We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists



122,000





Our authors are among the

TOP 1%





WEB OF SCIENCE

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

## Interested in publishing with us? Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected. For more information visit www.intechopen.com



## Load Time-Series Classification Based on Pattern Recognition Methods

George J. Tsekouras<sup>1-2</sup>, Anastasios D. Salis<sup>2</sup>, Maria A. Tsaroucha<sup>2</sup> and Irene S. Karanasiou<sup>2</sup> <sup>1</sup>Hellenic Naval Academy,

<sup>2</sup>School of Electrical and Computer Engineering, National Technical University of Athens, Greece

## 1. Introduction

### 1.1 Introduction to load time series classification

The formation of typical chronological load curves is an important tool of resolution of many problems in power systems, such as the short-term and medium-term load forecasting, the adaptation of customers' tariffs and the classification of electricity customers.

Classical indexes, like maximum power or load factor, can not describe the electricity behaviour of a customer or a power system thoroughly, as it can be comprehended from the example of Fig 1.1, where the customer of Fig 1.1(a) fatigues the power system's generators less than the customer of Fig 1.1(b) for the same peak load, load factor and power factor, because the number of the load demand changes are fewer. If an energy storage system is used, the second customer will need smaller battery system than the first one. These inferences cannot be drawn without the customers' load profiles.



Fig. 1.1. Indicative load curves of electricity consumers with the same max load, load factor and power factor

Source: Pattern Recognition Techniques, Technology and Applications, Book edited by: Peng-Yeng Yin, ISBN 978-953-7619-24-4, pp. 626, November 2008, I-Tech, Vienna, Austria

In the case of short-term load forecasting the use of the typical days decreases the mean absolute percentage error, especially for anomalous days (i.e. holiday periods) (Chicco et al., 2001; Lamedica et al., 1996). Through this segmentation the load forecasting models are not misled by the respective load curves of last days in which more weight is usually given. Similarly the formation of typical chronological load curves and the corresponding diachronic development can be used for medium-term load forecasting (Al - Hamadi & Soliman, 2005), so that the maintenance of the units and electric network, the fuel supply, the electrical energy imports/exports and the exploitation of the water reserves for hydrothermal scheduling can be implemented.

In a deregulated electricity market, each supplier wishes to identify his customers' electricity behaviour accurately, in order to provide them with satisfactory services at a low cost, recovering the energy and power cost and having a fair profit. So the classification of electricity customers is a necessary stage. At the same time, each consumer wants to know his electricity behaviour, in order to select the proper tariff or to apply energy efficiency measures successfully. Taking into consideration the demand-side bidding in competitive markets (Task VIII of IEA, 2002) the accurate estimation of the next day's load profile is a fundamental requirement for each large customer, so that it can find the way to minimize its electricity bill.

During the last years, a significant research effort has been focused on load curves classification regarding the short-term load forecasting of anomalous days and the clustering of the customers of the power systems. The clustering methods have been used so far are:

- the "modified follow the leader" (Chicco et al., 2003a; -, 2003b; -, 2004; -, 2006),
- the self-organizing map (Beccali et al., 2004; Chicco et al., 2004; -, 2006; Figueiredo et al., 2003; Lamedica et al. 1996; Verdu et al., 2003),
- the k-means (Chicco et al., 2006; Figueiredo et al., 2003),
- the average and Ward hierarchical methods (Chicco et al., 2004; -, 2006; Gerber et al., 2003) and
- the fuzzy k-means (Chicco et al., 2004; -, 2006; Gerber et al., 2003; -, 2004; -, 2005).

All the above methods generally belong to pattern recognition techniques (Theodoridis & Koutroumbas, 1999). Alternatively, classification problem can be solved by using data mining (Kitayama et al., 2003; Figueiredo et al., 2005), wavelet packet transformation (Petrescu & Scutariu, 2002), frequency-domain data (Carpaneto et al., 2006), stratified sampling (Chen et al., 1997). For the reduction of the size of the clustering input data set Sammon map, principal component analysis and curvilinear component analysis have been proposed (Chicco et al., 2006).

The respective adequacy measures that are commonly used are:

- the mean index adequacy (Chicco et al., 2003a; -, 2003b; -, 2004),
- the clustering dispersion indicator (Chicco et al., 2003a; -, 2003b; -, 2004; -, 2006),
- the similarity matrix indicator (Chicco et al., 2004),
- the Davies-Bouldin indicator (Beccali et al., 2004; Chicco et al., 2003; -, 2006; Gerbec et al., 2004; -, 2005),
- the modified Dunn index (Chicco et al., 2006),
- the scatter index (Chicco et al., 2006) and
- the mean square error (Gerbec et al., 2003; -, 2004; -, 2005).

In all cases analytical chronological load curves are required, which have resulted via suitable measurements or load surveys. The use of classification methods allow us to compress data information implementing the fundamental concepts of data mining and pattern recognition.

# 1.2 Why should load time-series classification be realized using unsupervised pattern recognition methods? What kind of problems are we going to meet?

According to R.O.Duda, P.E. Hart, D. G. Stock (Duda et al., 2001), it is known that "*pattern recognition*" is the act of taking in raw data and making an action based on the category of the pattern. Generally any method that incorporates information from training samples in the design of a classifier employs learning. There are three kinds of learning:

- *Supervised learning*, in which a category label or cost for each pattern is provided in a training set and the sum of these patterns should be minimized using methods based on Bayesian decision theory, maximum likelihood and Bayesian parameter estimation, multilayer neural networks, probabilistic neural network etc.
- *Unsupervised learning* or clustering, where there are no any a priori category labels for patterns and the pattern recognition system forms clusters sets of the input patterns. Different clustering algorithms, such as self organizing map, adaptive vector quantization etc, lead to different clusters.
- *Reinforcement learning*, where no desired category signal is given, but the only feedback is that the specified category is right or wrong, without saying why it is wrong.

In our case the clusters of the load time-series are unknown. We do not know either the clusters, or their number. The danger of an inappropriate representation is big. In order to realize the categorization of the load time-series and to select the proper number of clusters we are going to study the behaviour of the adequacy measures, which show us when the proper number of clusters is determined. Other basic notions in our approach are the following:

- 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 mono-dimensional SOM, in order to fit the classification needs;
- the comparison of the performance of the clustering algorithms for each one of the adequacy measures;
- the introduction of the ratio of within cluster sum of squares to between cluster variation, which is first presented for this kind of classification.

# 1.3 Mathematical modeling of clustering methods and adequacy measures 1.3.1 General introduction

We assume that the classification of daily load curves is necessary using the proper pattern recognition method. Generally *N* is defined as the population of the input vectors, which are going to be clustered. The  $\ell$  -th input vector is symbolized as follows:

$$\vec{x}_{\ell} = \left(x_{\ell 1}, x_{\ell 2}, \dots x_{\ell i}, \dots x_{\ell d}\right)^{T}$$
(1.1)

where *d* is its dimension, which equals to 96 or 24, if the load measurements are taken every 15 minutes or every hour respectively. The corresponding set of vectors is given by:

$$X = \{ \vec{x}_{\ell} : \ell = 1, ..., N \}$$
(1.2)

It is worth mentioning that  $x_{\ell_i}$  are normalized using the upper and lower values of all elements of the original input patterns set, aiming the achievement of the best possible results after the application of clustering methods.

Each classification process makes a partition of the initial N input vectors to M clusters, which can be the typical days of the under study customer (first example) or the customer classes (second example - the second stage of the proposed methodology of (Tsekouras et al., 2007)) or the typical days of the power system (third example). The *j*-th cluster has a representative, which is the respective load profile and is represented by the vector of *d* dimension:

$$\vec{w}_{j} = \left(w_{j1}, w_{j2}, \dots w_{ji}, \dots w_{jd}\right)^{T}$$
 (1.3)

The last vector also expresses the cluster's centre or the weight vector of neuron, if a clustering artificial neural network is used. In our case it is also called the *j*-th *class representative load diagram*. The corresponding set is the classes' set, which is defined by:

$$W = \{\vec{w}_k, k = 1, ..., M\}$$
(1.4)

The subset of input vectors  $\vec{x}_{\ell}$ , which belong to the *j*-th cluster, is  $\Omega_j$  and the respective population of load diagrams is  $N_j$ . More specifically  $\Omega_j$  is determined as follows:

$$\Omega_{j} = \left\{ \vec{x}_{\ell}, \ell = 1, \dots N \& \operatorname*{arg\,min}_{\forall k'} f\left(\vec{x}_{\ell}, \vec{w}_{k'}\right) \to j \right\}$$
(1.5)

where  $\Omega_j \subseteq X$  and  $\arg\min_{\forall k'} f(\vec{x}_{\ell}, \vec{w}_{k'})$  the corresponding criterion of classification of the *l*-th vector in the *j*-th cluster.

For the study and evaluation of classification algorithms the following distances' forms are defined:

1. the Euclidean distance between  $\ell_1$ ,  $\ell_2$  input vectors of the set X:

$$d\left(\vec{x}_{\ell_{1}}, \vec{x}_{\ell_{2}}\right) = \sqrt{\frac{1}{d} \sum_{i=1}^{d} \left(x_{\ell_{1}i} - x_{\ell_{2}i}\right)^{2}}$$
(1.6)

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

$$d\left(\vec{w}_{k},\Omega_{k}\right) = \sqrt{\frac{1}{N_{k}}\sum_{\vec{x}_{\ell}\in\Omega_{k}}d^{2}\left(\vec{w}_{k},\vec{x}_{\ell}\right)}$$
(1.7)

3. 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  $\Omega_j$  and for the subset *W*:

$$\hat{d}(\Omega_k) = \sqrt{\frac{1}{2N_k} \sum_{\vec{x}_\ell \in \Omega_k} d^2(\vec{x}_\ell, \Omega_k)}$$
(1.8)

$$\hat{d}(W) = \sqrt{\frac{1}{2M} \sum_{k=1}^{M} d^2(\vec{w}_k, W)}$$
(1.9)

#### 1.3.2 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 to make the load diagrams self explanatory. The definitions of these measures are the following:

1. *Mean square error or error function (J)* (Gerber et al., 2003), which expresses the distance of each vector from its cluster's centre with the same value of weight:

$$J = \frac{1}{N} \sum_{\ell=1}^{N} d^2 \left( \vec{x}_{\ell}, \vec{w}_{k:\vec{x}_{\ell} \in \Omega_k} \right)$$
(1.10)

2. *Mean index adequacy (MIA)* (Chicco et al., 2003a), which is defined as the average of the distances between each input vector assigned to the cluster and its centre:

$$MIA = \sqrt{\frac{1}{M} \sum_{k=1}^{M} d^2 \left( \vec{w}_k, \Omega_k \right)}$$
(1.11)

3. *Clustering dispersion indicator (CDI)* (Chicco et al., 2003a), 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:

$$CDI = \frac{1}{\hat{d}(W)} \sqrt{\frac{1}{M} \sum_{k=1}^{M} \hat{d}^{2}(\Omega_{k})}$$
(1.12)

4. *Similarity matrix indicator (SMI)* (Chicco et al., 2003b), 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:

$$SMI = \max_{p \mid q} \left\{ \left( 1 - \frac{1}{\ln\left[d\left(\vec{w}_p, \vec{w}_q\right)\right]} \right)^{-1} \right\} : p, q = 1, \dots, M$$
(1.13)

5. *Davies-Bouldin indicator (DBI)* (Davies & Bouldin., 1979), which represents the systemwide average of the similarity measures of each cluster with its most similar cluster:

$$DBI = \frac{1}{M} \sum_{k=1}^{M} \max_{p \neq q} \left\{ \frac{\hat{d}\left(\Omega_{p}\right) + \hat{d}\left(\Omega_{q}\right)}{d\left(\vec{w}_{p}, \vec{w}_{q}\right)} \right\} : p, q = 1, \dots, M$$
(1.14)

6. *Ratio of within cluster sum of squares to between cluster variation (WCBCR)* (Hand et al., 2001), which depends on the sum of the distance's square between each input vector and its cluster's representative vector, as well as the similarity of the clusters' centres:

$$WCBCR = \frac{\sum_{k=1}^{M} \sum_{\vec{x}_{\ell} \in \Omega_{k}} d^{2}\left(\vec{w}_{k}, \vec{x}_{\ell}\right)}{\sum_{1 \le q < p}^{M} d^{2}\left(\vec{w}_{p}, \vec{w}_{q}\right)}$$
(1.15)

The success of the different 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. (1.10)-(1.15), M is the number of the clusters without the dead ones.

