ELSEVIER

Electrical Power and Energy Systems 28 (2006) 13–20

ELECTRICAL POWER & ENERGY SYSTEMS

www.elsevier.com/locate/ijepes

Electricity customer classification using frequency–domain load pattern data

Enrico Carpaneto [a], Gianfranco Chicco [a,*], Roberto Napoli [a], Mircea Scutariu [b]

[a] *Dipartimento di Ingegneria Elettrica, Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy*
[b] *Electrica Muntenia Sud Distribution and Supply Company, Bucharest, Romania*

Received 19 August 2003; received in revised form 17 May 2005; accepted 11 August 2005

Abstract

In competitive electricity markets, electricity customer classification is becoming increasingly important, due to new degrees of freedom the electricity providers have been given in formulating dedicated tariff options for different customer classes. Several customer classification techniques have been proposed in the literature, in which the load patterns are typically represented by time–domain data. However, a good load pattern representation requires using several data for each customer, causing possible difficulties in storing a large amount of data in the electricity company's databases. In order to reduce the number of data to be stored for each customer, an original solution is proposed in this paper, based on post-processing the results of time–domain measurements to obtain a reduced set of data defined in the frequency domain. The new set of data is successively used in a customer classification procedure, e.g. a suitable clustering technique, whose adequacy can be assessed by means of properly defined indicators. This paper provides the mathematical background for the frequency–domain data definition and investigates on the effectiveness of the customer classification for different choices of the number of data to be stored. Results obtained on a set of customers belonging to a real distribution system are presented and discussed. These results show that the proposed representation is effective in reducing the number of data stored while maintaining a satisfactory level of classification adequacy.

Keywords: Electricity customer classification; Load patterns; Clustering; Frequency–domain data; Harmonics-based features

1. Introduction

The effects of electricity market restructuring include the increase in the number of market operators and the creation of new classes of electricity services. In several countries, changes have been made to the electricity regulation, on the basis of which the electricity providers now benefit from the possibility of designing different tariff structures to be applied to properly re-structured customer classes [1–3]. For the customer class definition, emergent solutions can be conveniently related to the effective behavior of the electricity customer consumption. This requires a more refined knowledge of the electrical load variability under different technical and economical conditions. The formulation of new tariff structures is essential to develop dedicated strategies aimed at scoring profits in the restructured market scenario.

The information required for electricity customer characterization include a number of *external features*, such as weather data, rated values of electrical quantities, type of activity and other commercial information, and specific *shape features* extracted from the load patterns. External features are typically used to make a preliminary customer partitioning into macro-categories. In addition, it is useful to introduce a distinction among *loading conditions* (e.g. summer and winter periods, with a possible inner partition into working days, Saturdays and Sundays for each period) classified on the basis of a statistical analysis of daily load patterns [4]. Any preliminary customer partitioning into macro-classes depending on external features and loading conditions can then lead to set with a reduced number of load patterns in each macro-class. Each set of load patterns is successively handled by using refined classification methods to form the customer classes.

The choice of the customers to perform effective load pattern survey is a critical aspect of the classification. The stratified sampling theory [5] provides the conceptual basis for identifying the number of customers to select in order to represent the characteristics of the customer population in terms of contract power and/or average power, taking into

* Corresponding author. Tel.: +39 011 564 7141; fax: +39 011 564 7199.
E-mail address: gianfranco.chicco@polito.it (G. Chicco).

0142-0615/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijepes.2005.08.017

account the constraint on the costs incurred for undertaking the survey. Applications of the stratified sampling approach to the customer selection are presented in [1,6]. The actual selection of the set of customers is then carried out by checking out the accessibility of the individual customers for carrying out measurements during a specified time period.

The customer classification addressed in this paper is specifically related to non-residential customers. The load patterns of the individual residential customers have not been taken into account among the representative load patterns. In fact, the load pattern of a residential customer highly depends on the specific customer behavior and lifestyle, that leads to a greatly diversified use of the electrical appliances during the day [7,8]. As such, assessing the representative load pattern from the measurements carried out on an individual residential customer would lead to a very large uncertainty in the results. Moreover, each distribution feeder supplies many residential customers at a time, thus the assessment of the load pattern for a residential customer aggregation could be more interesting than assessing the individual load patterns of the residential customers. In this case, it is possible to use a comprehensive and effective bottom-up approach [8,9] based on probabilistic techniques to combine electrical, demographic and social data of the customers to provide a synthesis of the aggregated residential load.

