMS) – and one to quantify how the different profiles of different clusters match with each other (cluster distinctiveness).

With respect to load profiling and relevant pre-processing, before feeding data to the clustering pipeline, the selection of the data features along with the timely resolution to be used are very important as they practically define what a load profile is. In this direction, within a large part of the literature the consumptions are normalized, that is they are divided by the total daily consumption. The purpose of this procedure is to cluster the daily consumption profile (load shape) and not to affect the measure of dissimilarity by the size of the load. The studies of Cao et al. [44], Lin et al. [45] and Kwac et al. [36] are some indicative examples. Subsequently, given that daily consumption profiles are highly influenced by the season, both in terms of total daily consumption and shape of the consumption curve, Cao et al. [44], Biswas and Abraham [46] and Smith et al. [35] divide the data into summer data and winter data. Due to the fact that the load curves of a house on weekdays typically differ from weekends or holidays, some researchers decide to only deal with everyday data [35,44]. McLoughlin et al. [43] and Kwac et al. [36] avoid such distinction. Lin et al. [45] create an average consumption profile for each load (with the consumption per hour being equal to their average consumption at that time within the study period) and cluster these profiles. This approach greatly reduces the amount of data (since in total there are as many profiles as the loads), but ignores day-to-day and seasonal patterns within a specific load. Another approach is clustering using all daily consumption profiles for all loads. This approach has the great advantage that it does not ignore the seasonality and periodicity that may occur in the consumption of a user, but it is more complex, since consumption profiles of the same load can be stored in different clusters, and requires a much larger volume of data (since for 1 year of data, there are 365 different profiles for each load). In [36,43], following the above tactic the authors consider that a load “belongs” to the cluster that contains the largest percentage of its daily curves.

### 1.5. Contribution

In this study, we propose an easy-to-use, clustering-based methodology, aimed at aggregator companies, for the extraction of general-purpose DR policies within flexible energy communities. Specifically, our contributions can be summarized as follows:

- We perform a targeted analysis within a relatively small, local distribution network based on real-life data collected by smart meters of an Italian utility company. Contrary to past studies which usually apply and validate a clustering methodology on a large dataset and a wider area [36,37, 43,46], we focus on developing a realistic DR framework that will be applied by an actual aggregation utility in the province of Terni in Italy.

- We apply clustering on daily load profiles as this approach incorporates the seasonality and periodicity that may occur in the consumption patterns of a specific prosumer. During the optimization of clustering algorithms, we employ the euclidean distance and the constrained DTW distance which allows for more effective edge matching at faster computation times. To the best of our knowledge, only Gao et al. [41] have employed constrained DTW in the electrical load clustering domain until the present. Additionally, we complement the constrained approach by a relaxation method which is inspired from the voice recognition domain, namely the Sakoe-Chiba radius [47]. Such relaxation technique is novel in the field of electrical load clustering.
- We introduce a novel evaluation metric for clustering algorithms, namely peak performance score (PPS). The PPS solves the shortcoming of the peak match score (PMS) [44], that is to penalize the false detection of an edge only when a sample has no edges at all.
- We employ and compare 3 different clustering algorithms – namely k-means, k-medoids, and hierarchical clustering – using various dissimilarity measures. We evaluate them using PPS and silhouette DTW score. However, to come up with an optimal clustering algorithm, we perform an additional empirical and visual evaluation that accounts for the specific needs of a realistic DR policy regarding the cluster number and separability.
- We describe the high-level behavioral characteristics of 14 different daily profile clusters within the energy community also investigating their capability for flexibility schemes. To this end we take into consideration three key factors: (i) daily load profile shapes, (i) load types, and (i) daily load profile entropy. To the best of our knowledge, no prior research study has attempted such a holistic investigation in the context of heterogeneous datasets that combine residential/commercial prosumers and demand/generation. For instance, Kwac et al. [36,48] and Biswas and Abraham [46] only treat residential loads while Biswas and Abraham [46] do not account for entropy at all within their analysis.
- We propose a formalization for discretizing the real values of cluster entropy by mapping them to a simplified qualitative scale. Such a scale can be useful and simple to use for EPES stakeholders that are interested in tailoring DR schemes that also account for load profile entropy within prosumer clusters.
- We provide targeted, general-purpose DR schemes for prosumers based on the flexibility characteristics of the cluster to which they belong as indicated by the majority of their electrical load/generation profiles. Adding up to the current approaches in DR, we account for all types of price-based DR programs documented in the literature – namely TOU, CPP, and RTP – offering a wide range of realistic options for the aggregator rather than entirely sticking to a specific one [37] or entirely neglecting the actual limitations of DR programs [46]. Finally, the allocation of prosumers to clusters, alongside the respective DR policies are static through time, therefore enabling their easy adoption by stakeholders.
- The overall proposed methodology in this study is easy-to-use and has low computational complexity as the feature space of the clustering models corresponds to the exact load/generation profiles of energy community participants. Additionally, it can cope with low volumes of training data and exhibits robustness to the presence of outliers and missing data, which is usually caused by the discontinuity of smart meter measurements, especially within small power systems.

## 2. Methodology

### 2.1. Data description

Initially, we discuss the exploratory data analysis and visual inspection that was conducted before applying clustering algorithms in order to gain a general overview of the dataset's characteristics and detect any data pre-processing issues, such as outliers and missing values. Data have been collected from 54 smart meters that have been decided by the aggregator to form the flexible energy community. The dataset ranges from October 2018 to June 2022 and the time series resolution is hourly. [Table A.1 \(Appendix A\)](#) lists the smart meters of the community to which the use case refer and on which flexibility analysis took place. It should be noted here that 51 smart meters are included in the analysis as 3 of them were discarded to ensure high data quality. [Fig. 2](#) illustrates the sum of production and consumption in the virtual network for each day in the dataset. We notice that reverse power flow (the parts of the graph that are negative) can be obtained at any time between 08:00 and 16:00. However, the phenomenon peaks in the period 11:00 to 14:00.

Regarding missing values, [Fig. 3](#) illustrates a heatmap that shows the number of samples for each month of the year, providing insight into the months with more missing data. [Fig. 4](#) depicts the number of samples per day of week and hour of the day for a random smart meter, as such graphs help to identify specific days or hours with fewer data points. Additionally, [Fig. 5](#) depicts a line chart of the active power of a randomly selected smart meter in time revealing the presence of several outliers. Before proceeding to data wrangling, such plots were created for each one of the 54 smart meters during the visual inspection of the dataset.

### 2.2. Data pre-processing

After identifying the periods where data were missing and outliers were present, we select the appropriate methods to address them as follows:

**Outlier detection** Among the metadata of the dataset, there was information on the nominal power of the smart meter indicated in each consumer's contract ([Table A.1](#)). Therefore all measured power values that exceeded the nominal power threshold were characterized as outliers and removed from the dataset resulting in missing values at the corresponding time period. Note here that a significant number of outliers were observed in our dataset due to inaccurate measurements that were often caused by misconfiguration of the smart meter devices (e.g. scaling inconsistencies).

**Missing data imputation** A significant number of missing values were present in our data set due to smart meter connectivity issues, alongside relevant maintenance activities (see [Fig. 3](#)). Therefore, advanced missing data imputation methodologies were required during our data pre-processing stage. For periods with missing values lasting less than 2 h during the day or more during the night (from 00:00 to 06:00), observations were filled in by evenly distributing the next available observation for cumulative energy consumption across the missing time intervals. For periods with missing values lasting more than 2 h and outside the 00:00 to 06:00 time range, the corresponding days were completely removed, since long interpolations in such cases are expected to distort the load curves and deteriorate the clustering performance. The method of interpolation was preferred against other methods (such as replacement with the historical mean) due to the scarcity of historical data and the presence of cumulative energy measurements from the smart meters which could easily provide inference on the in-between missing values. The limit of 2 h was chosen, as the larger the time window the less

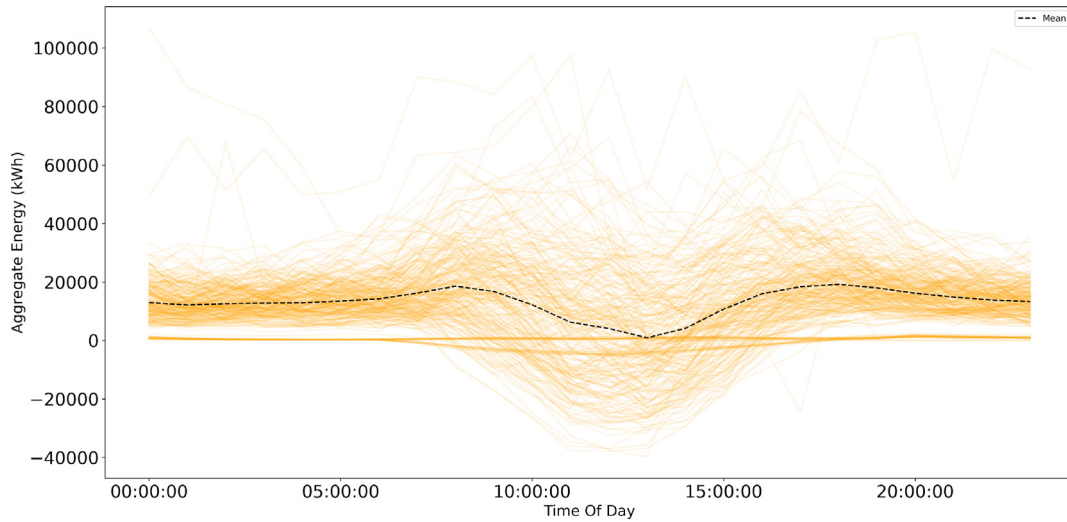


Fig. 2. Sum of production and demand within the energy community.

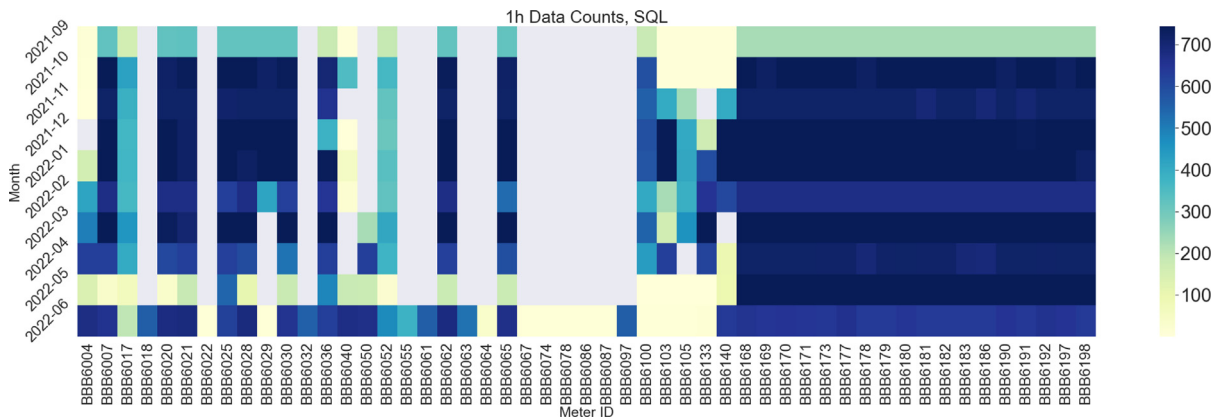


Fig. 3. Heatmap of number of samples per smart meter.

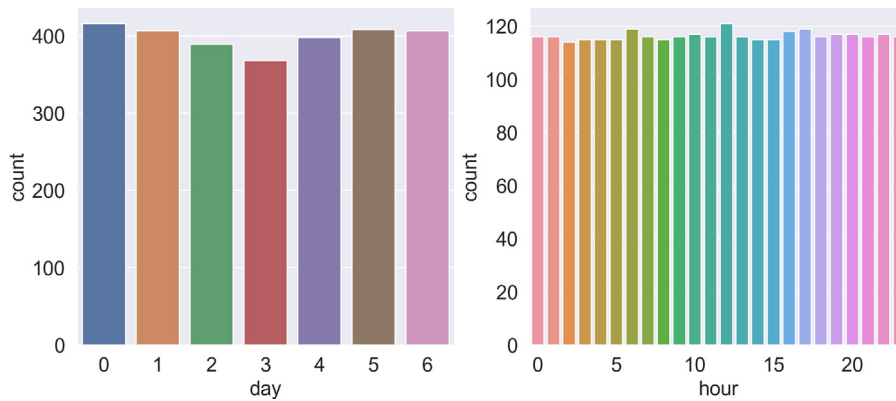


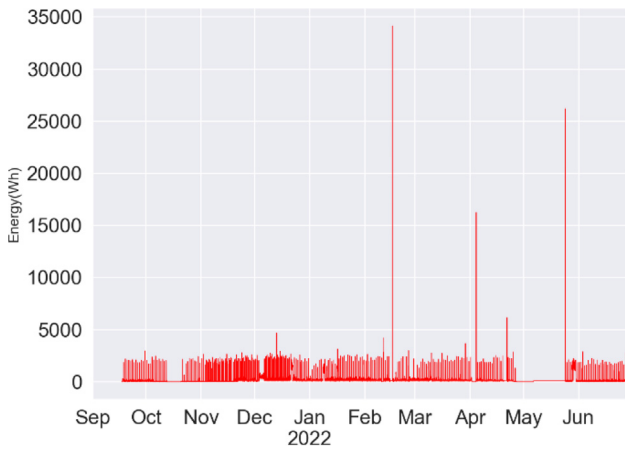
Fig. 4. Bar chart of number of samples per weekday and hour of the day for the smart meter BBB6133.

accurate the method is in restoring the precise shape of the curve. Nonetheless, night hours are an exception to this limit, as they are of low significance for DR programs while also load curves tend to be predictable or near to 0 during these hours. Applying this methodology prevented the complete removal from the dataset of days with significant information about daytime hours due to data gaps during the night. Note here that, given the daily duration of the created load profiles, time series continuity across

days was not a requirement for the clustering algorithm, hence leading to a setup that is robust to the presence of missing values and therefore outliers.

