ARTICLE IN PRESS

Renewable and Sustainable Energy Reviews xxx (xxxx) xxx

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

A comparative study of clustering techniques for electrical load pattern segmentation

Amin Rajabi [a], Mohsen Eskandari [a,*], Mojtaba Jabbari Ghadi [a], Li Li [a], Jiangfeng Zhang [a], Pierluigi Siano [b]

[a] Faculty of Engineering and Information Technology, University of Technology Sydney, PO Box 123, Broadway, NSW, 2007, Australia
[b] Department of Industrial Engineering, University of Salerno, Via Giovanni Paolo II, 132, Fisciano, SA, 84084, Italy

A R T I C L E I N F O

Keywords:
Smart grids
Smart meters
Load pattern
Data mining
Clustering algorithms
Comparative study

A B S T R A C T

Smart meters have been widely deployed in power networks since the last decade. This trend has resulted in an enormous volume of data being collected from the electricity customers. To gain benefits for various stakeholders in power systems, proper data mining techniques, such as clustering, need to be employed to extract the underlying patterns from energy consumptions. In this paper, a comparative study of different techniques for load pattern clustering is carried out. Different parameters of the methods that affect the clustering results are evaluated and the clustering algorithms are compared for two data sets. In addition, the two suitable and commonly used data size reduction techniques and feature definition/extraction methods for load pattern clustering are analysed. Furthermore, the existing studies on clustering of electricity customers are reviewed and the main results are highlighted. Finally, the future trends and major applications of clustering consumption patterns are outlined to inform industry practitioners and academic researchers to optimize smart meter operational use and effectiveness.

1. Introduction

Enhancement of the power networks using advanced metering infrastructure (AMI), measuring equipment, and smart devices is expected to restructure the existing power grids into a cyber-physical system. Such a system is not only able to carry power flow but can also transmit data for advanced measurement and control applications. The backbone of this cyber-physical system are smart meters and other sensory devices. Smart meters are specified with the sophisticated measurement, control and communication capabilities that they possess. Compared to a conventional energy meter, a smart meter includes measurement and calculation hardware, software, and communication capabilities that measures the energy consumption of a consumer and provides added information to the utility company [1,2]. In the future, smart meters will be in constant communication with distribution (data) management system (DMS) for providing online information and receiving commands [3].

It is projected that the total number of installed smart meters will reach 780 million in 2020 [4]. As some critical studies pointed out [5], despite the ongoing rollouts, many utilities are still unclear about the optimal route to extracting value from these large investments. In North America, the main priorities are to use smart metering information as a means to support outage management and increase grid reliability. On the other hand, European utilities are more focused on consumer-related capabilities. From the analytics point of view, the smart metering data is still mostly an underutilized area of value for existing deployments. However, it is becoming recognized as a strategic next step for many utilities. As the deployment of smart meters is increasing, the main question is how to utilize such a wealth of hourly or half-hourly measured data to gain benefits for various stakeholders in power systems.

The use of data mining techniques to analyse load data offers a variety of potentials within the power systems [6]. Clustering is a well-known unsupervised data mining technique for segmentation of a data set by assigning its objects to a set of clusters [7]. It has numerous applications in different fields such as market segmentation analysis, biology, and social network studies. In the power system domain, clustering techniques can be used to find similar patterns in electricity consumption behaviors of users. This will offer potential advantages for companies as it reveals characteristic customer load profiles within the heterogeneous population and enables utilities to gain better knowledge

* Corresponding author.
E-mail address: Mohsen.Eskandari@student.uts.edu.au (M. Eskandari).

https://doi.org/10.1016/j.rser.2019.109628
Received 3 January 2019; Received in revised form 21 August 2019; Accepted 24 November 2019

List of abbreviations

AMI	Advanced Metering Infrastructure
BIC	Bayesian Information Criterion
CER	Commission for Energy Regulation
CVI	Clustering Validity Index
DBI	Davies-Bouldin Indicator
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DFT	Discrete Fourier Transform
DR	Demand Response
DMS	Data Management System
DTW	Dynamic Time Warping
DWT	Discrete Wavelet Transform
EM	Expectation-Maximization
FCM	Fuzzy C-Means
GMM	Gaussian mixture model
HMM	Hidden Markov Model
ISODATA	Iterative Self-Organizing Data Analysis
MIA	Mean Index Adequacy
MSE	Mean Square Error
PAA	Piecewise Aggregate Approximation
PG&E	Pacific Gas and Energy Company
PCA	Principal Component Analysis
SAX	Symbolic Aggregate approXimation
SIL	Silhouette index
SOM	Self-Organizing Maps
TOU	Time of Use
WCBCR	The ratio of within-cluster sum of squares to between-cluster variation

of customers' electricity demand [8]. Furthermore, the results of customer segmentation can be used for numerous applications in power industry such as load forecasting, tariff design, and implementation of demand response (DR) programs.

This paper tries to utilize a cross-disciplinary approach spanning engineering and data science to present a comparative study of different techniques for clustering of load patterns. In this respect, this paper is opportune, because, despite the considerable changes in the area, there is no comprehensive study on the application of clustering techniques for power systems.

There are several features that distinguish the current work from the previous publications [2,9,10], mainly:

- Five major clustering techniques are introduced and the effects of their different parameters for load pattern clustering are analysed. Besides the well-studied clustering methods such as K-means, fuzzy c-means and hierarchical algorithms, clustering with the probabilistic and generative models and self-organizing maps (SOM) are also discussed.
- These clustering techniques and their applications in customer segmentation are compared.
- The use of data size reduction techniques and feature extraction/definition methods for load data are described and analysed.
- The future directions such as time series clustering and big data issues are illustrated and the possible applications of clustering in the power system domain are explained and categorized.
- An extensive review of the current literature is provided.

Section 2 provides a general overview of clustering concepts and load data clustering. The other sections are respectively dedicated to the clustering methods, discussions on the clustering parameters, case studies, data size reduction techniques, and applications and future trends. Finally, Section 8 concludes the paper.

2. Background

2.1. Clustering concepts

Clustering is an unsupervised data mining technique that enables the determination of intrinsic patterns in data sets. The main aim of clustering is to partition data instances (objects) of a data set into a number of groups (called clusters) which are as similar as possible. Objects belonging to a cluster are more similar to each other than to those in other clusters. Therefore, the goal is to achieve high intra-cluster similarity and low inter-cluster similarity.

Expressing mathematically, a data set S which has n records (observations) can be partitioned into a set of K clusters $C_1, C_2, \ldots, C_K$ that do not intersect (however, this assumption is sometimes violated when the soft clustering is applied) and the union of them is equal to the full data set as shown in (1):

$$S = \bigcup_{i=1}^{K} C_i \text{ and } C_i \cap C_j = \varnothing \quad \text{for } i \neq j \tag{1}$$

2.2. History of electricity customer clustering

In the power system domain, utilities and system operators are interested to classify the electricity users into distinct groups as it offers advantages in decision making and control of power network. However, this trend was naturally limited in power system studies as the system operators and researchers did not have access to the fine grained consumptions of the customers. Before the widespread availability of AMI data, little information about each household's consumption or energy use habits was available. The monthly usage of each household and some fixed information such as voltage level and nominal demand were the main sources of information for categorizing the households. On the other hand, utilities and researchers also conducted in-home surveys in order to evaluate the effect of various variables on consumption patterns. These variables could be categorized under one of these classes [11–13]: dwelling characteristics, demographics and socio-economic factors, habits of energy use such as consumption timing, attitudes toward energy use like level of concern toward energy conservation, knowledge about electricity consumption, and energy efficiency goals. Based on those available data, it was a common practice among distribution companies to define a set of classes and assign each customer to a specific class. For example, Pacific Gas and Energy Company (PG&E) segmented its residential customers into eleven clusters using a broad range of attitudinal and demographic variables [14]. In another attempt, 46 different customer class load profiles were defined by Finnish Electricity Association for categorizing customers in residential, agricultural, industrial, and services sectors [15,16].

Such approaches were inherently inaccurate as they did not have access to the real consumption data of customers. Installation of smart meters has fundamentally changed this situation. Nowadays, the fine-grained measurements are available in a large scale for tens of thousands to millions of users and moreover, they are accessible for successive years. As a result, the customer categorization can be achieved by implementing appropriate clustering techniques which are applied on the load data of the customers.

2.3. Stages of load pattern clustering

Stages of electricity customers' clustering are summarized and depicted in Fig. 1. These stages are as follows:

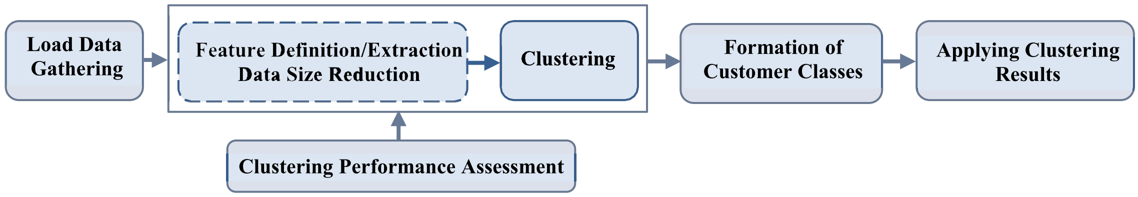

Fig. 1. Stages of load pattern clustering.

- ***Electricity consumption data gathering:*** The first step includes the collection of the consumption data of electricity customers. Like any other practical world data gathering, a pre-processing is needed to discover the missing (incomplete) or bad data in the data set. The missing data can be repaired at this stage by using different techniques like regression methods or can be handled by special ways as a part of clustering stage [7]. A customer's data may also be corrupted by the noise or by the occurrence of uncommon situations like anomalous days or unexpected failures. Consequently, elimination or replacement of bad data is another essential pre-processing step.
- ***Data size reduction/feature definition/feature extraction:*** In some cases, before the main clustering stage, the collected smart meter data are processed in some ways to reduce the scale of input data or to define more meaningful features for categorizing the customers. This preliminary stage can be categorized by feature definition (expert knowledge-based feature extraction), feature extraction, and data size reduction techniques. These concepts are described in more detail in Section 6.
- ***Clustering stage:*** Use of proper clustering techniques and accurate selection of parameters of clustering algorithms is vital in this stage; although, it depends on various factors such as the size of available data, the final goal of clustering, on-line or off-line clustering, the analytics and computational facilities, and user preferences. Sometimes more than one clustering method may be applied to the load patterns, and final results will later be compared to attain the best results. Furthermore, a combination of different clustering techniques is also possible to speed up the process or to obtain better outcomes [17].
- ***Clustering performance assessment:*** Since the clustering of a data set is an unsupervised process, it is not very clear how to assess the quality of the resulted clusters in an objective way [7]. Intuitively, a good clustering method must ensure that each cluster is compact and different clusters are widely separated from each other [18]. To evaluate the clustering results, various clustering validity indexes (CVIs) are used.
- ***Formation of customer classes:*** This stage represents the post-processing of the formed clusters, mostly based on the real-world scenarios. For example, the final number of clusters cannot be more than a certain number if the final goal of clustering is to define cluster-specific tariffs or to apply DR programs. So, the number of customer segments should be specified by the ultimate user like the retailer or DR aggregator. In this case, some clusters that have similar patterns may be consolidated [14].

In the next sections, 5 clustering methods and 2 indirect clustering algorithms for electricity customer segmentation are evaluated using proper CVIs. Furthermore, the applications of clustering for power systems are discussed.

3. Clustering algorithms

Many clustering algorithms are proposed in the data mining community, and for each method, different variations are developed. In the power system literature, some of these techniques have been applied to load patterns of customers (Table 1).

Table 1
Clustering methods.

Method	References
K-means	[10,14,19–31]
FCM	[10,19,20,22–24,28,32–34]
Hierarchical	[10,17,22–24,28,35,36]
SOM	[8,19,21,23–25,37–41]
Model-based approaches	[11,35,42,43]
Other methods	
K-medoids	[21,44]
Adaptive K-means	[17,45,46]
K-Shapes	[30]
Follow the leader	[10,20,47]
DBSCAN	[48,49]
ISODATA	[50]
Optimum-path forest	[51]
Fuzzy Gustafson-Kessel	[52]

In the following, the major clustering techniques studied in this paper are briefly illustrated and a review of the most important works from the literature is presented.

3.1. Distance-based methods

Distance-based methods are the most popular clustering algorithms since they are generally fast and easy to implement. These algorithms use similarity (or dissimilarity) measures to construct the clusters. As the main purpose of clustering is to group similar instances, defining proper measures that can numerically express the degree to which two objects are similar to or dissimilar from each other is required. The main types of similarity measures used in the literature can be categorized as [7,53]: 1) difference (distance)-based measures such as Minkowski distance (L_P-norm distance), Canberra distance, and Gower's coefficient, and 2) Correlation-based measures (similarity functions) such as cosine measure and Pearson's correlation measure.

Here, we confine the discussion to Minkowski measures as they are the most common measure used in the power system literature. These similarity measures try to calculate a distance value based on the differences between the features (attributes) of the two compared objects. If two load curves x_i and x_j are represented by h recordings, the Minkowski distance of order p between them can be calculated as follows:

$$d_{Mink,\ p} = \left(\left|x_{i,1} - x_{j,1}\right|^p + \left|x_{i,2} - x_{j,2}\right|^p + \ldots + \left|x_{i,h} - x_{j,h}\right|^p\right)^{1/p} \quad (2)$$

For $p = 1$ and $p = 2$ the L_P-norm distance is usually called the Manhattan distance (or city block distance) and Euclidean distance, respectively. Euclidean distance is by far the most widely used dissimilarity measure.

Two well-known and frequently used distance-based clustering algorithms are partitioning methods and hierarchical clustering methods which are well presented in the data mining literature.

3.1.1. K-center family

K-centers family including K-means, K-medians, and K-medoids are the most widely used partitioning clustering techniques. They do not create a tree structure to describe the groupings of data, but rather create a single level of clusters. They share the same basic operation principle which is outlined in Algorithm 1 [54,55].

Algorithm 1. K-centers clustering

Require: Number of clusters and cluster centers as follows:
•The number of clusters is predetermined (***k*** clusters).
•***k*** points are selected as the initial cluster centers.
Repeat:
1- Assign each instance to the closest center until ***k*** clusters are formed.
2- Recompute the center of each cluster based on all instances that belong to it.
Until: The convergence criterion is met.

K-means is a commonly used algorithm, which minimizes the square-error function, defined as:

$$E = \sum_{k=1}^{K} \sum_{x \in C_k} ||x - c_k||^2 \tag{3}$$

where K is the number of clusters and c_k is the center of kth cluster denoted by C_k.