Let us assume the representative daily load patterns of M customers are defined for a given macro-category in a specified loading condition. Each *representative* load pattern results from averaging the load patterns obtained by a set of measurements carried out in the same loading condition. A simple way to characterize the *shape features* of the mth representative load pattern, for $m=1,\ldots,M$, is to assume as features all or part of the power values of the representative load pattern in the time domain. In such a way, a set of H *direct shape features* is readily available without performing a load pattern post-processing. Assuming the maximum power value of the load pattern as *reference power*, the direct shape features can be normalized with respect to the reference power to fit the needs of the classification procedure adopted. Another way for characterizing a load pattern is the use of *indirect shape features*, such as comprehensive *shape factors*. Several shape factors restricted to the range (0,1) have been proposed in [3], each of them highlighting a particular aspect of the load pattern, such as the minimum to average power ratio, the average to maximum power ratio, and so on. However, there are several ways to define shape factors. Another possibility is resorting to a frequency–domain approach. Frequency–domain applications to load characterization are less usual than time–domain applications. Some recently proposed methods adopt a wavelet coefficients-based approach to characterize the load patterns [10] or perform customer classification by means of clustering techniques by using harmonics-based coefficients as input data [11].

This paper is a revised and extended version of [11]. An original technique is proposed to characterize the load patterns of each customer by using a *frequency–domain* approach. A novel application of clustering techniques using harmonics-based features is presented and discussed, showing its efficiency on a set of customer data belonging to a real system. The classification adequacy is assessed by means of different adequacy indicators. Assuming the day as fundamental period, a set of indirect shape features is defined by using amplitude and phase-related values of a given set of harmonics computed from the representative load patterns. The motivation for using this approach follows the definition of the representative load patterns as patterns characterizing the customers for any day belonging to a given loading condition. This characterization of the load patterns encompasses the definition of load profiles currently adopted in some countries [12,13]. In such a way, the representative load pattern can be seen as a periodic signal in the time interval of interest, for which harmonic analysis techniques apply.

This paper is structured as follows. Section 2 illustrates the characteristics of data sampling and the definition of the frequency–domain data. Section 3 recalls the basics of the clustering approach used to form the customer classes and introduces the indicators of clustering adequacy. Section 4 shows and discusses the results for a set of non-residential customers belonging to a distribution company. Section 5 contains the concluding remarks.

2. Frequency–domain data definition

Let us consider the time interval T_0 for performing load pattern sampling, corresponding to the fundamental frequency $f_0=1/T_0$. Data of the M customers considered are sampled inside the time interval T_0 with cadence T_{sam} (and sampling frequency $f_{sam}=1/T_{sam}$). According to the Nyquist's theorem, the maximum meaningful frequency that can be obtained from the sampled data is $f_{max}=f_{sam}/2$, resulting in the maximum meaningful harmonic order $h_{max}=f_{max}/f_0=T_0/(2T_{sam})$. Assuming a period T_0 of 1 day, Table 1 shows the maximum meaningful harmonic order that can be obtained from sampling data with different cadence (1 min, 15 min and 1 h). These results show that a sampling cadence of 1 h could be too long for collecting sufficient information to obtain meaningful results from harmonic analysis, while the 15-min sampling can be considered sufficiently fast for this purpose.