*Normalizing daily demand profiles* As the dataset included both small consumers (e.g. households) and large ones (e.g. university campus) and we intended the clustering process to be based only on the shape of daily profiles, we opt for normalizing them by dividing with the total daily consumption. This approach allowed





**Fig. 5.** Line chart of energy consumption for the smart meter BBB6133 for an extended period of time.

us to remove the distortion caused by the variability in consumption magnitudes among users and cluster exclusively the daily patterns and their shapes.

### 2.3. Clustering

Clustering was performed on the daily load curves of hourly resolution for the entire sample, instead of an average curve for each consumer as such aggregation does not permit the observation and analysis of the load patterns across different months, seasons, and years. Additionally, since the dataset contains only 51 smart meters, clustering on 51 samples would not be of interest. From now on we will refer to this type of clusters as load profile clusters. Similar methodologies have also been followed by other research studies in the past [36,37].

#### 2.3.1. Clustering algorithms

Regarding the selection of clustering algorithms, k-means was chosen as the golden industry standard and due to its low computational requirements. Additionally, we consider k-medoids, which generally handles noise well and has the additional advantage that its cluster centers are actual samples of the dataset rather than average values. Ultimately, we employ cumulative hierarchical algorithm with ward linkage because of its simplicity and compatibility with the selected distance/dissimilarity measures [38]. Note here that all clustering algorithms were executed for all cluster numbers ranging from 3 to 20 for both euclidean and constrained DTW distances. In the following paragraphs a brief description of each clustering method follows.

**k-means** The k-means algorithm [49] is a widely used unsupervised learning technique in machine learning and data mining. It is commonly used for clustering analysis in which a given dataset is divided into k clusters. The objective of the k-means algorithm is to minimize the sum of the squared distances between each data point and its assigned cluster centroid. Mathematically, this can be represented as in Eq. (2)

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2)$$

where J is the objective function, k is the number of clusters,  $C_i$  is the  $i_{th}$  cluster, x is a data point, and  $\mu_i$  is the centroid of the  $i_{th}$  cluster.

The k-means algorithm works by iteratively assigning each data point to the nearest centroid and updating the centroid of

each cluster based on the mean of all the data points in the cluster. This process continues until convergence is achieved. The algorithm is simple to implement and computationally efficient, making it suitable for large datasets. Several modifications and improvements to the k-means algorithm have been proposed over the years, including the use of different distance metrics, initialization methods, and convergence criteria.

**k-medoids** The k-medoids algorithm [50] is a popular clustering technique that is similar to the k-means algorithm but uses medoids instead of centroids as cluster representatives. The medoid of a cluster is defined as the data point that has the smallest average dissimilarity to all other points in the same cluster. Mathematically, the objective of the k-medoids algorithm is to minimize the sum of dissimilarities between each data point and its assigned medoid. This can be represented as in Eq. (3)

$$J = \sum_{i=1}^k \sum_{x \in C_i} d(x, m_i) \quad (3)$$

where J is the objective function, k is the number of clusters,  $C_i$  is the  $i_{th}$  cluster, x is a data point, and  $\mu_i$  is the medoid of the  $i_{th}$  cluster.  $d(x, \mu_i)$  represents the dissimilarity between the data point x and the medoid  $\mu_i$ .

The k-medoids algorithm works by iteratively assigning each data point to the nearest medoid and updating the medoid of each cluster based on the data point that has the smallest average dissimilarity to all other points in the same cluster. This process continues until convergence is achieved. The main advantage of k-medoids over the k-means algorithm is its robustness to outliers. However, it is more computationally expensive than the k-means algorithm, particularly for large datasets.

**Agglomerative hierarchical clustering** Agglomerative clustering is a hierarchical clustering technique that starts with each data point in its own cluster and iteratively merges the closest pairs of clusters until a single cluster containing all the data points is formed. The algorithm produces a dendrogram, which is a tree-like diagram that shows the hierarchical relationships between the clusters. The agglomerative clustering algorithm can be performed using different linkage criteria to measure the distance between clusters. Some common linkage criteria include:

- Single linkage: The distance between the closest pair of points in different clusters.
- Complete linkage: The distance between the farthest pair of points in different clusters.
- Average linkage: The average distance between all pairs of points in different clusters.

Mathematically, the agglomerative clustering algorithm can be represented as in Eq. (4)

$$d_{ij} = \text{linkage}(C_i, C_j) \quad (4)$$

where  $d_{ij}$  is the distance between clusters  $C_i$  and  $C_j$ , and linkage is the chosen linkage criterion.

The agglomerative clustering algorithm has several advantages, including its ability to handle non-linearly separable data and its interpretability through the dendrogram. However, it is computationally expensive for large datasets and can be sensitive to noise.

#### 2.3.2. Dissimilarity measures

The dissimilarity measures optimized within the training process for each one of the algorithms are as follows: (i) euclidean (ii) constrained DTW with Sakoe-Chiba radius equal to 1. The rationale behind the choice of those metrics – which is inspired from the



spoken word recognition domain – is the creation of clusters that are distinguished based on two characteristics: (i) the timing of the edges (peaks and valleys) of their profiles (e.g., two profiles that have a peak at 14.00 should be ideally in the same cluster), and (ii) the shape of their profiles (e.g., profiles with two peaks during the day should be ideally assigned in a different cluster than profiles with a single peak). The euclidean distance is good at capturing information about the former characteristic, but ignores the latter. Nonetheless, we decided to include it within our experiments given its low complexity and wide acceptance as a research and industry standard that is used by the majority of related research studies. On the other hand, simple DTW, when not explicitly constrained, tends to capture information about the shape of the profile rather than the timing of the peaks. We therefore employ a constrained distance measure that permits a certain relaxation in time. Therefore, it allows to capture information about the shape of the profile, but not to an extent that completely ignores time like the simple DTW. We implemented this idea using a constrained DTW version that computes the distance amongst time series points that are far apart from each other no more than a hourly time step. In practice this means that we set the Sakoe-Chiba radius equal to 1. We made this decision on the basis that a relaxation of one hour in time would be desirable, given that it allows to distinguish the load curve peaks according to the period of the day (morning, noon, afternoon, etc.) rather than the very specific hour of the day.

### 2.3.3. Evaluation of clustering algorithms and peak performance score

To compare the different options and select the outcome that best served the use case, we combine the mathematical evaluation of models through evaluation metrics with the visual inspection of the results. Visual inspection (which is useful to validate the clustering algorithm) indeed provides intuitive results i.e. that the extracted clusters significantly differ regarding the two aforementioned characteristics: the timing of the edges and the shape of the consumption curve. In general, the metrics considered to evaluate the clustering process are the following:

- Silhouette score [51]
- Silhouette score DTW (similar to silhouette score but calculated on DTW rather than euclidean distance)
- Davies–Bouldin validity index
- Peak match score
- Peak performance score (novel metric proposed within this study)

Note here that for the calculation of the metrics that identify peaks—such as peak match score (PMS) and peak performance score (PPS)—the peaks have been defined as local maxima with peak prominence larger than 0.2. The peak prominence is defined as the vertical distance from the lowest previous or next local minimum of the normalized consumption profile.

**Peak performance score** PMS is a good approach to quantify the performance of an edge-based electrical load clustering, i.e. to assign data samples to clusters that exhibit edges at the same time moments. However, the PMS has the disadvantage that it penalizes a false detection of an edge only when the sample has no edges at all [44]. To mitigate this issue, we define the PPS, which is inspired by PMS, but adapted to account for both the true detection and false detection of edges in all cases. The PPS is defined as the average of the individual  $m_i$  scores for the samples in the dataset as shown in Eq. (5).

$$PPS = \frac{1}{N} \sum_{i=1}^N m_i \quad (5)$$

where  $N$  is the number of data samples within the data set and  $m_i$  the individual score for each sample  $i$  that is in turn defined as shown in Eq. (6)

$$m_i = \begin{cases} 1, & \text{if } \sum_{k=1}^{24} l_i(k) = 0 \text{ and } \sum_{k=1}^{24} c_i(k) = 0 \\ \frac{\langle l_i, c_i \rangle}{\max\{\sum_{k=1}^{24} l_i(k), \sum_{k=1}^{24} c_i(k)\}}, & \text{otherwise} \end{cases} \quad (6)$$

where  $l_i$  a binary vector of 24-dimensions, one for each hour of the day, where 1 denotes the presence of a peak,  $c_i$ : the vector denoting the centroid of the respective cluster, and  $\langle l_i, c_i \rangle$ : the dot product of the two aforementioned quantities.

In contrast to the PMS, the PPS penalizes the false detection of an edge in all cases. If sample  $i$  has no edge but the cluster in which it is registered has one, then the individual score  $m_i$  becomes 0 and the overall score is reduced, as in the PMS. If the sample has edges but the cluster in which it is registered by the algorithm has more, then  $\sum_{k=1}^{24} c_i(k) > \sum_{k=1}^{24} l_i(k)$ , the denominator of  $m_i$  becomes  $\max\{\sum_{k=1}^{24} l_i(k), \sum_{k=1}^{24} c_i(k)\} \rightarrow \sum_{k=1}^{24} c_i(k)$ , the individual score  $m_i$  decreases, and so does the total score. This is not the case for the PMS. An example with 5 dimensions instead of 24 for simplicity: let  $l_i = [0, 0, 0, 1, 0]$  be the binary vector of the sample and  $c_i = [0, 1, 0, 1, 0]$  the center of the cluster in which it is assigned. In this case, for PMS  $m_i = 1$ , i.e. the algorithm rewards the method for correctly detecting the edge at position 3 but does not penalize it for incorrectly detecting an edge at position 1. In contrast, for PPS  $m_i = 0.5$ , i.e. the algorithm receives half of the reward for the true detection of one edge in the sample but also a false detection of a second edge.

PPS gets real values within  $[0, 1]$ . A value of 1, this means that the algorithm correctly detected all edges in the samples (i.e. all samples were placed in clusters that had edges at the same times as them) and that it did not make any false edge detections at all (no sample was placed in a cluster that had an edge at a time that the sample did not). In this case, the algorithm would be perfect in terms of clustering based on the alignment of edges through time. On the other hand, a value of 0 means that no true edge detections were made, and that all detections made were false. In this case, the algorithm does not cluster at all based on the edge timing. Note here that we employed the PPS with a time relaxation of one time step (i.e. one hour) regarding the “matching” of the edges as mentioned in Section 2.3.2.

### 2.3.4. Cluster extraction and characterization

After executing the experiments and selecting the optimal combination of algorithms and hyperparameters, we perform an analysis of the clusters aiming to distinguish those with the best potential to help achieve the optimization goals (minimization of reverse power flow and peak shifting/shaving) through tailored and realistic DR recommendations proposed to their pertaining prosumers. This analysis was done by considering three key factors that help decide whether a group of consumers is worth being included in a DR scheme:

1. **Daily profile shape:** We are concerned with consumers who can help (a) reduce reverse power flow, which can be done by increasing consumption during generation peak hours, that is mostly from 11.00 to 14.00 (Fig. 2), and (b) reduce consumption during load peak hours i.e. usually 17.00–19.00. We therefore search for groups of consumers (clusters) that have profile characteristics with a peak (a) just before 11.00 or just after 14.00, to encourage them to shift this peak to the 11.00–14.00 interval (e.g. if they usually cook at 15.00 to start cooking earlier, in the 11.00–14.00 interval) or (b) within 17.00–19.00, to encourage them to reduce consumption at these times, either by

canceling some tasks (peak shaving) or by scheduling them earlier or later in the day (peak shifting—e.g. shift the consumption of energy for domestic hot water earlier in the day, i.e. at noon, instead of when they get back from work)

2. **Load types:** The dataset contains both domestic and commercial loads. After identifying the clusters which are more suitable for DR, we proceed to the selection of the DR schemes that would best suit them. In this context, commercial loads are best suited for programs that follow constant billing policies (i.e. TOU) as they require strict scheduling of various electricity consuming tasks. Regarding residential loads, as their tasks are generally less critical and time-constrained, they can perform well in programs with occasional DR events (e.g. CPP schemes) or programs with dynamic pricing (RTP), provided that the proper physical infrastructure is in place so that load management can be done automatically (smart devices, controllers, etc.).
3. **Entropy:** Another characteristic that influences what is the most ideal type of plan for a consumer or group of consumers is entropy. As aforementioned, consumers with low entropy are best suited to incentive-based or predictable pricing (TOU) programs, while consumers with high entropy perform best in CPP or RTP programs. For the scope of this study, we measure the entropy for each load in the dataset, based on the Eq. (1), as well as the average entropy for each cluster. This information is useful in deriving specific DR recommendations.

