A New Classification Pattern Recognition Methodology for Power System Typical Load Profiles

G. J. TSEKOURAS[1,2], F.D. KANELLOS[1], V.T. KONTARGYRI[1],
I.S. KARANASIOU[1], A.D. SALIS[1], N. E. MASTORAKIS[2]

[1]School of Electrical and Computer Engineering
National Technical University of Athens
9 Heroon Polytechniou Street, Zografou, Athens

[2]Department of Electrical Engineering & Computer Science, Hellenic Naval Academy
Terma Hatzikyriakou, Piraeus

GREECE

Email: tsekouras_george_j@yahoo.gr, kanellos@mail.ntua.gr, vkont@central.ntua.gr,
ikaran@esd.ece.ntua.gr, anastasios.salis@gmail.com, mastor@wseas.org

Abstract: - In this paper a new pattern recognition methodology is described for the classification of the daily chronological load curves of power systems, in order to estimate their respective representative daily load profiles, which can be mainly used for load forecasting and feasibility studies of demand side management programs. It is based on pattern recognition methods, such as k-means, adaptive vector quantization, self-organized maps (SOM), fuzzy k-means and hierarchical clustering, which are theoretically described and properly adapted. The parameters of each clustering method are properly selected by an optimization process, which is separately applied for each one of six adequacy measures: the error function, the mean index adequacy, the clustering dispersion indicator, the similarity matrix indicator, the Davies-Bouldin indicator and the ratio of within cluster sum of squares to between cluster variation. This methodology is applied for the Greek power system, from which is proved that the separation between work days and non-work days for each season is not descriptive enough.

Key-Words: - Load profiles, clustering algorithms, adaptive vector quantization, fuzzy k-means, hierarchical clustering, k-means, self-organized maps, pattern recognition, adequacy measures

1 Introduction

In a deregulated electricity market, load profiles categorization can be useful to the power systems and their customers. The first ones can be used for load leveling, demand side management, fluctuation smoothing and load forecasting. It is an alternative solution for various problems of the conventional electrical power transmission and distribution systems, which can reduce the respective operational cost providing their customers with satisfactory services at low cost, particularly for special cases such as supply of discontinuous villages or autonomous islands [1], emergency power source [2], etc. The customers can participate into the competitive electricity market using demand side bidding mechanism [3] based on powerful technologies of the energy storage systems [4-5].

In order to carry out this classification, the chronological load curves per year, season or month can be used. During the last years, a significant research effort has been devoted to load curves classification, in order to solve the short-term load forecasting of anomalous days [6-7] and to cluster the customers of the power systems [8-13]. The clustering methods used so far are the self-organizing map [6-8], the "modified follow the leader" [8], the k-means [8], the fuzzy k-means [8-10] and the average and Ward hierarchical methods [8-10]. These methods generally belong to pattern recognition techniques [11]. Alternatively, the customers classification problem can be solved by using data mining [12], wavelet packet transformation [13], frequency-domain data [14], stratified sampling [15] etc. For the reduction of the size of the clustering input data set Sammon map, principal component analysis and curvilinear component analysis have been proposed [8]. The most commonly used respective adequacy measures are the mean index adequacy [9], the clustering dispersion indicator [8-9], the similarity matrix indicator [9], the Davies-Bouldin indicator [7-9], the modified Dunn index [8], the scatter index [8] and the mean square error [10].

The objective of this paper is to present a new methodology for the classification of the daily chronological load curves for power systems. Specifically, the respective load curves set is organized into well-defined and separate classes, in order to successfully describe the demand behavior of a power system. The proposed methodology compares the results obtained by certain clustering techniques (k-means with special weights initialization, adaptive vector quantization (AVQ), mono-dimensional and bi-dimensional self organizing maps (SOM), fuzzy k-means and seven hierarchical agglomerative clustering methods) using six adequacy measures (mean square error, mean index adequacy, clustering dispersion indicator, similarity matrix indicator, Davies-Bouldin indicator, ratio of within cluster sum of squares to between cluster variation). The basic aspects of this methodology are:

- the estimation of the typical days through the study period and the respective representative typical daily load profiles for the power system;
- the modification of the clustering techniques for this kind of classification problem, such as the appropriate weights initialization for the k-means and fuzzy k-means;
- the proper parameters calibration, such as the training rate of adaptive vector quantization, in order to fit the classification needs;
- the comparison of the clustering algorithms performance for each one of the adequacy measures;
- the selection of the proper adequacy measures.

Finally, the results of the application of the developed methodology are thoroughly presented for the Greek power system for the summer of the year 2000, while the respective ones are collectively registered for the seasons (summer-winter) and years of the time period 1985-2000. It is mentioned that the proposed methodology is applicable to any power system, leading to reliable results.

2 Proposed Pattern Recognition Methodology for the Classification of Load Curves of Power System

The classification of daily chronological load curves of power system is achieved by applying the pattern recognition methodology, as shown in Fig. 1.

The main steps are the following:

- *Data and features selection*: The active and reactive energy values are registered (in MWh and Mvarh) for each time period in steps of 1 hour. The daily chronological load curves are determined for the study period.
- *Data preprocessing:* The load diagrams are examined for normality, in order to modify or delete the values that are obviously wrong (*noise suppression*). If it is necessary, a preliminary execution of a pattern recognition algorithm is carried out, in order to track bad measurements or networks faults, which will reduce the number of the useful typical days for a constant number of clusters, if they are uncorrected.
- *Main application of pattern recognition methods*: For the load diagrams, a number of clustering algorithms (k-means, fuzzy k-means, adaptive vector quantization, self-organized map and hierarchical clustering) is applied. Each algorithm is trained for the set of load diagrams and evaluated according to six adequacy measures. The parameters of the algorithms are optimized, if necessary. The developed methodology uses the clustering methods that provide the most satisfactory results.

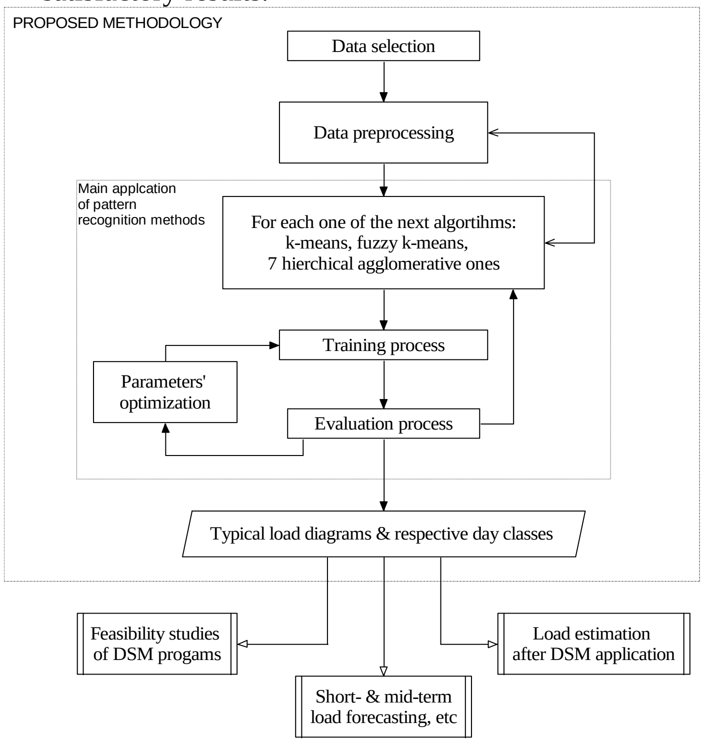


Fig. 1. Flow diagram of pattern recognition methodology for the classification of daily chronological load curves of power system

The results of the developed methodology can be used for power system short-term and mid-term load forecasting, energy trades, techno-economic studies of the energy efficiency and demand side management programs and the respective load estimation after the application of these programs.

3 Mathematical Modeling of Clustering Methods and Clustering Validity Assessment

3.1 General

In the case study of the chronological typical load curves of a power system a number of N analytical daily load curves is given. The main target is to determine the respective sets of days and load patterns. Generally N is defined as the population of the input vectors, which are going to be clustered. The $\vec{x}_\ell = \left(x_{\ell 1}, x_{\ell 2}, \ldots x_{\ell i}, \ldots x_{\ell d}\right)^T$ symbolizes the ℓ-th input vector and d its dimension, which equals to 24 (the load measurements are taken every hour). The corresponding set is given by $X = \{\vec{x}_\ell : \ell = 1,...,N\}$. It is worth mentioning that $x_{\ell i}$ are normalized using the higher and lower values of all elements of the original input patterns set, in order to have better results from the application of clustering methods.