#### 1.3.3 K-means

The *k*-means method is the simplest hard clustering method, which gives satisfactory results for compact clusters (Duda et al., 2001). The *k*-means clustering method groups the set of the N input vectors to M clusters using an iterative procedure. The respective steps of the algorithm are the follows:

a. Initialization of the weights of *M* clusters is determined. In the classic model a random choice among the input vectors is used (Chicco et al., 2006; Figueiredo et al., 2003). In the developed algorithm the  $w_{ji}$  of the *j*-th centre is initialized as:

$$w_{ji}^{(0)} = a + b \cdot (j-1)/(M-1)$$
(1.16)

where *a* and *b* are properly calibrated parameters. Alternatively the  $w_{ji}$  is initialized as:

where 
$$a_i = \min_{\forall j} (x_{ji})$$
 and  $b_i = \max_{\forall j} (x_{ji})$ . (1.17)

b. During epoch *t* for each training vector  $\vec{x}_{\ell}$  its Euclidean distances  $d(\vec{x}_{\ell}, \vec{w}_{j})$  are calculated for all centres. The  $\ell$ -th input vector is put in the set  $\Omega_{j}^{(t)}$ , for which the distance between  $\vec{x}_{\ell}$  and the respective centre is minimum, which means:

$$d\left(\vec{x}_{\ell}, \vec{w}_{k}\right) = \min_{\forall j} d\left(\vec{x}_{\ell}, \vec{w}_{j}\right)$$
(1.18)

c. When the entire training set is formed, the new weights of each centre are calculated as:

$$\vec{w}_{j}^{(t+1)} = \frac{1}{N_{j}^{(t)}} \sum_{\vec{x}_{\ell} \in \Omega_{j}^{(t)}} \vec{x}_{\ell}$$
(1.19)

where  $N_i^{(t)}$  is the population of the respective set  $\Omega_i^{(t)}$  during epoch *t*.

d. Next, the number of the epochs is increased by one. This process is repeated (return to step b) until the maximum number of epochs is used or weights do not significantly change  $(\left|\vec{w}_{j}^{(t)} - \vec{w}_{j}^{(t+1)}\right| < \varepsilon$ , where  $\varepsilon$  is the upper limit of weight change between sequential iterations). The algorithm's main purpose is to minimize the appropriate error function *J*. The main difference with the classic model is that the process is repeated for various pairs of (*a*,*b*). The best results for each adequacy measure are recorded for various pairs (*a*,*b*).

At the end of the execution of the algorithm the six adequacy measures are calculated, which are used for comparison reasons with the other clustering methods. The core of algorithm is executed from  $M_1$  to  $M_2$  neurons, because the necessary number of clusters is not known a priori, as it depends on the time period which is examined and the available number of patterns.

#### 1.3.4 Kohonen adaptive vector quantization - AVQ

This algorithm is a variation of the *k*-means method, which belongs to the unsupervised competitive one-layer neural networks. It classifies input vectors into clusters by using a competitive layer with a constant number of neurons. Practically in each step all clusters compete each other for the winning of a pattern. The winning cluster moves its centre to the direction of the pattern, while the rest clusters move their centres to the opposite direction (supervised classification) or remain stable (unsupervised classification).

Here, we will use the last unsupervised classification algorithm. The respective steps are the following:

- a. Initialization of the weights of *M* clusters is determined, where the weights of all clusters are equal to 0.5, that is  $w_{ji}^{(0)} = 0.5, \forall j, i$ .
- b. During epoch *t* each input vector  $\vec{x}_{t}$  is randomly presented and its respective Euclidean distances from every neuron are calculated. In the case of existence of bias factor  $\lambda$ , the respective minimization function is:

$$f_{winner\_neuron}\left(\vec{x}_{\ell}\right) = j : \min_{\forall j} \left( d\left(\vec{x}_{\ell}, \vec{w}_{j}\right) + \lambda \cdot N_{j} / N \right)$$
(1.20)

where  $N_j$  is the population of the respective set  $\Omega_j$  during epoch *t*-1.

The weights of the winning neuron (with the smallest distance) are updated as:

$$\vec{w}_{j}^{(t)}(n+1) = \vec{w}_{j}^{(t)}(n) + \eta(t) \cdot \left(\vec{x}_{\ell} - \vec{w}_{j}^{(t)}(n)\right)$$
(1.21)

where *n* is the number of input vectors, which have been presented during the current epoch, and  $\eta(t)$  is the learning rate according to:

$$\eta(t) = \eta_0 \cdot \exp\left(-\frac{t}{T_{\eta_0}}\right) > \eta_{\min}$$
(1.22)

where  $\eta_0$ ,  $\eta_{\min}$  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 (Kohonen, 1989; Haykin, 1994).

c. Next, the number of the epochs is increased by one. This process is repeated (return to step b) until either the maximum number of epochs is reached or the weights converge or the error function *J* does not improve, which means:

$$\left| \frac{J^{(t)} - J^{(t+1)}}{J^{(t)}} \right| < \varepsilon' \text{ for } t \ge T_{in}$$
(1.23)

where  $\varepsilon'$  is the upper limit of error function change between sequential iterations and the respective criterion is activated after  $T_{in}$  epochs.

The core of algorithm is executed for specific number of neurons and the respective parameters  $\eta_0$ ,  $\eta_{\min}$  and  $T_{\eta 0}$  are optimized for each adequacy measure separately. This process is repeated from  $M_1$  to  $M_2$  neurons.

#### 1.3.5 Fuzzy k-means

During the application of the k-mean or the adaptive vector quantization algorithm each pattern is assumed to be in exactly one cluster (hard clustering). In many cases the areas of two neighbour clusters are overlapped, so that there are not any valid qualitative results.

If we want to relax the condition of exclusive partition of an input pattern to one cluster, we should use fuzzy clustering techniques. Specifically, 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:

$$\sum_{j=1}^{M} u_{\ell j} = 1 \& 0 \le u_{\ell j} \le 1, \forall j$$
(1.24)

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

$$J_{fuzzy} = \frac{1}{N} \sum_{j=1}^{M} \sum_{\ell=1}^{N} u_{\ell j} \cdot d^2 \left( \vec{x}_{\ell}, \vec{w}_{j} \right)$$
(1.25)

The simplest algorithm is the fuzzy k-means clustering one, in which the respective steps are the following:

- a. Initialization of the weights of *M* clusters is determined. In the classic model a random choice among the input vectors is used (Chicco et al., 2006; Figueiredo et al., 2003). In the developed algorithm the  $w_{ji}$  of the *j*-th centre is initialized by eq. (1.16) or eq. (1.17).
- b. During epoch *t* for each training vector  $\vec{x}_{t}$  the membership factors are calculated for every cluster:

$$u_{\ell j}^{(t+1)} = \frac{1}{\sum_{k=1}^{M} \frac{d\left(\vec{x}_{\ell}, \vec{w}_{j}^{(t)}\right)}{d\left(\vec{x}_{\ell}, \vec{w}_{k}^{(t)}\right)}}$$
(1.26)

c. Afterwards the new weights of each centre are calculated as:

$$\vec{w}_{j}^{(t+1)} = \frac{\sum_{\ell=1}^{N} \left( u_{\ell j}^{(t+1)} \right)^{q} \cdot \vec{x}_{\ell}}{\sum_{\ell=1}^{N} \left( u_{\ell j}^{(t+1)} \right)^{q}}$$
(1.27)

where *q* is the *amount of fuzziness* in the range  $(1, \infty)$  which increases as fuzziness reduces.

d. Next, the number of the epochs is increased by one. This process is repeated (return to step b) until the maximum number of epochs is used or weights do not significantly change.

This process is repeated for different pairs of (a,b) and for different amounts of fuzziness. The best results for each adequacy measure are recorded for different pairs (a,b) and q.

#### 1.3.6 Self-organizing map - SOM

The Kohonen SOM (Kohonen, 1989; SOM Toolbox for MATLAB 5, 2000; Thang et al., 2003) is a topologically unsupervised neural network that projects a *d*-dimensional input data set into a reduced dimensional 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 neighbour units in inverse proportion of their distance. The neighbourhood size of each neuron shrinks progressively during the training process, starting with nearly the whole map and ending with the single neuron.

The process of algorithm is described by the following stages:

- *Initialization stage*. The weights of the neural network are initialized connecting the neurons of the input layer with the map neurons.
- *Competition stage.* For each input pattern the map neurons calculate the corresponding value of the competition function, where the neuron with the biggest value is the winner.
- *Collaboration stage*. The winner neuron determines the territorial area of topological neighbourhood, providing the subbase for the collaboration between the neighbouring neurons.
- *Weights' adaptation stage.* The neurons that belong in the winning neighbourhood adapt their weights of winner-neuron, so that its response will be strengthened during the presentation of a training input pattern.

The training of SOM is divided to 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,

369

- *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 the fine tuning one is happened after  $T_{s0}$  epochs.

More analytically, the respective steps of the SOM algorithm are the following:

The shape and the number of neurons of the SOM's grid are defined and the a. initialization of the respective weights is determined. Specifically, in the case of the mono-dimensional SOM the weights can be given by (a)  $w_{ki} = 0.5, \forall k, i$ , (b) the random initialization of each neuron's weight, (c) the random choice of the input vectors for each neuron. 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 by rectangular or hexagonal arrangement of neurons, as it is presented in Fig. 1.2. The population of the neurons is recommended to be  $5 \times \sqrt{N}$  to  $20 \times \sqrt{N}$  (Chicco et al., 2004; SOM Toolbox for MATLAB 5, 2000; Thang et al., 2003;). The height/width ratio  $M_1/M_2$  of the rectangular grid can be calculated as the ratio between the two major eigenvalues  $\lambda_1$ ,  $\lambda_2$  of the covariance matrix of the input vectors set (with  $\lambda_1 > \lambda_2$ ). The initialization of the neurons can be a linear combination of the respective eigenvectors  $\vec{e}_1$  and  $\vec{e}_2$  of the two major eigenvalues or can be equal to 0.5. It is reminded that the element  $s_{k_1,k_2}$  of the covariance matrix of the input vectors set is given by:

$$s_{k_1,k_2} = \sum_{\ell=1}^{N} \left( x_{k_1\ell} - \overline{x}_{k_1} \right) \cdot \left( x_{k_2\ell} - \overline{x}_{k_2} \right) / (N-1)$$
(1.28)

where  $\overline{x}_{k_1}$  is the mean value of the respective  $k_1$  dimension of all input patterns.

b. The SOM training commences 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 winning neuron *i*' that is closest to  $\vec{x}_{\ell}$  (competition stage). The *j*-th reference vector is updated (weights' adaptation stage) according to:

$$\vec{w}_{j}^{(t)}(n+1) = \vec{w}_{j}^{(t)}(n) + \eta(t) \cdot h_{i'j}(t) \cdot \left(\vec{x}_{\ell} - \vec{w}_{j}^{(t)}(n)\right)$$
(1.29)

where  $\eta(t)$  is the learning rate according to:

$$\eta(t) = \eta_0 \cdot \exp\left(-\frac{t}{T_{\eta_0}}\right) > \eta_{\min}$$
(1.30)

with  $\eta_0$ ,  $\eta_{\min}$  and  $T_{\eta_0}$  representing the initial value, the minimum value and the time parameter respectively. During the rough ordering phase  $\eta_r$ ,  $T_{\eta_0}$  are the initial value



x-axis

a. Rectangular arrangement



Fig. 1.2. Arrangement of bi-dimensional self-organized map 10x10

and the time parameter respectively, while during the fine tuning phase the respective values are  $\eta_f$ ,  $T_{\eta_0}$ . The  $h_{i'j}(t)$  is the neighbourhood 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  $\vec{x}_{\ell}$  (collaboration stage). In this case the Gauss function is proposed:

$$h_{i'j}\left(t\right) = \exp\left[-\frac{d_{i'j}^2}{2\cdot\sigma^2\left(t\right)}\right]$$
(1.31)

where  $d_{i'j} = \|\vec{r}_{i'} - \vec{r}_{j}\|$  is the respective distance between *i'* and *j* neurons,  $\vec{r}_{j} = (x_{j,}y_{j})$  are the respective co-ordinates in the grid,  $\sigma(t) = \sigma_{0} \cdot \exp(-t/T_{\sigma_{0}})$  is the decreasing neighbourhood radius function where  $\sigma_{0}$  and  $T_{\sigma_{0}}$  are the respective initial value and time parameter of the radius respectively.

c. Next, the number of the epochs is increased by one. This process is repeated (return to step b) until either the maximum number of epochs is reached or the index *I*<sub>s</sub> gets the minimum value (SOM Toolbox for MATLAB 5, 2000):

$$I_s(t) = J(t) + ADM(t) + TE(t)$$
(1.32)

where the quality measures of the optimum SOM are based on the quantization error *J* - given by (1.10)-, the topographic error *TE* and the average distortion measure *ADM*. The topographic error 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 neighbouring map units:

$$TE = \sum_{\ell=1}^{N} neighb\left(i_{1}^{\prime}, i_{2}^{\prime}\right) / N$$
(1.33)

where, for each input vector,  $neighb(i'_1, i'_2)$  equals to 1, if  $i'_1$  and  $i'_2$  neurons are not neighbours, either 0. The average distortion measure is given for the *t* epoch by:

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

$$(1.34)$$

This process is repeated for different parameters of  $\sigma_0$ ,  $\eta_f$ ,  $\eta_r$ ,  $T_{\eta 0}$ ,  $T_{\sigma 0}$  and  $T_{s_0}$ . Alternatively, the multiplicative factors  $\phi$  and  $\xi$  are introduced -without decreasing the generalization ability of the parameters' calibration:

$$T_{s_0} = \phi \cdot T_{\eta_0} \tag{1.35}$$

$$T_{\sigma_0} = \xi \cdot T_{\eta_0} / \ln \sigma_0 \tag{1.36}$$

The best results for each adequacy measure are recorded for different parameters  $\sigma_0$ ,  $\eta_f$ ,  $\eta_r$ ,  $T_{\eta_0}$ ,  $\phi$  and  $\xi$ .