Fuzzy c-means (FCM) is another popular method of K-centers family [56]. It is similar to K-means clustering, but each instance has a grade of membership to each cluster [23]. FCM minimizes the following objective function:

$$J_m = \sum_{l=1}^{N} \sum_{k=1}^{K} {\mu_{lk}}^m ||x_l - c_k||^2 \tag{4}$$

where N is the number of load curves (observations), μ_{lk} is the degree of membership of lth load curve in kth cluster, and m is the parameter that controls the amount of fuzziness.

In fuzzy clustering, each load curve does not belong to only one cluster. Instead, the degree of membership determines the amount of membership of each load curve to each cluster, where:

$$\sum_{k=1}^{K} \mu_{lk} = 1 \tag{5}$$

An observation is assigned to the cluster to which it has the maximum value of membership degree [34]. The membership degrees are updated in each step as:

$$\mu_{lk} = \left[\sum_{j=1}^{K} \left[\frac{||x_l - c_k||}{||x_l - c_j||} \right]^{\frac{2}{m-1}} \right]^{-1} \tag{6}$$

Fuzzy overlap refers to how fuzzy the boundaries between clusters are and can take a value above 1. The higher values of this parameter will result in fuzzier clusters.

3.1.2. Hierarchical clustering

Hierarchical clustering is a more flexible and deterministic algorithm than K-centers method. The hierarchical algorithm produces a tree or dendrogram by either agglomerative (bottom-up) or divisive (top-down) methods. In the agglomerative method, initially each instance is classified as a cluster and then clusters are merged iteratively to build a bottom-up hierarchy of the clusters until a single root cluster is reached. The divisive approach, on the other hand, starts with a single root cluster and splits it into subclusters continuously, generating a top-down hierarchy of clusters. Fig. 2 displays the hierarchical tree or dendrogram of an agglomerative clustering method. This formed hierarchy can be cut at any given level which allows obtaining the corresponding clusters. This is the main advantage of hierarchical algorithms that makes them considerably different from partitioning methods which require the number of clusters before starting the algorithm. Also, hierarchical clustering has fewer assumptions about the distribution of data. However, it should be noted that hierarchical clustering is generally more computationally expensive than K-means (time complexity of $O(n^3)$ where n is the number of observations compared to the linear complexity of K-means).

3.2. Self-organizing map

SOM is an unsupervised artificial neural network that projects the original input space to a reduced output space [23]. . It produces a graphical representation of the data which allows an easy evaluation of the results and grouping them into clusters by visual inspection [24,57]. The SOM consists of a grid containing $W_1 \times W_2$ map units (neurons). The original h-dimensional data vector is transformed to a (typically) bi-dimensional space where similar observations in the input space are mapped into nearby units. Each unit i is represented by a prototype vector $w_i = [w_{i1}, w_{i2}, \ldots, w_{ih}]$, which has the same dimension of input vectors (h). The number of units can vary from a few dozen up to several thousand [57]. Each unit is connected to adjacent units by a neighborhood relation, which determines the topology or structure of the map.

The SOM is trained iteratively. In each training step, a sample vector x from the data set is picked out randomly. The distance of this vector and all prototype vectors are calculated and the unit whose prototype vector is closest to x is selected as the best-matching unit (BMU) or winning unit:

$$||x - w_b|| = \min_i ||x - w_i|| \tag{7}$$

The learning algorithm updates the weight of the winning unit and also the weights of its adjacent units. The prototype vector of unit i is updated using the following equation:

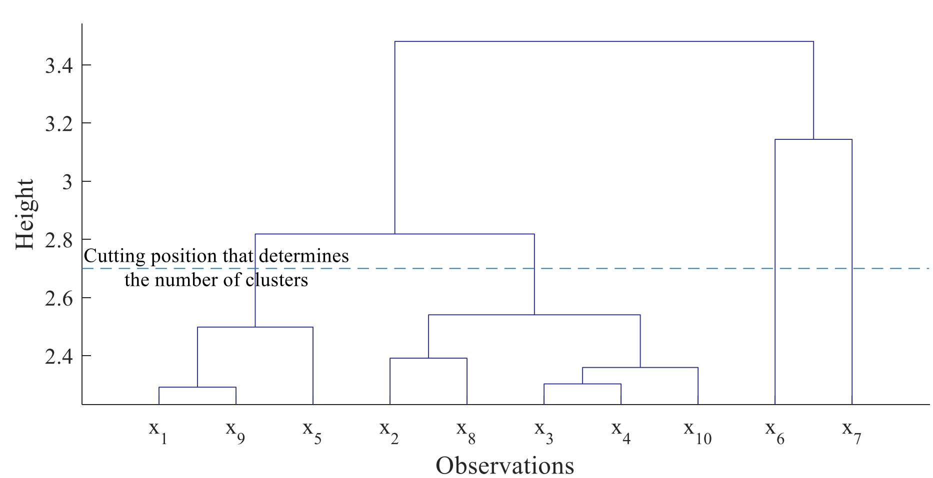

Fig. 2. Dendrogram of hierarchical clustering method.

$$w_i(t+1) = w_i(t) + \alpha(t)h_{bi}(t)[x(t) - w_i(t)] \tag{8}$$

where t represents time, $\alpha(t)$ is the learning rate or adaptation coefficient at time t, and $h_{bi}(t)$ is the neighborhood kernel (neighborhood symmetrical function) around the winner unit b. The units that are topologically close to the winning unit b are activated using $h_{bi}(t)$:

$$h_{bi}(t) = \exp\left(-\frac{\|r_b - r_i\|^2}{2\sigma_t{}^2(t)}\right) \tag{9}$$

where r_i represents the coordinates of unit i in the SOM grid and $\sigma(t)$ is the neighborhood radius function. Both $\alpha(t)$ and $\sigma(t)$ decrease monotonically with time. Therefore, the neighborhood size of each unit reduces in each training step and, finally, it ends with a single unit.

3.3. Probabilistic and generative models

In the model-based clustering, it is assumed that instances arise from a distribution that is a mixture of several components. The problem is to estimate the parameters of each component and identify which component produced each observation [54]. This process leads to the clustering of the data. In practice, the attention is mostly paid to parametric mixture models, where all the components are from the same family of distributions. Gaussian (normal) distributions are by far the most commonly used representation in the model-based clustering. In this case, the mixture model is the Gaussian mixture model (GMM), where components are Gaussian distributions with different means and variances.

The mathematical formulations of GMM are illustrated in the following [58–60].

Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a set of n observations. The variable $\mathbf{x}_i$ is assumed to be distributed according to a mixture of K components. The probability density function (or mixture distribution) of $\mathbf{x}_i$ can be written as:

$$p(\mathbf{x}_i|\boldsymbol{\theta}) = \sum_{m=1}^{K} \alpha_m p(\mathbf{x}_i|\boldsymbol{\theta}_m) \tag{10}$$

where $\alpha_1, \ldots, \alpha_K$ are the mixing probabilities and each θ_m is the set of parameters which define the mth component. $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K, \alpha_1, \ldots, \alpha_K\}$ is the complete set of parameters. α_m must satisfy:

$$\alpha_m \geq 0, \quad m = 1, \ldots, K, \quad and \quad \sum_{m=1}^{K} \alpha_m = 1 \tag{11}$$

In the Gaussian mixture model, each component is specified by the parameters of a multivariate Gaussian distribution:

$$p(\mathbf{x}_i|\boldsymbol{\theta}) = p(\mathbf{x}_i|\alpha, \mu, \Sigma) = \sum_{m=1}^{K} \alpha_m \mathcal{N}(\mathbf{x}_i|\mu_m, \Sigma_m) \tag{12}$$

where,

$$\mathcal{N}(\mathbf{x}_i|\mu_m, \Sigma_m) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}_i - \mu_m)^T \Sigma_m^{-1} (x_i - \mu_m)\right\}}{(2\pi)^{\frac{D}{2}} \ |\Sigma_m|^{\frac{1}{2}}} \tag{13}$$

For a D-dimensional vector x, μ and Σ are the D-dimensional mean vector and the $D \times D$ covariance matrix respectively. In the case of a single variable x, (13) is reduced to:

$$\mathcal{N}\left(\mathbf{x}_i|\mu_m, \sigma_m^2\right) = \frac{\exp\left\{-\frac{1}{2\sigma_m^2}(\mathbf{x}_i - \mu_m)^2\right\}}{\left(2\pi\sigma_m^2\right)^{\frac{1}{2}}} \tag{14}$$

where, μ and σ^2 are the mean and variance respectively.

Usually, Expectation-Maximization (EM) algorithm is utilized for parameters estimation of the model. Bayesian information criterion (BIC) and Akaike's information criterion (AIC) are the main criteria for choosing the best number of components (clusters) [42].

3.4. Literature review

K-center family methods are by far the most common approaches used in the literature. Ref. [29] utilizes the electricity consumption usage of 103 residential dwellings with the time resolution of 1 min. The data are firstly averaged over each hour to build up the hourly load profiles and a representative load profile is created for every home for each season of the year. K-means is applied to partition the dwellings into two clusters for each season. A similar procedure is followed in Ref. [26], where the data of just working days are used. An improved FCM is used in Ref. [32] to cluster the electricity consumption data of one month of 938 households in China. FCM is also applied along with the K-means and hierarchical algorithms to the consumption data of a group of South Korean high voltage customers [22].

Ref. [52] uses a fuzzy Gustafson-Kessel clustering for identification of non-technical losses. This clustering method can be seen as an extension of regular FCM in which Euclidean distance is replaced by a dissimilarity measure that results in hyperellipsoidal clusters. This method can provide greater flexibility for the shape of clusters.

Clustering of a set of LV substations in the United Kingdom is performed in Ref. [36] using a hierarchical algorithm. 15 different loading conditions are considered by dividing the year into 5 seasons and 3 types of days.

Clustering with SOM has been done in several studies. In Ref. [8] an SOM-based clustering of Finnish electricity consumers is presented. The aim is to introduce a visual data mining driven application to exemplify the potentials of real-time business intelligence for electricity companies. In Ref. [39], besides the annual electricity usage, various physical characteristics and property features are used for the clustering. On the other hand, an SOM-based methodology used in Ref. [38] to segregate customers based on three different sets of indices: information on the clients' climate areas, quantitative information extracted from daily load patterns, and quantitative and qualitative information obtained from questionnaires.

GMM models are recently used in some studies to cluster electricity customers. *Labeeuw et al.* [11] analyse the electricity demand of 58 households. They favour a GMM approach to K-means, FCM and hierarchical algorithms because of the smoothing effect of GMM and the need for data upsclaing. Moreover, GMM is used for segregating 3600 residential customers [42] and its performance is compared with K-means [43] and hierarchical and K-means methods [35]. A few other algorithms such as adaptive K-means [17,45,46], follow the leader [10,47], K-shapes [30], and density-based spatial clustering of applications with noise (DBSCAN) [48,49] have been also used in the literature for clustering of load data.

K-means method needs to determine the number of clusters before running the algorithm. Instead of trying out several candidate values for K, an adaptive K-means algorithm can be utilized to determine the final number of clusters during the cluster formation process [61]. This algorithm starts with an initial best guess $k = k_0$, but permits changing it on the go whenever it appears too large or too small for a given dataset [7]. Kwac *et al.* [17] proposed a clustering methodology which combines adaptive K-means and hierarchical clustering. Firstly, adaptive K-means is applied to segregate customers to a large number of clusters. In the next stage, a hierarchical clustering merges those clusters that are highly correlated.

Fahiman *et al.* [30] compare the performance of K-means with a newly introduced clustering method called K-shapes algorithm to cluster several thousands of dwellings. K-shapes considers the shape of time series during clustering rather than treating the observations as independent attributes. It consists of three main components [62]: 1) a shape-based distance measure which is based on a cross-correlation measure, 2) time series shape extraction which defines a centroid based on an optimization problem, 3) shape-based time series clustering which clusters time series data based on the last two steps. The authors claim that K-shapes significantly outperforms the K-means with respect to clustering accuracy.

DBSCAN technique clusters those observations which are closely packed together and specifies the data points in low-density regions as outliers. It is employed in Ref. [48] for clustering customers' load patterns and designing customized tariffs for each household based on its dominant load pattern. Ref. [49] uses an adaptive DBSCAN to find a typical consumption pattern in each season for each individual customer. K-means is then applied to group these typical load curves into several clusters.

Biclustering techniques are used in Ref. [63] to analyse the building consumption data. The biclustering allows simultaneous clustering of both the observations (buildings) and features (days). The proposed method obtains subgroups of buildings that exhibit a similar consumption pattern during a specific time period.

Furthermore, Markov model is used in Ref. [46] to capture the dynamics of the load data and transfer the large data set of load curves to some state transition matrices which are used for clustering. Ref. [64] suggests that when weather effects are accounted for, household consumption is solely based on the occupancy. Here, occupancy refers to socio-demographic factors and the lifestyle. A hidden Markov model (HMM) framework is utilized to infer the occupancy states from consumption data. Spectral clustering is used to segment the collection of HMMs.

4. Discussion on the algorithms

Each of the presented methods for the clustering has its advantages and disadvantages. In addition, different considerations need to be taken into account before applying the clustering algorithms. In this section the affecting parameters of each of the clustering methods are discussed.

4.1. K-center family

For K-centers methods various parameters including the number of clusters, initial centers, and the dissimilarity measure must be initially determined. Each of these parameters can affect the final outcomes of the clustering. Initial centers can be selected by a random fashion among the instances of the data set [23]. The random selection of cluster centers may affect the final cluster formations. CVI measures can be used to find the best choice for multiple runs of the clustering algorithm with different random initial centers [7]. In addition, numerous initialization methods are also proposed for the selection of the centers. Ref. [65] provides a thorough study of various initialization methods and compares their performance for real and synthetic datasets. Table 2 compares the main characteristics of the three main K-centers methods.

For FCM, the degree of fuzzy overlap needs to be decided. The selection of parameter m has been the subject of many studies in the data science literature [66–68]. These studies follow different approaches for the selection of optimal m and suggest different values and ranges for that. For example [68,69], propose the selection of m from the range of [1.5, 4] and [1.5, 2.5] respectively. The most frequently used and accepted value in various applications is $m = 2$ [69,70] which is also the suggested value in MATLAB software. In this paper, the value of m is selected based on the results of CVIs. FCM computation time is longer compared to K-means since the degrees of membership need to be updated at each step.

4.2. Hierarchical

In almost all the studies in the power system domain, the agglomerative approach is used as the preferred hierarchical method. For agglomerative methods, the formation of clusters is based on the similarity measures. Firstly, using a distance criterion, a similarity matrix D

Table 2
Characteristics of main methods of K-centers family.