Each classification process makes a partition of the initial N input vectors to M clusters. The j-th cluster has a representative, which is the respective load profile and is represented by the vector $\vec{w}_j = (w_{j1}, w_{j2}, ..., w_{ji}, ...w_{jd})^T$ of d dimension. The vector $\vec{w}_j$ expresses the cluster center. The corresponding set is the classes set, which is defined by $W = \{\vec{w}_k, k = 1,...M\}$. The subset of input vectors $\vec{x}_\ell$, which belong to the j-th cluster, is Ω_j and the respective population of load diagrams is N_j. For the study and evaluation of classification algorithms the following distance forms are defined:

a. the Euclidean distance between ℓ_1, ℓ_2 input vectors of the set X:

$$d(\vec{x}_{\ell_1}, \vec{x}_{\ell_2}) = \sqrt{\frac{1}{d}\sum_{i=1}^{d}(x_{\ell_1 i} - x_{\ell_2 i})^2} \quad (1)$$

b. the distance between the representative vector $\vec{w}_j$ of j-th cluster and the subset Ω_j, calculated as the geometric mean of the Euclidean distances between $\vec{w}_j$ and each member of Ω_j:

$$d(\vec{w}_j, \Omega_j) = \sqrt{\frac{1}{N_j}\sum_{\vec{x}_\ell \in \Omega_j} d^2(\vec{w}_j, \vec{x}_\ell)} \quad (2)$$

c. the infra-set mean distance of a set, defined as the geometric mean of the inter-distances between the members of the set, i.e. for the subset Ω_j:

$$\hat{d}(\Omega_j) = \sqrt{\frac{1}{2N_j}\sum_{\vec{x}_\ell \in \Omega_j} d^2(\vec{x}_\ell, \Omega_j)} \quad (3)$$

The basic characteristics of the three clustering methods being used are the following.

3.2 K-means model

The k-means clustering method groups the set of the N input vectors to M clusters using an iterative procedure. Initially the weights of the M clusters are determined. In the classical model a random choice among the input vectors is used [8], while in the developed algorithm the w_{ji} of the j-th center is initialized as:

$$w_{ji}^{(0)} = a + b \cdot (j-1)/(M-1) \quad (4)$$

where a and b are properly calibrated parameters. During epoch t for each training vector $\vec{x}_\ell$ its Euclidean distances $d(\vec{x}_\ell, \vec{w}_j)$ are calculated for all centers. The ℓ-th input vector is put in the set $\Omega_j^{(t)}$, for which the distance between $\vec{x}_\ell$ and the respective center is minimum. When the entire training set is formed, the new weights of each center are calculated as:

$$\vec{w}_j^{(t+1)} = \frac{1}{N_j^{(t)}}\sum_{\vec{x}_\ell \in \Omega_j^{(t)}} \vec{x}_\ell \quad (5)$$

where $N_j^{(t)}$ is the population of the respective set $\Omega_j^{(t)}$ during epoch t. This process is repeated until the maximum number of iterations is used or the variation of the weights is not significant. The algorithm's main purpose is to minimize the error function:

$$J = \frac{1}{N}\sum_{\ell=1}^{N} d^2(\vec{x}_\ell, \vec{w}_{k:\vec{x}_\ell \in \Omega_k}) \quad (6)$$

The main difference compared to the classical model is that the process is repeated for different pairs of (a,b). The best results for each adequacy measure are recorded for different pairs (a,b).

3.3 Fuzzy k-means

Each input vector $\vec{x}_\ell$ does not belong to only one cluster, but it participates to every j-th cluster by a membership factor $u_{\ell j}$, where:

$$\sum_{j=1}^{M} u_{\ell j} = 1, \; u_{\ell j} : 0 \le u_{\ell j} \le 1, \forall j \quad (7)$$

Theoretically, the membership factor gives more flexibility in the vector's distribution. During the iterations the following objective function is minimized:

$$J_{fuzzy} = \frac{1}{N}\sum_{j=1}^{M}\sum_{\ell=1}^{N} u_{\ell j} \cdot d^2(\vec{x}_\ell, \vec{w}_j) \quad (8)$$

The membership factors and the cluster centers are calculated in each epoch as:

$$u_{\ell j}^{(t+1)} = 1 \Bigg/ \sum_{k=1}^{M} \frac{d(\vec{x}_\ell, \vec{w}_j^{(t)})}{d(\vec{x}_\ell, \vec{w}_k^{(t)})} \quad (9)$$

$$\vec{w}_j^{(t+1)} = \left(\sum_{\ell=1}^{N} \left(u_{\ell j}^{(t+1)}\right)^q \cdot \vec{x}_\ell\right) \Bigg/ \sum_{\ell=1}^{N} \left(u_{\ell j}^{(t+1)}\right)^q \quad (10)$$

where q is the *amount of fuzziness* in the range $(1, \infty)$ which increases as fuzziness decreases. The weights of the clusters centers are initialized by (4), which is similar to the developed k-means.

3.4 Adaptive vector quantization

This algorithm is a variation of the k-means method, which belongs to the unsupervised one-layer neural networks. It classifies input vectors into clusters by using a competitive layer with a constant number of neurons. During epoch t each input vector $\vec{x}_\ell$ is randomly presented and its respective Euclidean distances from every neuron are calculated. The weights of the winning neuron (with the smallest distance) are updated as:

$$\vec{w}_j^{(t)}(n+1) = \vec{w}_j^{(t)}(n) + \eta(t) \cdot \left(\vec{x}_\ell - \vec{w}_j^{(t)}(n)\right) \quad (11)$$

where *n* is the number of input vectors, which have been presented during the current epoch, $w_{ji}^{(0)} = 0.5, \forall j, i$ and $\eta(t)$ is the learning rate according to:

$$\eta(t) = \eta_0 \cdot \exp\left(-t/T_{\eta 0}\right) > \eta_{\min} \quad (12)$$

where η_0, $\eta_{\min}$ and $T_{\eta 0}$ are the initial value, the minimum value and the time parameter respectively. The remaining neurons are unchangeable for $\vec{x}_\ell$, as introduced by the Kohonen winner-take-all learning rule [14-15]. This process is repeated until either the maximum number of epochs is reached or the weights converge or the appropriate error function is not improving.

3.5 Hierarchical agglomerative algorithms

Agglomerative algorithms are based on matrix theory [11]. The input is the $N \times N$ dissimilarity matrix P_0. At each level *t*, when two clusters are merged into one, the size of the dissimilarity matrix P_t becomes $(N-t)\times(N-t)$. Matrix P_t is obtained from P_{t-1} by deleting the two rows and columns that correspond to the merged clusters and adding a new row and a new column that contain the distances between the newly formed cluster and the old ones. The distance between the newly formed cluster C_q (the result of merging C_i and C_j) and an old cluster C_s is determined as:

$$\begin{aligned} d(C_q, C_s) &= a_i \cdot d(C_i, C_s) + a_j \cdot d(C_j, C_s) \\ &+ b \cdot d(C_i, C_j) + c \cdot \left| d(C_i, C_s) - d(C_j, C_s) \right| \end{aligned} \quad (13)$$

where a_i, a_j, b and c correspond to different choices of the dissimilarity measure. It is noted that in each level *t* the respective representative vectors are calculated by (4).

The basic algorithms, which are going to be used in our case, are:

- the *single link* algorithm (*SL*) -it is obtained from (13) for $a_i = a_j = 0.5$, $b = 0$ and $c = -0.5$:

$$d(C_q, C_s) = \min\left\{d(C_i, C_s), d(C_j, C_s)\right\} \quad (14)$$

- the *complete link* algorithm (*CL*) -it is obtained from (13) for $a_i = a_j = 0.5$, $b = 0$ and $c = 0.5$:

$$d(C_q, C_s) = \max\left\{d(C_i, C_s), d(C_j, C_s)\right\} \quad (15)$$

- the *unweighted pair group method average* algorithm (*UPGMA*):

$$d(C_q, C_s) = \left\{n_i \cdot d(C_i, C_s) + n_j \cdot d(C_j, C_s)\right\} / (n_i + n_j) \quad (16)$$

where n_i and n_j are the respective members populations of clusters C_i and C_j.

- the *weighted pair group method average* algorithm (*WPGMA*):

$$d(C_q, C_s) = 0.5 \cdot \left\{d(C_i, C_s) + d(C_j, C_s)\right\} \quad (17)$$

- the *unweighted pair group method centroid* algorithm (*UPGMC*):

$$d^{(1)}(C_q, C_s) = \frac{n_i \cdot d^{(1)}(C_i, C_s) + n_j \cdot d^{(1)}(C_j, C_s)}{n_i + n_j} - n_i \cdot n_j \cdot \frac{d^{(1)}(C_i, C_j)}{(n_i + n_j)^2} \quad (18)$$

where $d^{(1)}(C_q, C_s) = \left\|\vec{w}_q - \vec{w}_s\right\|^2$ and $\vec{w}_q$ is the representative center of the *q*-th cluster (eq. 5).