In the case of the bi-dimensional map, the immediate exploitation of the respective clusters is not a simple problem. We can exploit the map either through human vision or applying a second simple clustering method. According to Chicco et al., (2002), the simple k-mean method was used, while, here, the proposed k-mean method with initialization by eq. (1.16) is used. Practically, the neurons of the map sustain a new data compression from which the final classification of the input patterns is concluded.

### 1.3.7 Hierarchical agglomerative algorithms

Hierarchical algorithms have a different philosophy compared to the aforementioned algorithms. Instead of producing a single clustering, they produce a hierarchy of clustering.

Agglomerative algorithms are based on matrix theory (Theodoridis & Koutroumbas, 1999). 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:

$$d(C_q, C_s) = f(d(C_i, C_s), d(C_j, C_s), d(C_i, C_j))$$

$$(1.37)$$

Alternatively eq. (1.37) is written as:

$$d(C_q, C_s) = a_i \cdot d(C_i, C_s) + a_j \cdot d(C_j, C_s) + b \cdot d(C_i, C_j) + c \cdot \left| d(C_i, C_s) - d(C_j, C_s) \right| (1.38)$$

where  $a_i$ ,  $a_j$ , b and c correspond to different choices of the dissimilarity measure. The basic algorithms, which are going to be used in our case, are:

• the *single link* algorithm (*SL*) -it is obtained from (1.38) for  $a_i = a_j = 0.5$ , b = 0 and c = -0.5:

$$d(C_{q}, C_{s}) = \min\left\{d(C_{i}, C_{s}), d(C_{j}, C_{s})\right\} = \frac{1}{2} \cdot d(C_{i}, C_{s}) + \frac{1}{2} \cdot d(C_{j}, C_{s}) - \frac{1}{2} \cdot \left|d(C_{i}, C_{s}) - d(C_{j}, C_{s})\right|$$
(1.39)

• the *complete link* algorithm (*CL*) -it is obtained from (1.38) for  $a_i=a_j=0.5$ , b=0 and c=0.5:

$$d(C_{q},C_{s}) = \max\left\{d(C_{i},C_{s}),d(C_{j},C_{s})\right\} = \frac{1}{2} \cdot d(C_{i},C_{s}) + \frac{1}{2} \cdot d(C_{j},C_{s}) + \frac{1}{2} \cdot \left|d(C_{i},C_{s}) - d(C_{j},C_{s})\right| (1.40)$$

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

$$d\left(C_{q},C_{s}\right) = \frac{n_{i} \cdot d\left(C_{i},C_{s}\right) + n_{j} \cdot d\left(C_{j},C_{s}\right)}{n_{i} + n_{j}}$$
(1.50)

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*):

$$d\left(C_{q},C_{s}\right) = \frac{1}{2} \cdot \left\{ d\left(C_{i},C_{s}\right) + d\left(C_{j},C_{s}\right) \right\}$$

$$(1.42)$$

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

$$d^{(1)}(C_q, C_s) = \frac{\left(n_i \cdot d^{(1)}(C_i, C_s) + n_j \cdot d^{(1)}(C_j, C_s)\right)}{\left(n_i + n_j\right)} - n_i \cdot n_j \cdot \frac{d^{(1)}(C_i, C_j)}{\left(n_i + n_j\right)^2}$$
(1.43)

where  $d^{(1)}(C_q, C_s) = \|\vec{w}_q - \vec{w}_s\|^2$  and  $\vec{w}_q$  is the representative centre of the *q*-th cluster according to the following equation (which is similar to(1.39)):

$$\vec{w}_q = \frac{1}{n_q} \cdot \sum_{\vec{x}_v \in C_q} \vec{x}_v \tag{1.44}$$

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

$$d^{(1)}(C_q, C_s) = \frac{1}{2} \cdot \left( d^{(1)}(C_i, C_s) + d^{(1)}(C_j, C_s) \right) - \frac{1}{4} \cdot d^{(1)}(C_i, C_j)$$
(1.45)

• the Ward or minimum variance algorithm (WARD):  

$$d^{(2)}(C_q, C_s) = \frac{(n_i + n_s) \cdot d^{(2)}(C_i, C_s) + (n_j + n_s) \cdot d^{(2)}(C_j, C_s) - n_s \cdot d^{(2)}(C_i, C_j)}{(n_i + n_j + n_s)}$$
(1.46)

where:

$$d^{(2)}(C_{i},C_{j}) = \frac{n_{i} \cdot n_{j}}{n_{i} + n_{j}} \cdot d^{(1)}(C_{i},C_{j})$$
(1.47)

It is noted that in each level t the respective representative vectors are calculated by eq.(1.44).

- The respective steps of each algorithm are the following:
- a. *Initialization*: The set of the remaining patterns  $\Re_0$  for zero level (t = 0) is the set of the input vectors *X*. The similarity matrix  $P_0 = P(X)$  is determined. Afterwards *t* increases by one (t=t+1).
- b. During level *t* clusters  $C_i$  and  $C_j$  are found, for which the minimization criterion is satisfied  $d(C_i, C_j) = \min_{r,s=1,\dots,N,r \neq s} d(C_r, C_s)$ .
- c. Then clusters  $C_i$  and  $C_j$  are merged into a single cluster  $C_q$  and the set of the remaining patterns  $\Re_t$  is formed as:  $\Re_t = \left(\Re_{t-1} \{C_i, C_j\}\right) \cup \{C_q\}$ .
- d. The construction of the dissimilarity matrix  $P_t$  from  $P_{t-1}$  is realized by applying eq.(1.37).
- e. Next, the number of the levels is increased by one. This process is repeated (return to step b) until the remaining patterns  $\Re_{N-1}$  is formed and all input vectors are in the same and unique cluster.

It is mentioned that the number of iterations is determined from the beginning and it equals to the number of input vectors decreased by 1 (*N*-1).

# 1.4 A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers

#### 1.4.1 General description of the proposed methodology

The classification of daily chronological load curves of one customer is achieved by means of the pattern recognition methodology, as shown in Fig. 1.3.

The main steps are the following (Tsekouras et al., 2008):

374

a. *Data and features selection*: Using electronic meters, the active and reactive energy values are registered (in kWh and kvarh) for each time period in steps of 15 minutes, 1 hour, etc. The daily chronological load curves are determined for the study period.



Fig. 1.3. Flow diagram of pattern recognition methodology for the classification of daily chronological load curves of one large electricity customer

b. *Data preprocessing:* The load diagrams of the customer 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 remain uncorrected. In future, a filtering step can be added using principal component analysis, Sammon map, and curvilinear component analysis (Chicco et al., 2006), for the reduction of the load diagrams dimensions.

c. *Main application of pattern recognition methods*: For the load diagrams of the customer, a number of clustering algorithms (k-means, adaptive vector quantization, self organized map, fuzzy k-means 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 it is necessary. The developed methodology uses the clustering methods that provide the most satisfactory results. It should be noted that conventional methods, like statistical tools, supervised techniques, etc., cannot be used, because the classification of the typical days must be already known.

## **1.4.2 Application of the proposed methodology to a medium voltage customer** *1.4.2.1 General*

The developed methodology was analytically applied on one medium voltage industrial paper mill customer of the Greek distribution system. The data used are 15 minutes load values for a period of ten months in 2003. The respective set of the daily chronological curves has 301 members. Nine curves were rejected through data pre-processing, while the remaining 292 diagrams were used by the aforementioned clustering methods. The last diagrams are registered in Fig. 1.4 and Fig. 1.5, in which the load variability is also presented. The load behaviour is significantly decreased during holiday time. The mean load demand is 6656 kW and the peak load demand is 9469 kW during the period under study.



Fig. 1.4. Daily chronological 15-minutes load diagrams for a set of 292 days for the industrial medium voltage customer for each day (February – November 2003)



Fig. 1.5. Chronological 15-minutes load diagram for a set of 292 days for the industrial medium voltage customer for the time period under study (February-November 2003)

The main goal of the application of this methodology is the representation of the load behaviour of the customer with typical daily load chronological diagrams. This is achieved through the following steps:

- The calibration of the parameters of each clustering method is realized for every adequacy measure separately and the performance for different number of clusters is registered.
- The clustering models are compared to each other using the six adequacy measures, the behaviour of these measures is studied and the appropriate number of the clusters is defined.

The representative daily load chronological diagrams of the customer are calculated for the best clustering techniques and the proposed number of clusters.

#### 1.4.2.2 Application of the k-means

The proposed model of the *k*-means method (*k*-means-scenario 1 with the weights initialization based on eq.(1.16)) is executed for different pairs (*a*,*b*) from 2 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 checked. The best results for the 6 adequacy measures do not refer to the same pair (*a*,*b*). The second model of the *k*-means method (*k*-means-scenario 2) is based on eq. (1.17) for the weights initialization. The third model (*k*-means-scenario 3) is the classic one with the random choice of the input vectors during the centres' initialization. For the classic *k*-means model, 100 executions are carried out and the best results for each index are registered. In Fig. 1.6, it is obvious that the proposed k-means is superior to the other two scenarios of *k*-means. The superiority of the proposed model applies in all cases of neurons.

A second advantage comprises the convergence to the same results for the respective pairs (a,b), which cannot be reached using the classic model.



d. SMI indicator

378



Fig. 1.6. Adequacy measures for the k-means method for a set of 292 training patterns for 5 to 25 clusters (scenario 1: proposed method – weights initialization based on eq. (1.16)-, scenario 2: alternative method – weights initialization based on eq. (1.17)-, scenario 3: classic method)

In Fig. 1.7 the dead clusters for the proposed k-means method are presented for the six different adequacy measures. It is obvious that *WCBCR* presents the best behaviour, because the first dead cluster is presented when 23 clusters are required, while all other measures present dead clusters for smaller required clusters.



Fig. 1.7. Dead clusters for the proposed k-means method for the six different adequacy measures for a set of 292 training patterns for 5 to 25 clusters

It is mentioned that the maximum number of epochs is 200 for the three scenarios, the upper limit of the weight change between sequential iterations  $\mathcal{E}$  is 10<sup>-4</sup>. Practically the algorithm is always converged after at most 20-30 iterations.

## 1.4.2.3 Application of the adaptive vector quantization

During the application of the AVQ method with serial presentation and without bias factor the parameters  $\eta_0$ ,  $\eta_{\min}$  and  $T_{\eta 0}$  should be optimized. Specifically, the model is executed for  $\eta_0 = \{0.05, 0.1, ..., 0.9\}$  and  $T_{\eta 0} = \{500, 1000, ..., 5000\}$  from 2 to 25 clusters with  $\eta_{\min}$  stable (=10<sup>-5</sup>). Indicatively the adequacy measures of the AVQ method for 10 clusters are presented in Fig. 1.8, where the best results for each adequacy measure are presented for different areas of  $\eta_0$  and  $T_{\eta 0}$ .



a. J indicator



b. MIA indicator



e. DBI indicator



f. WCBCR indicator

Fig. 1.8. Adequacy measures for the AVQ method (with serial presentation and without bias factor) for a set of 292 training patterns for 10 clusters,  $\eta_0$ ={0.05, 0.1,..., 0.9}, T<sub> $\eta_0$ </sub>={500, 1000, ..., 5000}

The *J* indicator presents the best results for  $\eta_0 > 0.45$  and  $T_{\eta_0} \le 2000$ , while the *CDI* indicator has similar behaviour to the *J* one. The *MIA* and *WCBCR* indicators present their best results for  $\eta_0 > 0.85$  and  $T_{\eta_0} \le 2000$ , *DBI* and *SMI* indicators for  $\eta_0 \approx 0.45$ ,  $\forall T_{\eta_0}$ . For different number of clusters the pairs ( $\eta_0$ ,  $T_{\eta_0}$ ) for the best results are not the same, but the greater areas are similar, as it is presented for the *J* indicator for 8 and 15 clusters in Fig. 1.9 indicatively. Generally, as the number of clusters increases, so the respective behaviour of ( $\eta_0$ ,  $T_{\eta_0}$ ) is stabilized. The value of the parameter  $\eta_{\min}$  is not significant, but it helps towards the algorithm's convergence for a big number of epochs with the condition  $\eta_{\min}$  not having zero value. In this problem the proper values of this parameter are between 10<sup>-4</sup> and 10<sup>-6</sup>.

During the application of the AVQ method with random presentation and without bias factor the respective results are improved against the serial presentation having two disadvantages:

• the computing time increases by 10% and

• the convergence areas for pairs ( $\eta_0$ ,  $T_{\eta_0}$ ) have more unstable shape.

Indicatively, *J* and *WCBCR* measures are presented for 10 clusters in Fig. 1.10 improving the respective values in comparison to serial presentation from 0.259 to 0.250 and from 0.0083 to 0.0068 respectively.