Method	Calculation of center	Best dissimilarity measure	Disadvantages	Advantages
K-means	Center is calculated as the mean of members of the cluster	Euclidean	Not applicable to discrete attributes; Handling the data containing outliers; Handling asymmetrically distributed data; Cluster centers might not be similar to any instance.	Easy to implement and efficient.
K-medians	Center is selected as the median of members of the cluster	Manhattan	More costly to calculate; Cluster centers might not be similar to any instance.	More robust to asymmetric distributions and outliers; Not skewed so much by extremely large or small values.
K-medoids	Center is the cluster member that is the least dissimilar to other cluster members, on the average	Different measures can be used	More expensive computationally than K-means and K-medians.	Robust with respect to noise and outliers; Guarantees convergence.

is built in which $d_{ij} \in D$ shows the distance between the observation i and the observation j. In the next step, based on this similarity matrix, instances are grouped into clusters using a *linkage criterion*. The linkage is an evaluation function which indicates the best candidates for merging. Therefore, at each level the closest sets of clusters are merged until the final cluster (which contains all the observations) is obtained.

Some of the most important linkage criteria and their features are reported in Table 3 [7,54,60,71].

In single linkage, the similarity of two clusters is determined based on the similarity between their most similar members. On the other hand, in complete linkage the similarity of two clusters is measured as the similarity of their most dissimilar members. Average linkage method (sometimes called UPGMA which stands for "unweighted pair group method using arithmetic averages") diminishes the problems associated with single and complete linkage methods by considering the similarity between all pairs of instances present in both of the clusters. So, the average dissimilarity between instances from two clusters serves as the dissimilarity between the clusters. Another method called centroid linkage clustering computes the dissimilarity between the center for cluster i and the center for cluster j. In addition, linkages can be defined based on a specific quality criterion or objective function. The most famous one among these linkage methods is Ward criterion with the objective to minimize the total sum of squared dissimilarities between cluster members and cluster centers for all the clusters. In other words, for every two clusters C_i and C_j, Wards' criterion measures the increase in the value of sum of squared errors for the clustering obtained by merging them into $C_i \cup C_j$.

Likewise the dissimilarity measure, the choice of linkage can also have a significant impact on the final clustering outcomes.

4.3. SOM

The parameters of neural network such as the learning rate and the radius of the neighborhood might slightly affect the partitioning of the data set by SOM. In addition, the SOM results depend on the population of neurons as well as the topology or structure of the map. For N data points, the number of neurons is recommended to be between $5\times \sqrt{N}$ to $20 \times \sqrt{N}$ [24]. In addition, different topologies can be selected for the neural lattice. Traditionally, hexagonal or rectangular arrangements of neurons are chosen, in them internal neurons are bounded by six and four adjacent neurons respectively.

A visual inspection of the generated SOM map can give an initial idea of the number of clusters. This is particularly performed using the unified distance matrix (called U-matrix) that shows the distances between prototype vectors of adjacent units and can visualize the cluster structure of the SOM. However, this process does not guarantee the best results. Sometimes a two-level approach is used in which the prototypes are formed using the SOM and then, a clustering algorithm is applied on the prototypes to obtain the final clusters [25,56]. This is especially beneficial when the data set contains a large number of data points.

4.4. GMM

GMM is able to model both continuous and categorical data which is an advantage of this method over many clustering techniques such as K-means [42]. Interested readers are referred to Ref. [72] for technical explanations and examples of applying mixture models on mixed continuous and categorical variables. GMM requires the number of clusters to be specified before fitting the model. In addition, for applying GMM, the parameters of covariance matrix of each component need to be specified. The structure of the covariance defines the shape of a confidence ellipsoid over a cluster. The detailed technical discussion of the various covariance structures is beyond the scope of this paper. Interested readers are referred to Refs. [73,74] for practical implementations in Matlab and R programs, respectively. In this paper, different configurations of covariance matrices are examined and their effects on clustering results are studied. Firstly, two different structures for covariance matrices, which specify the cases with correlated and uncorrelated predictors, are considered. In the case studies, the former and latter cases are denoted as *full* and *diagonal*, respectively. Secondly, the effects of *shared* or *unshared* covariance matrices among all components are investigated. Each combination of these parameters defines the orientation and shape of ellipsoids. Since the appropriate covariance structure and number of components (clusters) are not known, the information criteria like AIC or BIC are used to compare different models. Lower values of AIC and BIC indicates better models with the most suitable parameters or the best number of components.

Furthermore, EM algorithm that fits the GMM is sensitive to initial conditions and might converge to a local optimum. To ensure global convergence is achieved, the algorithm can be run repeatedly with different initial conditions [42]. The initial component parameters can be decided in various ways, for example, in a random fashion or by applying a K-means clustering to choose a number of observations [75].

5. Application of clustering algorithms to the load curves of customers

In this section, the impacts of discussed parameters of presented algorithms on clustering of daily loads curves of electricity customers are discussed. The data is part of the smart metering trial carried out by Commission for Energy Regulation (CER) in Ireland [76]. In this project, the consumption of each customer was recorded every half-hour.

A pre-processing of data including the correction for daylight saving time changes in spring and autumn and exclusion of some special

Table 3
Linkage criteria for hierarchical clustering.

Linkage criterion	Description	Features
Single	$\min_{x \in C_i, y \in C_j} d(x,y)$	Neglects the overall cluster structure; Sensitive to noise and outliers; Capable of clustering non-elliptical shaped groups of data points; Not affected by the monotone transformations (like the logarithmic transformation) of the original data
Complete	$\max_{x \in C_i, y \in C_j} d(x,y)$	Obtains more compact shaped clusters; Sensitive to outliers
Average	$\frac{1}{n_i.n_j} \sum_{x \in C_i, y \in C_j} d(x,y)$	A compromise between single and complete linkages; Computationally expensive, especially for large datasets; Noise resistant
Centroid	$d(c_i, c_j)$	Does not have monotonic property i.e. a merged cluster might become closer to other clusters than its descendants which is usually undesirable.
Ward	$\sqrt{\frac{n_i.n_j}{n_i + n_j}} d(c_i, c_j)$	Not directly based on similarities between data points of the two clusters, instead works based on an objective function

C_i : ith cluster; c_i : center of cluster C_i; n_i: number of data points belonging to cluster i.$d(x,y)$ = distance between the objects x and y

Table 4
List of CVIs.

Cluster validity index	Descriptions	Rule
$MSE = \frac{1}{N}\left(\sum_{k=1}^{K}\sum_{x_i\in C_k} d^2(x_i, c_k)\right)$		min
$SIL = \frac{1}{K}\sum_{i=1}^{K}\zeta_i$, where, $\zeta_i = \frac{1}{n_i}\sum_{x_j\in C_i} s(j)$	$s(i) = \frac{\mathscr{B}(i) - \mathscr{A}(i)}{\max(\mathscr{B}(i), \mathscr{A}(i))}$, where, $\mathscr{A}(i) = \text{within} - \text{cluster mean distance} = \frac{1}{n_k - 1}\sum_{\substack{j\in C_k \\ j\neq i}} d(x_i, x_j)$, $\mathscr{B}(i) = \text{the smallest of mean distances to other clusters} = \min_{k^{'}\neq k}\left(\frac{1}{n_{k^{'}}}\sum_{j\in C_{k^{'}}} d(x_i, x_j)\right)$	max
$DBI = \frac{1}{K}\sum_{i=1}^{K} max_{j\neq i}\left\{\frac{\left[\frac{1}{n_i}\sum_{x\in C_i} d(x, c_i) + \frac{1}{n_j}\sum_{x\in C_j} d(x, c_j)\right]}{d(c_i, c_j)}\right\}$		min
$MIA = \sqrt{\frac{1}{K}\sum_{k=1}^{K} {d_{C_k}}^2}$	d_{C_k} = the distance between cluster center c_i and the member of the cluster $i = \sqrt{\frac{1}{n_k}\sum_{x_i\in C_k} d^2(x_i, c_k)}$	min
$WCBCR = \frac{\sum_{k=1}^{K}\sum_{x_i\in C_k} d^2(x_i, c_k)}{\sum_{1\leq i<k}^{K} d^2(c_i, c_k)}$		min
$Dunn = \frac{\min_{i\neq j} d_{ij}}{\max_i D_i}$	d_{ij} : the distance between the closest instances of two clusters (separation) D_i : the largest distance between two instances that belong to the cluster i (diameter)	max

N: Number of observations (load curves); K: number of clusters; c_i : center of cluster i; $d(x,y)$ = distance between the objects x and y; n_i: number of data points belonging to cluster i.

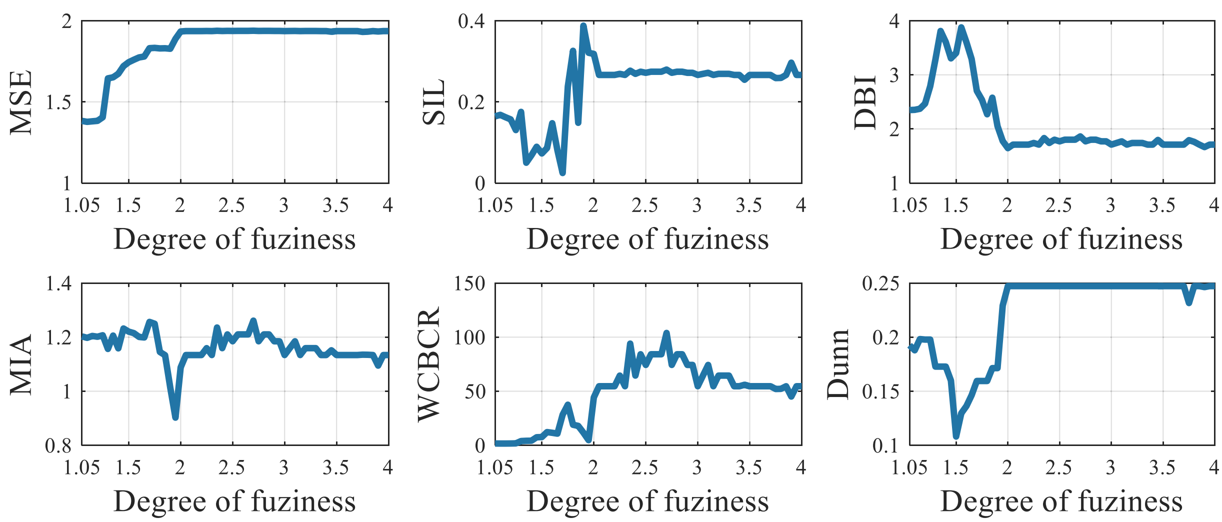

Fig. 3. Effect of fuzziness degree on the clustering results for FCM method.

holidays is initially performed. In the following, most of the case studies are carried out for clustering of 356 daily load curves of one residential customer (sections 5.2 to 5.6). In addition, to investigate the clustering of a large number of users, clustering techniques are also applied on a data set comprising more than 4000 customers (section 5.7). Since the aim is to cluster the daily load curves based on their shapes, each daily load curve is normalized based on the maximum consumption of that day. Without normalization of daily curves, the resulted clusters will only reflect load magnitudes. The majority of clustering studies mostly focus on the shape of load curves. The effect of customers' consumption values can be examined with other methods. For instance Ref. [17], clusters the customers based on their load shapes. It also fits a mixture of log-normal distributions to the daily consumption values of customers and based on that, divides them into heavy, moderate and light energy users. Some studies use the expert knowledge-based feature extraction to account for the consumption values. For example, in Ref. [42], the relative average power in each time period of the day over the entire year is calculated. In the next step, customers are clustered based on these new features. R software which is a well-known data analysis tool and Matlab packages are used for the analyses and simulations. Firstly, the effect of different parameters of each algorithm is investigated and then, the performances of clustering methods are compared.

5.1. Cluster validity indexes

As explained in the preceding sections, parameters of each clustering method and initial conditions affect the final results and hence, clustering outcomes should be evaluated considering a range of parameters and conditions. CVIs can be used to study various aspects of clustering results and to compare the methods. In the electricity customer categorization, the CVIs may be used for different purposes, mainly:

- To determine the suitable number of customer clusters [15,25].
- To compare the performance of different clustering techniques [10, 20,77].
- To investigate the effect of method parameters on clustering results [78,79].
- To evaluate the performance of clustering when some attributes (features) are added or removed [47,80].

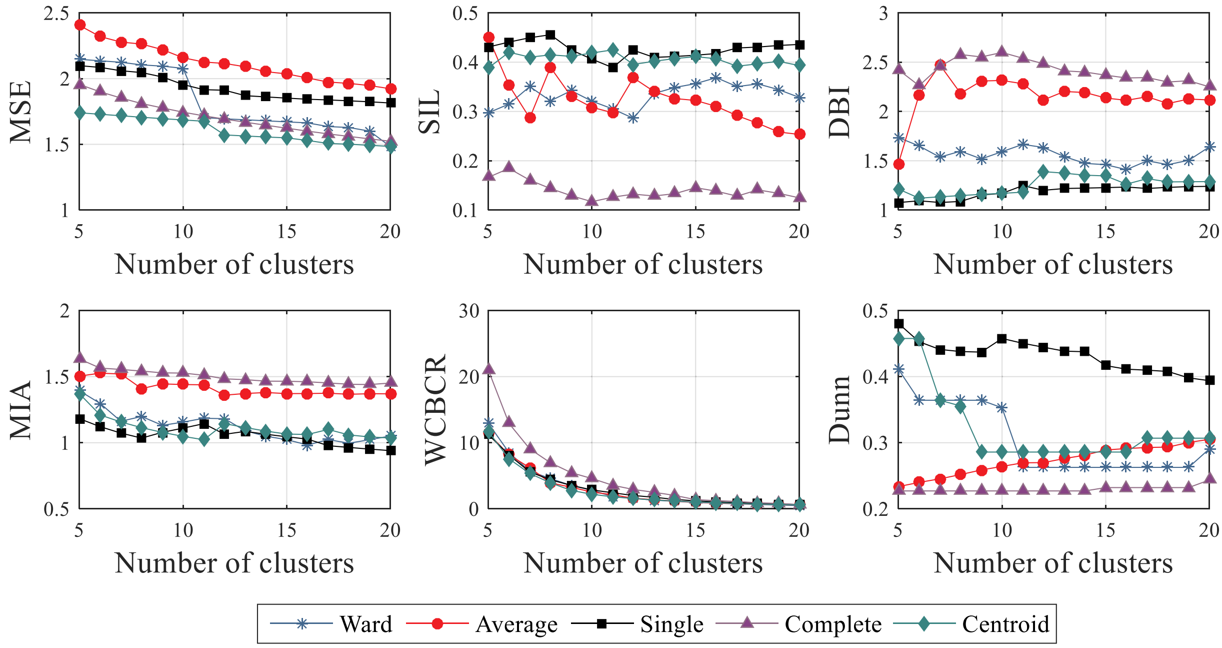

Fig. 4. Comparison of hierarchical algorithms.