- the *weighted pair group method centroid* algorithm (*WPGMC*):

$$d^{(1)}(C_q, C_s) = \frac{d^{(1)}(C_i, C_s) + d^{(1)}(C_j, C_s)}{2} - \frac{d^{(1)}(C_i, C_j)}{4} \quad (19)$$

- the *Ward or minimum variance* algorithm (*WARD*):

$$\begin{aligned} d^{(2)}(C_q, C_s) &= \frac{(n_i + n_s) \cdot d^{(2)}(C_i, C_s)}{(n_i + n_j + n_s)} \\ &+ \frac{(n_j + n_s) \cdot d^{(2)}(C_j, C_s) - n_s \cdot d^{(2)}(C_i, C_j)}{(n_i + n_j + n_s)} \end{aligned} \quad (20)$$

where $d^{(2)}(C_i, C_j) = (n_i \cdot n_j)/(n_i + n_j) \cdot d^{(1)}(C_i, C_j)$.

3.6 Self-Organized maps

The Kohonen SOM [16-19] is a topologically unsupervised neural network that projects a *d*-dimensional input data set into a reduced dimension space (usually a mono-dimensional or bi-dimensional map). It is composed of a predefined grid containing $M_1 \times M_2$ *d*-dimensional neurons $\vec{w}_k$, which are calculated by a competitive learning algorithm that updates not only the weights of the winning neuron, but also the weights of its neighbor units in inverse proportion of their distance. The neighborhood size of each neuron shrinks progressively during the training process, starting with nearly the whole map and ending with the single neuron. The training of SOM is divided into two phases:

- *rough ordering*, with high initial learning rate, large radius and small number of epochs, so that neurons are arranged into a structure which approximately displays the inherent characteristics of the input data,
- *fine tuning*, with small initial learning rate, small radius and higher number of training epochs,

in order to tune the final structure of the SOM.

The transition of the rough ordering phase to fine tuning one is happened after T_{s_0} epochs.

Once all vectors of neurons $\vec{w}_k$ have been initialized, the SOM training starts by first choosing an input vector $\vec{x}_\ell$, at *t* epoch, randomly from the input vectors' set. The Euclidean distances between the *n*-th presented input pattern $\vec{x}_\ell$ and all $\vec{w}_k$ are calculated, so as to determine the wining neuron i' that is closest to $\vec{x}_\ell$. The *j*-th reference vector is updated according to:

$$\vec{w}_j^{(t)}(n+1) = \vec{w}_j^{(t)}(n) + \eta(t) \cdot h_{i'j}(t) \cdot \left(\vec{x}_\ell - \vec{w}_j^{(t)}(n)\right) \quad (21)$$

where $\eta(t)$ is the learning rate according to eq. 12. During the rough ordering phase η_r, T_η are the initial value and the time parameter respectively, while during the fine tuning phase the respective values are η_f, T_η. The $h_{i'j}(t)$ is the neighborhood symmetrical function, that will activate the *j* neurons that are topologically close to the winning neuron i', according to their geometrical distance, who will learn from the same $\vec{x}_\ell$. In this case the Gauss function is proposed:

$$h_{i'j}(t) = \exp\left[-d_{i'j}^2 \Big/ \left(2 \cdot \sigma^2(t)\right)\right] \quad (22)$$

where $d_{i'j} = \left\|\vec{r}_{i'} - \vec{r}_j\right\|$ is the respective distance between i' and j neurons, $\vec{r}_j = \left(x_{j,} y_j\right)$ are the respective co-ordinates in the grid, $\sigma(t) = \sigma_0 \cdot \exp\left(-t/T_{\sigma_0}\right)$ is the decreasing neighborhood radius function where σ_0 and T_{σ_0} are the respective initial value and time parameter of the radius correspondingly.

The case studies deal with the matters of the shape of the map, the parameters calibration and the weights initialization. Especially, the multiplicative factors ϕ and ξ are introduced -without decreasing the generalization ability of the parameters calibration:

$$T_{s_0} = \phi \cdot T_{\eta_0} \quad (23) \qquad T_{\sigma_0} = \xi \cdot T_{\eta_0} / \ln \sigma_0 \quad (24)$$

The supposed best trained SOM is the one trained with a number of *t* epochs for which the following index *Is* get the minimum value [7]:

$$Is(t) = J(t) + ADM(t) + TE(t) \quad (25)$$

where $J(t)$ is the respective quantization error -given by eq. (6)-, $ADM(t)$ is the average distortion measure -given by eq. (26)- and $TE(t)$ is the topographic error which measures the distortion of the map as the percentage of input vectors for which the first i_1' and second i_2' winning neuron are not neighboring map units -given by eq. (27):

$$ADM(t) = \sum_{\ell=1}^{N} \sum_{j=1}^{M} h_{i' \to \vec{x}_\ell, j}(t) \cdot d^2\left(\vec{x}_\ell, \vec{w}_j\right) \Big/ N \quad (26)$$

$$TE = \sum_{\ell=1}^{N} neighb\left(i_1', i_2'\right) \Big/ N \quad (27)$$

It is noted that for each input vector $neighb\left(i_1', i_2'\right)$ equals to 1, if i_1' and i_2' neurons are not neighbors, either 0.

3.7 Adequacy measures

In order to evaluate the performance of the clustering algorithms and to compare them with each other, six different adequacy measures are applied. Their purpose is to obtain well-separated and compact clusters, in order to make the load curves self explanatory. The definitions of these measures are the following:

1. *Mean square error or error function (J)* [10] given by eq. 6.
2. *Mean index adequacy (MIA)* [9], which is defined as the average of the distances between each input vector assigned to the cluster and its center:

$$MIA = \sqrt{\frac{1}{M} \sum_{j=1}^{M} d^2\left(\vec{w}_j, \Omega_j\right)} \quad (27)$$

3. *Clustering dispersion indicator (CDI)* [9], which depends on the mean infra-set distance between the input vectors in the same cluster and inversely on the infra-set distance between the class representative load curves:

$$CDI = \sqrt{\frac{1}{M} \sum_{k=1}^{M} \hat{d}^2\left(\Omega_k\right)} \Big/ \hat{d}(W) \quad (28)$$

4. *Similarity matrix indicator (SMI)* [9], which is defined as the maximum off-diagonal element of the symmetrical similarity matrix, whose terms are calculated by using a logarithmic function of the Euclidean distance between any kind of class representative load curves:

$$SMI = \max_{p \rangle q} \left\{ \left(1 - 1 \Big/ \ln\left[d\left(\vec{w}_p, \vec{w}_q\right)\right]\right)^{-1} \right\} : p, q = 1, \ldots, M \quad (29)$$

5. *Davies-Bouldin indicator (DBI)* [20], which represents the system-wide average of the similarity measures of each cluster with its most similar cluster:

$$DBI = \frac{1}{M} \sum_{k=1}^{M} \max_{p \neq q} \left\{ \frac{\hat{d}\left(\Omega_p\right) + \hat{d}\left(\Omega_q\right)}{d\left(\vec{w}_p, \vec{w}_q\right)} \right\} : p, q = 1, \ldots, M \quad (30)$$

6. *Ratio of within cluster sum of squares to between cluster variation (WCBCR)* [21], which depends on the sum of the distance square between each input vector and its cluster representative vector, as well as the similarity of the clusters centres:

$$WCBCR = \sum_{k=1}^{M} \sum_{\vec{x}_\ell \in \Omega_k} d^2\left(\vec{w}_k, \vec{x}_\ell\right) \Big/ \sum_{1 \le q < p}^{M} d^2\left(\vec{w}_p, \vec{w}_q\right) \quad (32)$$

The success of the various algorithms for the same final number of clusters is expressed by having small values of the adequacy measures. By increasing the number of clusters all the measures decrease, except of the similarity matrix indicator. An additional adequacy measure could be the number of the *dead* clusters, for which the sets are empty. It is intended to minimize this number. It is noted that in eq. (6), (29) - (32), M is the number of the clusters without the dead ones.