If the bias factor is used with values between 10<sup>-3</sup> and 10, the respective results are not practically improved (there is a slight improvement of the forth significant digit for each adequacy measure). Since the computing time is increased by 50%, we propose not to use the bias factor.



Fig. 1.9. *J* adequacy measures for the AVQ method (with serial presentation and without bias factor) for a set of 292 training patterns for 8 clusters and 15 clusters,  $\eta_0$ ={0.05, 0.1,..., 0.9},  $T_{\eta_0}$ ={500, 1000, ..., 5000}

It is mentioned that the maximum number of epochs is 10000, the upper limit of the weight change between sequential iterations  $\varepsilon$  and the upper limit of the error function change between sequential iterations  $\varepsilon'$  are 10<sup>-4</sup>. The algorithm usually converges after a few hundreds epochs after  $T_{\eta 0}$  epochs.



Fig. 1.10. *J* and WCBCR adequacy measures for the AVQ method (with random presentation, without bias factor) for a set of 292 training patterns for 10 clusters,  $\eta_0$ ={0.05, 0.1,..., 0.9}, T<sub> $\eta_0$ </sub>={500, 1000, ..., 5000}

In Fig. 1.11 the two basic scenarios of the AVQ method (serial and random presentation without the bias factor) are presented for all adequacy measures. The method with the random presentation is slightly superior to the other one for all adequacy measures for 9 clusters and above, except the *SMI* and *DBI* indicators.



Fig. 1.11. Adequacy measures for the AVQ method for a set of 292 training patterns for 5 to 25 clusters (serial presentation and random presentation without bias factor)

The use of the AVQ algorithm with random presentation without the bias factor is proposed, even if there is a small computing time increment against the algorithm with serial presentation.

#### 1.4.2.4 Application of the fuzzy k-means

In the fuzzy k-means algorithm the results of all adequacy measures (except *J*) improve as the amount of fuzziness increases, as shown in Fig. 1.12, where the six adequacy measures are presented for different number of clusters and for three cases of q={2,4,6}.

It is noted that the initialization of the respective weights is similar to the proposed k-means. The maximum number of epochs is 500 for the three scenarios and the upper limit of the weight change between sequential iterations  $\varepsilon$  is 10<sup>-4</sup>. Practically the algorithm is always converged after at most 400 iterations.



Fig. 1.12. Adequacy measures for the fuzzy k-means method for a set of 292 training patterns, for 5 to 25 clusters and q=2, 4, 6

### 1.4.2.5 Application of hierarchical agglomerative algorithms

In the case of the seven hierarchical models the best results are given by the *WARD* model for *J* and *CDI* adequacy measures and by the *UPGMA* model for *MIA*, *SMI*, *WCBCR* indicators. For the *Davies-Bouldin* indicator there are significant variances, according to Fig.1.13. It should be mentioned that there are not any other parameters for calibration, such as maximum number of iterations etc.



Fig. 1.13. Adequacy measures for the 7 hierarchical clustering algorithms for a set of 292 training patterns for 5 to 25 clusters

## 1.4.2.6 Application of mono-dimensional self-organizing maps

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 are:

• to stop the training process of the optimum SOM. In this case the target is to minimize the index *Is* (eq.(1.32)), which combines the quality measures of the quantization error given by eq.(1.4), the topographic error given by eq. (1.33) and the average distortion measure error given by eq.(1.34). In Fig. 1.14, the normalized values of these four indices are registered for the case of a 10x1 SOM for the chronological load curves of the industrial customer under study. Generally, it is noticed that the convergence is

completed after 0.5÷2.0·  $T_{\eta_0}$  epochs during fine tuning phase, when  $T_{\eta_0}$  has big values (≥1000 epochs).



Fig. 1.14. Quality normalized measures of the quantization error (QE), the topographic error (TE), the average distortion measure error (ADM) and the index (Is) for the monodimensional SOM with 10 clusters,  $\eta_r = 0.1$ ,  $\eta_f = 0.001$ ,  $T_{\eta_0} = 1000$ ,  $\sigma_0 = 10$ ,  $T_{s_0} = T_{\eta_0}$ ,  $T_{\sigma 0} = T_{\eta_0} / \ln \sigma_0$  in the case of a set of 292 training patterns of the industrial customer under study

- the proper initial value of the neighbourhood radius  $\sigma_0$ . The radius follows the decreasing power of the neighbourhood radius function  $\sigma(t) = \sigma_0 \cdot \exp(-t/T_{\sigma 0})$ , which has the advantage to act on all neurons of the map with decreasing weights according to the respective distances of the winning neuron. On the contrary, the linear radius function does not change the weights of those neurons with distances from the winning neuron larger than  $\sigma(t)$ . The computational time of the last one is significantly smaller than the power function. In Fig. 1.15 the effects of the initial radius  $\sigma_0$  on the adequacy measures are registered. It is noticed that the neural network's performance is improved, if the initial radius is increased  $\sigma_0$ , especially for  $T_{\eta_0} \leq 2000$ .
- the proper values of the multiplicative factor φ between T<sub>s<sub>0</sub></sub> (epochs of the rough ordering phase) and T<sub>η<sub>0</sub></sub> (time parameter of learning rate). In Fig. 1.16 the adequacy measures with respect to φ and T<sub>η<sub>0</sub></sub> are presented as indicative examples, from which it is concluded that the best behaviour of *J*, *CDI*, *SMI*, *DBI* indicators is registered for T<sub>η</sub> ≥ 800 and φ =1, while of *MIA*, *WCBCR* ones for φ =2 respectively.



Fig. 1.15. Normalized adequacy measures with respect to the initial radius  $\sigma_0$  for the monodimensional SOM with 10 clusters,  $\eta_r = 0.1$ ,  $\eta_f = 0.001$ ,  $T_{\eta_0} = 1000$ ,  $T_{s_0} = T_{\eta_0}$ ,  $T_{\sigma 0} = T_{\eta_0} / \ln \sigma_0$  in the case of a set of 292 training patterns of the industrial customer under study



a. J indicator



d. SMI indicator

390





2000

Fig. 1.16. Adequacy measures with respect to  $\phi = \{1, 2, 3, 4, 5\}$  and  $T_{\eta_0} = \{200, 400, ..., 2000\}$  for the mono-dimensional SOM with 10 clusters,  $\eta_r = 0.1$ ,  $\eta_f = 0.001$ ,  $\sigma_0 = 10$ ,

 $T_{\sigma_0} = T_{\eta_0} / \ln \sigma_0$  in the case of a set of 292 training patterns for the industrial customer under study

• the proper values of the multiplicative factor  $\xi$  between  $T_{\sigma_0}$  (time parameter of neighbourhood radius) and  $T_{\eta_0}$ . In Fig. 1.17 the adequacy measures with respect to  $\xi$  and  $T_{\eta_0}$  are presented as indicative examples, from which it is concluded that the best behaviour of *J*, *CDI* indicators is registered for  $T_{\eta} \ge 1000$  and  $\xi = \{0.2, .., 1\}$ , of *SMI*, *DBI* ones for  $T_{\eta} \ge 1000$  and  $\xi = \{0.2, 0.4\}$  respectively.



c. CDI indicator



f. WCBCR indicator

Fig. 1.17. Adequacy measures with respect to  $\xi = \{0.2, 0.4, \dots, 1.0\}$  and  $T_{\eta_0} = \{500, 1000, \dots, 3000\}$  for the mono-dimensional SOM with 10 clusters,  $\eta_r = 0.1$ ,  $\eta_f = 0.001$ ,  $\sigma_0 = 10$ ,  $T_{\sigma_0} = T_{\eta_0} / \ln \sigma_0$  in the case of a set of 292 training patterns for the industrial customer under study

- the proper values of the learning rate  $\eta_r$  during the rough ordering phase. In Fig. 1.18 the adequacy measures with respect to  $\eta_r$  and  $T_{\eta_0}$  are presented as indicative examples, from which it is concluded that the best behaviour for all indicators is registered for  $T_{\eta} \ge 1000$  and  $0.1 \le \eta_0 \le 0.15$ . Especially for *MIA*, *WCBCR* indicators the best results are succeeded for  $1000 \le T_{\eta} \le 1500$  and  $0.2 \le \eta_0 \le 0.4$  with big variations.
- the learning rate  $\eta_f$  during the fine tuning phase. From the results of the performed study, it is derived that the proper value of the parameter  $\eta_f$  must be smaller than 20% of the initial value of the learning rate  $\eta_r$  and between 10<sup>-3</sup> and 10<sup>-4</sup>. If  $\eta_f$  is increased, the behaviour of *J*, *CDI*, *SMI*, *DBI*, *WCBCR* indicators is improved whereas that of *MIA* is worsened.



a. J indicator



b. MIA indicator



e. DBI indicator

Initial value of learning rate

0.35

4000

0.4

395


#### f. WCBCR indicator

Fig. 1.18. Adequacy measures with respect to  $\eta_r = \{0.05, 0.10, ..., 0.4\}$  and  $T_{\eta_0} = \{500, 1000, ..., 4000\}$  for the mono-dimensional SOM with 10 clusters,  $\eta_f = 0.001$ ,  $\sigma_0 = 10$ ,  $T_{s_0} = T_{\eta_0}$ ,  $T_{\sigma_0} = T_{\eta_0} / \ln \sigma_0$  in the case of a set of 292 training patterns for the industrial customer under study

the initialization of the weights of the neurons. Three cases were examined: (a) w<sub>ki</sub> = 0.5, ∀k, i, (b) the random initialization of each neuron's weight, (c) the random choice of the input vectors for each neuron. The best training behaviour was presented in case (a).

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

## 1.4.2.7 Application of bi-dimensional self-organizing maps

In the case of the bi-dimensional SOM, the shape, the population of neurons and their respective arrangement are issues to be solved –beyond the optimization of parameters, which is considered during the training process of the mono-dimensional SOM.

The rectangular shape of the map is defined with rectangular or hexagonal arrangement of neurons. The population of the last ones is recommended to be  $5 \times \sqrt{N}$  to  $20 \times \sqrt{N}$  (SOM Toolbox for MATLAB 5, 2000; Thang et al., 2003; Chicco et al., 2004). In the case of the industrial customer, a set of 292 vectors was given. The map can have  $85 \ (\cong 5 \times \sqrt{292})$  to  $342 \ (\cong 20 \times \sqrt{292})$  neurons. The respective square maps can be 9x9 to 19x19. Using the ratio between the two major eigenvalues, the respective ratio is 27.31 (=4.399/0.161) and the proposed grids can be 55x2 and 82x3. In Table 1.1 the quality indices are presented for different grids, arrangements of neurons, weights' initialization. The best result for the index *Is* is given for the square grid 19x19. It is noted that the initialization of the neurons can be a linear combination of the respective eigenvectors of the two major eigenvalues (scenario of initialization 1) or can be equal to 0.5 (scenario of initialization 2).

396

Load Time-Series Classification Based on Pattern Recognition Methods

2D SOM- neurons population	Arrangement – weights initialization	Total epochs - t	Is(t)	ADM(t)	TE(t)	J(t)	Calibration of $T_{\eta} - \xi - \phi - \eta_r - \eta_f - \sigma_0$
9x9=81	Rect(2)	3600	0.2832	0.08853	0.10616	0.08853	1500-0.4-2-0.15-0.001-9
9x9=81	Hex(2)	2400	0.3245	0.08372	0.15753	0.08372	1000-1.0-2-0.20-0.001-9
10x10=100	Rect(2)	3600	0.2208	0.06757	0.08562	0.06757	1500-0.6-2-0.40-0.001-10
10x10=100	Hex(2)	2400	0.2354	0.07147	0.09247	0.07147	1000-1.0-2-0.25-0.001-10
12x12=144	Rect(2)	1100	0.2371	0.05176	0.13356	0.05176	500-1.0-2-0.15-0.001-12
12x12=144	Hex(2)	4400	0.2679	0.05003	0.16781	0.05003	2000-1.0-2-0.05-0.001-12
14x14=196	Rect(2)	3600	0.2306	0.04167	0.14726	0.04167	1500-1.0-2-0.05-0.001-14
14x14=196	Hex(2)	2200	0.2245	0.03521	0.15410	0.03521	500-1.0-2-0.05-0.001-16
16x16=256	Rect(2)	2200	0.1356	0.02840	0.07877	0.02840	1000-1.0-2-0.10-0.001-16
16x16=256	Hex(2)	3300	0.1678	0.02909	0.10959	0.02909	1500-1.0-2-0.10-0.001-16
19x19=361	Rect(2)	4400	0.0970	0.01254	0.07192	0.01254	2000-1.0-2-0.15-0.001-19
19x19=361	Hex(2)	4400	0.1267	0.01538	0.09589	0.01538	2000-1.0-2-0.05-0.001-19
55x2=110	Rect(2)	1200	0.3504	0.05532	0.23973	0.05532	500-1.0-2-0.15-0.001-55
55x2=110	Rect(1)	1200	0.3503	0.05532	0.23973	0.05532	500-1.0-2-0.15-0.001-55
82x3=246	Rect(2)	1100	0.2040	0.01982	0.16438	0.01982	500-1.0-2-0.30-0.001-82

Table 1.1. Quality Indices for Different Cases of Bi-Dimensional SOM

The type of the arrangement and the weights initialization do not affect the respective results significantly. Practically, the clusters of the bi-dimensional map cannot be directly exploited because of the size and the location of the neurons into the grid, as shown in Fig. 1.19. This problem is solved through the application of a basic classification method (e.g. the



a. Bi-dimensional SOM 10x10 with rectangular arrangement



a. Bi-dimensional SOM 10x10 with hexagonal arrangement



Fig. 1.19. Different cases of bi-dimensional SOM after the application of the proposed kmeans method at the neurons of SOM

proposed k-means) for the neurons of the bi-dimensional SOM (Chicco et al., 2004). The adequacy measures are calculated using the load daily chronological curves of the neurons which form the respective clusters of the basic classification method. In Table 1.2, the adequacy measures of the aforementioned maps are presented, using the proposed k-means method for 10 final clusters.