### 3. Results and discussion

#### 3.1. Evaluation and selection of clustering models

As already mentioned in Section 2.3.3, various evaluation metrics were considered while conducting the clustering experiments, therefore contributing to the rational selection of the model hyperparameters. However, the final evaluation process for the selection of the best-performing clustering setup focuses on two time-relaxed metrics (1 hour relaxation), namely the PPS (custom metric that is tailored to the needs of our task) and the silhouette score DTW (well-established metric within the literature) to ensure the interpretability of our results.

Using MLflow [52] as the logging platform for the machine learning experiments, we took advantage of its visualization capabilities to create the diagram of Fig. 6, which depicts the evaluation of our experiments based on the relaxed PPS (y-axis) and silhouette score DTW (x-axis) for various combinations of hyperparameters. We are mostly interested in the combinations with good performance on both metrics, namely those in the top right region of said figure. Table 1 lists the hyperparameters and evaluation metrics of those best performing models that went through the final visual and empirical evaluation which eventually led to the selection of the optimal model.

As a relatively large number of clusters is desired to have high detail for the DR recommendations we were left with models 4, 7, 8. After further visual inspection, some clusters in the k-medoids models looked quite similar to each other, leading to the exclusion of this setup. Regarding model 4, the clusters were all well-separated from each other concerning the timing of the main edges and the load curve shapes. Model 4 also obtained by far the highest PPS value. Therefore, we opted for this algorithmic setup as the optimal and promoted it to the following stages of our analysis. Fig. 7 depicts the load curves of the clusters obtained by the optimal algorithm, while Fig. 8 depicts the percentage of total samples covered by each cluster. Finally, Fig. 9 illustrates, in much more detail, the memberships of profiles originating from each smart meter across the extracted clusters. This information can be valuable for the aggregator by allowing them to directly assign prosumers to clusters based on their membership as we also attempt in the following stages of the present study.

**Table 1**

Properties of the experiments that fall within the circle of optimal evaluation metric values (PPS and Silhouette DTW). The best performing model is 4 (marked in bold) as it serves the empirical evaluation requirements, namely big cluster number and intuitive cluster separability.

Model id	Algorithm	Distance measure	Number of clusters	PPS	Silhouette DTW
1	k-means	DTW	6	0.652	0.219
2	k-means	DTW	8	0.634	0.277
3	k-means	DTW	9	0.616	0.270
<b>4</b>	<b>k-means</b>	<b>DTW</b>	<b>14</b>	<b>0.689</b>	<b>0.256</b>
5	k-means	Euclidean	8	0.613	0.284
6	k-means	Euclidean	9	0.622	0.287
7	k-medoids	DTW	13	0.677	0.248
8	k-medoids	DTW	14	0.670	0.249

#### 3.2. Cluster extraction and characterization

In this section we discuss the extraction and assessment of the DR potential of the load profile clusters on the shape of their average profile, and the use of the loads assigned to them. To further validate the optimal clustering algorithm, the entropy calculations and visualizations are also included confirming the valuable and intuitive contribution of the clustering process based on the current literature for DR. Regarding training times, the whole set of the clustering experiments (3 algorithms, 18 trials for each algorithm to account for cluster numbers ranging from 3 to 20) lasted approximately 300 min on a MacBook air laptop with Apple M1 CPU and 8 GB RAM.

##### 3.2.1. Cluster analysis based on load profile shapes and types

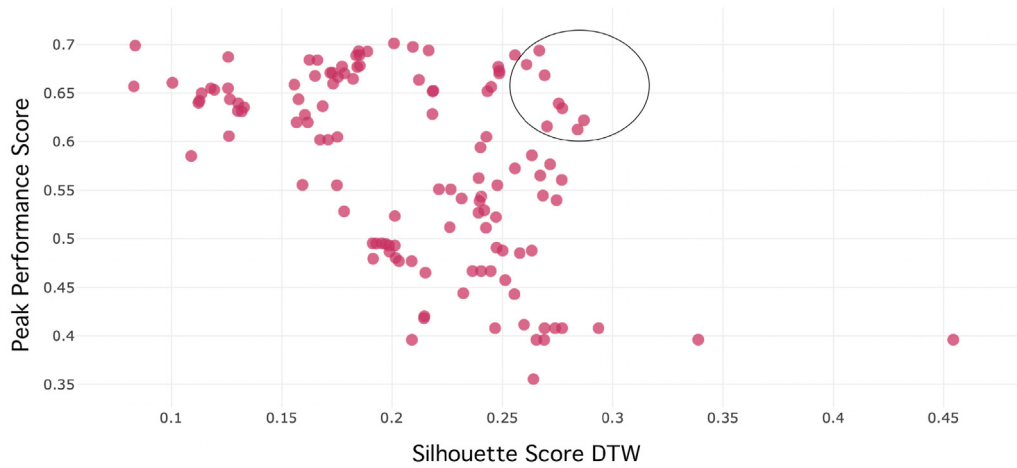
In this part, we dive into the characteristics of the extracted load profile clusters, with respect to flexibility and potential for DR. Note here that mainly the average load profiles and load types are taken into consideration. Apart from Fig. 7 which illustrates the cluster shapes, some individual visualizations with more detailed information on load shapes and load type distributions for each cluster are included in Appendix B.

**Load profile Cluster 0 – Generation** Cluster 0 contains the daily profiles of installations used only for power generation, since if there is any consumption it is very small and most probably originates from small auxiliary loads. The active hours of generation coincide with the hours of sunshine, and therefore it becomes obvious that the loads are PV installations (see Fig. B.1). Given that PV generation is stochastic and cannot be controlled, such loads cannot be useful in providing flexibility to the grid.

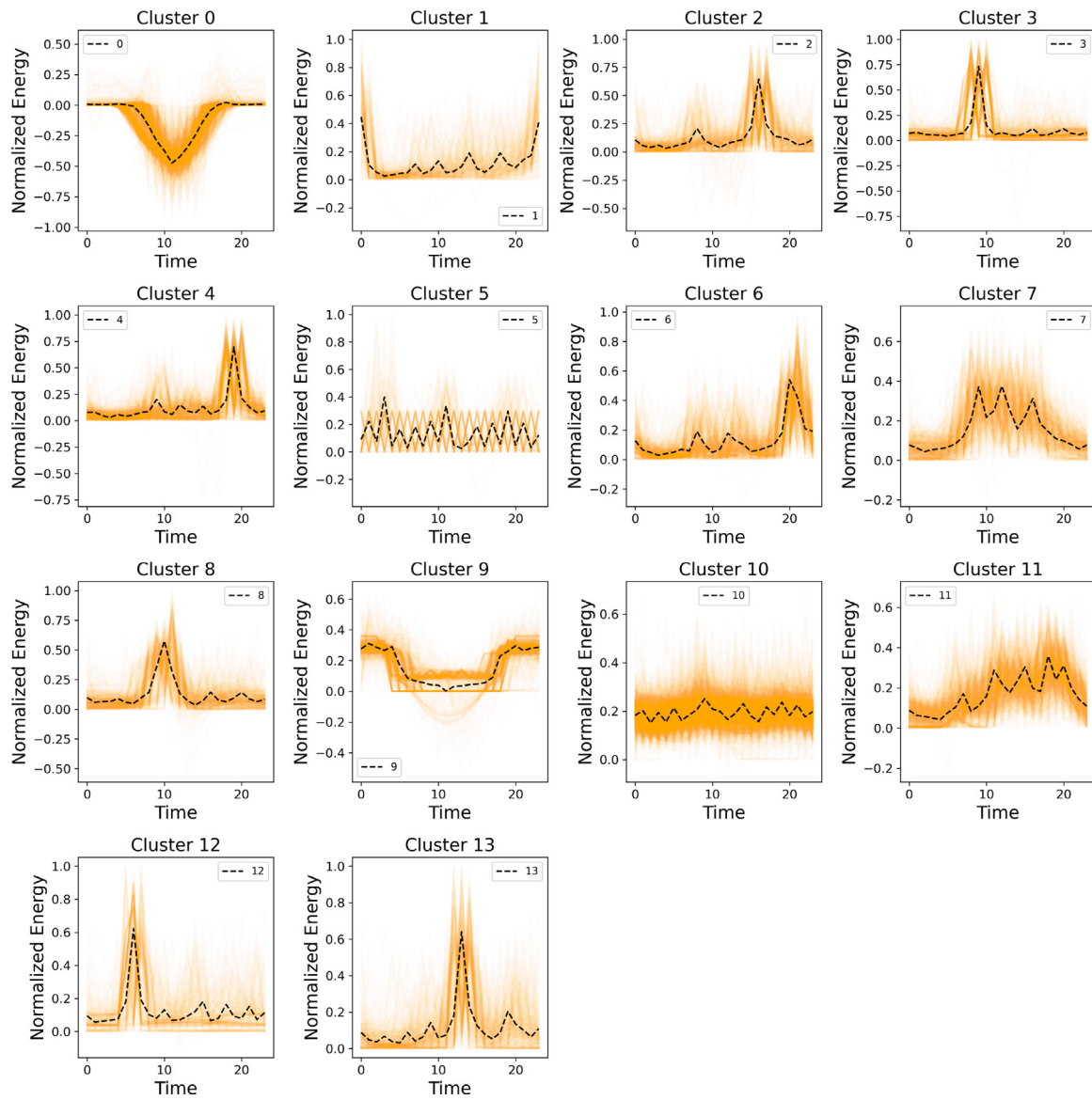
**Load profile Cluster 1 – Residential loads with a late-night peak** Cluster 1 mainly involves consumption from residential users (88%) which peaks around midnight. Since there is usually no congestion in the network at midnight (Fig. 7), there is no interest for loads belonging to this cluster in being included in DR schemes.

**Load profile Cluster 2 – Residential loads with a late afternoon peak** Load profiles belonging to cluster 2 mainly correspond to residential users (73%) and are characterized by a peak in the early afternoon, between 15:00 and 17:00. It would be advantageous for the network to shift this peak earlier, towards midday hours, in order to help both to reduce reverse power flow and to decongest the distribution network at times when demand peaks occur (from 17:00 and onwards).

**Load profile Cluster 3 & 8 – Mixed loads with morning peak** We group clusters 3 and cluster 8 together, because they have a similar shape with a peak in the morning hours (3 from 8.00 to 10.00 and 8 from 9.00 to 11.00). There is a difference in the



**Fig. 6.** Combined diagram of PPS and Silhouette DTW metrics for several clustering experiments. The circle denotes the experiments that combine maximum values for both metrics.



**Fig. 7.** The cluster shapes obtained for k-means algorithm, DTW distance measure and  $k = 14$ . The dashed lines denote the center of each cluster, and the orange lines denote samples that form the cluster. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



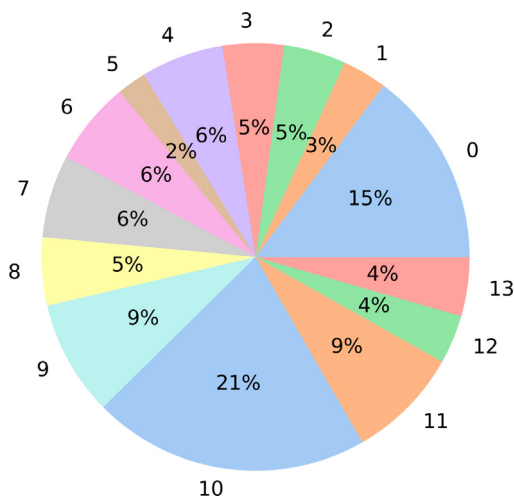


Fig. 8. Pie chart of percentages of total samples covered by the samples of each cluster.

type of loads belonging to them, with cluster 8 containing more residential loads (61% vs. 29%) as well as the load from the public pool. As mentioned previously, the hours 11:00–14:00 are when the reverse power flow peaks. For this reason, loads belonging to clusters 3 and 8 can contribute to reducing the phenomenon if they shift the peak in their consumption later, between 11:00 and 14:00.