4 Application of the Proposed Methodology

4.1 Case study

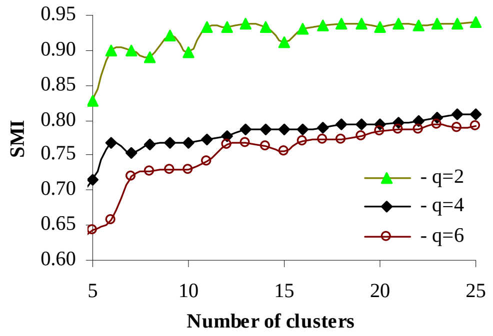
The developed methodology is applied on the Greek power system, analytically for the summer of the year 2000 and concisely for the period of years 1985-2002 per epoch and per year. The data used are hourly load values for the respective period, which is divided into two epochs: summer (from April to September) and winter (from October to March of the next year). In the case of the summer of the year 2000, the respective set of the daily chronological curves has 183 members, from which none is rejected through data pre-processing. In the next paragraphs the application of each clustering method is analyzed.

4.2 Application of the k-means

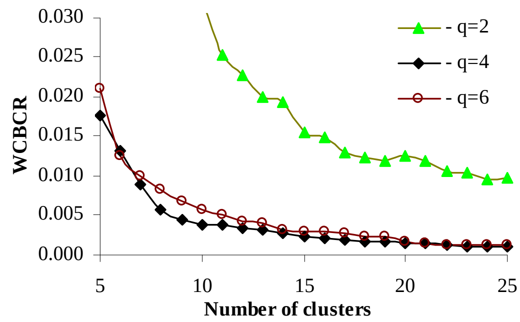
The proposed model of the *k*-means method is executed for different pairs (a,b) from 5 to 25 clusters, where a={0.1,0.11,…,0.45} and $a+b$={0.54,0.55,…,0.9}. For each cluster 1332 different pairs (a,b) are examined. The best results for the six adequacy measures do not refer to the same pair (a,b) –as it is presented in Table 1. The alternative model is the classical one with the random choice of the input vectors during the centers' initialization. For the classical *k*-means model 100 executions are carried out and the best results for each index are registered. The superiority of the proposed model applies in all cases of neurons, while a second advantage is the convergence to the same results for the respective pairs (a,b), which can not be achieved using the classical model.

4.3 Application of the fuzzy k-means

In the fuzzy *k*-means algorithm the results of the adequacy measures depend on the amount of fuzziness increment. In Fig. 2 *SMI* and *WCBCR* adequacy measures are indicatively presented for different number of clusters for three cases of q={2,4,6}. The best results are given by q=4 for *J*, *MIA* , *CDI* and *WCBCR* adequacy measures, by q=6 for *SMI* and *DBI* indicators. It is noted that the initialization of the respective weights is similar to the proposed k-means.



a. *SMI* indicator (similar to *DBI*)



b. *WCBCR* indicator (similar to *J*, *MIA* & *CDI*)

Fig. 2. *SMI* and *WCBCR* for the fuzzy k-means method for the set of 183 load curves of the summer of the year 2000 with q=2, 4, 6 for 5 to 25 clusters

4.4 Application of hierarchical agglomerative algorithms

In the case of the seven hierarchical models the best results are given by the *WARD* model for *J*, by the *UPGMC* model for *MIA*, by the *WPGMA* model for *CDI*, by the *UPGMC* and *UPGMA* models for *SMI*, by the *UPGMC* and *WPGMC* models for *DBI*, by the *UPGMC*, *UPGMA*, *WPGMC* and *WPGMA* models for *WCBCR* adequacy measure, according to Fig. 3.

4.5 Application of adaptive vector quantization

The initial value η_0, the minimum value $\eta_{\min}$ and the time parameter $T_{\eta 0}$ of learning rate must be properly calibrated. For example in Fig. 4 the sensitivity of the mean index adequacy *MIA* to the η_0 and $T_{\eta 0}$ parameters is presented for 90 experiments. The best results of the adequacy measures are not given for the same η_0 and $T_{\eta 0}$, according to the results of Table 1. The $\eta_{\min}$ value does not practically improve the neural network's behavior assuming that it ranges between 10^{-5} and 10^{-6}.

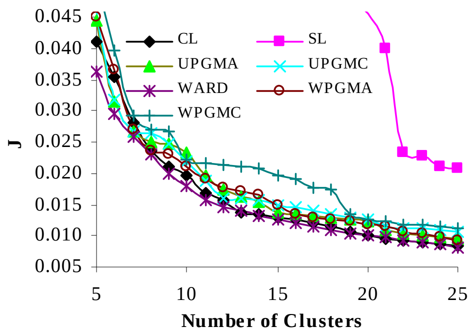
4.6 Application of self-organized map

Although the SOM algorithm is theoretically well defined, there are several issues that need to be solved for the effective training of SOM. The major problems for the mono-dimensional SOM are:

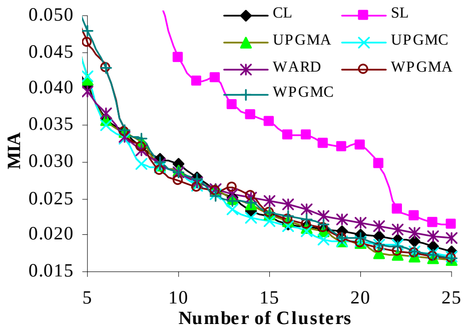
♦ the proper termination of the SOM's training process, which is solved by minimizing the index *Is* (eq. 27).

♦ the proper calibration of (a) the initial value of the neighborhood radius σ_0, (b) the multiplicative factor ϕ between T_{s_0} (epochs of the rough ordering phase) and T_{η_0} (time parameter of learning rate), (c) the multiplicative factor ξ between T_{σ_0} (time parameter of neighborhood radius) and T_{η_0}, (d) the proper initial values of the learning rate η_r and η_f during the rough ordering phase and the fine tuning phase respectively. These are suitably selected through extended research of the parameters' values, like the aforementioned one for the parameters η_0 and T_{η_0} of the AVQ method.

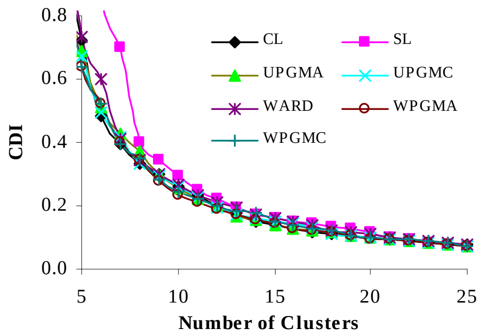
♦ the proper initialization of the weights of the neurons. Three cases are examined through preliminary executions: (a) $w_{ki} = 0.5, \forall k, i$ (which finally presents the best behaviour) (b) the random initialization of each neuron's weight, (c) the random choice of the input vectors for each neuron.



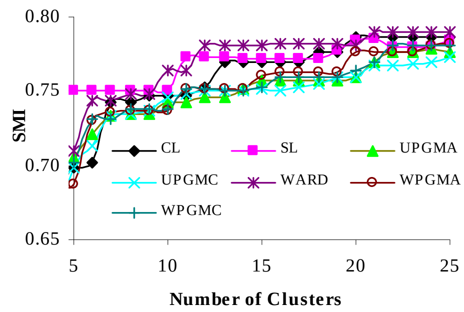
a. *J* indicator



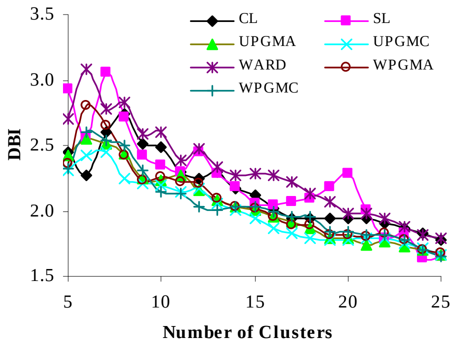
b. *MIA* indicator



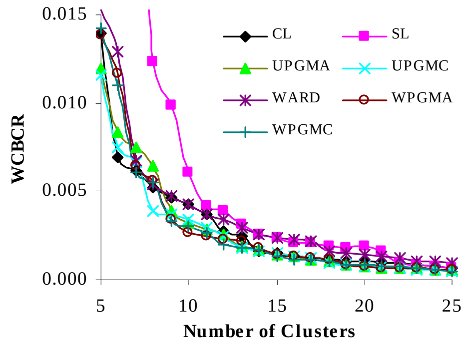
c. *CDI* indicator



d. *SMI* indicator



e. *DBI* indicator



f. *WCBCR* indicator

Fig. 3. Adequacy measures for the 7 hierarchical clustering algorithms for the set of 183 load curves of the summer of the year 2000 for 5 to 25 clusters