For the industrial customer, the best results of the application of the k-means method to the neurons of the SOM are given for the maps with the ratio between the two major eigenvalues of the covariance matrix of the input vectors set (see Table 1.2). The respective clusters are also more compact than the ones of the square maps, as it can be seen in Fig. 1.19.

Load Time-Series Classification Based on Pattern Recognition Methods

2D COM	A man a am	A deguary Measure								
2D SOM-	Arrangement			Аиедиис	y ivieusure					
neurons	-weights	T	λΛΙΔ	CDI	SMI	DRI	WCBCR			
population	initialization	J	111171	CDI	5111	DDI	(*10-3)			
9x9=81	Rect (2)	0.309430	0.064420	0.365331	0.668309	2.01767	8.4570			
9x9=81	Hex. – (2)	0.284911	0.064730	0.358870	0.666259	1.91128	9.1605			
10x10=100	Rect (2)	0.269351	0.066746	0.358512	0.661059	1.76758	8.7586			
10x10=100	Hex. – (2)	0.285281	0.062488	0.369192	0.682404	1.81237	8.4091			
12x12=144	Rect (2)	0.267056	0.065584	0.353097	0.671649	1.74857	9.2029			
12x12=144	Hex. – (2)	0.268810	0.070409	0.351914	0.661194	1.85451	9.4168			
14x14=196	Rect (2)	0.272213	0.066418	0.360227	0.662015	1.68854	8.8566			
14x14=196	Hex. – (2)	0.273781	0.069001	0.361199	0.666046	1.75621	10.0756			
16x16=256	Rect (2)	0.267521	0.066970	0.349423	0.669469	1.88224	9.2304			
16x16=256	Hex (2)	0.268528	0.068710	0.364127	0.660511	1.69430	9.4152			
19x19=361	Rect (2)	0.266128	0.065389	0.351903	0.682931	1.85560	8.5386			
19x19=361	Hex (2)	0.267087	0.067618	0.343808	0.660950	1.70171	8.8486			
55x2=110	Rect (2)	0.262634	0.060581	0.345677	0.654891	1.68728	7.7872			
82x3=246	Rect (2)	0.258002	0.063284	0.334516	0.681566	1.75790	8.1426			

Table 1.2. Adequacy Indices for 10 Clusters – Typical Load Chronological Curves of the Industrial Customer Using Proposed K-Means Method at the Second Classification Level for Different Cases of Bi-Dimensional SOM

1.4.2.8 Comparison of clustering models & adequacy indicators

In Fig. 1.20, the best results for each clustering method (proposed k-means, fuzzy k-means, adaptive vector quantization, self-organized maps and hierarchical algorithms) are depicted. The proposed k-means model has the smallest values for the *MIA*, *CDI*, *DBI* and *WCBCR* indicators. The *WARD* algorithm presents the best behaviour for the mean square error *J*, the unweighted pair group method average algorithm (*UPGMA*) and the bi-dimensional SOM (with the application of the proposed k-means at the second level) for the *SMI* indicator. The proposed k-means model has similar behaviour to the *WARD* algorithm for the *J* indicator and to the *UPGMA* algorithm for the *WCBCR*. All indicators -except DBI-exhibit improved performance, as the number of clusters is increased.

In Table 1.3 the results of the best clustering methods are presented for 10 clusters with the respective parameters, which is the finally proposed size of the typical days for that customer. The optimized parameters for the mono- dimensional and bi- dimensional self-organized maps have been analyzed in §1.4.2.5 and §1.4.2.6 respectively. The proposed k-means method gives the best results for the *MIA*, *CDI*, *DBI* and *WCBCR* indicators (for different pairs of (a,b)), while the adaptive vector quantization should be used for *J* indicator and the bi-dimensional self organized map using proposed k-means for classification in a second level for *SMI* indicator.

Observing the number of dead clusters for the under study models (Fig. 1.20.g) the behaviour of *DBI* and *SMI* indicators for bi-dimensional SOM and k-means emerges a significant variability. For the above reasons the proposed indicators are *MIA* and *WCBCR*. Studying the number of dead clusters for the proposed k-means model (Fig. 1.7), it is obvious that the use of *WCBCR* indicator is slightly superior to the use of *MIA* and *J* indicators. It is also noted, that the basic theoretical advantage of the *WCBCR* indicator is the fact 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.





Fig. 1.20. The best results of each clustering method for the set of 292 training patterns of a medium voltage industrial customer for 5 to 25 clusters

The improvement of the adequacy indicators is significant for the first 10 clusters. After this point, the behaviour of the most indicators is gradually stabilized. It can also be estimated graphically by using the rule of the "knee" (Gerbec et al., 2004; -, 2005), as shown in Fig. 1.21. If this knee is not clearly shown, the tangents are drawn estimating the knee for 10 clusters for the current case study.

After having taken into consideration that the ratio of the computational training time for the under study methods is 0.05:1:22:24:36:50 (hierarchical: proposed k-means: adaptive vector quantization: mono-dimensional SOM: fuzzy k-means for q=6: bi-dimensional SOM), the use of the hierarchical and k-means models is proposed. It is mentioned that the necessary computational training time for the proposed k-means method is approximately one hour for Pentium 4, 1.7 GHz, 768 MB.



Fig. 1.21. Indicative estimation of the necessary clusters for the typical load daily chronological curves of a medium voltage industrial customer, using the proposed k-means model with the *WCBCR* adequacy measure

Consequently, the proposed k-means model with *WCBCR* adequacy measure is suggested for the description of the load behaviour of the analyzed paper-mill medium voltage customer. More generally, the *WCBCR* indicator should be used because of its aforementioned basic theoretical advantage. But for completeness reasons we will examine all adequacy measures for all models for other 93 customers in §1.4.2.10. Before this step the results for the analyzed paper-mill medium voltage customer will be presented.

1.4.2.9 Representative daily load chronological diagrams of a paper-mill medium voltage customer

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 Tables 1.4 and in Fig. 1.22. This number of clusters is qualitatively satisfied.

The retailer and the head engineer of the under study industry can observe the customer's daily demand behaviour during the year based on the respective load curves. Specifically, cluster 1 represents holidays, clusters 2 and 4 the days of the re-operation of the industry, cluster 3 the days of stopping the operation of the industry, cluster 5 a day with partial internal power fault, cluster 6 the workdays with one of the two lines for production in operation, clusters 7 and 8 the workdays for which one of the two lines for production is out of operation for few hours, clusters 9 and 10 the usual workdays, where every 8 hours there is a small variance because of the workers' change.

Mothods Paramotors			Adequae	cy Measure		
Methous -1 arameters	J	MIA	CDI	SMI	DBI	WCBCR
Proposed k-means (scenario 1)	0.2527	0.05828	0.3239	0.6711	1.6515	0.006679
<i>a</i> parameter	0.10	0.19	0.35	0.17	0.18	0.11
<i>b</i> parameter	0.77	0.35	0.55	0.48	0.37	0.60
K-means (scenario 2)	0.2537	0.06782	0.3601	0.7311	2.5603	0.007760
Classic k-means (scenario 3)	0.2538	0.06435	0.3419	0.7306	2.4173	0.006716
AVQ	0.2496	0.06472	0.3537	0.7160	2.1884	0.006886
$\eta_0$ parameter	0.80	0.85	0.85	0.75	0.60	0.70
$T_{\eta 0}$ parameter	500	1500	500	2000	500	500
Fuzzy k-means (q=6)	0.3575	0.07144	0.3697	0.7635	3.2559	0.007153
<i>a</i> parameter	0.31	0.27	0.18	0.13	0.10	0.10
<i>b</i> parameter	0.49	0.31	0.36	0.54	0.59	0.46
CL	0.2973	0.07271	0.3977	0.7427	2.9928	0.010052
SL	0.7027	0.09644	0.5049	0.6798	2.4855	0.015696
UPGMA	0.3127	0.06297	0.4008	0.6494	2.1714	0.006684
UPGMC	0.4147	0.07656	0.4346	0.6494	2.1198	0.009908
WARD	0.2538	0.06804	0.3723	0.7296	2.7334	0.008399
WPGMA	0.3296	0.06807	0.4112	0.6781	2.4764	0.007991
WPGMC	0.5747	0.07386	0.4665	0.6900	2.7903	0.008884
Mono-dimensional SOM	0.2607	0.07189	0.3903	0.7588	3.8325	0.009631
Bi-dimensional SOM 55x2						
using proposed k-means for	0.2623	0.06059	0.3456	0.6549	1.6873	0.007787
classification in a second level						
<i>a</i> parameter of k-means	0.15	0.24	0.36	0.44	0.22	0.22
<i>b</i> parameter of k-means	0.69	0.28	0.51	0.11	0.30	0.30

Table 1.3. Comparison of the Best Clustering Models for 10 Clusters for the Medium Voltage Industrial Customer

Load ductor	Da	ıy (1 fo	r Mon	day, 2 f	for Tue	esday e	tc.)	Days
Load Cluster	1	2	3	4	5	6	7	per cluster
	4	4	4	4	3	4	4	27
2	1	_0	07	0	2	0	0	3
3	1	0	-0	0	0	_1	0	
4	0	1	1	0	0	0	0	2
5	0	1	0	0	0	0	0	1
6	4	2	4	4	3	2	2	21
7	6	4	3	4	4	2	0	23
8	2	1	1	1	2	2	1	10
9	9	11	16	11	11	11	13	82
10	14	17	12	18	16	21	22	120
	F	Total n	umber	of day	rs unde	er stud	V	292

Table 1.4. Results of the Proposed k-means Model with optimization to *WCBCR* Adequacy Measure for 10 clusters for the Medium Voltage Industrial Customer



Fig. 1.22. Typical daily chronological load curves for the medium voltage industrial customer using proposed k-means model with optimization to *WCBCR* adequacy measure

The daily load diagrams are well identified using the k-means clustering method and the *WCBCR* indicator, as it is indicatively presented in Fig. 1.23, where the typical load curve of cluster 10 (which represents the most populated day and the day with peak load simultaneously) along with the 120 measured clustered load curves are shown. It is obvious from table 1.4 that the number of the days for each representative cluster of this customer is not influenced by the day of the week.



Fig. 1.23. Daily chronological load curve of cluster 10 for the MV industrial customer (bold line) along with its 120 clustered measured curves (thin lines) using the proposed k-means clustering method and the *WCBCR* indicator

#### 1.4.2.10 Application of the Proposed Methodology to a Set of Medium Voltage Customers

The same process was repeated for 93 more medium voltage customers of the Greek power distribution system, with load curves qualitatively described by using 8-12 clusters for each customer. The scope of this application is the representation of the comparison of the clustering algorithms and the adequacy measures for more than one customer. The performance of these methods is presented in Table 1.5 and in Fig. 1.24, through the indication of the number of customers which achieves the best value of adequacy measure.

It is evident, by observing Fig. 1.24, in which a comparison of the algorithms is depicted, 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* measure and the adaptive vector quantization for *SMI*, *DBI* indicators. It can be noticed that the other two k-means models show the worst performance in adequacy measures.

Mathada			Adequac	y Measure		
Methods	J	MIA	CDI	SMI	DBI	WCBCR
Proposed k-means (scenario 1)	6	28	85	16	13	37
K-means (scenario 2)	0	0	0	0	0	0
Classic k-means (scenario 3)	0	0	0	0	0	0
AVQ	2	2	0	32	47	1
Fuzzy k-means (q=6)	0	0	0	0	1	6
	0		0	0	0	0
SL	0	5	0	2	4	4
UPGMA	0	15	0	3	7	12
UPGMC	0	20	0	22	6	18
WARD	16	0	0	0	0	1
WPGMA	0	13	0	0	3	8
WPGMC	0	6	0	2	1	7
Mono-dimensional SOM	5	1	0	0	0	0
Bi-dimensional SOM using proposed k-means for classification in a second level	65	3	9	17	12	0

Table 1.5. Comparison of the Clustering Models for the Set of 94 MV Customers for 10 clusters



Fig. 1.24. Population of customers with the best value of adequacy measure, with respect to different clustering models for the set of the 94 medium voltage customers of the Greek Power Distribution System

In practice, the proposed k-means model and hierarchical ones should be used, as they lead to the best results compared to the other models, especially for the *WCBCR* indicator.