**Load profile Cluster 4 & 6 – Residential loads with evening peak** Clusters 4 and 6 contain consumption profiles with a peak in the late afternoon/early evening (from 18:00 to 22:00) and mainly consist of residential loads (93% and 96% respectively). This is expected since evening consumption increases mainly in households that are empty during working hours and fill up when residents return from work. Since the demand within said distribution network often peaks in the evening, the loads belonging to these clusters are of significant interest for DR schemes aimed at reducing consumption during these hours through load shifting (earlier or later, depending on the needs of the network and the constraints of each user).

**Load profile Cluster 5 – Residential loads with mild, uniformly distributed load peaks** Cluster 5 corresponds to loads that do not have one large peak during the day, but several smaller peaks of similar size. It contains profiles of residential users and a large part of the profiles of an EV charging station (28% of the charger's consumption belongs to the cluster). Given that the cluster is not characterized by a specific peak, it is more difficult to draw conclusions about how the loads belonging to it could be used in a DR scheme. To be able to draw certain conclusion, a further analysis of each individual load is required.

**Load profile Cluster 7 – Commercial and public loads with peaks at working hours** Cluster 7 consists of daily profiles that have high consumption during working hours (approximately from 7:00 to 19:00), with no sharp peak like other clusters, but many smaller peaks (the center of the cluster has 3 peaks). The cluster members could help reduce the reverse power flow effect by increasing consumption during the hours when generation is high, but also reduce network congestion in the late afternoon by shifting consumption from these hours a little earlier.

**Load profile Cluster 9 – Consumption and generation** Cluster 9 involves mixed loads in terms of consumption/generation, i.e. installations that have both energy consumption and energy generation. This can be observed from the way the total consumption

decreases during sunny hours (6:00 to 20:00). Because of that, it is difficult to draw conclusions about how the consumption of loads can change for the goals of a DR scheme. A good potential approach for this cluster would be the further division of each daily profile into separate consumption and generation profiles.

**Load profile Cluster 10 – Inactive loads and university campus** This cluster does not exhibit peaks and probably corresponds to days when the load is not used, e.g. when a house is empty or a business is closed. Something worth noting is that most of the local university's consumption also belongs to this cluster, not because the university is empty, but because its consumption does not peak at any time during the day. On the contrary it is relatively stable, with a slight increase at times when there is more teaching activity. As there are no peaks during the day, it is not clear how the loads belonging to this cluster could help in a DR scheme. However, given that the university has very high consumption, even small relative changes in its consumption during the critical hours of the day would be very helpful in meeting the goals of such a program.

**Load profile Cluster 11 – Residential loads with multiple peaks in late afternoon** Cluster 11 is similar to cluster 7 in that it does not have one sharp peak but many smaller peaks during the day. One difference is that the profiles of 7 seem to have higher consumption in the morning hours, while the profiles of 11 have higher consumption in the evening. Therefore, it seems rational that the majority of load profiles in cluster 7 are owned by corporations (50%), while those in cluster 11 are owned by residential consumers (81%). Similar to cluster 7, cluster 11 could be used to address both reverse power flow and congestion during peak hours with its residential nature denoting more flexibility opportunities.

**Load profile Cluster 12 – Mixed loads with early morning peak** Cluster 12 contains loads from companies or residential users that peak in the early morning (5:00 to 8:00 am). Since this early in the morning there is usually no congestion in the network and the times when the load shedding problem occurs are relatively far away, there is no reason for loads belonging to these clusters to be included in a DR scheme.

**Load profile Cluster 13 – Mostly residential with midday peak** The daily profiles belonging to this cluster originate mainly from residential users (73%) and are characterized by a midday peak (from 12:00 to 14:00). Since these are the times when reverse power flow peaks, consumers belonging to this cluster are already contributing to solve the problem and should not be encouraged to change their behavior through a DR scheme.

### 3.2.2. Entropy analysis

In the diagram of Fig. 10 the smart meters of the dataset are arranged in ascending order according to their entropy. Additionally, the chart of Fig. 11 shows the average entropy for each one of the 14 clusters. The average entropy of each cluster is defined as the weighted average of the entropies of the prosumers belonging to that cluster, with the weight of each consumer being equal to the percentage of daily profiles of this consumer within the cluster.

The following can be observed:

- The clusters with the lowest entropy are those containing mixed loads in terms of consumption/generation or pure generation (clusters 0 and 9). This is expected as energy generation profiles exhibited limited variability (the main



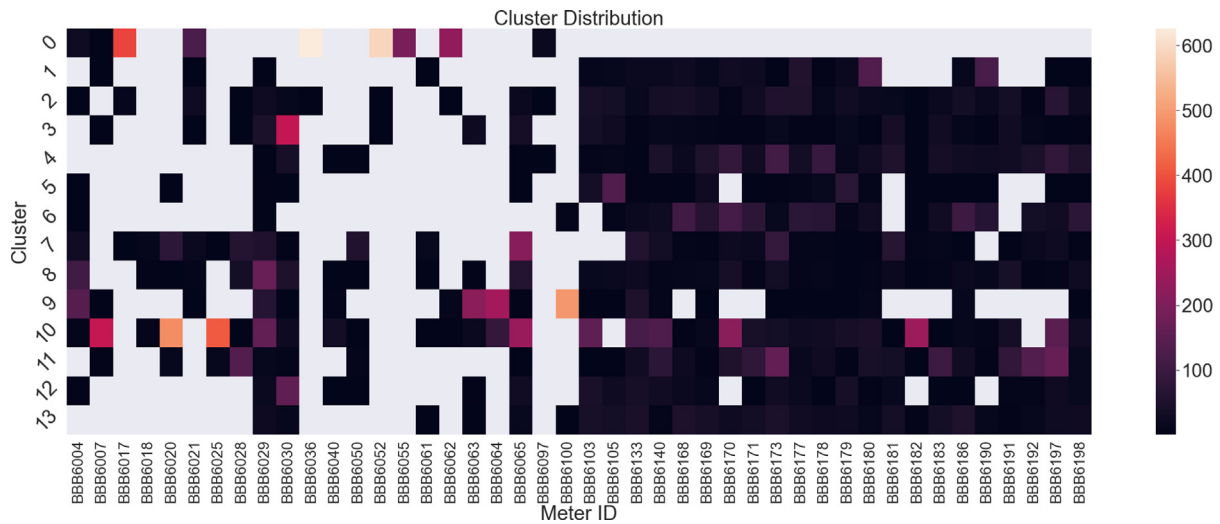


Fig. 9. Distribution of smart meter load profiles across clusters.

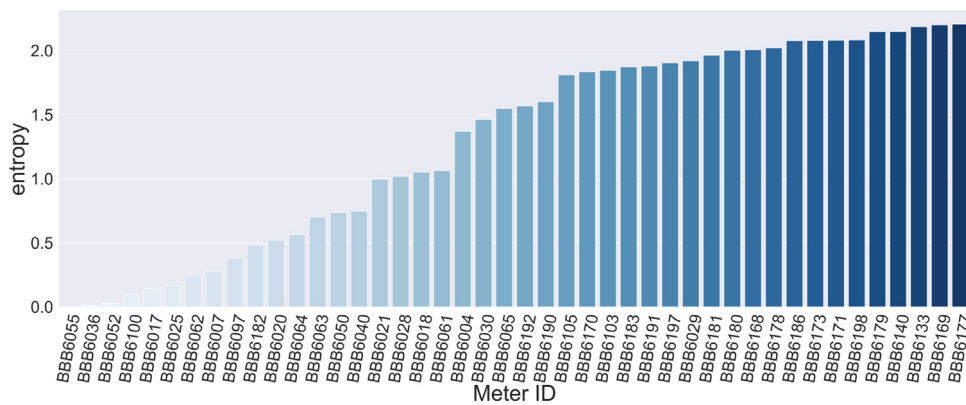


Fig. 10. Entropy of each smart meter in the dataset. Smart meters are sorted by increasing entropy.

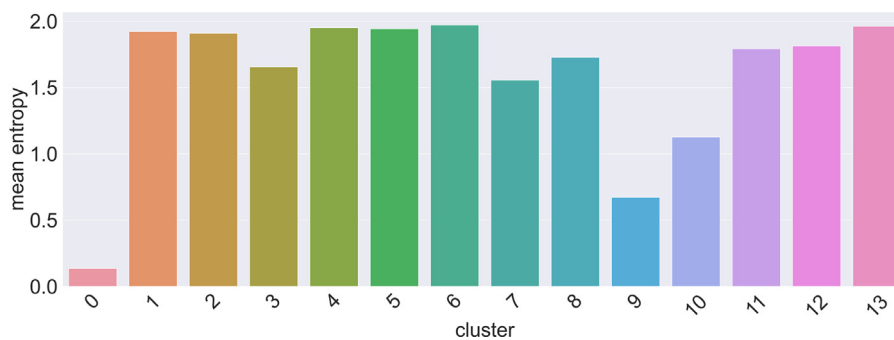


Fig. 11. Average entropy within each load profile cluster.

shape of the generation curve is almost the same every day: a valley during sunny hours).

- The statement that clusters composed mostly of corporate consumption have a lower average entropy than clusters composed mostly of residential consumption is confirmed. Indicatively, regarding the 5 clusters with the highest average entropy (1, 4, 5, 6, 13), all of them contain a larger number of residential users compared to commercial ones. The opposite can be observed for the 5 clusters with the lowest average entropy (0,3,7,9,10).

### 3.3. Discussion on personalized demand response schemes

In this section we discuss on the personalized recommendations for the extracted prosumer clusters that can help solve the problems of reverse power flow and congestion during peak hours. As from the previous steps of our analysis the clusters that have the potential to provide flexibility to the grid are 2, 3, 4, 6, 7, 8 and 11. Note here that load profile clusters until this point refer to the clustering of the daily load curves and not on the average consumption of each individual prosumer. Therefore, the load curves pertaining to an individual prosumer may switch

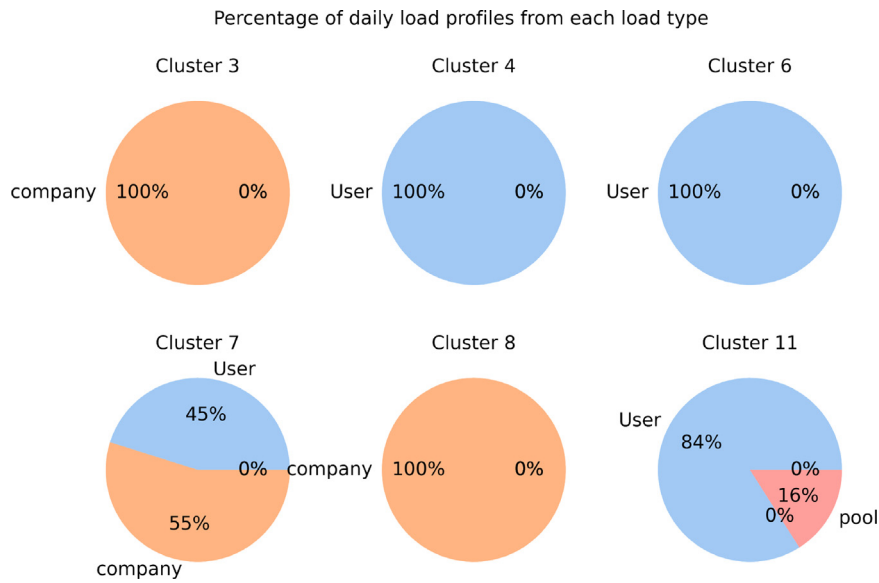


Fig. 12. The new distributions of load type across prosumer clusters that are eligible for DR.

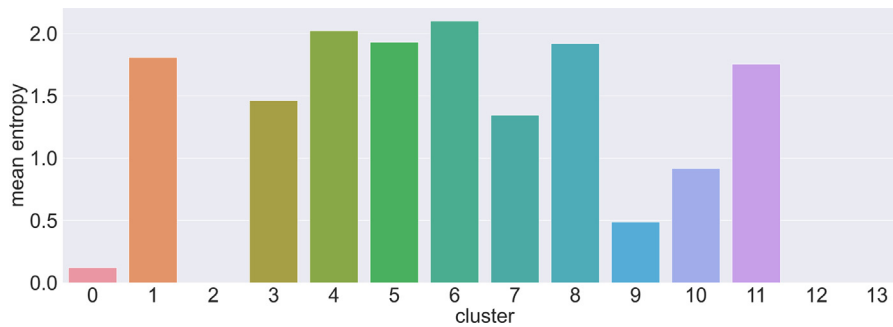


Fig. 13. Average entropy within each prosumer cluster.

clusters among different days. In this context, to be able to provide personalized DR recommendations, we assign each prosumer (identified by their smart meter ID) to the cluster that contains the majority of their daily load curves similar to McLoughlin et al. [43] and Kwac et al. [36]. This assumption leads to an additional classification that has already been referred to as “prosumer cluster” that differs from the previously mentioned profile cluster in that each prosumer is assumed exclusively to a unique cluster rather than partially pertaining to more than one. Therefore, the distribution of load types alongside the average entropy within each cluster differ from what we saw in the previous section. For further details on the results of the new classification of smart meters the reader is referred to Table C.1 (Appendix C).