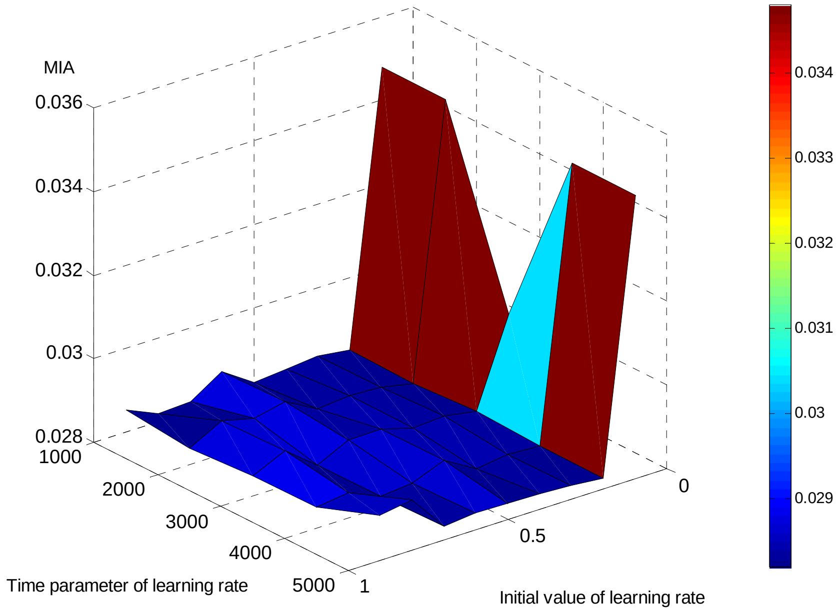


Fig. 4. *MIA* for the AVQ method for the set of 183 load curves of the summer of the year 2000 for 10 neurons, $\eta_0 = \{0.1, 0.15, ..., 0.9\}$, $T_{\eta 0} = \{500, 1000, ...5000\}$, $\eta_{min} = 5 \cdot 10^{-6}$

The optimization process for the mono-dimensional SOM parameters is repeated for any population of clusters.

In the case of the bi-dimensional SOM the additional issues that must be solved, are the shape, the population of neurons and their respective arrangement. The rectangular shape of the map is defined with rectangular or hexagonal arrangement of neurons. It must be mentioned that the two kinds of arrangement practically give the same results. The population of the neurons is recommended to be $5\times\sqrt{N}$ to $20\times\sqrt{N}$ [18-19]. The height / width ratio M_1/M_2 of the rectangular grid can be calculated as the ratio between the two major eigenvalues of the covariance matrix of the input vectors set. The initialization of the neurons can be a linear combination of the respective eigenvectors of the two major eigenvalues or can be equal to 0.5 giving equivalent results substantially.

In the case of the set of 183 load curves for the summer of the year 2000 the map can have 67 ($\cong 5\times\sqrt{183}$) to 270 ($\cong 20\times\sqrt{183}$) neurons. Using the ratio between the two major eigenvalues the respective value is 22.739 (=0.26423/0.01162) and the proposed grids can be 46x2 and 68x3. Practically the clusters of the bi-dimensional map can not be directly exploited because of the size and the location of the neurons into the grid (see Fig. 5).

This problem is solved by the application of a basic classification method -like the proposed k-means- for the neurons of the bi-dimensional SOM [7].

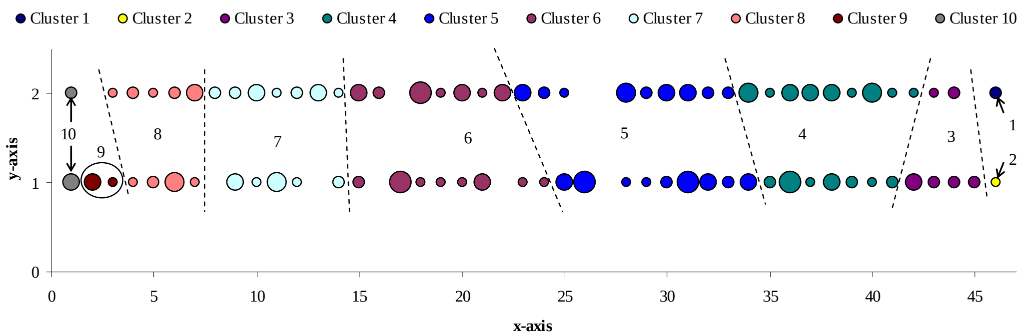
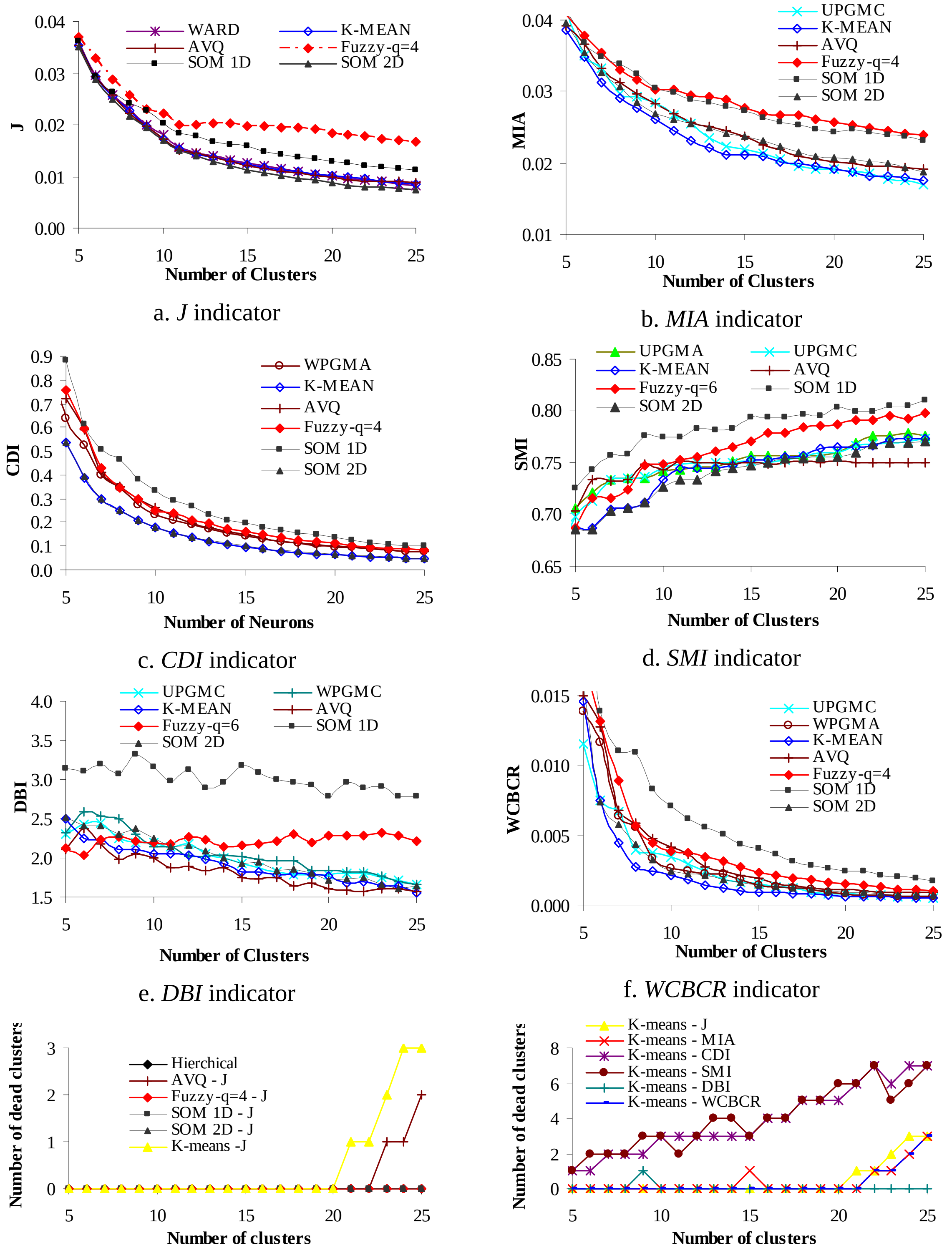


Fig. 5. 46x2 SOM after the application of the proposed k-means method at the neurons of SOM for the set of 183 load curves of the summer of the year 2000 for 10 neurons

The adequacy measures are calculated using the load curves of the neurons which form the respective clusters of the basic classification method and the best results are given by the 46x2 grid for all adequacy measures for different pairs (*a*,*b*) of the k-means method.