#### 1.4.3 Usefulness of the application of the proposed methodology

The results of the developed methodology can be used for:

- the proper selection of an adequate tariff by the customer or the recommendation of a tariff from the supplier,
- the settlement of the customer's bills in the case of energy and power bought from more than one suppliers,
- the feasibility studies of the energy efficiency and demand side management measures, which are proper for the customer,
- the customer's short-term and mid-term load forecasting, load estimation after the application of demand side management programs, in which the customer as well as the suppliers are interested,
- the selection of the representative chronological load diagram of the customer by choosing the type of typical day (such as the most populated day, the day with the peak demand load or with the maximum demand energy, etc), which will be used for the customers' classification by the suppliers.

# 1.5 A two-stage pattern recognition of load curves for classification of electricity customers

#### **1.5.1 General description of the proposed two-stage methodology**

Based on the pattern recognition methodology for the classification of the daily load curves of a customer a two-stage methodology, which has been developed for the classification of electricity customers, is presented by Tsekouras et al., 2007. In the first stage, typical chronological load curves of various customers are estimated using pattern recognition methods and their results are compared using the six adequacy measures, as in the case of the first paradigm. In the second stage, classification of customers is performed by the same methods and measures, along with the representative load patterns of customers being obtained from the first stage. The flow chart of the proposed methodology is shown in Fig. 1.25, while its basic steps are the following:

- a. *Data and features selection* (same to (*a*) step of the methodology of §1.4).
- b. *Customers' clustering using a priori indices*: Customers can be characterized by their geographical region, voltage level (high, medium, low), economic activity, installed power, contracted energy, power factor, etc. These indices are not necessarily related to the load curves according to the experience of the power distribution company. They can be used however for the pre-classification of customers. It is mentioned that the load curves of each customer are normalized using the respective minimum and maximum loads of the period under study.
- c. *Data preprocessing* (same to (*b*) step of the methodology of §1.4).
- d. *Typical load curves clustering for each customer –First stage application of pattern recognition methods*: For each customer, a number of clustering algorithms (k-means, adaptive vector quantization, fuzzy k-means, self-organized maps and hierarchical clustering) is applied. Each algorithm is trained for the set of load curves 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. This process is repeated for the total set of customers under study. Special customers, such as seasonal ones (e.g. oil-press industry, small seaside hotels) are identified. Practically, it is the (c) step of the main application of the pattern recognition methods of the methodology of §1.4.



Fig. 1.25. Flow chart of two stage pattern recognition methodology for the classification of customers

- e. *Selection of typical chronological load curves for customers*: The typical load curves of customers that will be used for the final clustering are selected by choosing the type of typical day (such as the most populated day, the day with the peak demand load or with the maximum demand energy, etc). It is possible to omit the customer's typical load curves clustering, if the user wishes to compare the customer's behaviour in specific days, such as the day of system peak load, the mean July workday, etc. However, the customers' behaviour is not entirely representative for the period under study. It is noticed that special customers can be handled separately.
- f. *Clustering of customers Second stage application of pattern recognition methods*: The clustering methods are applied for the set of the customer's representative load curves. After algorithms' calibration, the clusters of customers and the respective classes representative load curves are formed.

### 1.5.2 Application of the two-stage methodology to a set of medium voltage customers

#### 1.5.2.1 General

For the application of the proposed methodology a set of 94 medium voltage customers of the Greek power distribution system is used. It should be noticed that larger customer sets coming from different power distribution systems can be handled applying the same procedure and the expected results might be better. However, only the set of 94 customers is available.

Firstly, the first stage of the proposed methodology is realized, which has been already presented in § 1.4.2.1 -§ 1.4.2.9 (analytically for one customer) and § 1.4.2.10 (synoptically for the set of the 94 medium voltage customers). Next, the second stage is implemented. The characteristic customer's typical day can be either the most populated day of the customer or the day with the peak load demand (independently of the best number of clusters for each individual customer). Here, two case studies are presented: the first with the most populated day of each customer and the second with the peak load demand. In both cases the representative load curve for each customer is obtained by the clustering method that shows the best results for the adequacy measure being used (here is *WCBCR*). The clustering methods are applied for the set of the representative load curve for each customer using *WCBCR* as adequacy measure because of its theoretical advantage (see §1.4.2.8).

### 1.5.2.2 Case study I: the most populated day of each customer

For this case study the most populated day of each customer is used. For example the respective cluster is the 10<sup>th</sup> one for the industrial customer in Table 1.2. Fig. 1.26 shows the best results of each clustering method by using the *WCBCR* measure. The developed k-means and UPGMC models are proved to be the best ones, as it is also registered in Table 1.6. The respective number of clusters is determined by using the rule of the "knee" (see §1.4.2.8) finding that the necessary number of clusters is 12.

The results of clustering for 12 clusters using the UPGMC model with the optimization of the *WCBCR* measure are presented in Table 1.7 and in Fig. 1.27. Practically eighty-nine customers form seven main clusters (it is proposed empirically the number of the clusters to be between 2 and  $\sqrt{89} \approx 9$  (Figueiredo et al., 2005)), while the remaining five customers

show specific unique characteristics among the members of the set of the 94 customers (respective individual clusters 2, 4, 5, 9, 10).

Each customer class presents its separate behaviour. Specifically, customers of cluster 1 have stable load demand equal to approximately 10% of the respective normalized peak load

(=1.0). Similarly customers of cluster 7 and cluster 12 have stable load demand equal to approximately 45% and 80% of the respective normalized peak load respectively. Cluster 3 has the most customers (40 from 94), whose load behaviour is characteristic: gradual load increment from 18% to 45% of normalized peak load from 6:00 to 10:00, a small variation at 12:00, afterwards a slow load reduction from 14:00 to 24:00.

Load demand of customers of cluster 8 has a rapid increment at 8:00 (from 40% to 70% of normalized peak load), it remains stable until 20:00, then it has a slow reduction until 23:00 (receiving 40% of normalized peak load). This cluster has mainly industrial and commercial customers. On the contrary, cluster 11 has only industrial customers with similar load behaviour (load demand has a rapid increment at 6:00 from 50% to 80% of maximum peak load, it remains stable until 22:00, then it has a rapid reduction to 50% of maximum peak load, while it is obvious that there is a small variation approximately at 14:00). The separate customers of clusters 4, 5 and 9 have similar load behaviour with the customers of clusters 8 and 11, but they have some special characteristics, such as different hours of load increment etc.

The two commercial customers of cluster 6 and the separate customer of cluster 2 have maximum load demand during early night hours (completely opposite behaviour from the other customers).

The obtained representative curves provide useful information about the load demand of the customers' clusters throughout the year. It is obvious that the a priori index of customer's activity is not representative for load curves, which is also confirmed by (Chicco et al., 2003a; -, 2003b; -, 2004; -, 2006; Figueiredo et al., 2003). This can not be generalized since it may vary within countries and distribution companies depending on the respective data of customers (Gerbec et al., 2003; -, 2004; -, 2005). But the proposed methodology can be applied directly to the respective set of customers, in order to study their respective load behaviour. The same process can be repeated for all other adequacy measures. The number of the clusters being used can also be selected according to the desirable precision and the relative improvement of the respective measure.



Fig. 1.26. *WCBCR* measure of the best fitting clustering methods for 5 to 25 neurons for the training patterns set of 94 medium voltage customers for the most populated day

Methods -Parameters			Adequac	cy Measure		
Withous 4 drameters	J	MIA	CDI	SMI	DBI	WCBCR
Proposed k-means	0.3840	0.04950	0.2064	0.6732	1.6694	0.004100
AVQ	0.3601	0.05453	0.2215	0.6736	1.2345	0.004877
Fuzzy k-means (q=6)	0.5751	0.06559	0.2656	0.7411	2.1210	0.007577
CL	0.4058	0.05484	0.2291	0.7010	1.7056	0.004926
SL	1.2718	0.08737	0.3421	0.7050	2.7942	0.011694
UPGMA	0.4956	0.05070	0.2442	0.6664	1.6341	0.004008
UPGMC	0.5462	0.04696	0.2528	0.6593	1.8610	0.003315
WARD	0.3728	0.05369	0.2349	0.6984	1.7817	0.005258
WPGMA	0.4573	0.05367	0.2288	0.6768	1.6965	0.004452
WPGMC	0.4617	0.05579	0.2301	0.6752	1.6779	0.004712
Mono-dimensional SOM	0.4163	0.06330	0.2694	0.7184	2.0013	0.008752
Bi-dimensional SOM 14x3 using proposed k-means for classification in a second level	0.3265	0.05920	0.2157	0.6676	1.7604	0.006599

Table 1.6. Comparison of the Best Clustering Models for 12 Clusters for the set of 94 medium voltage customers using the most populated day based on WCBCR adequacy measure of the 1<sup>st</sup> stage

Load	Activit industi	ty of customer rial, 3: public s	r (1: commer services, 4:tr	rcial, 2: raction)	Customers
cluster -	1	2	3	4	– per cluster
1	12	6	3	1	22
2	1	0	0	0	1
3	26	9	3	2	40
4	0	1	0	0	
5		0-	0	0	
6	2 7	07	0	0	2
7	5	3	1	0	9
8	6	2	0	0	8
9	0	1	0	0	1
10	0	1	0	0	1
11	0	3	0	0	3
12	0	5	0	0	5
Total	53	31	7	3	94

Table 1.7. Results of the UPGMC Model with optimization to WCBCR measure for 12 clusters for a set of 94 Customers using the most populated day





Fig. 1.27. Normalized representative chronological load curves of typical classes from the classification of a set of 94 medium voltage customers (as derived from the load curves of the most populated typical day of each customer) for the Greek power distribution system to 12 clusters using UPGMC model with the optimization of the *WCBCR* measure

#### 1.5.2.3 Case study II: the day with peak load demand

For this case study the day with the respective peak load demand of each customer is used. For example the respective cluster is the 10<sup>th</sup> one for the industrial customer (see Fig. 1.22), which is also characterised by the most populated day (this congruency rarely happens). Fig. 1.28 shows the best results of each clustering method by using the *WCBCR* measure. The WPGMA and UPGMC models are proved to be the best ones, as it is also registered in Table 1.8. Using the rule of the "knee" the necessary number of clusters is between 9 and 15 choosing finally 12 (the position of the knee is not clear).

The results of clustering for 12 clusters using the WPGMA model with the optimization of the *WCBCR* measure are presented in Table 1.9 and in Fig. 1.29. Practically eighty-nine customers form seven main clusters, as it has been already happened for the most populated day, but it is an accidental occasion. The respective representative load curves are more abrupt and sharp than the ones of the most populated day.

Specifically, customers of cluster 11 (with 20 customers from 94) have stable load demand equal to approximately 80% of the respective normalized peak load. The load behaviour of cluster 10, which has the most customers (36 from 94), presents a gradual load increment

from 30% to 85% of normalized peak load from 6:00 to 11:00, afterwards a slow load reduction from 13:00 to 24:00. Both of these clusters present small variations.

The load behaviour of clusters 3, 4, 8 and 9 presents larger variations than the respective one of clusters 10 and 11. Customers of cluster 3 have a gradual load increment from 10% to 70% of normalized peak load from 6:00 to 15:00, afterwards a slow load reduction from 15:00 to 24:00. Customers of cluster 4 have a rapid load increment from 10% to 90% of normalized peak load at 8:00, then their load remains stable practically until 16:00, afterwards their load reduces sharply at 16:00. Load demand of customers' cluster 9 has a gradual increment from 10% to 90% of normalized peak load from 8:00 to 13:00, it remains stable until 20:00, then it has a rapid reduction until 24:00.



Fig. 1.28. *WCBCR* measure of the best fitting clustering methods for 5 to 25 neurons for the training patterns set of 94 medium voltage customers for the day with the peak load demand

Mathada Paramatara			Adequae	cy Measure		
Methods -1 arameters	J	MIA	CDI	SMI	DBI	WCBCR
Proposed k-means	1.1754	0.1043	0.3434	0.6591	2.1421	0.02033
AVQ	1.1304	0.1043	0.4389	0.6557	2.0546	0.02222
Fuzzy k-means (q=6)	1.6999	0.1282	0.4963	0.7073	3.3089	0.03387
	1.2138	0.1105	0.4663	0.6752	2.7041	0.02702
SL	2.5647	0.1256	0.5420	0.6373	3.3553	0.02870
UPGMA	1.3303	0.0989	0.4219	0.6584	2.5309	0.01657
UPGMC	1.6186	0.0951	0.4435	0.6223	2.7116	0.01430
WARD	1.1696	0.1145	0.4739	0.6828	2.7926	0.03122
WPGMA	1.4141	0.0953	0.4031	0.6377	2.5680	0.01327
WPGMC	1.7250	0.1040	0.4274	0.6303	2.9650	0.01536
Mono-dimensional SOM	1.1682	0.1139	0.4981	0.6788	2.7897	0.03480
Bi-dimensional SOM 14x3 using proposed k-means for classification in a second level	1.0373	0.1023	0.3475	0.6500	2.1731	0.02067

Table 1.8. Comparison of the Best Clustering Models for 12 Clusters for the set of 94 medium voltage customers using the day with the peak load demand based on WCBCR adequacy measure of the 1<sup>st</sup> stage

Load	Activity industri	of custome al, 3: public	er (1: commer services, 4:tr	cial, 2: action)	Customers
cluster	1	2	3	4	— per cluster
1	0	1	0	0	1
2	1	0	1	0	2
3	8	1	1	3	13
4	3	4	0	0	7
5			0	0	
6	0	17	0	0	
7	0	$\neg$	0	0	
8	6	0	0	0	6
9	2	2	1	0	5
10	25	8	3	0	36
11	8	11	1	0	20
12	0	1	0	0	1
Total	53	31	7	3	94

Table 1.9. Results of the UPGMC Model with optimization to WCBCR measure for 12 clusters for a set of 94 Customers using the day with the peak load demand







Fig. 1.29. Normalized representative chronological load curves of typical classes from the classification of a set of 94 medium voltage customers (as derived from the load curves of the days with the peak load of each customer) for the Greek power distribution system to 12 clusters using WPGMA model with the optimization of the *WCBCR* measure

Cluster 8 presents an opposite load behaviour against clusters 3, 4 and 9, because the maximum load demand is achieved during early night hours (70% of peak load), while during the rest day load varies from 30% to 50% of peak load.