### 3.3.1. Properties of prosumer clusters

The properties of the newly formed clusters are as follows:

- **Load types:** In the diagram of Fig. 12 the distribution of load types within each cluster are depicted. Note here that this time clusters 3 and 8 are composed purely of commercial load profiles and clusters 4 and 6 of residential profiles. Cluster 7 contains load profiles from both residential users and companies, while cluster 11 has profiles from residential loads and from the public pool.
- **Average cluster entropy:** the graph of Fig. 13 shows the average entropy for each of the newly formed prosumer clusters. Clusters 2, 12 and 13 have no more entropy values as after the reassignment of consumers among the clusters no consumer belonged to them.

### 3.3.2. Discretization of entropy

There are several suggestions in the literature for types of programs that are appropriate for low-entropy and high-entropy consumers [38]. However, there is no objective standard for deciding which entropy values are considered high and low. This could only be done by comparing the relative entropy amongst prosumer clusters. To this end, we divide the range of values by 0.5 and come up with 5 equally distributed discrete intervals assigning them labels as follows: very low (between 0 and 0.5), low (between 0.5 and 1), average (between 1 and 1.5), high (between 1.5 and 2) and very high (greater than 2) entropy. This labeling procedure is demonstrated graphically in Fig. 14.

### 3.3.3. Proposed DR policy

Table 2 lists the recommended DR schemes per prosumer cluster based on the results of the previous analysis. More precisely, the proposed DR schemes are as follows:

- DR scheme 1 – TOU: As we wish to motivate consumers to increase consumption from 11.00 to 14.00 we need to set the price at low levels during those hours and comparatively higher earlier (8.00 to 11.00) and later (14.00 to 17.00). The inverse policy should be followed for peak hours (17.00 to 19.00), i.e., the price should be high during those hours and comparatively lower earlier (14.00 to 17.00) and later (20.00 to 00.00). To achieve these goals, we can propose a program with the high-level pricing scheme of Table 3. As already mentioned, TOU programs are best suited to low entropy

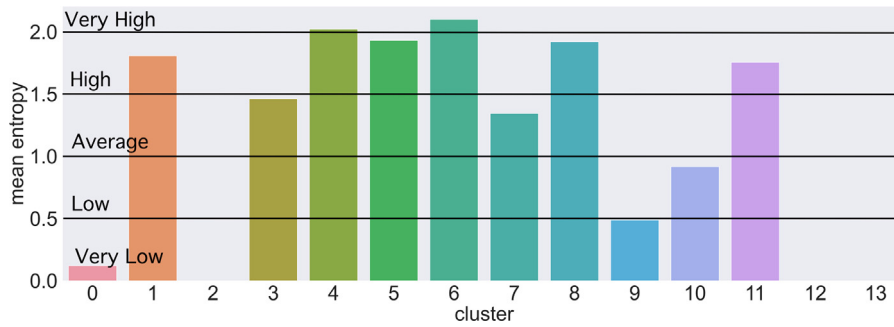


Fig. 14. Demonstration of the methodology for mapping the real entropy values to the categorical/qualitative space.

**Table 2**  
DR scheme recommendations based on the characteristics (load shape, load type, entropy) of each prosumer cluster.

Cluster	Contributes to reverse power flow	High demand during peak hours	Entropy	Load type	Proposed DR scheme
3	✓		Average	Commercial	1
4		✓	Very high	Residential	1,2,3
6		✓	Very high	Residential	1,2,3
7	✓	✓	Average	Mixed	1
8	✓		High	Commercial	1,3
11	✓	✓	High	Residential + public pool	1,2,3

**Table 3**  
Proposed pricing scheme for TOU program.

Time of day	Price
00.00–7.00	Low
7.00–11.00	Moderate
11.00–14.00	Low
14.00–17.00	Moderate
17.00–19.30	High
19.30–24.00	Moderate

consumers and commercial loads. Therefore, the program is recommended for clusters 1 and 7. However, as it is simple to implement and it requires less investment in equipment and person-hours to operate, it could also be applied to clusters with residential consumers and higher entropy.

- DR scheme 2 – CPP: DR events in CPP programs are occasional but during the events, they yield more significant results (greater change in consumer behavior) than TOU programs. A CPP program for the present use case could be activated 20–30 times a year when network problems are expected to be severe (very high reverse power flow or very high peak in the evening hours). The aggregator could alert consumers of DR events a day in advance, announcing the energy price for the event period and suggesting what they can do to alleviate the problem and increase their profit (e.g., proactively heating domestic water or houses earlier than the event if a reduction in overall consumption is required). Of course, as discussed previously, CPP programs are better suited to high entropy and residential consumers. This is why they are proposed for clusters 4, 6, and 11.
- DR scheme 3 – RTP: Generally, an RTP program could adjust the price of energy according to the needs of the grid more efficiently than TOU and CPP programs. However, as already discussed, in order for consumers to react to prices by varying their consumption, they will need special equipment (EMC) that allows them to schedule or automatically handle different loads, such as smart appliances and controllers. In the absence of such devices, it would probably be preferable to implement one of the programs 1 and 2. However, if there is willingness to invest in such equipment, an RTP program could perform well, particularly for clusters with high

entropy (4, 6, 8, and 11) without causing severe user dissatisfaction through load shifting. Obviously, the aggregator’s mechanism that manages the energy should be designed to generate low values when reverse power flow is expected and high values when congestion is forecasted in the local distribution network. Table 3 could also serve as the starting point for an RTP program and later be enriched with higher resolution intervals alongside ad hoc modifications triggered by real-time signals.

**4. Conclusions**

In this study we conducted a clustering analysis on prosumer clusters within a distribution network in the province of Terni in Italy. Three clustering techniques were compared – k-means, k-medoids, agglomerative – to the end of segmenting a small flexible energy community of 54 prosumers that is managed by a municipal aggregator company. The clustering methods were applied on daily consumption/generation profiles of prosumers allowing to better investigate their behavioral patterns across time. To evaluate the model performance we followed both mathematical and empirical methods. In this context, we proposed the peak performance score (PPS) as a novel evaluation metric that can be utilized for effective clustering including but not limited to electrical load clustering and the EPES domain. The k-means with constrained DTW distance matrix and 14 clusters was the best performing setup resulting to high values of both the PPS (0.689) and silhouette DTW (0.256) evaluation metrics. At the same time, the relatively big number of clusters is useful to produce more targeted and detailed DR schemes for prosumers. Subsequently, 14 load profile clusters were derived and characterized regarding their capacity for flexibility and DR taking into account their daily load profile shapes, load distribution (residential vs commercial), and intra-cluster entropy. Note here that this cluster analysis further confirmed the effectiveness of the clustering algorithm in generating meaningful flexibility clusters as they were separated adequately and intuitively regarding their load distribution (residential/commercial, generation/consumption) and also in agreement with the current literature regarding their entropy levels.

Furthermore, a DR policy was proposed for the municipal company/aggregator that manages the smart energy community

taking into account all types of DR programs that can be found in the current literature, namely TOU, CPP, and RTP. Such an approach results a wide range of realistic, price-based DR approaches for the grid stakeholders, tailored on the needs and requirements of their prosumer groups. To propose a personalized DR scheme to a specific prosumer cluster, 4 key decision criteria were considered: (i) the capability for peak shaving and peak shifting during late afternoon, (ii) the ability to maximize consumption during generation peak hours (to prevent reverse power flow at the primary substation), (iii) the entropy of the cluster demonstrating its prosumers' tolerance to change, and (iv) the load type. Briefly, TOU programs are appropriate for all load types as they are the simplest form of DR but are strongly recommended for low entropy clusters/prosumers such as commercial loads. CPP schemes are mainly recommended for residential users that are usually linked with high entropy and tolerance for behavioral changes, while RTP programs are also recommended for high entropy users provided the availability of special equipment for automated energy management. Specifically, to achieve the objectives of the aggregator – namely the minimization of reverse power flow during generation peaks and peak shaving/shifting during demand peaks – a pricing scheme is proposed that:

- i. encourages consumption through low prices during the hours of generation peaks as well as during the night;
- ii. discourages consumption through high prices during late afternoon when load peaks are observed. This pricing scheme can form the starting point both for TOU and RTP programs in the future.

The key findings of this study will be leveraged by the aggregator company in Italy for the development of a DR policy for the said energy community. Nonetheless, the methodological approach can be easily adapted to the needs of other energy communities as well as their respective datasets, therefore effectively serving EPES stakeholders such as TSOs, suppliers, and aggregators that intend to enter the emerging field of DSM.

## 5. Future work

With respect to future work, the deployment in production of the presented machine learning pipeline as an application tailored for aggregators is investigated. Said application will dynamically assess the flexibility and DR potential of each prosumer (e.g. day-by-day). Specifically, the composition of the application is envisaged as follows:

1. A forecasting service, that will generate daily day-ahead forecasts for the consumption and/or generation of each participant in the DR program. The service will incorporate both past load and generation values, alongside historical and live forecasts of climatic factors, such as solar radiation and temperature, as said factors are known to improve the performance forecasting models. This service will also provide an aggregated forecast that will decide on the presence or not of reverse power flow in the primary substation. A similar forecasting service has already been proposed by Pelekis et al. [7]. Note here that larger and continuous datasets are required here to ensure the high accuracy of forecasting algorithms.
2. The proposed initial clustering service within this paper, omitting the subsequent static assignment of prosumers to clusters. At this stage, the served clustering model will classify each forecast into one of the flexibility clusters.
3. A pricing engine that will generate recommendations on how the aggregator can motivate consumers (based on the cluster to which each forecast belongs and the objectives of the DR program) to alter their consumption pattern compared to their forecasted load profile.

4. A projection engine that will modify the conventional forecasted load profiles for end-users based on their likelihood to adopt the proposed DR recommendations. This engine will also assess the impact of the adopted DR policies on the optimization objectives (minimization of reverse power flow, peak shifting/shaving).

In the same context, an additional step of the current analysis would be a post-hoc investigation of the impact of key external factors, such as the season of the year, and the day of week (mostly focusing on weekdays, weekends, and holidays), and external temperature on the initial allocation of prosumers to clusters, as such features are known to significantly affect consumption patterns [9]. Solar radiation could be also considered given its strong linkage with PV generation profiles. Note here that such an approach has been beyond the scope of the present study as we eventually opted for a static assignment of prosumers to clusters aiming at a general-purpose and easy-to-use DR framework for aggregators rather than a dynamic day-by-day policy that would rely on load and generation forecasts.

Moreover, we encourage future research on the development of detailed DR schemes for flexible energy communities proposing specific energy prices. In this purpose, it is necessary to obtain the current pricing schemes within the region of interest as a baseline and thereafter decide on specific modifications that will provide the prosumers with sufficient monetary motivation to change their electrical energy consumption habits. These modifications will be promoted to the customers via the DR scheme (TOU, CPP, RTP) that best fits the cluster to which they belong. To the end of forming complete pricing policies reinforcement learning techniques are of high interest as they can result to programmatical agents that will decide on TOU or RTP pricing policies based on simulations with historical data [53–56].

Additionally to selecting the appropriate method for the creation of DR schemes, it is of utmost importance that prior to their application in production, there is the possibility to evaluate their expected impact and effectiveness. This step is necessary in all cases where observational data from real-life demand response (e.g. [37]) are unavailable. To this end, it is necessary to be capable of generating a projection of how DR recommendations will actually affect cluster/consumer behavior, namely load consumption curves. This step is currently under investigation and is left as future work. In this direction, the concepts of demand and elasticity functions are being investigated in an effort to address this issue [53,57,58].

Regarding the privacy concerns of DR programs, several decentralized/distributed approaches have been recently proposed in the literature [59,60], including federated learning setups [56]. Such methodologies effectively ensure and protect the participants' privacy by minimizing the required data exchanges and avoiding centralized storage. Note here, however, that due to the measuring nature of the existing smart meters in our case study, and the absence of edge/fog computing technologies [61] at the demand side, such practices have been beyond the scope of the present work. In this direction, considering the potential installation of such devices in the near future, the adoption of similar setups is left as future work.

## CRedit authorship contribution statement

**Sotiris Pelekis:** Conceptualization, Methodology, Visualization, Software, Writing – original draft. **Angelos Pipergias:** Methodology, Software, Data curation. **Evangelos Karakolis:** Writing – original draft, Writing – review & editing. **Spiros Mouzakitis:** Project administration. **Francesca Santori:** Validation. **Mohammad Ghoreishi:** Writing – review & editing. **Dimitris Askounis:** Funding acquisition, Supervision.