4.7 Comparison of clustering models & adequacy indicators

In Fig. 6 the best results achieved by each clustering method (proposed k-means, fuzzy k-means, adaptive vector quantization, hierarchical algorithms and self-organized maps) are depicted.



a. *J* indicator

b. *MIA* indicator

c. *CDI* indicator

d. *SMI* indicator

e. *DBI* indicator

f. *WCBCR* indicator

g. Dead clusters for the basic clustering methods

h. Dead clusters for proposed k-means method

Fig. 6. The best results of each clustering method for the set of 183 load curves of the summer of the year 2000 for 5 to 25 clusters

The proposed k-means model has the smallest values for the *MIA* and *WCBCR* indicators, the bi-dimensional SOM (with the application of the proposed k-means at the second level) for the *J* and *SMI* indicator and the adaptive vector quantization for *DBI* indicator. The proposed k-means model and the bi-dimensional SOM give equivalent results for the *CDI* indicator. All indicators -except DBI- exhibit improved performance, as the number of clusters is increased. Observing the number of dead clusters for the proposed k-means model (Fig. 6.h) it is obvious that the use of *WCBCR* indicator is slightly superior to *MIA* and *J* indicators. It is also noted that the basic theoretical advantage of the *WCBCR* indicator is that it combines the distances of the input vectors from the representative clusters and the distances between clusters, covering also the *J* and *CDI* characteristics. The behavior of *DBI* and *SMI* indicators for different clustering techniques appears significant variability. For the above reasons the proposed indicator is *WCBCR*.

The improvement of the adequacy indicators is significant until 10 clusters. After this value the behavior of the most indicators is gradually stabilized. It can also be estimated graphically by using the rule of the "knee", which gives values between 8 to 10 clusters (see Fig. 7). In Table 1 the results of the best clustering methods are presented for 10 clusters, which is the finally proposed size of the typical days for this case.

Having also taken into consideration that the analogy of the computational training time for the under study methods is 0.05:1:24:28:36:50 (hierarchical: proposed k-means: mono-dimensional SOM: AVQ: fuzzy k-means: bi-dimensional SOM), the use of the hierarchical and k-means models is proposed. It is mentioned that the computational training time for the proposed k-means method is approximately 20 minutes for a Pentium 4, 1.7 GHz, 768 MB.

TABLE 1
COMPARISON OF THE BEST CLUSTERING MODELS FOR 10 CLUSTERS FOR THE SET OF 183 LOAD CURVES OF THE SUMMER OF THE YEAR 2000 FOR THE GREEK POWER SYSTEM

Methods -Parameters	Adequacy Measure					
	J	*MIA*	*CDI*	*SMI*	*DBI*	*WCBCR*
Proposed k-means	0.01729	*0.02262*	*0.1778*	0.7331	2.0606	*0.002142*
a- b parameters	0.26 – 0.39	0.15-0.44	0.45-0.45	0.10-0.78	0.15-0.43	0.14-0.61
Classical k-means	0.01934	0.02434	0.1935	0.7549	2.7517	0.002346
AVQ	0.01723	0.02819	0.2615	0.7431	*1.9973*	0.004145
$\eta_0-\eta_{\min}-T_{\eta_0}$ parameters	0.5-5x10^{-7}-1000	0.4-5x10^{-7}-4000	0.5-5x10^{-7}-5000	0.4-5x10^{-7}-2000	0.8-5x10^{-7}-1000	0.4-5x10^{-7}-4000
Fuzzy k-means	0.02208	0.03036	0.25328	0.7482	2.1936	0.003894
q- a- b parameters	4-0.22-0.46	4-0.18-0.62	4-0.18-0.70	6-0.12-0.62	6-0.14-0.74	4-0.18-0.62
CL	0.01960	0.02974	0.2636	0.7465	2.4849	0.004233
SL	0.06249	0.04435	0.2950	0.7503	2.3509	0.006103
UPGMA	0.02334	0.02885	0.2544	*0.7423*	2.2401	0.003186
UPGMC	0.02200	0.02847	0.2603	*0.7455*	2.1934	0.003412
WARD	0.01801	0.02858	0.2645	0.7635	2.5964	0.004227
WPGMA	0.02094	0.02743	0.2330	0.7373	2.2638	0.002619
WPGMC	0.02227	0.02863	0.2418	0.7378	2.1498	0.003008
Mono-dimensional SOM	0.02024	0.03043	0.3366	0.7752	3.1656	0.007126
$\sigma_0-\phi-\xi-\eta_f-\eta_r-T_{\eta_0}$ parameters	10-1.0-0.6-0.15-10^{-3}-1500	10-2.0-0.2-0.10-10^{-3}-1750	10-1.0-0.6-0.15-10^{-3}-1500	10-1.0-0.6-0.15-10^{-3}-1500	10-1.0-0.6-0.10-10^{-3}-1500	10-2.0-0.2-0.10-10^{-3}-1750
2D SOM 46x2 using proposed k-means for classification in a 2nd level	*0.01685*	0.02697	0.1785	*0.7271*	2.2572	0.002459
$\sigma_0-\phi-\xi-\eta_f-\eta_r-T_{\eta_0}$ -a- b parameters	46-1.0-1.0-0.30-10^{-3}-500-0.28-0.36	46-1.0-1.0-0.30-10^{-3}-500-0.15-0.58	46-1.0-1.0-0.30-10^{-3}-500-0.44-0.46	46-1.0-0.2-0.20-10^{-3}-500-0.10-0.77	46-1.0-0.2-0.20-10^{-3}-500-0.44-0.25	46-1.0-1.0-0.30-10^{-3}-500-0.15-0.58

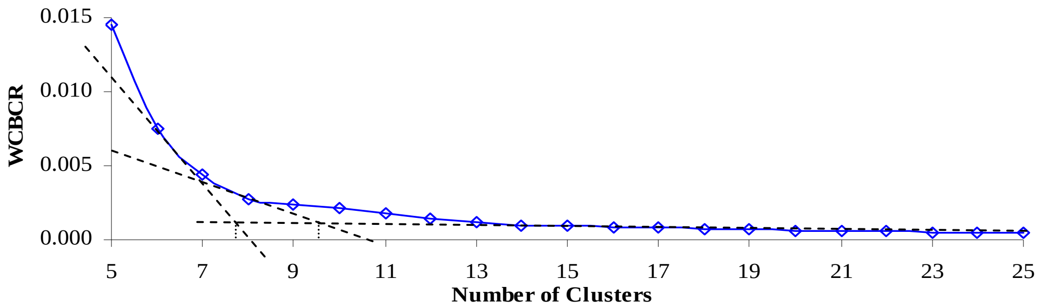


Fig. 7. Indicative estimation of the necessary clusters for the typical load daily chronological curves of the summer of the year 2000 for the *WCBCR* adequacy indicator

4.7 Representative daily load curves of the summer of the year 2000 for the Greek power system

The results of the respective clustering for 10 clusters using the proposed k-means model with the optimization of the *WCBCR* indicator are presented in Table 2 (total number of days per cluster & number of days per cluster & per day of week - Monday, Tuesday, etc-), Table 3 (calendar of the summer of the year 2000 with the kind of cluster) and Fig. 8 (representative load curves per cluster). Additionally, in Fig. 8 the confidence limits of the variations (mean value ± standard deviation) are presented and this has a probability of occurrence equal to 68.27% assuming a normal distribution.

TABLE 2
RESULTS OF THE PROPOSED K-MEANS MODEL WITH OPTIMIZATION TO *WCBCR* FOR 10 CLUSTERS FOR A SET OF 183 LOAD CURVES OF THE SUMMER OF THE YEAR 2000 FOR THE GREEK POWER SYSTEM

Load cluster	Day (1 for Monday, 2 for Tuesday etc.)							Days per cluster
	1	2	3	4	5	6	7	
1	0	0	0	0	0	0	1	1
2	1	0	0	0	1	0	0	2
3	0	1	0	0	0	2	13	16
4	9	8	9	8	7	12	2	55
5	4	3	2	3	4	4	8	28
6	4	6	6	4	3	7	1	31
7	4	3	4	6	6	0	1	24
8	4	3	2	3	3	2	0	17
9	0	2	3	1	2	0	0	8
10	0	0	0	1	0	0	0	1

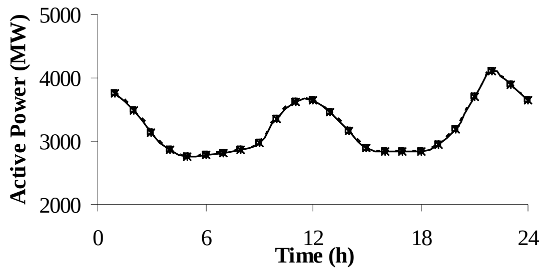
Specifically, the cluster 1 represents Easter, the cluster 2 Holy Friday and Monday after Easter, the cluster 3 the Sundays of April, May, early June and September, Holy Saturday and Labor day. The cluster 4 contains the workdays of very low demand (during April, early May and September) with normal temperatures (22-28^{o}C) and Saturdays of April, May, early June and September, while the cluster 5 includes the workdays of low demand and Sundays of high peak load demand during the hot summer days. The cluster 6 represents the workdays of medium peak load demand and Saturdays of high peak load demand, while the clusters 7 to 10 mainly involves workdays with gradually increasing peak load demand.