The rest seven customers are represented by six clusters (the clusters 1, 5, 6, 7, 12 contain one customer each and the cluster 2 contains two ones), where its customer presents unique characteristics for its chronological typical load curve.

It is mentioned that the type of typical day (such as the most populated day etc) is defined by the user according to his needs.

#### 1.5.3 Usefulness of the application of the proposed two-stage methodology

The results of the developed methodology can be used either for each customer separately or for a set of customers. The results of the first stage are the respective ones of the typical chronological load curves of each customer of §1.4.3.

- The results of the second stage can be used as important input information for:
- the adaptation of tariffs for each customer class from the suppliers,
- the adaptation of tariffs for ancillary services of the reactive demand on behalf of the distribution or transmission operator, if the respective representative curves of reactive load are calculated,
- the feasibility studies of the energy efficiency and demand side management measures, which are proper for each customer class (extraordinary useful for the suppliers, in order to smooth their respective daily load demand curve),
- the short-term and mid-term load forecasting for the customer classes, for which the suppliers, the system operator and the regulatory energy authority are interested.

# 1.6 A pattern recognition methodology for power system load profiles for applications of demand side management programs

#### 1.6.1 General description of the proposed methodology

Based on the pattern recognition methodology for the classification of the daily load curves of a customer a similar pattern recognition methodology can be used for the classification of

daily chronological load curves of power system, as shown in Fig. 1.30. The main steps are the following:

- a. *Data and features selection* (same to (*a*) step of the methodology of §1.4). The active and reactive energy values are registered (in MWh and Mvarh) for each time period in steps of 1 hour.
- b. *Data pre-processing* (same to (*b*) step of the methodology of §1.4).
- c. *Main application of pattern recognition methods* (same to (c) step of the methodology of §1.4).

As we can see this methodology is quite similar to one of §1.4. This can lead us to propose the extension of the application of this methodology for the classification of similar timeseries curves, such as daily chronological temperatures curves, etc (Tsekouras, 2006).



Fig. 1.30. Flow diagram of pattern recognition methodology for the classification of daily chronological load curves of power system

#### 1.6.2 Application of the Greek power system

## 1.6.2.1 General

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 following next paragraphs the application of each clustering method is analyzed.

#### 1.6.2.2 Application of the k-means

The proposed model of the k-means method is executed for different pairs (a,b) from 2 to 25 clusters, where  $a=\{0.1,0.11,...,0.45\}$  and  $a+b=\{0.54,0.55,...,0.9\}$ , as in the case of § 1.4.2.2. The best results for the six adequacy measures do not refer to the same pair (a,b) –as it is presented in Table 1.10 for 10 clusters. The alternative model is the classic one with the random choice of the input vectors during the centres' initialization. For the classic k-means model 100 executions are carried out and the best results for each index are registered. The superiority of the proposed model is the fact that applies in all above cases of neurons and that it converges to the same results for the respective pairs (a,b), which can not be achieved using the classic model.

## 1.6.2.3 Application of the adaptive vector quantization

The initial value  $\eta_0$ , the minimum value  $\eta_{\min}$  and the time parameter  $T_{\eta 0}$  of learning rate are properly calibrated. The best results of the adequacy measures are given for different pairs of ( $\eta_0$ ,  $T_{\eta 0}$ ), according to the results of Table 1.10 for 10 clusters. The  $\eta_{\min}$  value does not practically improve the neural network's behaviour assuming that it ranges between 10<sup>-5</sup> and 10<sup>-6</sup>.



a. SMI indicator (similar to DBI)

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

Fig. 1.31. *SMI* and *WCBCR* for the fuzzy k-means method for the set of 183 load curves of the summer of the year 2000 for the Greek power system with q=2, 4, 6 for 5 to 25 clusters

1.6.2.4 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. 1.31 *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.



Fig. 1.32. Adequacy measures for the 7 hierarchical clustering algorithms for the set of 183 load curves of the summer of the year 2000 for the Greek power system for 5 to 25 clusters

### 1.6.2.5 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. 1.32.

1.6.2.6 Application of mono-dimensional self-organizing maps

The main problems during the training of the mono-dimensional SOM are:

- the proper termination of the SOM's training process, which is solved by minimizing the index  $I_s$  (eq.(1.32)),
- the proper calibration of (a) the initial value of the neighbourhood radius σ<sub>0</sub>, (b) the multiplicative factor φ between T<sub>s0</sub> (epochs of the rough ordering phase) and T<sub>η0</sub> (time parameter of learning rate), (c) the multiplicative factor ξ between T<sub>σ0</sub> (time parameter of neighbourhood radius) and T<sub>η0</sub>, (d) the proper initial values of the learning rate η<sub>r</sub> and η<sub>f</sub> during the rough ordering phase and the fine tuning phase respectively.

the proper initialization of the weights of the neurons.

The optimization process for the mono-dimensional SOM parameters is similar to that one of §1.4.2.6 and it is repeated for any population of clusters

#### 1.6.2.7 Application of bi-dimensional self-organizing maps

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. 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 (see Fig. 1.33) and 68x3.



Fig. 1.33. 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 the Greek power system for 10 neurons

Because of the size and the location of the neurons in the grid, the clusters of the bidimensional map cannot be directly exploited and the proposed k-means is applied for the neurons of the bi-dimensional SOM, as it has already happened in §1.4.2.7. The adequacy measures are calculated using the load curves of the neurons which form the respective clusters of the proposed k-means 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.

#### 1.6.2.8 Comparison of clustering models & adequacy indicators

In Fig. 1.34 the best results achieved by each clustering method are depicted. 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.

By observing the number of dead clusters for the proposed k-means model (Fig. 1.34.h) it is obvious that the use of *WCBCR* indicator is slightly superior to *MIA* and *J* indicators. Taking into consideration the basic theoretical advantage of the *WCBCR* indicator and the significant variability of the behaviour of *DBI* and *SMI* indicators for different clustering techniques the *WCBCR* indicator is proposed to be used.





Fig. 1.34. The best results of each clustering method for the set of 183 load curves of the summer of the year 2000 for the Greek power system for 5 to 25 clusters

The improvement of the adequacy indicators is significant until 10 clusters. After this value the behaviour 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. 1.35). In Table 1.10 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.

Taking into consideration that the ratio 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.

For this case study (load daily chronological curves of the summer of the year 2000 for the Greek power system) the proposed k-means model with the *WCBCR* adequacy indicator is going to be used.

## 1.6.2.9 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 1.11 and in Fig. 1.36 respectively.



Fig. 1.35. Indicative estimation of the necessary clusters for the typical load daily chronological curves of the summer of the year 2000 for the Greek power system for the *WCBCR* adequacy indicator

Methode -Parameters			Adequa	cy Measur	е	
	J	MIA	CDI	SMI	DBI	WCBCR
Proposed k-means	0.01729	0.02262	0.1778	0.7331	2.0606	0.002142
<i>a-b</i> 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
Classic 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
n - n - T parameters	0.5-5x10-	0.4-5x10-	0.5-5x10-	0.4-5x10-	0.8-5x10-	0.4-5x10-7-
$\eta_0 \eta_{\min} \eta_0$ parameters	7-1000	7-4000	7-5000	7-2000	7-1000	4000
Fuzzy k-means	0.02208	0.03036	0.25328	0.7482	2.1936	0.003894
a-a-b parameters	4-0.22-	4-0.18-	4-0.18-	6-0.12-	6-0.14-	4-0 18-0 62
<i>q u o parameters</i>	0.46	0.62	0.70	0.62	0.74	4 0.10 0.02
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
$\sigma - \phi - \xi - n - n - T$	10-1.0-	10-2.0-	10-1.0-	10-1.0-	10-1.0-	10-2.0-0.2-
$\sigma_0 \varphi \varsigma \eta_f \eta_r \eta_0$	0.6-0.15-	0.2-0.10-	0.6-0.15-	0.6-0.15-	0.6-0.10-	0.10-10-3-
parameters	10-3-1500	10-3-1750	10-3-1500	10-3-1500	10-3-1500	1750
2D SOM 46x2 using proposed k-						
means for classification in a 2 <sup>nd</sup>	0.01685	0.02697	0.1785	0.7271	2.2572	0.002459
level						
	46-1.0-	46-1.0-	46-1.0-	46-1.0-	46-1.0-	46-1.0-1.0-
$\sigma_{_0}$ – $\phi$ – $\xi$ – $\eta_{_f}$ – $\eta_{_r}$ – $T_{\eta_{_0}}$ -a- $b$	1.0-0.30-	1.0-0.30-	1.0-0.30-	0.2-0.20-	0.2-0.20-	0.30-10-3-
parameters	10-3-500-	10-3-500-	10-3-500-	10-3-500-	10-3-500-	500-0.15-
Pursuitetere	0.28-0.36	0.15-0.58	0.44-0.46	0.10-0.77	0.44-0.25	0.58

Table 1.10. 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

Load cluster	Da	y (1 fo	r Mono	day, 2 f	for Tue	esday e	tc.)	Days
Load Cluster	1	2	3	4	5	6	7	per cluster
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

Table 1.11. 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





Fig. 1.36. Typical daily chronological load curves for the set of 183 curves of the summer of the year 2000 for the Greek power system using proposed k-means model with optimization to *WCBCR* 

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

As we can notice the separation between work days and non-work days for each season is not so much descriptive for the load behaviour of a power system, as we have proved that 8 to 10 clusters are needed.

# 1.6.2.10 Application of the Proposed Methodology for the Greek Power System Per Seasons and Per Years for the time period 1985-2002

The same process is repeated for the summer (April–September) and the winter (October-March) periods for the years between 1985 and 2002. The load curves of each season are qualitatively described by using 8-10 clusters. The performance of these methods is presented in Table 1.12 by indicating the number of seasons that achieves the best value of adequacy measure respectively.

Methods			Adequac	y Measure		
Wiethous	J	MIA	CDI	SMI	DBI	WCBCR
Proposed k-means	1	24	31	7	12	29
Classic k-means	0	0	0	0	0	0
AVQ	2	0	0	7	16	0
Fuzzy k-means	0	-0	0	0	0	_1
	0	0	0	0	0	0
SL	0	0	0	0	0	0
UPGMA	0	3	0	0	0	1
UPGMC	0	7	0	2	5	3
WARD	0	0	0	0	0	0
WPGMA	0	0	0	0	1	0
WPGMC	0	3	0	0	2	2
Mono-dimensional SOM	0	0	0	0	0	0
Bi-dimensional SOM using proposed k-means for classification in a second level	34	0	6	21	1	1

Table 1.12. Comparison of the Clustering Models for the Sets of Load Curves of the Greek Power System per Season for the time period 1985-2002

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 1.13 by indicating the number of years which achieves 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*, *DBI* and *WCBC*R measures, the bi-dimensional SOM model using proposed k-means for classification in a second level for *J* indicator and the UPGMC algorithm for *SMI* index.

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 behaviour of the Greek power system per year, the 20<sup>th</sup> 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.

#### 1.6.3 Usefulness of the application of the proposed methodology

The results of the second stage can be used as important input information for:

Load Time-Series Classification Based on Pattern Recognition Methods

- 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.