**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.

**Data availability**

Data will be made available on request.

**Acknowledgments**

This work has been funded by the European Union’s Horizon 2020 research and innovation program under the I-ENERGY project, grant agreement No. 101016508.

**Appendix A. Specifications of the flexible energy community**

See Table A.1.

**Appendix B. Load profile clustering: cluster shapes and distributions**

See Figs. B.1–B.14.

**Table A.1**

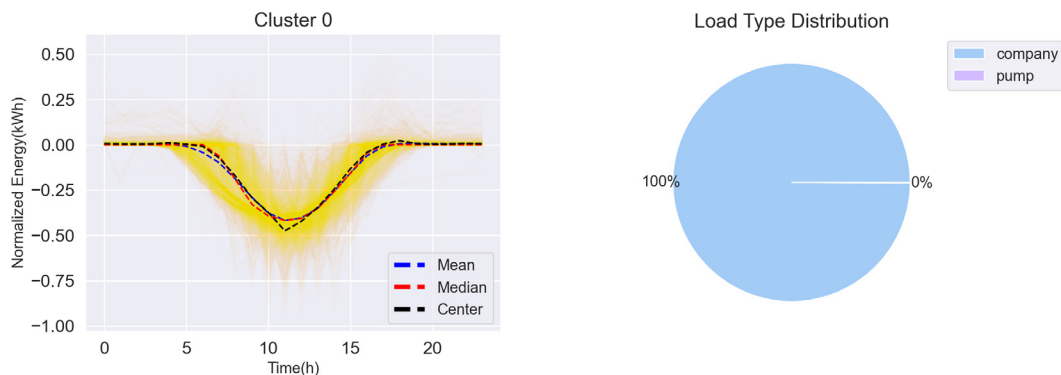
Description of smart meters of the energy community. The smart meters are accompanied by contractual consumption and production powers alongside the type of the load. The smart meters with missing descriptions have been annotated with “–”.

id	Contractual power (kW)	Production (kW)	Type
BBB6004	40	10.8	–
BBB6007	–	–	Pump
BBB6017	1.5	31	Company
BBB6018	137	155	Company
BBB6020	100	0	University
BBB6021	35	50	Company
BBB6022	95	0	Pump
BBB6025	7.7	19.3	Company
BBB6028	125	0	Pool
BBB6029	60	0	Company

(continued)

**Table A.1 (continued).**

BBB6030	150	0	Company
BBB6032	165	0	Pump
BBB6036	19.3	19.3	Company
BBB6040	51	34.6	Company
BBB6048	30	49.9	Company
BBB6050	100	0	Company
BBB6051	35	34.7	Company
BBB6052	1.5	89.9	Company
BBB6055	1.5	175.5	Company
BBB6061	80	0	Company
BBB6062	46	49.3	Company
BBB6063	100	0	Company
BBB6064	200	0	Company
BBB6065	125	0	Company
BBB6067	70	0	–
BBB6071	–	–	–
BBB6074	33	49	–
BBB6078	76	50	Company
BBB6086	–	–	–
BBB6087	100	0	–
BBB6097	70	108.2	–
BBB6100	–	1	Substation
BBB6103	22	0	Electric vehicle charging station
BBB6105	22	0	Electric vehicle charging station
BBB6133	–	–	Company
BBB6140	–	–	–
BBB6168	3	0	Household
BBB6169	3	0	Household
BBB6170	3	0	Household
BBB6171	3	0	Household
BBB6173	3	0	Household
BBB6177	3	0	Household
BBB6178	3	0	Household
BBB6179	3	0	Household
BBB6180	3	0	Household
BBB6181	3	0	Household
BBB6182	3	0	Household
BBB6183	3	0	Household
BBB6186	3	0	Household
BBB6190	3	0	Household
BBB6191	3	0	Household
BBB6192	3	0	Household
BBB6197	3	0	Household
BBB6198	3	0	Household



**Fig. B.1.** Load curves and distribution of load types for cluster 0.

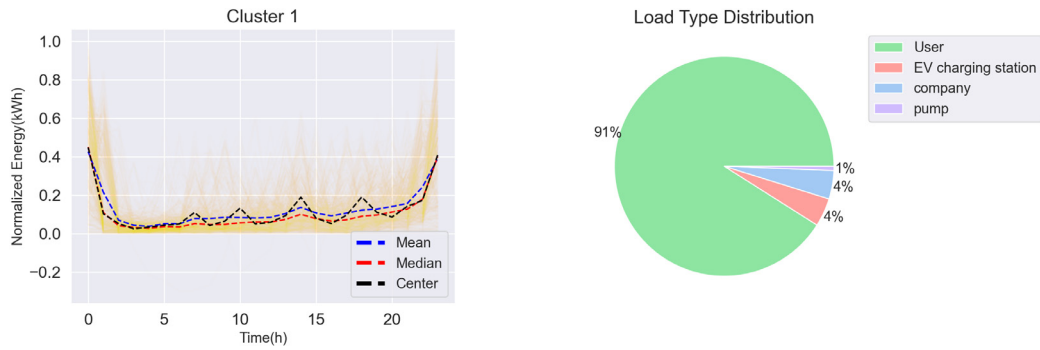


Fig. B.2. Load curves and distribution of load types for cluster 1.

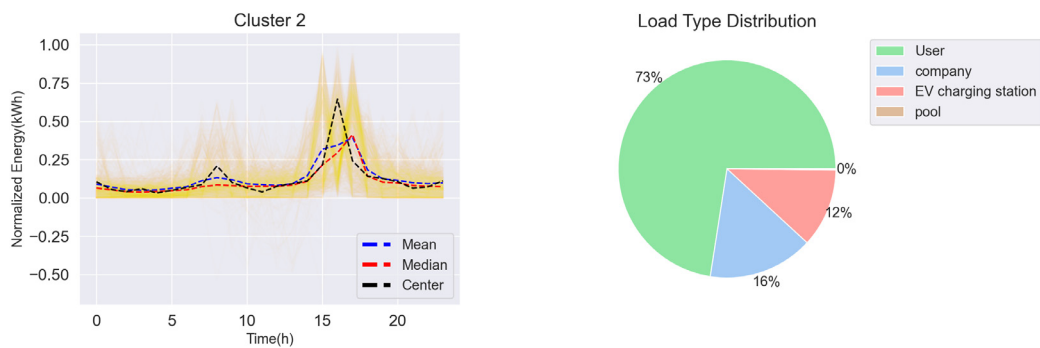


Fig. B.3. Load curves and distribution of load types for cluster 2.

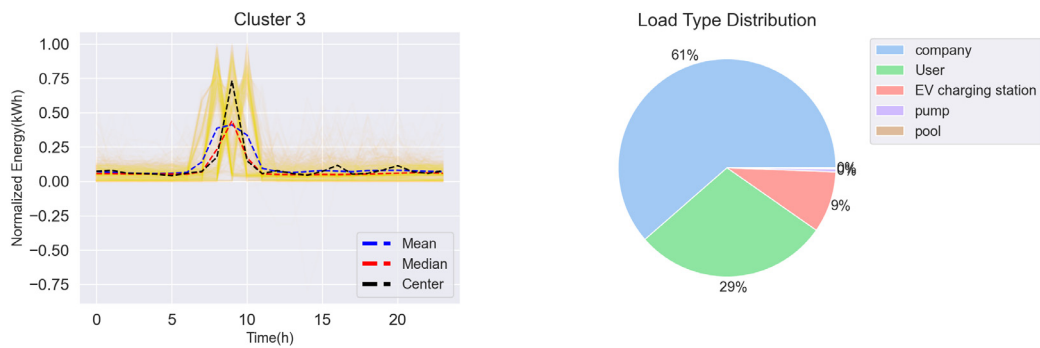


Fig. B.4. Load curves and distribution of load types for cluster 3.

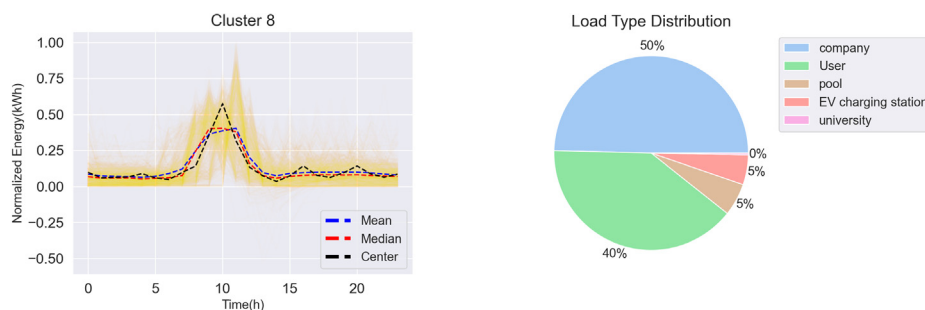
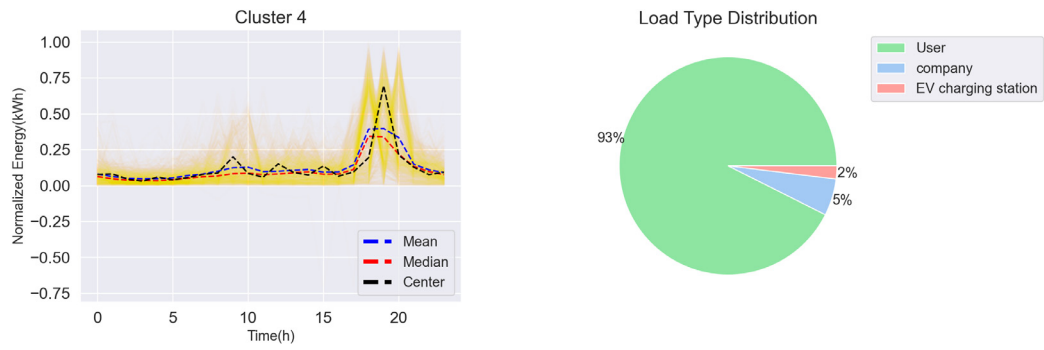
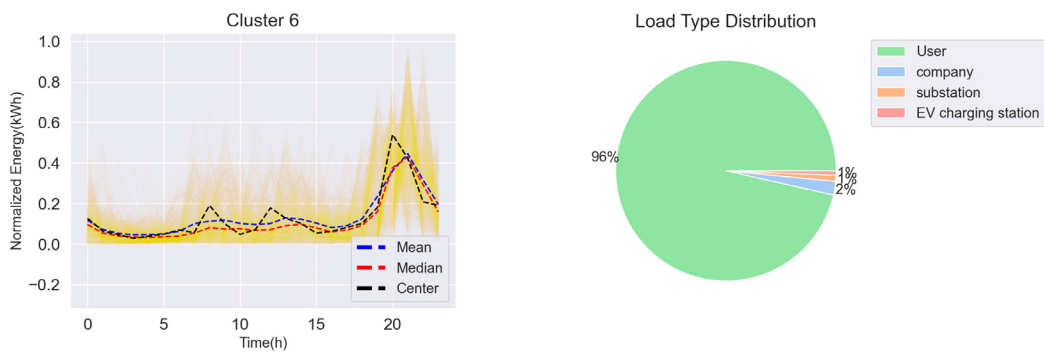


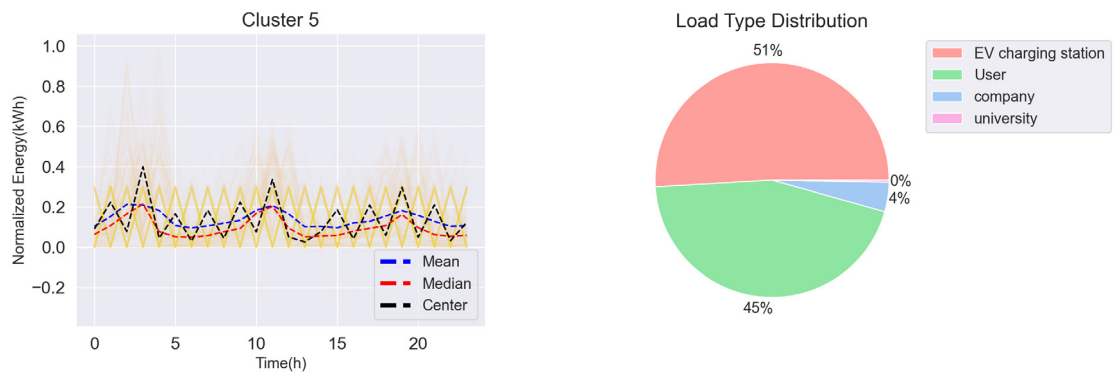
Fig. B.5. Load curves and distribution of load types for cluster 8.