The sampled representative load patterns are subject to harmonic analysis, computing amplitude and phase of the harmonics by means of the Discrete Fourier Transform (DFT), taking values up to the harmonic order h_{max}. Information on amplitude and phase of the harmonic components is then used to extract a set of harmonics-based features. The objective is to extract a reduced set of features from the whole set of harmonic amplitudes and phase-related values resulting from the DFT

Table 1
Maximum meaningful harmonic order h_{max} for $T_0=1440$ min (1 day) and different data sampling

T_{sam} (min)	h_{max}
1	720
15	48
60	12

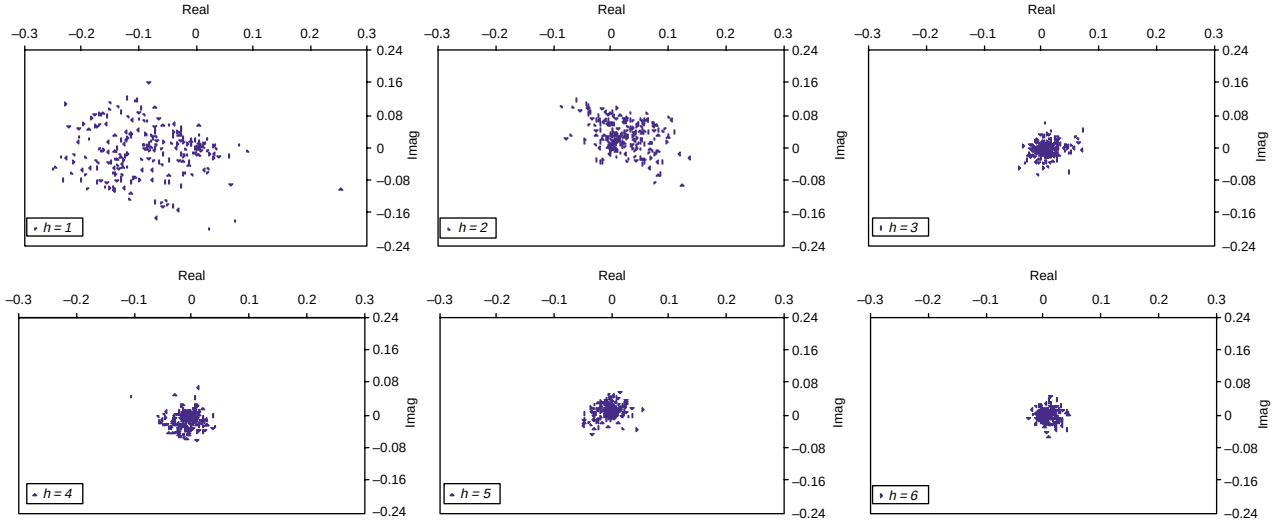

Fig. 1. Dispersion of the harmonic components for six harmonic orders.

calculations. Effective choice of this reduced set of features requires solving specific problems, in particular concerning the harmonic order ranking and the representation of the phase information. These aspects are detailed in the sequel.

2.1. Harmonic order ranking

The assessment of the most significant set of harmonics has been addressed by evaluating the dispersion of the information for each harmonic amplitude $h=1,\ldots,h_{\max}$ using the indicator

$$\varsigma_h = \frac{1}{M}\sum_{i=1}^{M}|A_{hi}-\bar{A}_h| \tag{1}$$

representing the average value of the absolute deviations of the harmonic amplitudes of each customer with respect to their average value $\bar{A}_h$. Harmonic orders have then been ranked in decreasing order of the indicator ς_h. Phase values are not used in ranking the harmonic orders, since the phase variability spreads over a wide range for all the harmonic orders $h>0$. The phase variability is illustrated in Fig. 1, showing on the complex plane the harmonic components extracted from the time–domain load pattern samples of the real-case example presented in Section 4.

2.2. Representation of the phase information

The treatment of the phase information presents some critical aspects, concerning the need for finding out a consistent representation, in order to avoid discontinuities in the phase representation due to periodicity and ensuring that similar phases are represented by similar values of the phase-related features. Since a unique function cannot take into account all the needs of the phase angle representation, the phase-related values corresponding to the phase angles α_h are represented by using the two functions $(1-\cos\alpha_h)/2$ and $(1-\sin\alpha_h)/2$, for $h=1,\ldots,h_{\max}$. This definition leads to normalize the phase angle information in the range (0,1). However, amplitudes and phase-related values should not be considered as totally independent features, otherwise there could be large values of the phase-related features in the presence of very small values of the corresponding amplitudes, leading to unreasonable amplification of the impact of the phase information for harmonic orders having relatively low amplitude. This would result in unjustifiably significant differences among the features used for classifying the subjects. In order to take this aspect into account, phase-related values are no longer represented in the normalized form and are defined in the most appropriate form described in Section 2.3.