4.8 Application of the Proposed Methodology for the Greek Power System Per Season and Per Year for the time period 1985-2002

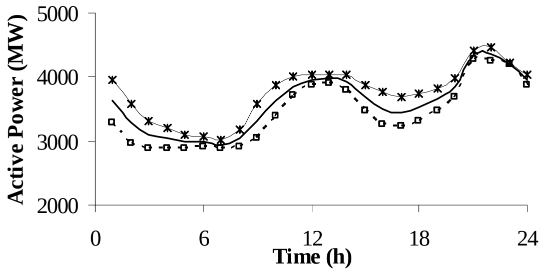
The same process is repeated for the summers (April–September) and the winters (October-March) for years 1985-2002. The load curves of each season are qualitatively described by using 8-10 clusters. The performance of these methods is presented in Table 4 by indicating the number of seasons which achieve the best value of adequacy measure respectively.

The comparison of the algorithms shows that the developed k-means method achieves a better performance for *MIA*, *CDI* and *WCBCR* measures, the bi-dimensional SOM model using proposed k-means for classification in a second level for *J* and *SMI* indicators and the adaptive vector quantization for *DBI* adequacy measure.

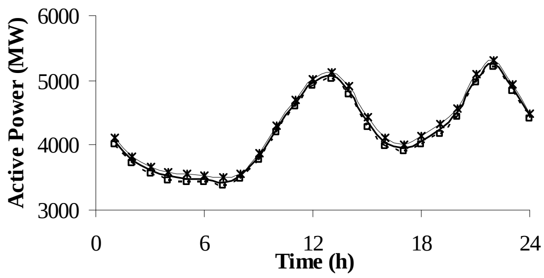
The methodology is also applied for each year during the period 1985-2002, where the load curves are qualitatively described by using 15-20 clusters. The respective performance is presented in Table 5 by indicating the number of years which achieves the best value of adequacy measure respectively. By comparing the algorithms it is obvious that the developed k-means method achieves a better performance for *MIA*, *CDI*, *DBI* and *WCBCR* measures, the 2-D SOM model using the proposed k-means for classification in a second level for *J* indicator and the unweighted pair group method centroid algorithm for *SMI* index.



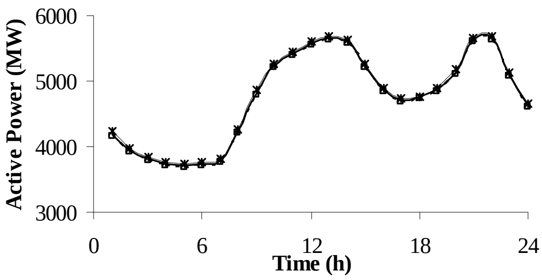
a. Cluster 1



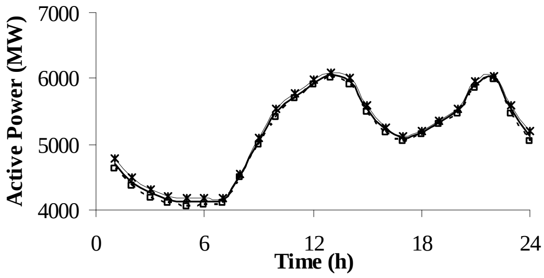
b. Cluster 2



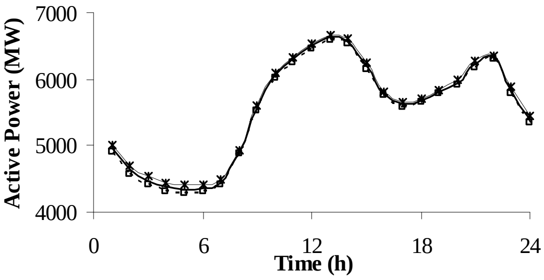
c. Cluster 3



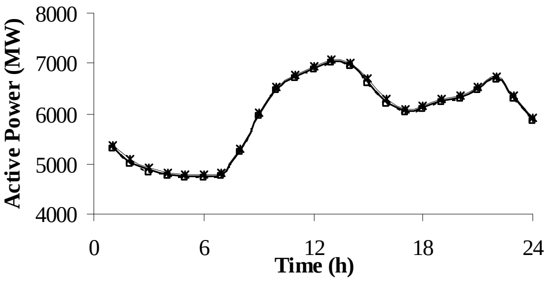
d. Cluster 4



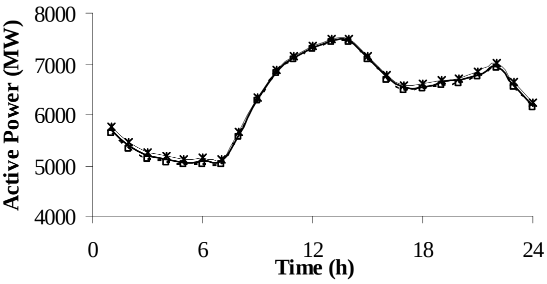
e. Cluster 5



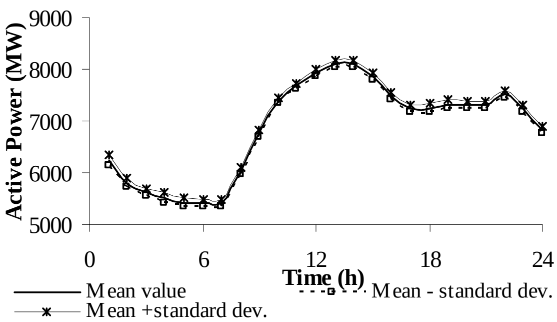
f. Cluster 6



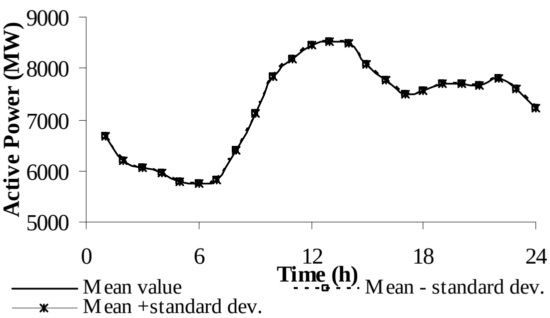
g. Cluster 7



h. Cluster 8



i. Cluster 9



k. Cluster 10

Fig. 8. Typical daily chronological load curves for the set of 183 curves of the summer of the year 2000 for the Greek power system using the proposed k-means model with optimization to *WCBCR*

TABLE 3

CALENDAR FOR 10 CLUSTERS FOR A SET OF 183 LOAD CURVES OF THE SUMMER OF THE YEAR 2000 FOR THE GREEK POWER SYSTEM

Month	April		May		June		July		August		September	
Day of Month	Day of week	Kind of cluster	Day of week	Kind of cluster	Day of week	Kind of cluster	Day of week	Kind of cluster	Day of week	Kind of cluster	Day of week	Kind of cluster
1	6	4	1	2	4	6	6	6	2	7	5	6
2	7	3	2	3	5	5	7	5	3	7	6	5
3	1	4	3	4	6	4	1	8	4	7	7	4
4	2	4	4	4	7	3	2	9	5	7	1	6
5	3	4	5	4	1	5	3	9	6	6	2	6
6	4	4	6	3	2	6	4	10	7	5	3	5
7	5	4	7	3	3	6	5	9	1	7	4	5
8	6	4	1	4	4	7	6	8	2	7	5	5
9	7	3	2	4	5	6	7	7	3	7	6	4
10	1	4	3	4	6	5	1	8	4	7	7	3
11	2	4	4	4	7	4	2	9	5	7	1	5
12	3	4	5	4	1	6	3	9	6	5	2	5
13	4	4	6	4	2	6	4	8	7	5	3	5
14	5	4	7	3	3	7	5	8	1	6	4	5
15	6	4	1	4	4	8	6	6	2	4	5	5
16	7	3	2	4	5	8	7	5	3	6	6	4
17	1	4	3	4	6	5	1	7	4	7	7	3
18	2	4	4	4	7	3	2	8	5	7	1	5
19	3	4	5	4	1	4	3	8	6	6	2	6
20	4	4	6	4	2	6	4	7	7	5	3	6
21	5	4	7	3	3	6	5	7	1	7	4	6
22	6	4	1	4	4	6	6	6	2	8	5	6
23	7	3	2	5	5	7	7	5	3	8	6	4
24	1	4	3	4	6	6	1	7	4	8	7	3
25	2	4	4	5	7	5	2	8	5	8	1	4
26	3	4	5	5	1	8	3	9	6	6	2	4
27	4	4	6	4	2	7	4	9	7	5	3	4
28	5	2	7	3	3	7	5	9	1	6	4	4
29	6	3	1	5	4	7	6	8	2	6	5	4
30	7	1	2	5	5	7	7	6	3	6	6	4
31			3	6			1	8	4	6		