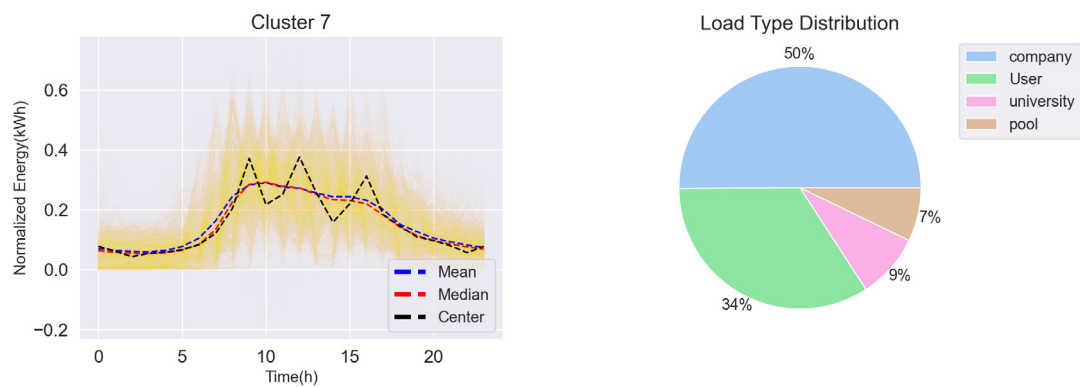
**Fig. B.6.** Load curves and distribution of load types for cluster 4.



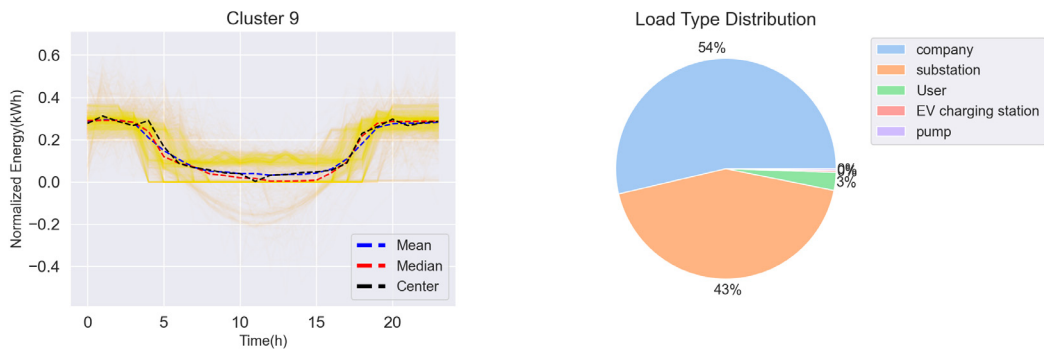
**Fig. B.7.** Load curves and distribution of load types for cluster 6.



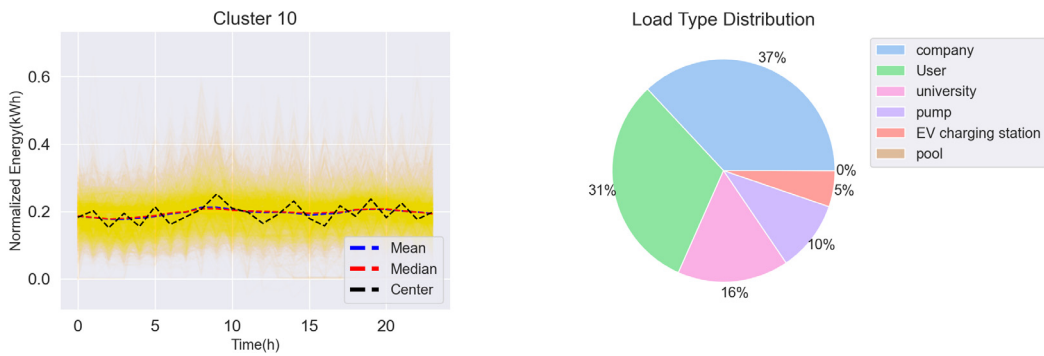
**Fig. B.8.** Load curves and distribution of load types for cluster 5.



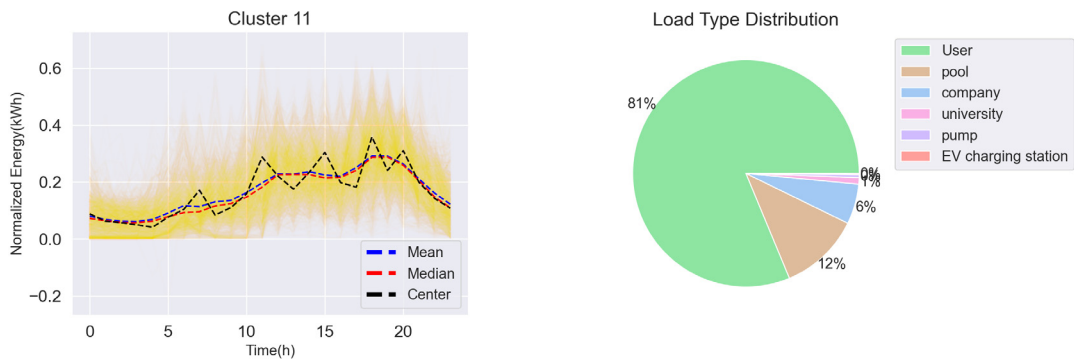
**Fig. B.9.** Load curves and distribution of load types for cluster 7.



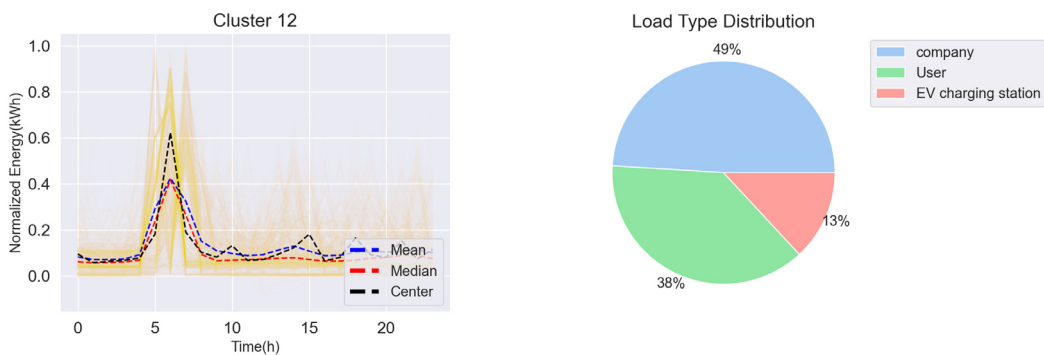
**Fig. B.10.** Load curves and distribution of load types for cluster 9.



**Fig. B.11.** Load curves and distribution of load types for cluster 10.



**Fig. B.12.** Load curves and distribution of load types for cluster 11.



**Fig. B.13.** Load curves and distribution of load types for cluster 12.



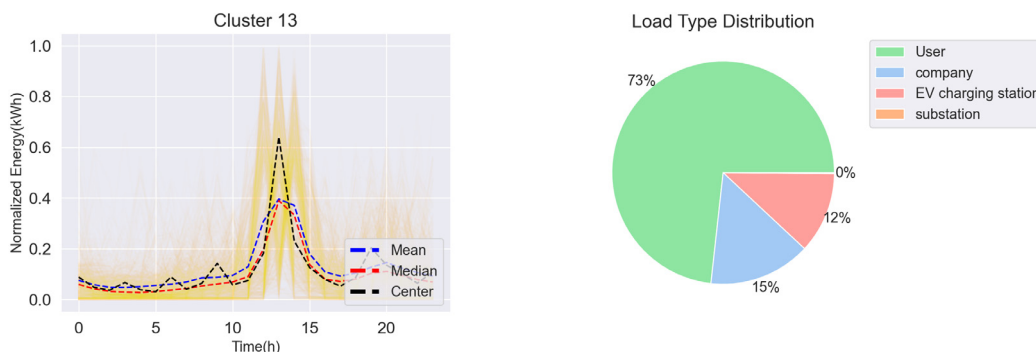


Fig. B.14. Load curves and distribution of load types for cluster 13.

Table C.1

The assignment of the community's smart meters to the new "prosumer clusters".

id	Cluster
BBB6004	9
BBB6007	10
BBB6017	0
BBB6018	7
BBB6020	10
BBB6021	0
BBB6025	10
BBB6028	11
BBB6029	8
BBB6030	3
BBB6036	0
BBB6040	10
BBB6050	7
BBB6052	0
BBB6055	0
BBB6061	7
BBB6062	0
BBB6063	9
BBB6064	9
BBB6065	10
BBB6097	0
BBB6100	9
BBB6103	10
BBB6105	5
BBB6133	10
BBB6140	10
BBB6168	6
BBB6169	6
BBB6170	10
BBB6171	11
BBB6173	11
BBB6177	6
BBB6178	4
BBB6179	5
BBB6180	1
BBB6181	7
BBB6182	10
BBB6183	11
BBB6186	6
BBB6190	1
BBB6191	11
BBB6192	11
BBB6197	11
BBB6198	6

Appendix C. Prosumer clustering labels

See Table C.1.

References

[1] A. Lewandowska-Bernat, U. Desideri, Opportunities of power-to-gas technology in different energy systems architectures, *Appl. Energy* 228 (2018) 57–67, <http://dx.doi.org/10.1016/j.apenergy.2018.06.001>.

[2] A. Varone, M. Ferrari, Power to liquid and power to gas: An option for the german energiewende, *Renew. Energy Rev.* 45 (2015) 207–218, <http://dx.doi.org/10.1016/j.rser.2015.01.049>.

[3] E. Reihani, M. Motalleb, R. Ghorbani, L. Saad Saoud, Load peak shaving and power smoothing of a distribution grid with high renewable energy penetration, *Renew. Energy* 86 (2016) 1372–1379, <http://dx.doi.org/10.1016/j.renene.2015.09.050>.

[4] M.P. Johnson, A. Bar-Noy, O. Liu, Y. Feng, Energy peak shaving with local storage, *Sustain. Comput. Inform. Syst.* 1 (3) (2011) 177–188, <http://dx.doi.org/10.1016/j.suscocom.2011.05.001>.

[5] E. Karakolis, S. Pelekis, S. Mouzakitis, O. Markaki, K. Papapostolou, G. Korbakis, J. Psarras, Artificial intelligence for next generation energy services across europe - the I-nergy project, in: *ES 2022: 20th International Conference E-Society 2022*, Lisbon, 2022, pp. 61–68, URL: <https://cordis.europa.eu/project/id/101016508>.

[6] E. Karakolis, S. Pelekis, S. Mouzakitis, G. Korpapakis, V. Michalakopoulos, J. Psarras, The I-nergy reference architecture for the provision of next generation energy services through artificial intelligence, in: *ES 2023 : 21st International Conference E-Society 2023*, 2023.

[7] S. Pelekis, E. Karakolis, T. Pountridis, G. Korpapakis, G. Lampropoulos, S. Mouzakitis, D. Askounis, DeepTSF: Codeless machine learning operations for time series forecasting, (2023) <http://dx.doi.org/10.48550/arXiv.2308.00709>.

[8] S. Pelekis, E. Karakolis, F. Silva, V. Schoinas, S. Mouzakitis, G. Korpapakis, N. Amaro, J. Psarras, In search of deep learning architectures for load forecasting: A comparative analysis and the impact of the Covid-19 pandemic on model performance, in: *2022 13th International Conference on Information, Intelligence, Systems & Applications, IISA, IEEE*, 2022, pp. 1–8, <http://dx.doi.org/10.1109/IISA56318.2022.9904363>.

[9] S. Pelekis, I.-K. Seisopoulos, E. Spiliotis, T. Pountridis, E. Karakolis, S. Mouzakitis, D. Askounis, A comparative assessment of deep learning models for day-ahead load forecasting: Investigating key accuracy drivers, 2023, <http://dx.doi.org/10.48550/arXiv.2302.12168>, [arXiv:2302.12168](https://arxiv.org/abs/2302.12168).

[10] E. Karakolis, K. Alexakis, P. Kapsalis, S. Mouzakitis, J. Psarras, An end-to-end approach for scalable real time anomaly detection in smart buildings, in: *13th International Conference on Information, Intelligence, Systems and Applications, IISA 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, <http://dx.doi.org/10.1109/IISA56318.2022.9904410>.

[11] T. Bragatto, M. Cresta, C. Kalogiros, M. Paulucci, F. Santori, M. Sanduleac, V.S. Di Carlo, Innovative tools for demand response strategies: A real-life experience, in: *Proceedings - 2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe, EEEIC/I and CPS Europe 2019*, Institute of Electrical and Electronics Engineers Inc., 2019, <http://dx.doi.org/10.1109/EEEIC.2019.8783584>.