2.3. Feature selection

The selection of harmonics-based features is carried out by including the amplitude of the zero-order harmonic (proportional to the average power) and suitable values of amplitude and phase-related values of a number of top-ranked harmonic orders. The first n harmonic amplitudes are extracted from the harmonic order ranking in descending order of the indicator ς_h introduced in (1). Let us define the set Θ_n containing the n top-ranked harmonic orders involved in the formation of the set of features under consideration. For example, Θ_0 contains only the top-ranked harmonic order, while Θ_{10} contains the first 10 harmonic orders, listed according to their descending ranking order.

Phase-related variables, not used in ranking the harmonic orders, are used as features for clustering purposes to avoid grouping together customer patterns whose harmonics have similar amplitudes but significantly different phases. In order to avoid direct use of normalized phases as features, the phase-related features for a given harmonic order are defined by multiplying the normalized phase value to a weight factor. This weight factor is the ratio between the corresponding harmonic order amplitude and the RMS value determined from harmonic amplitudes of all harmonic orders in the set Θ_n. This allows linking the phase information to the significance of the corresponding harmonic amplitude, attenuating the impact of

harmonic orders with low amplitude but high phase dispersion. Hence, the two phase-related features for the hth harmonic order belonging to the set Θ_n are

$$\psi'_h = \frac{1-\cos(\alpha_h)}{2} \frac{A_h}{\sqrt{\sum_{j\in\Theta_n} A_j^2}} \tag{2}$$

and

$$\psi''_h = \frac{1-\sin(\alpha_h)}{2} \frac{A_h}{\sqrt{\sum_{j\in\Theta_n} A_j^2}} \tag{3}$$

Various sets of features are then defined on the basis of the harmonic order ranking, including phase-related features. The only exception is the phase of the zero-order harmonic, always null, which is never included in the feature sets. Let us call 'Set Hn', with $n \geq 0$, the set of features

$$\mathbf{f}_n = \{(A_k, \psi'_k, \psi''_k), (k \neq 0) \cap (k \in \Theta_n)\} \cup \{(A_k), (k = 0) \cap (k \in \Theta_n)\} \tag{4}$$

For a given value of n, the set of data characterizing the representative daily load pattern of customer $m = 1,\ldots,M$ is

$$\mathbf{f}_n^{(m)} = \{f_j^{(m)}, j = 1, \ldots, n\} \tag{5}$$

and the whole set of data, for all customers, is

$$\boldsymbol{F}_n = \{\mathbf{f}_n^{(m)}, m = 1, \ldots, M\} \tag{6}$$

3. Load pattern clustering and adequacy assessment

3.1. Clustering algorithm

The above-defined harmonics-based features is used for customer classification, by running a clustering algorithm to group the load patterns on the basis of their distinguishing features. Several clustering techniques are available to perform this task [14,15]. A modified follow-the-leader procedure is used in this paper. This procedure is a slight modification of the algorithm proposed in [16], which does not require initialization of the number of clusters. The modification refers to the use of a variance-based criterion to weight the input data [3]. This procedure has been shown to be numerically effective, compared to other clustering approaches, using time–domain data as features [14,15]. The clustering process is driven by a distance threshold and consists in an iterative reassignment of customer data to the clusters until a stable formation occurs. More specifically, a first cycle groups together the load patterns whose modified Euclidean distance does not exceed the distance threshold, automatically fixing the final number of clusters and computing their centroids. The successive cycles reassign the load patterns to the existing clusters on the basis of the minimum modified Euclidean distance with respect to the existing cluster centroids, updating the *in* and the *left* cluster centroids each time a load pattern changes cluster. The process stops when a stable cluster formation is reached.

3.2. Clustering adequacy indices

In order to evaluate the effectiveness of choosing a set of features based on harmonic analysis, a criterion for comparing the clustering results obtained with different sets of features is needed.