Methods	Adequacy Measure					
		MIA	CDI	SMI	DBI	WCBCR
Proposed k-means	0	8	18	0	13	14
Classic k-means	0	0	0	0	0	0
AVQ	1	0	0	1	3	0
Fuzzy k-means	0	0	0	0	0	0
CL	0	0	0	0	0	0
SL	0	0	0	0	0	0
UPGMA	0	1	0	0	0	0
UPGMC	0	6	0	14	1	2
WARD	0	0	0	0	0	0
WPGMA	0	1	0	0	0	0
WPGMC	0	1	0	0	0	2
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

Table 1.13. Comparison of the Clustering Models for the Sets of Load Curves of the Greek Power System per Year for the time period 1985-2002

## **1.7 Conclusions**

In this chapter pattern recognition methodologies for the study of the load time series were presented. Specifically, the first methodology deals with the classification of the daily chronological load curves of each large electricity customer, in order to estimate his typical days and his respective representative daily load profiles. It is based on classical pattern recognition methods, such as k-means, hierarchical agglomerative clustering, Kohonen adaptive vector quantization, mono-dimensional and bi-dimensional self-organized maps and fuzzy k-means. 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 latter are the mean square error, the mean index adequacy, the clustering dispersion indicator, the similarity matrix, the Davies-Bouldin indicator and the ratio of within cluster

sum of squares to between cluster variation. Some pattern recognition methods, such as kmeans, were properly modified, in order to achieve better values for the adequacy measures. The results can be used for the load forecasting of each consumer, the choice of the proper tariffs and the feasibility studies of demand side management programs. This methodology is in detail applied for one medium voltage paper mill industrial customer and synoptically for a set of 94 medium voltage customers of the Greek power distribution system, although it is applicable to any power system. From this execution the basic conclusions are:

- The daily chronological load of each large customer for a year can be classified to 8÷12 clusters satisfactorily (in the special case of seasonal customers, like small seaside hotels, oil-press industry, less clusters are needed).
- The ratio of within cluster sum of squares to between cluster variation (*WCBCR*) is proposed as the most suitable adequacy measure, because of (a) the presentation of the minimum dead clusters with respect to the desired number of clusters against the other adequacy measures and (b) its basic theoretical advantage, which is the combination of the distances of the input vectors from the representative clusters and the distances between clusters, covering the characteristics of the mean square error (*J*) and the mean index adequacy (*CDI*) simultaneously.
- The proposed k-means method and the hierarchical agglomerative methods (especially the unweighted pair group method average (*UPGMA*) & the unweighted pair group method centroid (*UPGMC*)) present the best results for the set of 94 medium voltage customers of the Greek power distribution system with respect to the *WCBCR* adequacy measure taking into consideration the computational training time.

Secondly, a two-stage methodology developed for the classification of electricity customers is presented. In the first stage, typical chronological load curves of various customers are estimated using pattern recognition methods and their results are compared using six adequacy measures, as it has already happened in the first case. In the second stage, classification of customers is performed by the same methods and measures, together with the representative load patterns of customers being obtained from the first stage. The basic contribution of this methodology is that its first stage enables the modification of the representative day, such as the most populated day, and avoids the a priori definition of a single day or the "mean" day of a specific time period (as it is suggested by previously published methodologies (Chicco et al., 2002; -,2003a; -,2003b; -, 2004; -,2006; Figueiredo et al., 2003; -, 2005; Gerbec et al., 2003; -,2004;-,2005)). The results of the second stage provide valuable information for electricity suppliers in competitive energy markets. The developed methodology was applied on the aforementioned set of 94 customers. From this execution the basic conclusions are:

- The representative clusters of the customers classes can be approximately 10÷15 for a set of 94 customers.
- The ratio of within cluster sum of squares to between cluster variation (*WCBCR*) is proposed as the most suitable adequacy measure for the same reasons for which it was also proposed in the first stage.

- The proposed k-means method and the hierarchical agglomerative methods (especially the weighted pair group method average (*WPGMA*) & the unweighted pair group method centroid (*UPGMC*)) present the best results for the classification of the second stage with respect to the *WCBCR* adequacy measure independently from the kind of the typical day which was examined (the most populated one and the day with the peak load).
- The a priori index of customer's activity is not representative for the classification of the load curves, which is also confirmed by (Chicco et al., 2002; -,2003a; -,2003b; -, 2004; -, 2006; Figueiredo et al., 2003; -, 2005). This can not be generalized since it may vary within countries and distribution companies, depending on the respective data of customers (Gerbec et al., 2003;-,2004;-,2005).

Finally, the pattern recognition methodology for the classification of the daily chronological load curves of the Greek power system is presented, in order to estimate their respective representative daily load profiles, which can be used for load forecasting and the feasibility studies of demand side management programs. Practically it is the same one with the first methodology or with the first stage of the two-stage methodology. It has been applied for the Greek power system for the period of years 1985-2002 per season (summer & winter) and per year, and from its execution the main conclusions are:

- The daily chronological load curves of the Greek power system for a season can be classified to 8÷10 clusters, which proves that the separation to workdays and no-workdays is not satisfactory. For a year the necessary clusters should be 15÷20.
- The ratio of within cluster sum of squares to between cluster variation (*WCBCR*) is proposed as the most suitable adequacy measure for the same reasons for which it was also proposed in the first methodology.
- The proposed k-means method and the hierarchical agglomerative methods (especially the weighted pair group method centroid (*WPGMC*) & the unweighted pair group method centroid (*UPGMC*)) present the best results for the classification of the load curves with respect to the *WCBCR* adequacy measure.

At the end, it should be mentioned that the basic contributions of the aforementioned methodologies are:

- The use of a set of pattern recognition methods, whose parameters are optimized properly for each adequacy measure separately, in order to use that method which gives the best results for the respective adequacy measure.
- The use of the ratio of within cluster sum of squares to between cluster variation (*WCBCR*) for this kind of methodologies for the first time.

These pattern recognition methodologies can be used for the classification of similar time-series curves, such as daily chronological temperatures curves, etc (Tsekouras, 2006).

## 1.8 References

Al-Hamadi, H. M. & Soliman, S.A. (2006). Long-term/mid-term electric load forecasting based on short-term correlation and annual growth. *Electric Power Systems Research*, Vol. 74, 2005, pp.353-361.

- Beccali, M.; Cellura, M.; Lo Brano, V. & Marvuglia, A. (2004). Forecasting daily urban electric load profiles using artificial neural networks. *Energy Conversion and Management*, Vol. 45, 2004, pp. 2879-2900.
- Carpaneto, E.; Chicco, G.; Napoli, R. & Scutariu, M. (2006). Electricity customer classification using frequency-domain load pattern data. *Electrical Power and Energy Systems*, Vol. 28, 2006, pp. 13-20.
- Chen, C. S.; Hwang, J. C. & Huang, C. W. (1997). Application of load survey systems to proper tariff design. *IEEE Transactions on Power Systems*, Vol. 12, No. 4, November 1997, pp. 1746-1751.
- Chicco, G.; Napoli, R. & Piglione, F. (2001). Load pattern clustering for short-term load forecasting of anomalous days. *IEEE Porto Power Tech Conference*, September 10-13, 2001, Porto, Portugal, p.6.
- Chicco, G.; Napoli, R.; Piglione, F.; Scutariu, M.; Postolache, P. & Toader, C. (2002). Options to classify electricity customers. *Med Power* 2002, November 4-6, 2002, Athens, Greece.
- Chicco, G.; Napoli, R.; Postolache, P.; Scutariu, M. & Toader, C. (2003a). Customer characterization for improving the tariff offer. *IEEE Transactions on Power Systems*, Vol. 18, No. 1, February 2003, pp. 381-387.
- Chicco, G.; Napoli, R. & Piglione, F. (2003b). Application of clustering algorithms and self organising maps to classify electricity customers. *IEEE Power Tech Conference*, June 23-26, 2003, Bologna, Italy.
- Chicco, G.; Napoli, R.; Piglione, F.; Postolache, P.; Scutariu, M. & Toader, C. (2004). Load pattern-based classification of electricity customers. *IEEE Transactions on Power Systems*, Vol. 19, No. 2, May 2004, pp. 1232-1239.
- Chicco, G.; Napoli, R. & Piglione, F. (2006). Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, Vol. 21, No. 2, May 2006, pp. 933-940.
- Davies, D.L. & Bouldin, D.W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, April 1979, pp. 224-227.
- Duda, R.O.; Hart, P.E. & Stork, D.G.(2001). *Pattern Classification*. A Wiley-Interscience Publication, 1<sup>st</sup> edition.
- Figueiredo, V.; Duarte, F. J.; Rodrigues, F.; Vale, Z.; Ramos, C. & Gouveia, J. B.(2003). Electric customer characterization by clustering. *ISAP 2003*, Lemnos, Greece, August 2003.
- Figueiredo, V.; Rodrigues, F.; Vale, Z. & Gouveia, J. B.(2005). 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.
- Gerbec, D.; Gasperic, S. & Gubina, F. (2003). Determination and allocation of typical load profiles to the eligible consumers. *IEEE Power Tech Conference*, Bologna, Italy, June 23-26, 2003.
- Gerbec, D.; Gasperic, S.; Smon, I. & Gubina, F. (2004). Determining the load profiles of consumers based on fuzzy logic and probability neural networks, *IEE*

*Proceedings on Generation, Transmission & Distribution,* Vol. 151, May 2004, pp. 395-400.

- Gerbec, D.; Gasperic, S.; Smon, I. & Gubina, F. (2005). Allocation of the load profiles to consumers using probabilistic neural networks. *IEEE Transactions on Power Systems*, Vol. 20, No. 2, May 2005, pp. 548-555.
- Hand, D.; Manilla, H. & Smyth, P.(2001). *Principles of data mining*. The M.I.T. Press, Cambridge, Massachusetts, London, England.
- Haykin, S. (1994). *Neural Networks, A Comprehensive Foundation*. Englewood Cliffs, NJ: Prentice Hall.
- Kitayama, M.; Matsubara, R. & Izui, Y. (2002). Application of data mining to customer profile analysis in the power electric industry. *IEEE PES Winter Meeting 2002*, New York, USA.
- Kohonen, T. (1989). *Self–organization and Associative Memory*. New York: Springer-Verlag, 3<sup>nd</sup> edition.
- Lamedica, R.; Prudenzi, A.; Sforna, M.; Caciotta, M. & Orsolini Cencelli, V (1996). A neural network based technique for short-term forecasting of anomalous load periods. *IEEE Transactions on Power Systems*, Vol. 11, No. 4, November 1996, pp. 1749-1756.
- Petrescu, M. & Scutariu, M. (2002). Load diagram characterization by means of wavelet packet transformation. 2<sup>nd</sup> Balkan Conference, Belgrade, Yugoslavia, June 2002, pp. 15-19.
- SOM Toolbox for MATLAB 5 (2000). Helsinki, Finland: Helsinki Univ. Technology.
- Task VIII of International Energy Agency (2002). 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.*
- Thang, K. F.; Aggarwal, R. K.; McGrail, A. J. & Esp, D. G.(2003). 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.
- Theodoridis, S. & Koutroumbas, K. (1999). *Pattern Recognition*, Academic Press, New York, 1<sup>st</sup> edition,.
- Tsekouras, G. J. (2006). Contribution to short-term and mid-term load and energy forecasting for power systems based on pattern recognition methods. Ph.D. Thesis, School of Electrical and Computer Engineering, NTUA, Greece, July 2006 (in Greek).
- Tsekouras, G. J.; Hatziargyriou, N.D. & Dialynas, E.N. (2007). Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers. *IEEE Transactions on Power Systems*, Vol. 22, No. 3, August 2007, pp. 1120-1128.
- Tsekouras, G. J.; Kotoulas, P. B.; Tsirekis, C.D.; Dialynas, E.N. & Hatziargyriou, N.D. (2008). A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers. *Electric Power Systems Research*, Vol.78, 2008, pp.1491-1510.
Verdu, S. V.; Garcia, M. O.; Senabre, C.; Marin, A. G. & Garcia Franco F. J. (2003). Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps. *IEEE Transactions on Power Systems*, Vol. 18, No. 1, February 2003, pp. 381-387.



www.intechopen.com



Pattern Recognition Techniques, Technology and Applications Edited by Peng-Yeng Yin

ISBN 978-953-7619-24-4 Hard cover, 626 pages Publisher InTech Published online 01, November, 2008 Published in print edition November, 2008

A wealth of advanced pattern recognition algorithms are emerging from the interdiscipline between technologies of effective visual features and the human-brain cognition process. Effective visual features are made possible through the rapid developments in appropriate sensor equipments, novel filter designs, and viable information processing architectures. While the understanding of human-brain cognition process broadens the way in which the computer can perform pattern recognition tasks. The present book is intended to collect representative researches around the globe focusing on low-level vision, filter design, features and image descriptors, data mining and analysis, and biologically inspired algorithms. The 27 chapters coved in this book disclose recent advances and new ideas in promoting the techniques, technology and applications of pattern recognition.

## How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

George J. Tsekouras, Anastasios D. Salis, Maria A. Tsaroucha and Irene S. Karanasiou (2008). Load Time-Series Classification Based on Pattern Recognition Methods, Pattern Recognition Techniques, Technology and Applications, Peng-Yeng Yin (Ed.), ISBN: 978-953-7619-24-4, InTech, Available from: http://www.intechopen.com/books/pattern\_recognition\_techniques\_technology\_and\_applications/load\_timeseries\_classification\_based\_on\_pattern\_recognition\_methods

## INTECH

open science | open minds

## InTech Europe

University Campus STeP Ri Slavka Krautzeka 83/A 51000 Rijeka, Croatia Phone: +385 (51) 770 447 Fax: +385 (51) 686 166 www.intechopen.com

## InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai No.65, Yan An Road (West), Shanghai, 200040, China 中国上海市延安西路65号上海国际贵都大饭店办公楼405单元 Phone: +86-21-62489820 Fax: +86-21-62489821 © 2008 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the <u>Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License</u>, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.