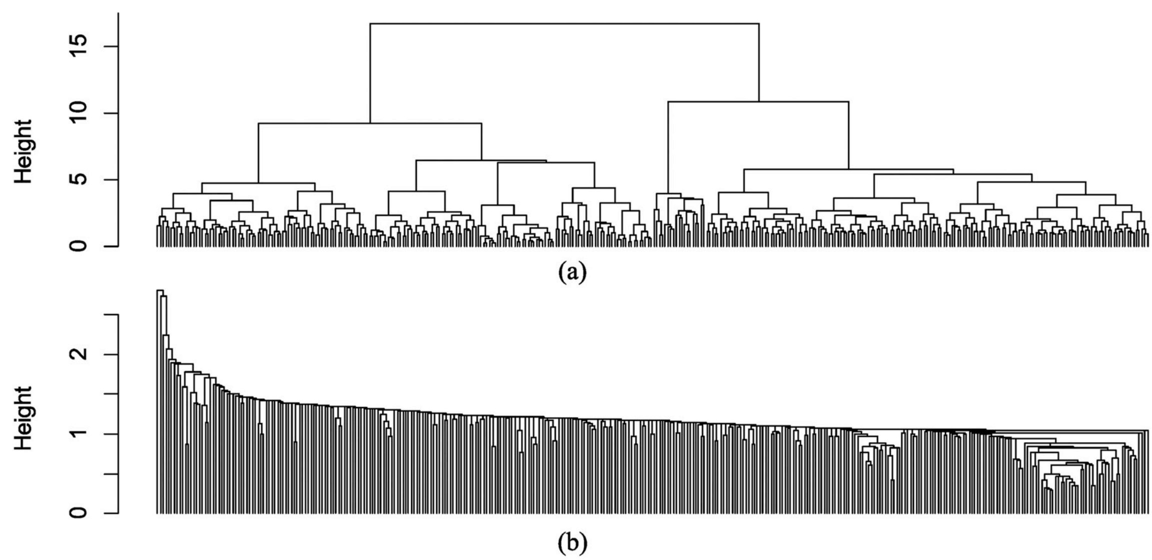

Fig. 5. Dendrograms of (a) ward method and (b) single method.

In this paper, the comparison of parameters and methods are conducted based on 6 different CVIs: mean square error (MSE), Silhouette index (SIL), Davies-Bouldin indicator (DBI), mean index adequacy (MIA), the ratio of within-cluster sum of squares to between-cluster variation (WCBCR), and Dunn index. The definitions of these CVIs are given in Table 4. The *rule* in this table refers to the interpretation of the CVIs for choosing among the results. For example, the minimum value of DBI indicates the best result. Besides these CVIs, AIC is utilized for evaluating the results of GMM method.

5.2. Fuzzy c-means

For FCM, the main parameter is the fuzziness degree, characterised by parameter $m$in Eq. (4). Fig. 3 shows the effect of this parameter on clustering results where the number of clusters is fixed at 10. The value of m is changed from 1.05 to 4 at the steps of 0.05. Since the initial centers are selected randomly, the clustering results slightly change in each execution of the method. Thus, for each value of fuzziness degree, the clustering is carried out ten different times and the outcomes are averaged for each CVI. As this figure shows, the CVIs indicate that the best results happen at around 1.9 to 2.

5.3. Hierarchical clustering

Here, 5 different hierarchical methods with different linkage criteria are compared for varying number of clusters and the results are displayed in Fig. 4. The single and centroid linkage models are selected as the best models by all CVIs except for MSE. Ward linkage also shows good performance having relatively low values for DBI, MIA, and WCBCR, and high values for SIL and Dunn. However, further inspection of the clusters shows that single and centroid methods assign most of the daily load curves to only one cluster. It can be observed by the

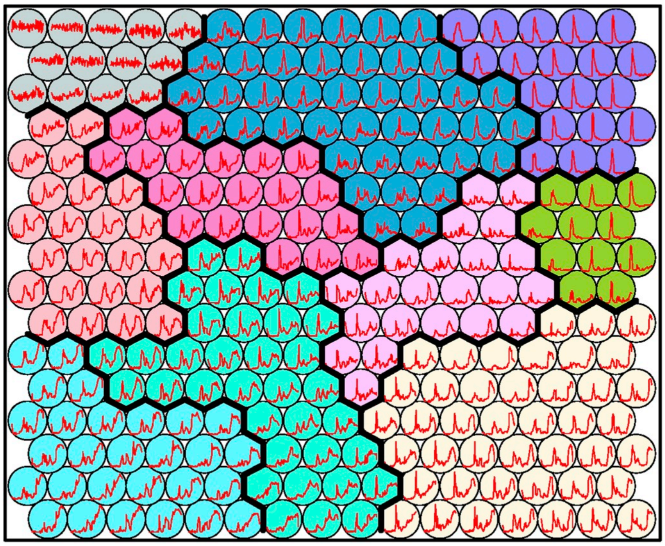

Fig. 6. a 16×16 SOM grid and the corresponding clusters after applying the hierarchical method.

dendrograms of the single and ward methods as shown in Fig. 5. For this reason, ward method which well separates load curves into different clusters is preferred.

5.4. SOM

To cluster the load curves, the SOM in conjunction with a hierarchical clustering method is used. The parameters under study are the neurons population and topology of the neural lattice. For 356 load patterns, the grid size changed accordingly from 10×10 ($5 \times \sqrt{356} \approx 94$) to 20×20 ($20 \times \sqrt{356} \approx 377$). The width, W_1, and height, W_2, of the grid are assumed to have the equal size. For each grid size, the effects of hexagonal and rectangular topologies are studied.

Fig. 6 displays a sample 16×16 SOM grid which is divided into 10 clusters after applying the hierarchical algorithm. To compare different configurations, the values of CVI indexes are calculated in each case as shown in Fig. 7. In this case, superior results can be observed for the grid size 18×18 with hexagonal topology.

5.5. GMM

Generally, GMM produces the best results when the number of variables is limited. Since the load curves have 48 variables (half-hour

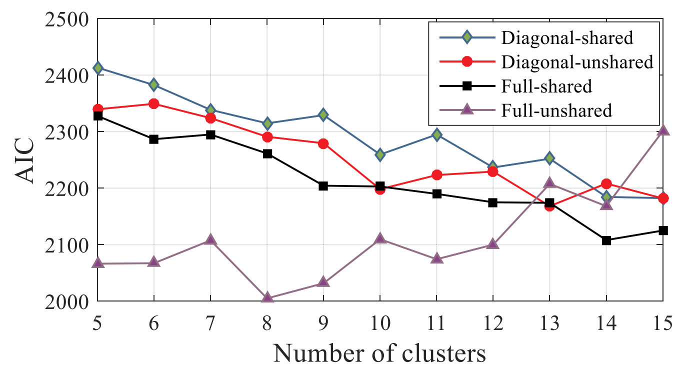

Fig. 8. Effects of parameters of covariance matrix on the GMM clustering.

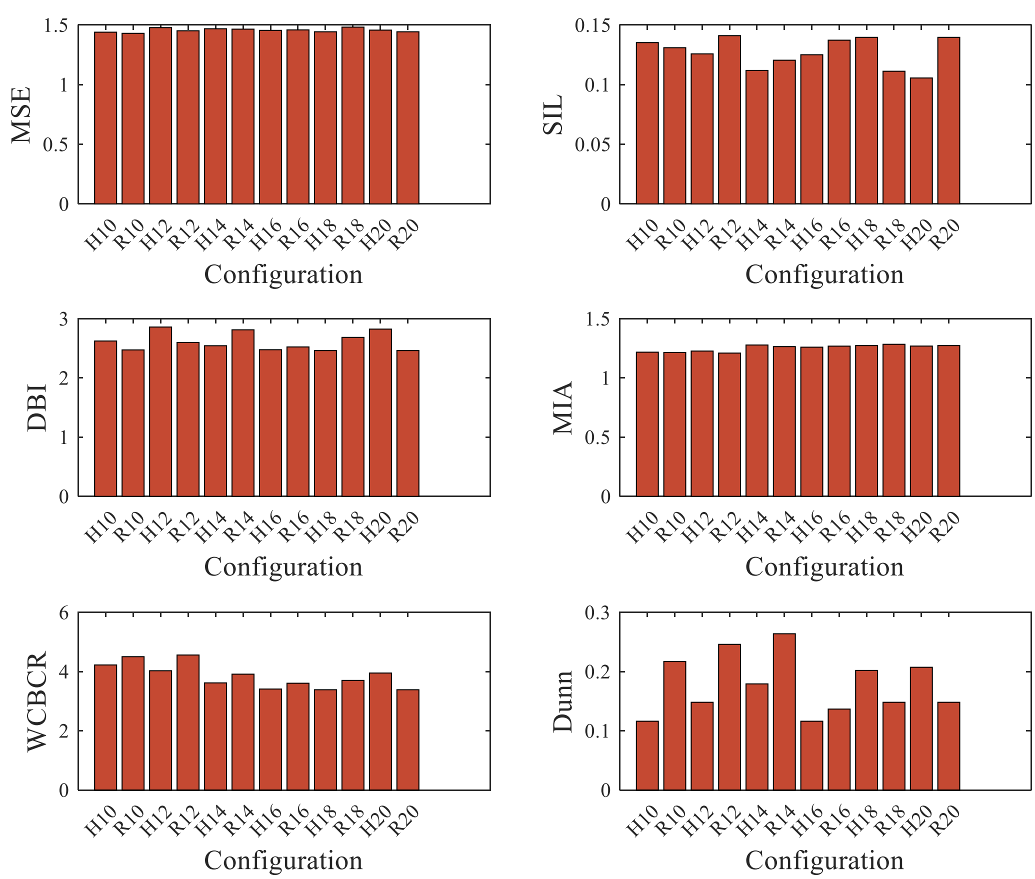

Fig. 7. Effect of grid size and topology on the two-level clustering of load curves using SOM and hierarchical method (R: Rectangular, H: Hexagonal).

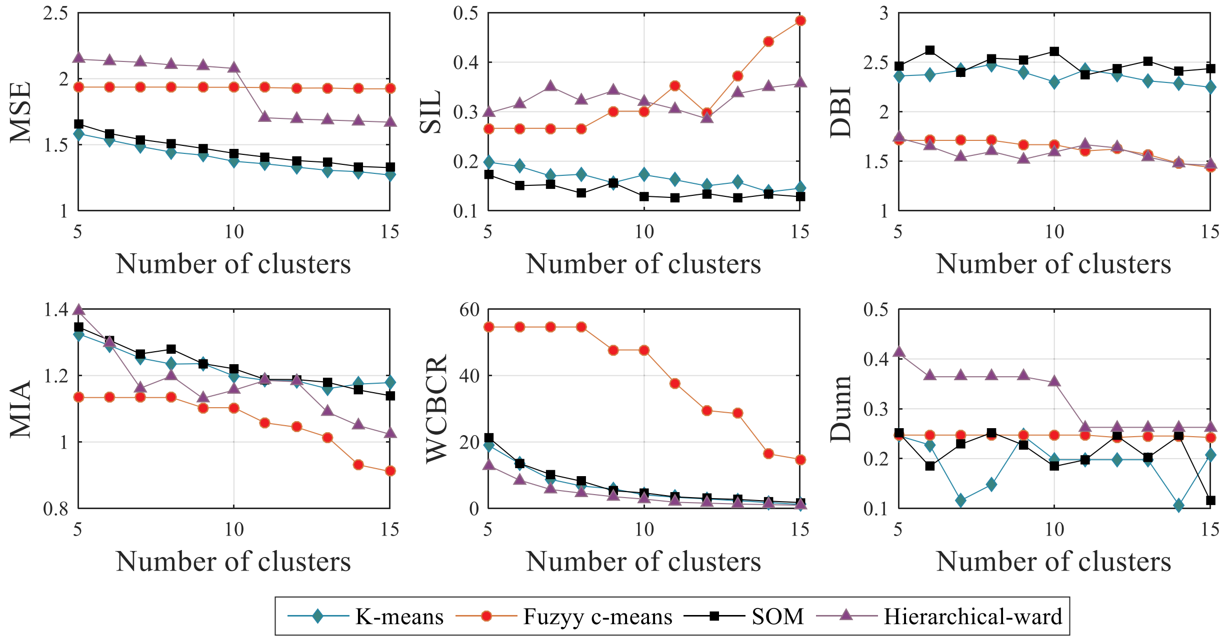

Fig. 9. The CVI values for four different clustering algorithms.

recordings) applying GMM might not lead to the promising results. To this end, prior to GMM clustering, use of an indirect clustering approach could be advantageous. For instance, in Ref. [42] the authors apply GMM on a set of features which are extracted from the load data.

Here, the principal component analysis (PCA) is used to reduce the size of the input data. Therefore, each load pattern is represented by a limited number of components. The selection of the best number of components is described in the next sections. GMM is applied to the PCA components and the impacts of parameters of covariance matrix (full vs. diagonal and shared vs. unshared matrixes) on final results are investigated using AIC as depicted in Fig. 8. The lowest values of AIC occur for full-unshared method. By increasing the number of clusters, diagonal-shared and full-shared observe a decreasing trend while the AIC for full-unshared increases gradually. The results also suggest the best segmentations with 8 clusters.

5.6. Comparing clustering methods

In the previous sections, some of the most important parameters of different clustering methods are illustrated and their effects on clustering results were analysed. In this section, four major clustering algorithms including K-mean, FCM, hierarchical, and SOM are compared and the formed clusters are analysed. The final aim is to determine the algorithm which can better reveal the various patterns of consumption behavior and form clusters that are more compact and well separated from each other.

The parameters of the methods are selected based on the analysis in previous sections. Correspondingly, the fuzziness degree is set to 1.9 for FCM, hierarchical clustering with ward linkage is chosen, and SOM is performed for a grid size of 18×18 with hexagonal topology. Fig. 9 shows the CVIs for the selected algorithms.

It can be observed that hierarchical algorithm shows superior results for this special data set which contains the daily load patterns of a customer. Furthermore, five CVIs indicate SOM as the worst clustering algorithm for this case study. For this specific value of fuzziness degree, the obtained results show the good performance for fuzzy clustering. However, it should be noted that in some cases, the results of FCM are sensitive to small changes in degree of fuzziness. Therefore, while using FCM clustering, it is necessary to study various fuzziness degrees for different number of clusters.

Based on the CVI values for all the algorithms, it can be seen that the optimum number of clusters falls into the range of 8–10 clusters. Eight clusters appear to produces the satisfactory results since adding more clusters does not improve the results significantly. This finding is in accordance with the GMM clustering outcomes. Generally, the final number of clusters is decided based on the pre-defined objectives and needs. In practice, the outcomes of electricity customer clustering will be used by the utilities for improving different applications such as demand response programs and tariff design. Therefore, typically, the number of clusters cannot be very large.