TABLE 4

COMPARISON OF THE CLUSTERING MODELS FOR THE SETS OF LOAD CURVES OF THE GREEK POWER SYSTEM PER SEASON FOR THE TIME PERIOD 1985-2002

Methods	*Adequacy Measure*					
	J	*MIA*	*CDI*	*SMI*	*DBI*	*WCBCR*
Proposed k-means	1	*24*	*31*	7	12	*29*
Classical k-means	0	0	0	0	0	0
AVQ	2	0	0	7	*16*	0
Fuzzy k-means	0	0	0	0	0	1
UPGMA	0	3	0	0	0	1
UPGMC	0	7	0	2	5	3
WPGMA	0	0	0	0	1	0
WPGMC	0	3	0	0	2	2
CL / SL / WARD / Mono-dimensional SOM	0	0	0	0	0	0
Bi-dimensional SOM using proposed k-means for classification in a 2^{nd} level	*34*	0	6	*21*	1	1

TABLE 5
COMPARISON OF THE CLUSTERING MODELS FOR THE SETS OF LOAD CURVES OF THE GREEK POWER SYSTEM PER YEAR FOR THE TIME PERIOD 1985-2002

Methods	*Adequacy Measure*					
	J	*MIA*	*CDI*	*SMI*	*DBI*	*WCBCR*
Proposed k-means	0	*8*	*18*	0	*13*	*14*
Classical k-means	0	0	0	0	0	0
AVQ	1	0	0	1	3	0
UPGMA	0	1	0	0	0	0
UPGMC	0	6	0	*14*	1	2
WPGMA	0	1	0	0	0	0
WPGMC	0	1	0	0	0	2
Fuzzy k-means / CL/ SL / WARD / Mono-dimensional SOM	0	0	0	0	0	0
Bi-dimensional SOM using proposed k-means for classification in a second level	*17*	1	0	3	1	0

The main disadvantage of the load curves classification per year is that each cluster does not contain the same family of days during the time period under study. I.e. if 20 clusters are selected to represent the load demand behavior of the Greek power system per year, the 20th cluster will contain the workdays with the highest peak load demand of the winter for the years 1985-1992 and that of summer for the rest years. In order to avoid this problem, the classification per season is proposed.

5 Conclusions

This paper presents an efficient pattern recognition methodology for the study of the load demand behavior of power systems. The unsupervised clustering methods can be applied, such as the k-means, fuzzy k-means, adaptive vector quantization (AVQ), mono-dimensional and bi-dimensional self organizing maps (SOM) and hierarchical methods. The performance of these methods is evaluated by six adequacy measures: mean square error, mean index adequacy, clustering dispersion indicator, similarity matrix indicator, Davies-Bouldin indicator, the ratio of within cluster sum of squares to between cluster variation. Finally the representative daily load diagrams along with the respective populations per each typical day are calculated. This information is valuable for the electric companies, because it facilitates the load forecasting and the techno-economic studies of demand side management programs.

By applying the proposed methodology to the Greek power system the classification per season is suggested, where 8 to 10 clusters are necessary in order to describe satisfactory the daily load curves of each season (describing the year with two seasons -winter and summer). It is practically impossible to describe the load curves satisfactory dividing the respective days into work days and non-work days, as it has been done until now. It is also concluded that, generally, the optimal clustering technique is the developed k-means, while the optimal adequacy measure is the ratio of within cluster sum of squares to between cluster variation.

The proposed methodology is applicable to any power system, either to active power, or reactive power, in any time period (day, week, etc.) and time step (15 minutes, 1 hour, etc.), leading to reliable results.

Acknowledgements

The authors want to express their sincere gratitude to C.A. Anastasopoulos, D. Voumboulakis and P. Eustathiou from PPC for the supply of all the necessary data for the training of the model and M. Tsaroucha and S. Gialabidis for their contribution on the original version of this paper.

References:

[1] PacifiCorp buys time with storage. *Energy prospects*, Vol. 37, 2004, p. 5, (http:// www.vrbpower.com)

[2] J. Kondoh, I. Ishii, H. Yamaguchi, A. Murata, K. Otani, K. Sakuta, N. Higuvhi, S. Sekine, M. Kamimoto. Electrical energy storage systems for energy networks. *Energy Conversion & Management*, Vol. 41, 2000, pp. 1863-1874.

[3] Task VI of International Energy Agency, Demand Side Management Programme, *Research Report No. 1*, October 1998, *Research Report No. 2*, July 1999, *Research Report No. 3*, August 2000.

[4] Task VIII of International Energy Agency, Demand-Side Bidding in a Competitive Electricity Market, *Research Report No. 1, Ver. 2, January 2002,* with title "Market participants' views towards and experiences with Demand Side Bidding".

[5] T. Sels, C. Dragu, T. Van Craenenbroeck, R. Belmans. New Energy Storage Devices for an Improved Load Managing on Distribution Level. *IEEE Power Tech Conference, Porto, Portugal*, 10th-13th September, p.6.

[6] R. Lamedica, A. Prudenzi, M. Sforna, M. Caciotta, V.O. Cencelli. A neural network based technique for short-term forecasting of anomalous load periods. *IEEE Transactions on Power Systems,* Vol. 11, No. 3, August 1996, pp. 1749-1756.

[7] M. Beccali, M. Cellura, V. Lo Brano, A. Marvuglia. Forecasting daily urban electric load profiles using artificial neural networks. *Energy Conversion and Management,* Vol. 45, 2004, pp. 2879-2900.

[8] G.Chicco, R. Napoli, F. Piglione. Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems,* Vol. 21, No. 2, May 2006, pp. 933-940.

[9] G. Chicco, R. Napoli, F. Piglione. Application of clustering algorithms and self organizing maps to classify electricity customers. *Presented at the IEEE Power Tech Conference, Bologna, Italy,* June 23-26, 2003.

[10] D. Gerbec, S. Gasperic, F. Gubina. Determination and allocation of typical load profiles to the eligible consumers. *IEEE Power Tech Conference, Bologna, Italy,* June 23-26, 2003.

[11] S. Theodoridis, K. Koutroumbas, *Pattern Recognition*, 1st Edition, Academic Press, New York, 1999.

[12] V. Figueiredo, F. Rodrigues, Z. Vale, J. B. Gouveia. An electric energy consumer characterization framework based on data mining techniques. *IEEE Transactions on Power Systems,* Vol. 20, No. 2, May 2005, pp. 596-602.

[13] M. Petrescu, M. Scutariu. Load diagram characterization by means of wavelet packet transformation. *2nd Balkan Conference, Belgrade, Yugoslavia*, June 2002, pp. 15-19

[14] E. Carpaneto, G. Chicco, R. Napoli, M. Scutariou. Electricity customer classification using frequency-domain load pattern data. *Electrical Power and Energy Syst.* Vol. 28, No. 1, 2006, pp. 13-20.

[15] C. S. Chen, J. C. Hwang, C. W. Huang. Application of load survey systems to proper tariff design. *IEEE Transactions on Power Systems,* Vol. 12, No. 4, November 1997, pp. 1746-1751.

[16] S. Haykin. *Neural Networks: A Comprehensive Foundation*, Prentice Hall, NJ, 1994.

[17] T. Kohonen. *Self–organization and Associative Memory*, New York: Springer-Verlag, 1989.

[18] K.F. Thang, R.K. Aggarwal, A.J. McGrail, D.G. Esp. Analysis of power transformer dissolved gas data using the self-organizing map. *IEEE Transactions on Power Delivery,* Vol. 18, No. 4, October 2003, pp. 1241-1248.

[19] *SOM Toolbox for MATLAB 5*. Helsinki, Finland: Helsinki Univ. Technology, 2000.

[20] D.L. Davies, D. W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 2, 1979, pp. 224-227.

[21] D. Hand, H. Manilla, P. Smyth. *Principles of data mining*, The M.I.T. Press, Cambridge, Massachusetts, London, England, 2001.