[12] A. Geri, F.M. Gatta, M. MacCioni, J. Dell'olmo, F. Carere, M.A. Bucarelli, P. Poursoltan, N. Hadifar, M. Paulucci, Distributed generation monitoring: a cost-effective raspberry pi-based device, in: *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, <http://dx.doi.org/10.1109/IRASET52964.2022.9737768>.

[13] GDPR, Regulation (EU) 2016/679 of the European parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation), 2016, URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.

[14] Keycloak, Keycloak, 2023, URL: <https://www.keycloak.org/>.

[15] R.G. Junker, A.G. Azar, R.A. Lopes, K.B. Lindberg, G. Reynders, R. Relan, H. Madsen, Characterizing the energy flexibility of buildings and districts, *Appl. Energy* 225 (2018) 175–182, <http://dx.doi.org/10.1016/j.apenergy.2018.05.037>.

- [16] I. Antonopoulos, V. Robu, B. Couraud, D. Kirli, S. Norbu, A. Kiprakis, D. Flynn, S. Elizondo-Gonzalez, S. Wattam, Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review, 2020, <http://dx.doi.org/10.1016/j.rser.2020.109899>.
- [17] S. Kiliccote, S. Lanzisera, A. Liao, O. Schetrit, M.A. Piette, S. Kiliccote, S. Lanzisera, A. Liao, O. Schetrit, M.A. Piette, DR: Controlling small loads over the internet, in: ACEEE Summer Study on Energy Efficiency Buildings, 2014.
- [18] J. Eto, J. Nelson-Hoffman, C. Torres, S. Hirth, B. Yinger, Demand response spinning reserve demonstration, 2007, <http://dx.doi.org/10.1109/HICSS.2012.39>.
- [19] J. MacDonald, S. Kiliccote, J. Boch, J. Chen, R. Nawy, Commercial building loads providing ancillary services in PJM, in: 2014 ACEEE Summer Study on Energy Efficiency in Buildings, 2014.
- [20] R. Yin, E.C. Kara, Y. Li, N. DeForest, K. Wang, T. Yong, M. Stadler, Quantifying flexibility of commercial and residential loads for demand response using setpoint changes, *Appl. Energy* 177 (2016) 149–164, <http://dx.doi.org/10.1016/j.apenergy.2016.05.090>.
- [21] R. Ahmadihangar, T. Häring, A. Rosin, T. Korötko, J. Martins, Residential load forecasting for flexibility prediction using machine learning-based regression model, in: 2019 IEEE International Conference on Environment and Electrical Engineering, Institute of Electrical and Electronics Engineers Inc., 2019, <http://dx.doi.org/10.1109/IEEEIC.2019.8783634>.
- [22] European Commission communication, et al., Delivering the internal electricity market and making the most of public intervention, Brussels 5 (2013) 16.
- [23] J.S. Vardakas, N. Zorba, C.V. Verikoukis, A survey on demand response programs in smart grids: Pricing methods and optimization algorithms, *IEEE Commun. Surv. Tutor.* 17 (1) (2015) 152–178, <http://dx.doi.org/10.1109/COMST.2014.2341586>.
- [24] X. Yan, Y. Ozturk, Z. Hu, Y. Song, A review on price-driven residential demand response, *Renew. Sustain. Energy Rev.* 96 (2018) 411–419, <http://dx.doi.org/10.1016/j.rser.2018.08.003>.
- [25] M.L. Nicolson, M.J. Fell, G.M. Huebner, Consumer demand for time of use electricity tariffs: A systematized review of the empirical evidence, 2018, <http://dx.doi.org/10.1016/j.rser.2018.08.040>.
- [26] S. Shao, T. Zhang, M. Pipattanasomporn, S. Rahman, Impact of TOU rates on distribution load shapes in a smart grid with PHEV penetration, in: IEEE/PES Transmission and Distribution Conference, IEEE, 2010, <http://dx.doi.org/10.1109/TDC.2010.5484336>.
- [27] S. Gottwalt, W. Ketter, C. Block, J. Collins, C. Weinhardt, Demand side management—a simulation of household behavior under variable prices, *Energy Policy* 39 (12) (2011) 8163–8174, <http://dx.doi.org/10.1016/j.enpol.2011.10.016>.
- [28] G.R. Newsham, B.G. Bowker, The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: A review, *Energy Policy* 38 (7) (2010) 3289–3296, <http://dx.doi.org/10.1016/j.enpol.2010.01.027>.
- [29] M.B. Rosenzweig, H. Fraser, J. Falk, S.P. Voll, Market power and demand responsiveness: Letting customers protect themselves, *Electr. J.* 16 (4) (2003) 11–23, [http://dx.doi.org/10.1016/S1040-6190\(03\)00048-4](http://dx.doi.org/10.1016/S1040-6190(03)00048-4).
- [30] K. Herter, Residential implementation of critical-peak pricing of electricity, *Energy Policy* 35 (4) (2007) 2121–2130, <http://dx.doi.org/10.1016/j.enpol.2006.06.019>.
- [31] A. Faruqui, S. Sergici, Household response to dynamic pricing of electricity: A survey of 15 experiments, *J. Regul. Econ.* 38 (2) (2010) 193–225, <http://dx.doi.org/10.1007/S11149-010-9127-Y/METRICKS>.
- [32] H. Allcott, A. Abadie, S. Athey, E. Budish, D. Fudenberg, M. Greenstone, B. Hogan, S. Joyce, C. Knittel, E. Mansur, E. Muehlegger, L. Katz, S. Mullainathan, P. Niehaus, C. Nosko, A. Pakes, D. Rapson, R. Stavins, D. Joe Bowring, H. Haas, E. Krawiec, M. Thompson, Real Time Pricing and Electricity Markets, Citeseer, 2009, URL: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=484174e34d4ff1b52373fe057e324595d102d315>.
- [33] D. Hammerstrom, R. Ambrosio, T. Carlon, Pacific Northwest Grid-wide Testbed Demonstration Projects. Part I, Technical Report, Olympic Peninsula Project, 2007.
- [34] L. Lutzenhiser, C. Authors, H. Chappells, M. Gossard, M. Moezzi, J. Peters, P. Stern, H. Wilhite, E. Vine, Behavioral assumptions underlying california residential sector energy efficiency programs, 2009.
- [35] B.A. Smith, Wong, Rajagopal, A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting use of segmentation schemes within the utility industry: A review, in: 2012 ACEEE Summer Study on Energy Efficiency in Buildings, Stanford, 2012.
- [36] J. Kwac, J. Flora, R. Rajagopal, Household energy consumption segmentation using hourly data, *IEEE Trans. Smart Grid* 5 (1) (2014) 420–430, <http://dx.doi.org/10.1109/TSG.2013.2278477>.
- [37] D. Zhou, M. Balandat, C. Tomlin, Residential demand response targeting using machine learning with observational data, in: 2016 IEEE 55th Conference on Decision and Control, CDC 2016, Institute of Electrical and Electronics Engineers Inc., 2016, pp. 6663–6668, <http://dx.doi.org/10.48550/arxiv.1607.00595>.
- [38] C. Si, S. Xu, C. Wan, D. Chen, W. Cui, J. Zhao, Electric load clustering in smart grid: Methodologies, applications, and future trends, *J. Mod. Power Syst. Clean Energy* 9 (2) (2021) 237–252, <http://dx.doi.org/10.35833/MPCE.2020.000472>.
- [39] F. Iglesias, W. Kastner, Analysis of similarity measures in times series clustering for the discovery of building energy patterns, *Energies* 6 (2) (2013) 579–597, <http://dx.doi.org/10.3390/en6020579>.
- [40] W. Choi, J. Cho, S. Lee, Y. Jung, Fast constrained dynamic time warping for similarity measure of time series data, *IEEE Access* 8 (2020) 222841–222858, <http://dx.doi.org/10.1109/ACCESS.2020.3043839>.
- [41] M. Gao, T. Gong, R. Lin, H. Zou, A power load clustering method based on limited DTW algorithm, in: Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2019, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 253–256, <http://dx.doi.org/10.1109/ITNEC.2019.8729249>.
- [42] E. Eskandarnia, H.M. Al-Ammal, R. Ksantini, An embedded deep-clustering-based load profiling framework, *Sustainable Cities Soc.* 78 (2022) 103618, <http://dx.doi.org/10.1016/j.scs.2021.103618>.
- [43] F. McLoughlin, A. Duffy, M. Conlon, A clustering approach to domestic electricity load profile characterisation using smart metering data, *Appl. Energy* 141 (2015) 190–199, <http://dx.doi.org/10.1016/j.apenergy.2014.12.039>.
- [44] H.-A. Cao, C. Beckel, T. Staake, Are domestic load profiles stable over time? An attempt to identify target households for demand side management campaigns, in: IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society, 2013.
- [45] S. Lin, F. Li, E. Tian, Y. Fu, D. Li, Clustering load profiles for demand response applications, *IEEE Trans. Smart Grid* 10 (2) (2019) 1599–1607, <http://dx.doi.org/10.1109/TSG.2017.2773573>.
- [46] S. Biswas, S.A. Abraham, Identification of suitable consumer groups for participation in demand response programs, *IEEE Power Energy Soc. (2019)*.
- [47] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, *IEEE Trans. Acoust. Speech Signal Process.* 26 (1) (1978) 43–49, <http://dx.doi.org/10.1109/TASSP.1978.1163055>.
- [48] J. Kwac, J. Flora, R. Rajagopal, Lifestyle segmentation based on energy consumption data, *IEEE Trans. Smart Grid* 9 (4) (2018) 2409–2418, <http://dx.doi.org/10.1109/TSG.2016.2611600>.
- [49] D. Arthur, S. Vassilvitskii, K-means++: The advantages of careful seeding, in: SODA 2007: The Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2007, pp. 1027–1035.
- [50] W.S. Sarle, L. Kaufman, P.J. Rousseeuw, Finding groups in data: An introduction to cluster analysis, *J. Amer. Statist. Assoc.* 86 (415) (1991) 830, <http://dx.doi.org/10.2307/2290430>.
- [51] P.J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.* 20 (C) (1987) 53–65, [http://dx.doi.org/10.1016/0377-0427\(87\)90125-7](http://dx.doi.org/10.1016/0377-0427(87)90125-7).
- [52] S. Alla, S.K. Adari, Beginning MLOps with MLFlow, Apress Berkeley, CA, 2021, <http://dx.doi.org/10.1007/978-1-4842-6549-9>.
- [53] B.G. Kim, Y. Zhang, M. Van Der Schaar, J.W. Lee, Dynamic pricing and energy consumption scheduling with reinforcement learning, *IEEE Trans. Smart Grid* 7 (5) (2016) 2187–2198, <http://dx.doi.org/10.1109/TSG.2015.2495145>.
- [54] A. Ghasemkhani, L. Yang, Reinforcement learning based pricing for demand response, in: 2018 IEEE International Conference on Communications Workshops, ICC Workshops 2018 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 1–6, <http://dx.doi.org/10.1109/ICCW.2018.8403783>.
- [55] S. Bahrami, Y.C. Chen, V.W. Wong, Deep reinforcement learning for direct load control in distribution networks, in: IEEE Power and Energy Society General Meeting 2020-August, IEEE Computer Society, 2020, <http://dx.doi.org/10.1109/PESGM41954.2020.9281703>.
- [56] S. Bahrami, Y.C. Chen, V.W. Wong, Deep reinforcement learning for demand response in distribution networks, *IEEE Trans. Smart Grid* 12 (2) (2021) 1496–1506, <http://dx.doi.org/10.1109/TSG.2020.3037066>.
- [57] R. Lu, S.H. Hong, X. Zhang, A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach, *Appl. Energy* 220 (2018) 220–230, <http://dx.doi.org/10.1016/j.apenergy.2018.03.072>.
- [58] R. Lu, S.H. Hong, X. Zhang, X. Ye, W.S. Song, A perspective on reinforcement learning in price-based demand response for smart grid, in: Proceedings - 2017 International Conference on Computational Science and Computational Intelligence, CSCI 2017, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 1822–1823, <http://dx.doi.org/10.1109/CSCI.2017.327>.
- [59] S. Fan, Q. Ai, L. Piao, Bargaining-based cooperative energy trading for distribution company and demand response, *Appl. Energy* 226 (2018) 469–482, <http://dx.doi.org/10.1016/j.apenergy.2018.05.095>.

- [60] S. Mohsen Hosseini, R. Carli, J. Jantzen, M. Dotoli, Multi-block ADMM approach for decentralized demand response of energy communities with flexible loads and shared energy storage system, in: 2022 30th Mediterranean Conference on Control and Automation, MED 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 67–72, <http://dx.doi.org/10.1109/MED54222.2022.9837173>.
- [61] J. Violos, S. Pelekis, A. Berdelis, S. Tsanakas, K. Tserpes, T. Varvarigou, Predicting visitor distribution for large events in smart cities, in: 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), 2019, pp. 1–8, <http://dx.doi.org/10.1109/BIGCOMP.2019.8679181>.