Fig. 10 and Fig. 11 display the final clusters which are formed by each method when the number of clusters is set to 9. The center of each cluster is shown by the red line and is computed by averaging on the load patterns belonging to the cluster. It can be observed that clustering can reveal various distinct consumption patterns among the daily load curves of the customer. Particularly, the following patterns are distinguishable (Here, the K-means results are examined. The analysis is similar for the other methods.):

- Morning peak (cluster #1)
- Mid-day peak (cluster #5)
- Morning and afternoon peaks (cluster #4)
- Morning and night peaks (cluster #3)
- Morning and late night peak (cluster #8)
- Late night peak (cluster #9)
- Variable consumption pattern (cluster #6)

Cluster #7 resembles to cluster #1; however, its corresponding peak has less magnitude and happens at earlier hours. Moreover, cluster #2 characterizes the high consumption during midnight and a local peak at around 10 a.m.

5.7. Clustering of a large number of electricity customers

Electricity companies desire to segregate their huge number of customers into certain classes based on the daily load patterns. However, as we noted in the previous sections, the daily load patterns of a certain customer might change significantly from a day to another day. This makes the clustering of customers challenging. To overcome this problem, one common approach is to cluster the customers based on their

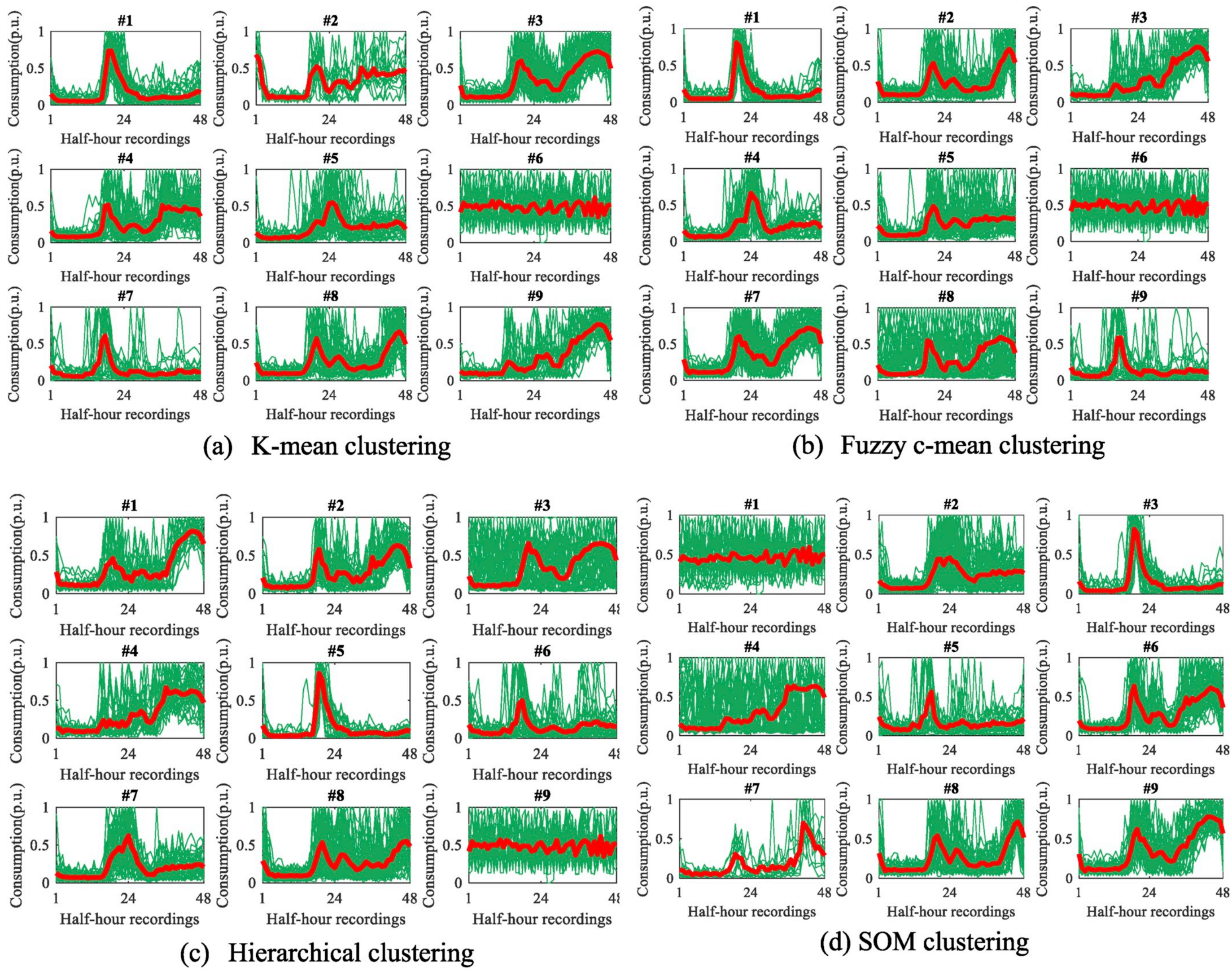

Fig. 10. Final clusters of 4 different clustering algorithms.

representative load patterns (RLPs). For this purpose, initially, a set of different *loading conditions* are defined based on the user preferences, climate conditions, and other affecting parameters. Loading condition refers to the type of the day and the season that the data are recorded. The daily load data of each customer in a specific loading condition can be organized to represent the customer's consumption by means of just one load pattern [10,81]. To this end, the daily load patterns are combined instant-by-instant based on a statistical criterion like mean or median to create a representative load diagram. Finally, the normalized RLP of each customer can be made by normalizing the original representative load diagram with respect to a reference power which is usually assumed as the maximum consumption [23]. This allows the clustering of those customers with similar load shapes into a class, regardless of the actual quantities of consumptions.

Here, the analysed data set comprises load data of 4141 customers over a year. Since the customers usually have different consumption behavior on the weekends compared with weekdays, the data set is divided into weekdays and weekends (two loading conditions). Fig. 12 and Fig. 13 show the final clusters (obtained by a hierarchical algorithm) for weekdays and weekends, respectively. The number of RLPs which belong to each cluster is also displayed in these figures. In order to identify various consumption patterns among customers, a sufficiently big number of clusters is selected.

It can be seen that the difference between the weekday and weekend consumption behavior is significant. For weekday clusters, generally a small peak happens in the morning and the major peak occurs in the evening and nights. Specially, this pattern is clearly visible for clusters #8, # 12, #1, and # 4 that have the highest number of RLPs and totally account for around 40% of load shapes. On the other hand, weekend clusters and particularly, the clusters with the highest number of members i.e. clusters #4, #12, #1, and #6 have a late peak around midday or early afternoon. Furthermore, it is also noticeable that the magnitude of the afternoon peak is higher or equal of the night peaks. In addition it can be observed that the consumption level is higher compared with the weekday consumption. Such a difference among usage behavior is predictable since, in the weekends, the residents usually wake up late and spend most of the day in the home while in the weekdays they leave their homes in early mornings and are outside the home for most of the day.

5.8. Method comparisons based on the computation time

In this section, the processing time of the algorithms are compared for the clustering of the larger dataset. Tables 5 and 6 show the results for the clustering methods and CVI measures, respectively. The results are the average values obtained by running the algorithms 10 different times. The simulations are carried out on a personal computer Intel® CORETMi5 with processors clocking at 2.3 GHz and 8 GB of RAM.

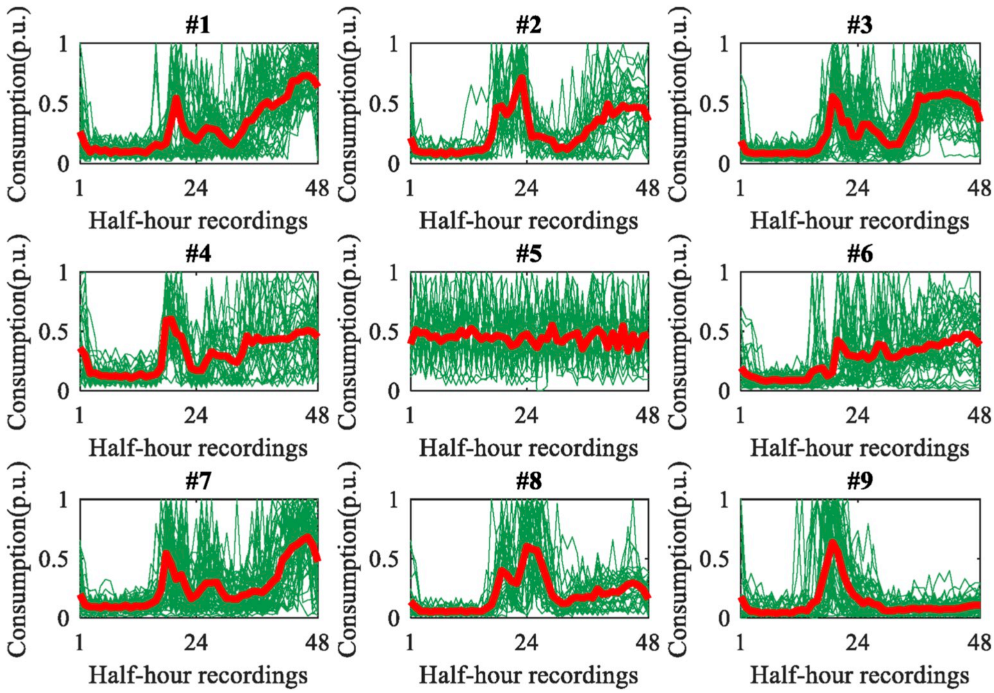

Fig. 11. GMM clustering results.

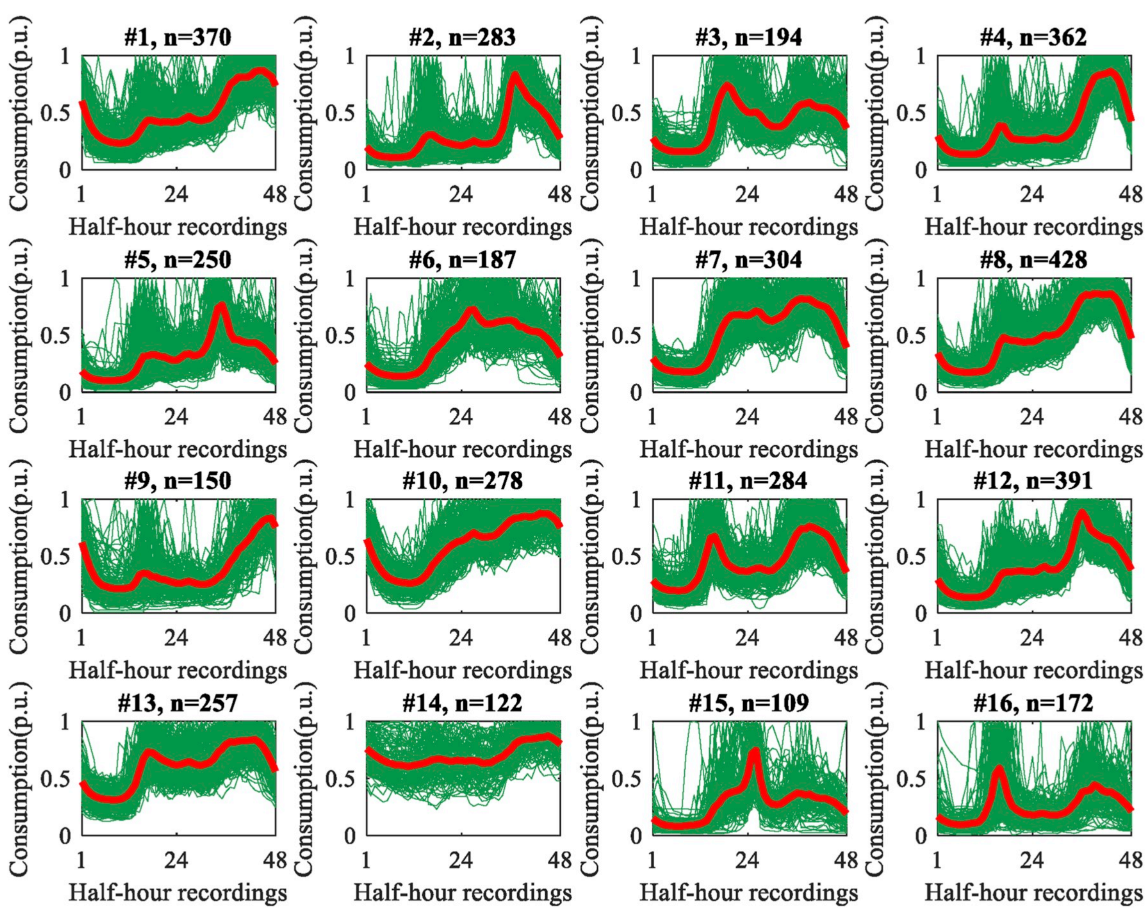

Fig. 12. Clusters of the weekday RLPs of 4141 customers.

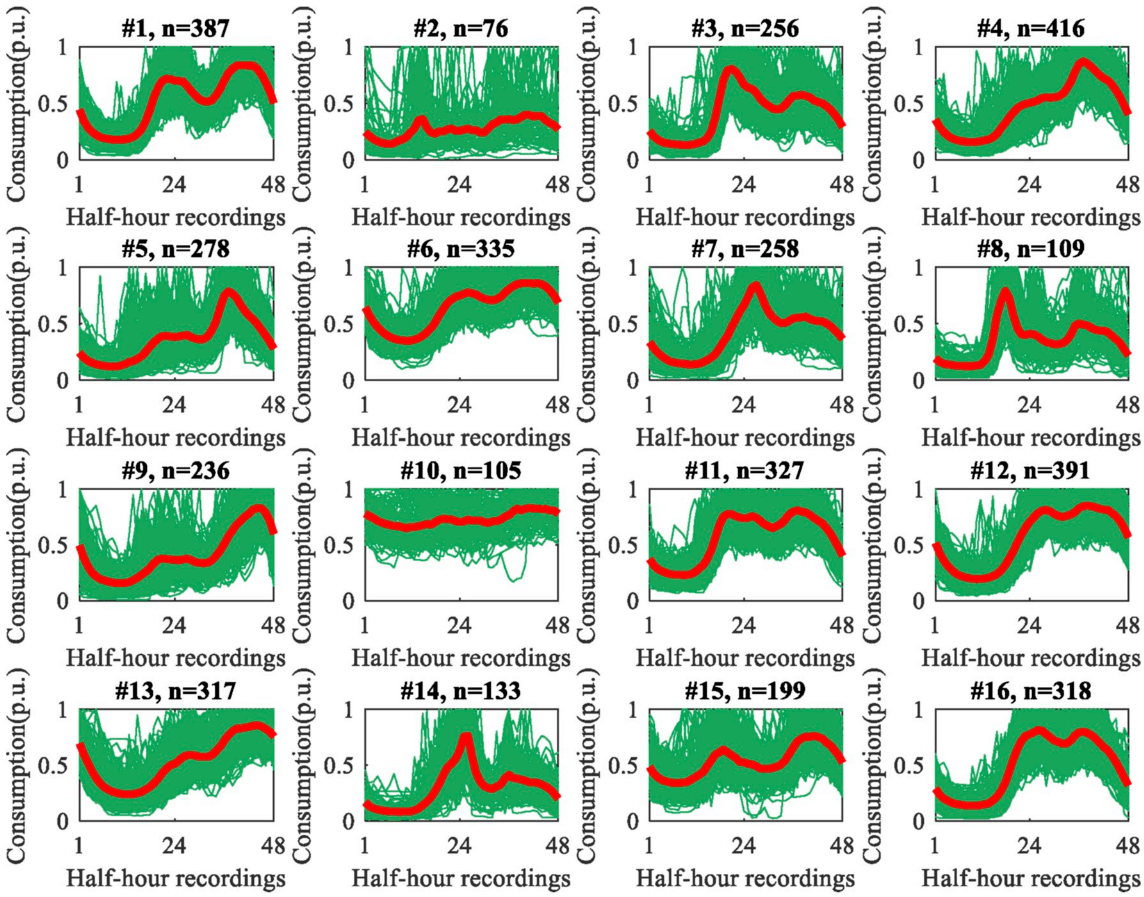

Fig. 13. Clusters of the weekend RLPs of 4141 customers.

Table 5
Comparison of processing time for different clustering algorithms.

Clustering method	Number of clusters										
	5	6	7	8	9	10	11	12	13	14	15
K-means	0.128	0.138	0.156	0.235	0.172	0.293	0.227	0.285	0.311	0.265	0.302
FCM-1.2	4.038	9.688	4.762	8.688	8.002	10.390	10.423	15.780	19.909	32.676	25.552
FCM-2	6.343	10.218	9.708	14.455	15.533	14.379	17.375	24.225	17.756	19.410	26.478
FCM-3	0.599	0.614	0.718	0.687	0.693	0.788	0.691	0.709	0.770	0.806	0.825
FCM-4	0.348	0.406	0.369	0.381	0.441	0.457	0.401	0.418	0.389	0.391	0.372
H-Single	0.631	0.598	0.597	0.631	0.623	0.605	0.618	0.597	0.590	0.607	0.602
H-Complete	0.601	0.561	0.572	0.639	0.587	0.582	0.622	0.578	0.593	0.564	0.599
H-Average	0.634	0.603	0.609	0.644	0.631	0.616	0.665	0.594	0.604	0.638	0.635
H-Centroid	0.954	0.943	0.930	1.053	0.986	0.970	1.038	0.934	0.913	0.944	0.981
H-Ward	0.619	0.611	0.610	0.686	0.617	0.610	0.633	0.611	0.625	0.604	0.643
SOM-H12	5.598	5.539	5.556	5.776	5.839	5.652	5.710	5.760	5.763	5.601	5.482
SOM-R12	5.617	5.655	5.675	5.869	5.946	5.697	5.851	5.847	5.730	5.692	5.571
SOM-H18	12.062	12.354	12.293	12.579	12.382	12.630	12.583	12.433	12.488	11.979	12.066
SOM-R18	12.129	12.069	12.337	12.218	12.686	12.341	12.527	12.689	12.422	12.060	12.178

Table 6
Comparison of processing time for different CVIs.

CVI	Number of clusters										
	5	6	7	8	9	10	11	12	13	14	15
MSE	0.017	0.012	0.010	0.011	0.011	0.013	0.012	0.015	0.011	0.011	0.011
SIL	6.916	6.796	6.885	7.017	7.712	7.324	7.192	7.660	7.403	7.314	7.469
DBI	0.033	0.024	0.017	0.018	0.018	0.020	0.019	0.019	0.018	0.020	0.019
MIA	0.022	0.012	0.013	0.013	0.013	0.013	0.013	0.014	0.012	0.013	0.015
WCBCR	0.027	0.017	0.015	0.013	0.019	0.014	0.013	0.012	0.012	0.012	0.014
Dunn	0.420	0.392	0.402	0.394	0.424	0.409	0.399	0.406	0.399	0.404	0.402

Table 7
Feature definition/extraction and data size reduction techniques.

Method	Reference
Feature definition	[15,38,42,47,82,83]
Feature extraction	DFT [28,80]:
	WT [84–87]:
Data size reduction	PCA [20,33,37,43,77]:
	SAX [46,77]:
	Other [20,77]:

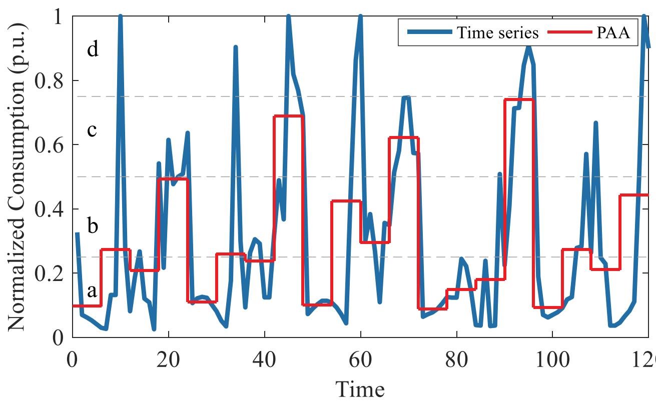

Fig. 14. PAA discretization of consumption data and SAX representation of PAA values (SAX word "*abababacabbcaaacabab*").

It should be noted that the computation time for the algorithms depends on the initialization parameters. For example, in K-means, the method for selection of centers greatly affects the processing time. In addition, if the centers are selected randomly, the time is multiplied by the number of iterations which is needed to repeat the clustering using the new initial centers. Here, the results are reported for only one random selection of centers.

FCM calculation time depends on the fuzziness parameter. In Table 5, the processing time for $m = 1.2$, 2, 3 and 4 are reported. It is observed that by increasing the amount of fuzziness degree, the calculation time increases at first and then experiences a steady decrease.

For the hierarchical method, the calculation time only slightly differs for different cluster numbers since the main processing time is dedicated to creating the dendrogram which can later be cut at any level. Centroid linkage method has the longest time followed by average and Ward methods.

As expected, the time for training the SOM model is relatively high when the number of observations is large. Also, it can be seen that by increasing the grid size from 12×12 to 18×18 the computation time almost doubles.

For CVI measures, MSE, DBI, WCBCR, and MIA calculate the distances between each cluster center with the members of that cluster. Among them, DBI considers all the pairwise combination of clusters which makes it more complicated. Silhouette and Dunn indexes require the calculation of inter-cluster and intra-cluster distances between all the objects. Silhouette has the most complicated formulation which makes it more computationally expensive for large datasets. Table 6 also shows that the Silhouette and Dunn indexes represent the longest computation time and the processing time of Silhouette is much higher compared to other CVIs.

6. Preliminary stages before the clustering

In the previous sections, we studied the major clustering algorithms and their applications. However, it should be noted that the volume of recorded electricity consumptions is enormous. Setting the data sampling resolution of smart meters to 1 h, 15 min, and 1 min results in 24, 96, 1440 records per day respectively. This clearly shows the effect of sampling rate on the dimensionality of time series data. Specifically, analysing these massive sets of data could be a challenging task for electrical utilities. Therefore, data size reduction and feature definition/extraction methods are examined in the literature to reduce the size of load data sets (Table 7). The proper use of these techniques can reduce the input data of clustering algorithms, save computation time, and produce features that are suitable for a specific application. However, in some cases, the new features cannot reflect the daily and intraday consumption behavior of the customers and the variations of the load pattern over time. In addition, it might be hard to interpret the output features of these techniques or to attribute a physical meaning to them. Therefore, the proper selection of these methods depends on the knowledge of experts and the objectives which are expected from the clustering process. In the following, firstly, two major data size reduction methods are introduced and their relevant parameters are analysed and then, the applications of feature definition/extraction methods are briefly discussed.

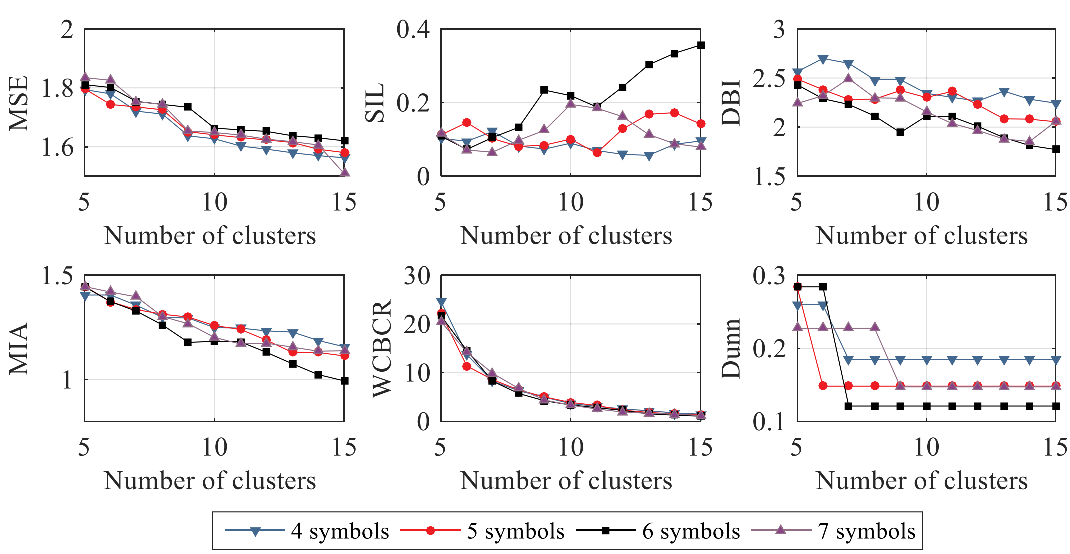

Fig. 15. Performance of a combined clustering of SAX and hierarchical for different number of clusters and alphabet size.

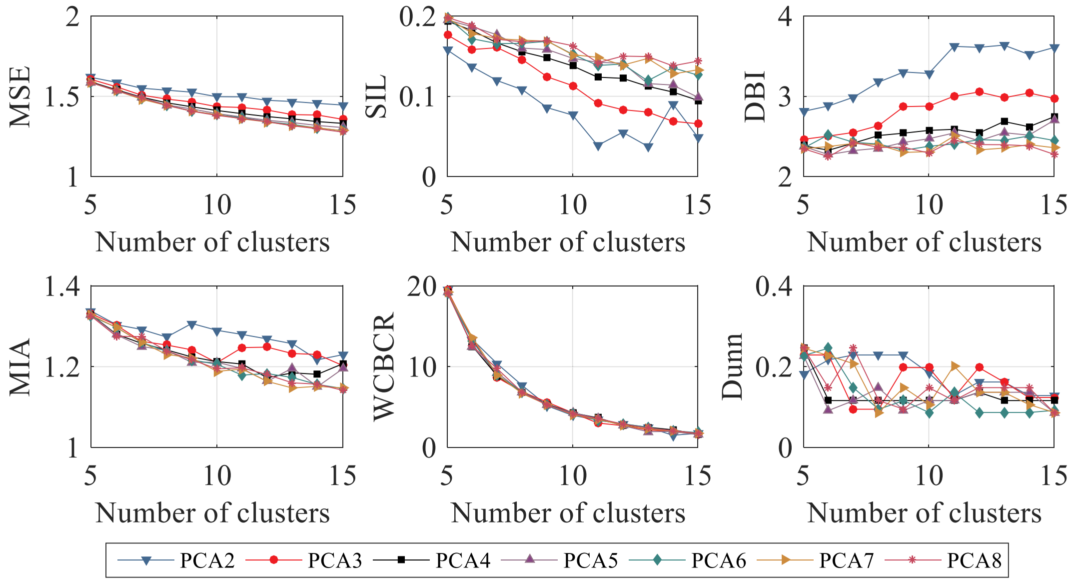

Fig. 16. Performance of a combined clustering of PCA and K-mean for different number of clusters and PCs.

6.1. Data size reduction methods

The first approach uses data size reduction techniques to obtain a reduced data set from the primary data set. Symbolic aggregate approximation (SAX) and PCA are among the popular methods for such data reduction.

SAX is a data size reduction technique which transforms a numeric time series into symbolic strings. The algorithm consists of two steps: i) transforming the original time series into a piecewise aggregate approximation (PAA) representation and ii) symbolizing the PAA data into a discrete string [88]. In this regard, firstly, it is needed to divide the time axis into several intervals. For example, for electricity customers, these time periods can be determined based on the periods of household activities and consumption changes during the day. The PAA technique replaces the amplitude values falling in the same time interval with their mean values. In the next step, the amplitude axis is partitioned into Q intervals and a suitable alphabet of symbols is used for the univocal representation of each range [77]. In the next step, SAX representation (SAX word) of the load curve can be made based on the intervals that PAA values fall into. Fig. 14 clarifies the application of SAX method on a household consumption data.

After transforming the load curves into SAX words, a suitable clustering method can be used to cluster them. Considering that the SAX entries are categorical data, certain clustering algorithms such as K-means cannot be used for clustering. Other algorithms such as K-modes, hierarchical, and DBSCAN are suitable for classifying the load curves into clusters. For clustering the SAX representations, it is necessary to define a proper distance measure to evaluate the similarity of two SAX words. Usually, a distance measure called MINIDIST (based on the work in Ref. [88]) is used for calculating the distance among SAX words. However, other modified versions of this distance measure are also introduced, which can be more suitable for clustering purposes [89].

The partitioning of the time axis and the number of symbols (the number of partitions on the amplitude axis) affect the performance of SAX and final clustering results. These effects can be evaluated using CVIs. Fig. 15 depicts the application of SAX and a hierarchical algorithm on the customer data set in which the SAX method with 6 alphabets shows superior results. In this case, the partitions on the amplitude axis are equal while the time intervals on the time axis have different sizes which are selected according to the usual consumption habits of customers.

SAX is performed in Refs. [46,77] to reduce the scale of the data set. In Ref. [46] the time domain breakpoints are determined by taking into account the regular routine of customers and the implementation of time-of-use (TOU) tariffs. The amplitude breakpoints, on the other hand, are determined by the quantiles of the statistical distribution of amplitudes in the whole data set. Ref. [77] highlights the application of SAX with a hierarchical clustering. It proposes a specific partitioning of the time axis based on the cumulative distribution function of the representative load pattern variations in time.

The fundamental idea of PCA is to reduce the dimensionality of a data set consisting of a large number of possibly correlated variables while retaining as much as possible of the variation present in the data set. This is achieved by an orthogonal transformation that converts the data to a new set of variables called principal components (PCs) which are uncorrelated. It should be noticed that, in PCA analysis, we assume that data is mean-centered. It is an important assumption because accordingly all the projections will have zero mean too.

This transforms the data to a new coordinate system such that the first few PCs retain most of the variation present in all of the original variables [90]. Therefore, the greatest variance by any projection of the data becomes the first coordinate (the first component), the second greatest variance the second coordinate, and so on [91]. Often the number of PCs needed to sufficiently represent the original data is quite small and this makes PCA a suitable tool for dimensionality reduction.

The application of PCA method to the customer data set for different number of PCA components is carried out as shown in Fig. 16. Most of the variance is explained by the first six components and its value does not change meaningfully after around 10 PCs. The adequate number of PCs and the suitable number of clusters can be acquired by CVIs as displayed in this figure. By increasing the number of PCs from 2 to 4, the results improve significantly. However, no considerable change can be observed for more PCs. In this case, the number of final PCs can be selected as 5 or 6.

PCA is applied in several studies to characterise customers' consumptions. In Ref. [43] PCA is employed to understand and visualize measured consumption data. K-means clustering is then used to cluster the data set based on the first four PCs. In Ref. [37], using PCA, 48 half-hour load data are converted to a few PCs and an SOM strategy is applied to reveal a number of distinct behavioral components, for example, high consumptions vs. low consumption. Moreover, in Ref. [33], PCA is used to reduce the dimensionality and to detect the existence of seasonality in load curves.

6.2. Feature definition (expert knowledge-based feature extraction)

Each customer load profile might be represented by a limited number of features. In feature definition approaches some features are defined and employed by the experts based on the specific applications. In Ref. [15] the authors define seven features and extract them from the raw data including the mean, standard deviation, skewness, kurtosis, chaos, energy, and periodicity. Ref. [47] defines a set of shape indicators, for example, daily average load to maximum load factor, to characterise the load patterns. Haben *et al.* [42] divide each day into four time periods, overnight, breakfast, daytime, and evening periods. Using the consumption values in these periods, seven attributes are defined for each customer. In Ref. [38] different variables are derived from the hourly measured energy consumption of customers such as the number of consumption peaks, hourly average consumption, and maximum consumption per day. Also, a regression analysis is adopted in Ref. [83] which gives eight regression coefficients for the electric load pattern of any customer. These coefficients are different for each customer and are used for the clustering purpose. The proper clustering methods can be applied on these features to distinguish customer classes.

6.3. Feature extraction

Feature extraction techniques can also be employed to extract certain features from the load data using techniques such as frequency domain analysis [92], discrete Fourier transform (DFT) [80], and wavelet transform (WT). DFT is used in Ref. [80] to transform time-domain measurements to the frequency domain. Based on the acquired information on amplitude and phase of the harmonic components, a set of features is defined which is used to cluster customers. Ref. [85] proposes two approaches based on WT for clustering one-year load data of a group of French electricity customers. The first method employs discrete WT for feature extraction and K-means algorithm for clustering. This approach is very fast and allows the elimination of non-informative features. On the other hand, the second approach is to cluster using a continuous WT and partitioning around medoid algorithm and can result in more refined clusters. Ref. [84,86] define a clustering strategy by combining an individual signal pre-processing by wavelet denoising, a dimensionality reduction step by wavelet compression, and a hierarchical clustering algorithm which is applied to a suitably chosen set of wavelet coefficients.

7. Applications and future trends

7.1. Applications

The smart metering concept is seen as an essential part of the future smart grids, providing invaluable data which can be used for the improvement of the electrical network operation. Clustering, as a suitable data mining tool, is able to facilitate various applications in the power system domain which, in turn, can contribute significantly to sustainable power grids.

Basically, the outcomes of clustering give general insights into the energy behavioral use of customers which can be beneficial for operation and management of power systems. Besides this obvious advantage, identifying the classes of customers with similar characteristics might be used in more sophisticated ways.

One of the most important applications of clustering of electricity users is to design suitable tariffs for different customers based on the classes that they belong to. As different customers show different load patterns, clustering can help to design cluster-specific tariff structures which can result in the reduction of peak load [47,48].

In addition, clustering of customers to different classes is a promising way for DR program targeting and customer engagement [11]. For instance, if households whose peak demand corresponds to the total system peak are identified, they may be good prospects for recruiting for DR programs [14].

Other studies have investigated the cluster-based load forecasting in which the customers are firstly divided into classes with similar consumption behaviors and then, the load is forecasted for each cluster of customers separately [45,86]. These works can be categorized according to their clustering techniques and the applied prediction methods.

The goal of classification is to classify observations into a set of predefined classes or categories. In power systems, classification can be used to assign new customers or the customers without smart meters to the classes that are previously formed by the clustering process [26,36]. Furthermore, the use of clustering and classification techniques can help to the detection of non-technical losses [93,94].

Surveys provide a lot of information regarding physical characteristics of the dwellings and various data regarding households' socio-economic situations. For an individual dwelling, it is hard to evaluate the correlation between these attributes and its consumption. On the other hand, clustering can reveal those possible correlations between household features and energy usage, as customers with specific attributes usually belong to the same cluster [21,29].

7.2. Future trends

The changes which are gradually happening in power systems and the advancements in data mining techniques will affect customer segmentation in various ways. The improvements in algorithms of time series clustering, advancements in the parallel, distributed, and on-line clustering, and introduction of other novel technologies in smart grids such as cloud computing [95,96] will have a great impact on clustering of electricity customers.

In recent years, new methods have been evolved and applied for clustering of customers. One of these approaches is *time series clustering*. A time series is defined as a series of data points indexed in time order. These data can be the values of a quantity obtained at successive times, often with an equal interval between them [54,97]. The measured data by smart meters also represent a time series data. The same goals that are set for all other clustering applications are also applicable to the clustering of time series data, however, the nature of time series data poses unique challenges for applying any efficient clustering algorithm.

For time series clustering, use of *dynamic time warping (DTW)* [98] as a similarity measure can be beneficial. The Minkowski similarity measures such as Euclidean distance are only defined for series of equal length and are sensitive to scale and time shifts [99]. They also reflect similarity in time by performing a one-to-one mapping between the data instances of the time series under comparison. On the other hand, DTW distance reflects similarity in shape by performing a one-to-many mapping, hence allowing time shifting, and thus matches similar shapes even if they have a time-phase difference [54]. However, one should notice that calculating DTW is computationally expensive [100].

In Ref. [44] the clustering results of K-means and K-medoids with different distance measures are compared. While the former utilizes Euclidean distance as the similarity measure, the latter uses DTW. The results show the advantage of DTW metric for the clustering.

In spite of current achievements, the sheer quantity of data from smart meters poses challenges for traditional data analysis tools of utility companies. In order to deal with this "big data", new infrastructure and tools are required. Big data is usually characterised by three main features [101]: volume, variety, and velocity. Companies in the energy sector, facing this challenge of big data, need to implement more powerful analysis tools to extract value from the collected data. In this respect, leading companies have started working on data science solutions for power systems to control and monitor the network, and increase their profits [102,103].

A few studies in the literature have addressed the "*dynamic*"/"*online*" clustering of load data and the problem of big data. Dynamic clustering of time series data is considered in Refs. [104,105] to deal with the dynamic evolution of the consumption data through time. The presented framework in Ref. [104] for dynamic clustering of load curves compares the performance of K-means and FCM algorithms with different similarity measures including the Euclidean distance, the Pearson correlation coefficient, and another measure called Hausdorff distance.

Ref. [45] proposes an online clustering method for high dimensional time series data. It applies an adaptive K-means algorithm and performs analysis of clustering based on an online algorithm. The principle behind this online time series clustering is a *batch divide-and-conquer* scheme in which the clustering is applied on chunks of data points and once the entire data set is scanned, it combines the results to find the final clustering. Moreover, to tackle the problem of big data, a fully distributed clustering framework is introduced in Ref. [46]. The procedure starts with dividing the data set into k parts and applying an adaptive K-means to each individual part to obtain the cluster centers. Then, these cluster centers are selected as the inputs to another clustering algorithm to obtain the global clustering results.

In another approach, a novel encoding engine based on an artificial neural network is developed [106] which encodes and clusters load profiles in real-time by a distributed approach. The advantage of this neural network based auto-encoder is that it does not need to know anything a priori about the input, nor use any fixed distance metrics like Euclidean distance.

Deep learning-based clustering methods are other novel trends in the clustering of smart meter data. In deep learning, multiple layers are used to extract higher-level features from raw input. These methods can be divided into two categories: two-stage approaches and integrated approaches [107]. While the former performs the feature extraction and clustering in two stages, the latter combines the representation learning process and the clustering stage into one model. A probabilistic baseline estimation framework is proposed in Ref. [107] for DR applications. It employs a deep embedded clustering which is able to extract new features and forms the clusters jointly. A combination of deep neural networks and K-shape clustering is used in Ref. [108] for load forecasting. Ryu *et al.* [109] propose a joint deep learning and clustering process that captures daily and seasonal variations. Deep learning techniques are used in other studies for example, for identifying the socio-demographic information from the load data [110] and designing incentive DR programs [111].

In addition to above-mentioned progresses, with the advancements in smart meter technologies, DMS tools, and data transfer standards and protocols, fine grained electricity data with shorter time resolutions can be made available. The immediate impact will be on the real-time operation of power networks. Due to the fundamental limitations, most of the clustering studies consider the offline data of customers. However, due to these advancements, on-line monitoring and real-time management of power systems will be achievable. Possible applications include very short term load forecasting and dynamic demand response. For instance, system operators and DR aggregators will be able to analyse the load consumption data at very short time scales to forecast the electricity demand and to initiate DR programs such as load curtailments.

8. Conclusion

In this paper, we comprehensively explored the clustering of electricity customers according to their daily load patterns. The primary aim is to detect different consumption patterns which, subsequently, can be used for improving the other applications in the power system domain. Firstly, the past trends of customer segmentation and the stages of customer clustering were presented. In the next step, the major clustering algorithms were introduced and the main parameters of them are discussed. The case studies were performed to show the effect of these parameters and to compare different clustering methods. Furthermore, the applications of cluster validity indexes were described. In another section, some of the most important data size reduction techniques were illustrated and the parameters of them were evaluated. Finally, the future trends and applications were discussed in detail. In each part, an extensive review of the literature was provided to support the discussions.

Acknowledgments

This research is supported by an Australian Government Research Training Program Scholarship.

References

[1] Depuru SSSR, Wang L, Devabhaktuni V. Smart meters for power grid: challenges, issues, advantages and status. Renew Sustain Energy Rev 2011;15:2736–42.
[2] Yildiz B, Bilbao JI, Dore J, Sproul AB. Recent advances in the analysis of residential electricity consumption and applications of smart meter data. Appl Energy 2017;208:402–27.
[3] Barai GR, Krishnan S, Venkatesh B. Smart metering and functionalities of smart meters in smart grid-a review. Electrical power and energy conference (EPEC). IEEE; 2015. p. 138–45. IEEE; 2015.
[4] Leiva J, Palacios A, Aguado JA. Smart metering trends, implications and necessities: a policy review. Renew Sustain Energy Rev 2016;55:227–33.
[5] Realizing the Full Potential of Smart Metering. Accenture's digitally enabled grid program. Available online: https://wwwaccenturecom/us-en/~/media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Industries_9/Accenture-Smart-Metering-Report-Digitally-Enabled-Gridpdf.
[6] Flath C, Nicolay D, Conte T, van Dinther C, Filipova-Neumann L. Cluster analysis of smart metering data. Bus Inf Syst Eng 2012;4:31–9.
[7] Cichosz P. Data mining algorithms: explained using R. John Wiley & Sons; 2014.
[8] Liu H, Yao Z, Eklund T, Back B. From smart meter data to pricing intelligence–visual data mining towards real-time BI. Turku Centre for Comput Sci, Finland, TUCS Tech Rep No 2012;1035.
[9] Yang S-l, Shen C. A review of electric load classification in smart grid environment. Renew Sustain Energy Rev 2013;24:103–10.
[10] Chicco G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. Energy 2012;42:68–80.
[11] Labeeuw W, Stragier J, Deconinck G. Potential of active demand reduction with residential wet appliances: a case study for Belgium. IEEE Trans. Smart Grid 2015;6:315–23.
[12] Lutzenhiser L. Behavioral assumptions underlying California residential energy efficiency programs. CIEE Energy & Behavior Program; 2009.
[13] Dethman L, Thomley D. Comparison of segmentation plans for residential customers. Energy Trust; 2009.
[14] Smith BA, Wong J, Rajagopal R. A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting. ACEEE summer study on energy efficiency in Buildings 2012. p. 374-386.
[15] Räsänen T, Kolehmainen M. Feature-based clustering for electricity use time series data. International conference on adaptive and natural computing algorithms. Springer; 2009. p. 401–12.
[16] Stephen B, Mutanen AJ, Galloway S, Burt G, Järventausta P. Enhanced load profiling for residential network customers. IEEE Trans Power Deliv 2014;29:88–96.
[17] Kwac J, Flora J, Rajagopal R. Household energy consumption segmentation using hourly data. IEEE Trans. Smart Grid 2014;5:420–30.
[18] Wang Y, Chen Q, Kang C, Zhang M, Wang K, Zhao Y. Load profiling and its application to demand response: a review. Tsinghua Sci Technol 2015;20:117–29.
[19] Zhang T, Zhang G, Lu J, Feng X, Yang W. A new index and classification approach for load pattern analysis of large electricity customers. IEEE Trans Power Syst 2012;27:153–60.
[20] Chicco G, Napoli R, Piglione F, Postolache P, Scutariu M, Toader C. Comparisons among clustering techniques for electricity customer classification. IEEE Trans Power Syst PWRS 2006;21:933.
[21] McLoughlin F, Duffy A, Conlon M. A clustering approach to domestic electricity load profile characterisation using smart metering data. Appl Energy 2015;141:190–9.
[22] Kim Y-I, Ko J-M, Choi S-H. Methods for generating TLPs (typical load profiles) for smart grid-based energy programs. Computational intelligence applications in smart grid (CIASG), 2011 IEEE symposium on. IEEE; 2011. p. 1–6.
[23] Chicco G, Napoli R, Piglione F. Application of clustering algorithms and self organising maps to classify electricity customers. Power tech conference proceedings, vol. 1. Bologna: IEEE; 2003. p. 7. IEEE; 2003.
[24] Tsekouras G, Kotoulas P, Tsirekis C, Dialynas E, Hatziargyriou N. A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers. Electr Power Syst Res 2008;78:1494–510.
[25] Figueiredo V, Rodrigues F, Vale Z, Gouveia JB. An electric energy consumer characterization framework based on data mining techniques. IEEE Trans Power Syst 2005;20:596–602.
[26] Viegas JL, Vieira SM, Melício R, Mendes V, Sousa JM. Classification of new electricity customers based on surveys and smart metering data. Energy 2016;107:804–17.

[27] Ramos S, Duarte JM, Duarte FJ, Vale Z. A data-mining-based methodology to support MV electricity customers' characterization. Energy Build 2015;91:16–25.
[28] Ozawa A, Furusato R, Yoshida Y. Determining the relationship between a household's lifestyle and its electricity consumption in Japan by analyzing measured electric load profiles. Energy Build 2016;119:200–10.
[29] Rhodes JD, Cole WJ, Upshaw CR, Edgar TF, Webber ME. Clustering analysis of residential electricity demand profiles. Appl Energy 2014;135:461–71.
[30] Fahiman F, Erfani SM, Rajasegarar S, Palaniswami M, Leckie C. Improving load forecasting based on deep learning and K-shape clustering. Neural networks (IJCNN), 2017 international joint conference on. IEEE; 2017. p. 4134–41.
[31] Zhou D, Balandat M, Tomlin C. Residential demand response targeting using machine learning with observational data. In: Decision and control (CDC), 2016 IEEE 55th conference on. IEEE; 2016. p. 6663–8.
[32] Zhou K, Yang S, Shao Z. Household monthly electricity consumption pattern mining: a fuzzy clustering-based model and a case study. J Clean Prod 2017;141: 900–8.
[33] Ferreira AM, Cavalcante CA, Fontes CH, Marambio JE. A new method for pattern recognition in load profiles to support decision-making in the management of the electric sector. Int J Electr Power Energy Syst 2013;53:824–31.
[34] Zhou K, Yang C, Shen J. Discovering residential electricity consumption patterns through smart-meter data mining: a case study from China. Util Policy 2017;44: 73–84.
[35] Crow M. Clustering-based methodology for optimal residential time of use design structure. In: North American power symposium (NAPS). IEEE; 2014. p. 1–6.
[36] Li R, Gu C, Li F, Shaddick G, Dale M. Development of low voltage network templates—Part I: substation clustering and classification. IEEE Trans Power Syst 2015;30:3036–44.
[37] Motlagh O, Foliente G, Grozev G. Knowledge-mining the Australian smart grid smart city data: a statistical-neural approach to demand-response analysis. Planning support systems and smart cities. Springer; 2015. p. 189–207.
[38] Sanchez IB, Espinós ID, Sarrión LM, López AQ, Burgos IN. Clients segmentation according to their domestic energy consumption by the use of self-organizing maps. Proc 6th Int Conf Eur Energy Market (EEM'09): IEEE; May 2009:1–6.
[39] Räsänen T, Ruuskanen J, Kolehmainen M. Reducing energy consumption by using self-organizing maps to create more personalized electricity use information. Appl Energy 2008;85:830–40.
[40] Mc Loughlin F, Duffy A, Conlon M. Analysing domestic electricity smart metering data using self organising maps. Integration of Renewables into the Distribution Grid. CIRED 2012 Workshop: IET; 2012. p. 1–4.
[41] Verdú SV, Garcia MO, Senabre C, Marín AG, Franco FJG. Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps. IEEE Trans Power Syst 2006;21:1672–82.
[42] Haben S, Singleton C, Grindrod P. Analysis and clustering of residential customers energy behavioral demand using smart meter data. IEEE Trans. Smart Grid 2016; 7:136–44.
[43] Koivisto M, Heine P, Mellin I, Lehtonen M. Clustering of connection points and load modeling in distribution systems. IEEE Trans Power Syst 2013;28:1255–65.
[44] Teeraratkul T, O'neilly D, Lallz S. Condensed representation and individual prediction of consumer demand. Smart Energy Grid Engineering 2016:11–6. IEEE: IEEE; 2016.
[45] Pal R, Chelmis C, Frincu M, Prasanna V. Time series clustering for demand response, an online algorithmic approach. Online: wwwcsuscedu/assets/007/ 93954pdf.
[46] Wang Y, Chen Q, Kang C, Xia Q. Clustering of electricity consumption behavior dynamics toward big data applications. IEEE Trans. Smart Grid 2016;7:2437–47.
[47] Chicco G, Napoli R, Postolache P, Scutariu M, Toader C. Customer characterization options for improving the tariff offer. IEEE Trans Power Syst 2003;18:381–7.
[48] Yang J, Zhao J, Wen F, Dong Z. A model of customizing electricity retail prices based on load profile clustering analysis. IEEE Trans. Smart Grid 2018.
[49] Wang F, Li K, Duić N, Mi Z, Hodge B-M, Shafie-khah M, et al. Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns. Energy Convers Manag 2018;171: 839–54.
[50] Mutanen A, Ruska M, Repo S, Jarventausta P. Customer classification and load profiling method for distribution systems. IEEE Trans Power Deliv 2011;26: 1755–63.
[51] Júnior LAP, Ramos CCO, Rodrigues D, Pereira DR, de Souza AN, da Costa KAP, et al. Unsupervised non-technical losses identification through optimum-path forest. Electr Power Syst Res 2016;140:413–23.
[52] Viegas JL, Esteves PR, Vieira SM. Clustering-based novelty detection for identification of non-technical losses. Int J Electr Power Energy Syst 2018;101: 301–10.
[53] Maimon O, Rokach L. Data mining and knowledge discovery handbook. Springer-Verlag New York, Inc.; 2005.
[54] Aggarwal CC, Reddy CK. Data clustering: algorithms and applications. Chapman and Hall/CRC; 2013.
[55] Prahastono I, King D, Ozveren C. A review of electricity load profile classification methods. In: Universities power engineering conference, 2007 UPEC 2007 42nd international. IEEE; 2007. p. 1187–91.
[56] Bezdek JC, Ehrlich R, Full W. FCM: the fuzzy c-means clustering algorithm. Comput Geosci 1984;10:191–203.
[57] Vesanto J, Alhoniemi E. Clustering of the self-organizing map. IEEE Trans Neural Netw 2000;11:586–600.
[58] Figueiredo MAT, Jain AK. Unsupervised learning of finite mixture models. IEEE Trans Pattern Anal Mach Intell 2002;24:381–96.
[59] Aggarwal CC, Reddy CK. Data clustering. 2014.
[60] Gan G, Ma C, Wu J. Data clustering: theory, algorithms, and applications. Siam; 2007.
[61] Bhatia SK. Adaptive K-means clustering. Proc Int Florida Artif Intell Res Soc Conf2004.
[62] Paparrizos J, Gravano L. k-shape: efficient and accurate clustering of time series. Proceedings of the 2015 ACM SIGMOD international conference on management of data. ACM; 2015. p. 1855–70.
[63] Divina F, Goméz Vela AF, García Torres M. Biclustering of smart building electric energy consumption data9. Applied Sciences; 2019.
[64] Albert A, Rajagopal R. Smart meter driven segmentation: what your consumption says about you. IEEE Trans Power Syst 2013;28:4019–30.
[65] Celebi ME, Kingravi HA, Vela PA. A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert Syst Appl 2013;40:200–10.
[66] Zhou K, Fu C, Yang S. Fuzziness parameter selection in fuzzy c-means: the perspective of cluster validation. Sci China Inf Sci 2014;57:1–8.
[67] Ozkan I, Turksen I. Upper and lower values for the level of fuzziness in FCM. Fuzzy logic. Springer; 2007. p. 99–112.
[68] Wu K-L. Analysis of parameter selections for fuzzy c-means. Pattern Recognit 2012;45:407–15.
[69] Pal NR, Bezdek JC. On cluster validity for the fuzzy c-means model. IEEE Trans Fuzzy Syst 1995;3:370–9.
[70] Hall LO, Bensaid AM, Clarke LP, Velthuizen RP, Silbiger MS, Bezdek JC. A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. IEEE Trans Neural Netw 1992;3:672–82.
[71] Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? J Classif 2014;31:274–95.
[72] McLachlan G, Peel D. Finite mixture models. John Wiley & Sons; 2004.
[73] Matlab help. Online: https://au.mathworks.com/help/stats/clustering-using-gaussian-mixture-models.html.
[74] Fraley C, Raftery AE. MCLUST version 3: an R package for normal mixture modeling and model-based clustering. WASHINGTON UNIV SEATTLE DEPT OF STATISTICS; 2006.
[75] McLachlan G, Peel D. Finite mixture models, willey series in probability and statistics. New York: John Wiley & Sons; 2000.
[76] Commission for Energy Regulation (CER)- Electricity Smart Metering Customer Behaviour Trials (CBT)- Findings Report [Online], http://www.cer.ie/docs/000340/cer11080(a)(i).pdf.
[77] Notaristefano A, Chicco G, Piglione F. Data size reduction with symbolic aggregate approximation for electrical load pattern grouping. IET Gener, Transm Distrib 2013;7:108–17.
[78] Xu T-S, Chiang H-D, Liu G-Y, Tan C-W. Hierarchical K-means method for clustering large-scale advanced metering infrastructure data. IEEE Trans Power Deliv 2017;32:609–16.
[79] Panapakidis IP, Alexiadis MC, Papagiannis GK. Deriving the optimal number of clusters in the electricity consumer segmentation procedure. European energy market (EEM), 2013 10th international conference on the. IEEE; 2013. p. 1–8.
[80] Carpaneto E, Chicco G, Napoli R, Scutariu M. Electricity customer classification using frequency–domain load pattern data. Int J Electr Power Energy Syst 2006; 28:13–20.
[81] Luo X, Hong T, Chen Y, Piette MA. Electric load shape benchmarking for small- and medium-sized commercial buildings. Appl Energy 2017;204:715–25.
[82] Dent I, Craig T, Aickelin U, Rodden T. An approach for assessing clustering of households by electricity usage. 2012.
[83] Shahzadeh A, Khosravi A, Nahavandi S. Improving load forecast accuracy by clustering consumers using smart meter data. Neural networks (IJCNN), 2015 international joint conference on. IEEE; 2015. p. 1–7.
[84] Misiti M, Misiti Y, Oppenheim G, Poggi J-M. Clustering signals using wavelets. Proceedings of the 9th international work conference on Artificial neural networks. Springer-Verlag; 2007. p. 514–21.
[85] Antoniadis A, Brossat X, Cugliari J, Poggi J-M. Clustering functional data using wavelets. Int J Wavelets, Multiresolut Inf Process 2013;11:1350003.
[86] Misiti M, Misiti Y, Oppenheim G, Poggi J-M. Optimized clusters for disaggregated electricity load forecasting. Rev Stat J 2010;8:105–24.
[87] Alzate C, Sinn M. Improved electricity load forecasting via kernel spectral clustering of smart meters. Data mining (ICDM), 2013 IEEE 13th international conference on. IEEE; 2013. p. 943–8.
[88] Lin J, Keogh E, Lonardi S, Chiu B. A symbolic representation of time series, with implications for streaming algorithms. Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery. ACM; 2003. p. 2–11.
[89] Aghabozorgi S, Wah TY. Clustering of large time series datasets. Intell Data Anal 2014;18:793–817.
[90] Jolliffe IT. Principal component analysis and factor Analysis. Principal component analysis. Springer; 1986. p. 115–28.
[91] Abdi H, Williams LJ. Principal component analysis. Wiley Interdiscipl. Rev.: Comput Stat 2010;2:433–59.
[92] Zhong S, Tam K-S. Hierarchical classification of load profiles based on their characteristic attributes in frequency domain. IEEE Trans Power Syst 2015;30: 2434–41.
[93] Angelos EWS, Saavedra OR, Cortés OAC, de Souza AN. Detection and identification of abnormalities in customer consumptions in power distribution systems. IEEE Trans Power Deliv 2011;26:2436–42.
[94] Han SY, No J, Shin J-H, Joo Y. Conditional abnormality detection based on AMI data mining. IET Generation. Transm Distrib 2016;10:3010–6.

[95] Bera S, Misra S, Rodrigues JJ. Cloud computing applications for smart grid: a survey. IEEE Trans Parallel Distrib Syst 2015;26:1477–94.
[96] Yigit M, Gungor VC, Baktir S. Cloud computing for smart grid applications. Comput Network 2014;70:312–29.
[97] Fu T-c. A review on time series data mining. Eng Appl Artif Intell 2011;24:164–81.
[98] Liao TW. Clustering of time series data—a survey. Pattern Recognit 2005;38:1857–74.
[99] Aghabozorgi S, Shirkhorshidi AS, Wah TY. Time-series clustering–A decade review. Inf Syst 2015;53:16–38.
[100] Sarda-Espinosa A. Comparing time-series clustering algorithms in R using the dtwclust package. Online: https://cranr-projectorg/web/packages/dtwclust/vignettes/dtwclustpdf.
[101] Laney D. 3-D data management: controlling data volume, velocity and variety. Gartner: Technical report; 2001.
[102] Molina-Solana M, Ros M, Ruiz MD, Gómez-Romero J, Martin-Bautista M. Data science for building energy management: a review. Renew Sustain Energy Rev 2017;70:598–609.
[103] Teradata. Siemens and Teradata form global strategic partnership for big data in the utility sector. Online: https://wwwteradatacom/Press-Releases/2013/Siemens-and-Teradata-form-global-strategic-pa.
[104] Benítez I, Díez J-L, Quijano A, Delgado I. Dynamic clustering of residential electricity consumption time series data based on Hausdorff distance. Electr Power Syst Res 2016;140:517–26.
[105] Benítez I, Quijano A, Díez J-L, Delgado I. Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. Int J Electr Power Energy Syst 2014;55:437–48.
[106] Varga ED, Beretka SF, Noce C, Sapienza G. Robust real-time load profile encoding and classification framework for efficient power systems operation. IEEE Trans Power Syst 2015;30:1897–904.
[107] Sun M, Wang Y, Teng F, Ye Y, Strbac G, Kang C. Clustering-based residential baseline estimation: a probabilistic perspective. IEEE Trans. Smart Grid 2019.
[108] Jarábek T, Laurinec P, Lucká M. Energy load forecast using S2S deep neural networks with k-Shape clustering. IEEE 14th international scientific conference on Informatics 2017. 2017. p. 140–5.
[109] Ryu S, Choi H, Lee H, Kim H, Wong VW. Residential load profile clustering via deep convolutional autoencoder. IEEE international conference on communications, control, and computing technologies for smart grids (SmartGridComm). IEEE; 2018. p. 1–6. 2018.
[110] Wang Y, Chen Q, Gan D, Yang J, Kirschen DS, Kang C. Deep learning-based socio-demographic information identification from smart meter data. IEEE Trans. Smart Grid 2018;10:2593–602.
[111] Lu R, Hong SH. Incentive-based demand response for smart grid with reinforcement learning and deep neural network. Appl Energy 2019;236:937–49.