Let us assume the clustering process, performed by using the features included in a vector called $\mathbf{y}$, results in forming K customer classes. Let us introduce the subsets $\boldsymbol{L}^{(k)}$ containing the initial (time–domain) data of the load patterns belonging to class $k = 1,\ldots,K$. Let us further introduce the set $\boldsymbol{R} = \{\mathbf{r}^{(k)}, k = 1, \ldots, K\}$ of the class representative load patterns, where $\mathbf{r}^{(k)}$ is obtained by computing the weighted average of the time–domain load patterns belonging to the subset $\boldsymbol{L}^{(k)}$, assuming the reference power of each load pattern as weighting factor.

Defining the adequacy indices requires some notion of distance. Let us consider the distance $d(\mathbf{r}^{(k)},\boldsymbol{L}^{(k)})$ between a representative load pattern $\mathbf{r}^{(k)}$ and the subset $\boldsymbol{L}^{(k)}$, defined as the geometric mean of the distances between $\mathbf{r}^{(k)}$ and each member of $\boldsymbol{L}^{(k)}$, and the infra-set mean distance $\hat{d}(\boldsymbol{L}^{(k)})$, defined as the geometric mean of the distances between the members of $\boldsymbol{L}^{(k)}$. On the basis of these distances, the following adequacy indicators have been used:

(a) the Mean Index Adequacy (MIA), depending on the distance among each class representative load pattern and the load patterns falling into the corresponding cluster [3]:

$$\mathrm{MIA} = \sqrt{\frac{1}{K}\sum_{k=1}^{K} d^2(\mathbf{r}^{(k)}, \boldsymbol{L}^{(k)})} \tag{7}$$

(b) the Clustering Dispersion Indicator (CDI), depending on the infra-set distances between the load patterns in the same cluster and between the class representative load patterns [3]:

$$\mathrm{CDI} = \frac{1}{\hat{d}(\boldsymbol{R})} \sqrt{\frac{1}{K}\sum_{k=1}^{K} \hat{d}^2(\boldsymbol{L}^{(k)})} \tag{8}$$

the Scatter Index (SI), obtained by a slight modification of the index defined in [13]:

$$\mathrm{SI} = \left(\sum_{m=1}^{M} d^2(\mathbf{l}^{(m)}, \mathbf{p})\right)\left(\sum_{k=1}^{K} d^2(\mathbf{r}^{(k)}, \mathbf{p})\right)^{-1} \tag{9}$$

where $\mathbf{p}$ is the pooled scatter

$$\mathbf{p} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{l}^{(m)} \tag{10}$$

and the Davies – Bouldin Index [17], representing the system-wide average of the similarity measures of each cluster with its most similar cluster, written in its

Euclidean form, for $i, j = 1,\ldots,K$:

$$\mathrm{DBI} = \frac{1}{K}\sum_{k=1}^{K}\max_{i\neq j}\left\{\frac{\hat{d}(\boldsymbol{L}^{(i)}) + \hat{d}(\boldsymbol{L}^{(j)})}{d(\mathbf{r}^{(i)}, \mathbf{r}^{(j)})}\right\} \tag{11}$$

A common characteristic of these indices is that smaller index values correspond to higher adequacy.

4. Application to a real distribution system

4.1. Data definition and clustering

The results of the proposed approach are illustrated by using a set of customers supplied by the Romanian distribution and supply company *Electrica*. Application of the stratified sampling approach led to the selection of 232 non-residential customers, whose load patterns have been recorded at locations spread all over the country over 3-week time intervals. According to the results shown in Table 1, the 15-min data sampling has been considered, so that the daily load pattern of each customer contains 96 time–domain samples. For each customer, bad data detection has been performed in such a way to ensure that the load patterns used for customer classification correspond to normal operating conditions. For this purpose, load data corresponding to anomalous days (e.g. bank holidays occurring at weekdays), strikes, unexpected events or failures have been eliminated from the analysis. The effects of failures or abnormal conditions have been detected by identifying the quarters of hours at which the average RMS voltage was outside the acceptable range (90–110% of the rated voltage) and the corresponding load values have been eliminated. The representative load patterns have then been obtained by averaging the load values corresponding to the same time quarter of hour of the weekdays (bad data excluded) and scaling each pattern to a normalized form, in such a way that each representative load pattern has a unity maximum value.

At the first step of the proposed approach, harmonic analysis of each load pattern is performed by using the Discrete Fourier Transform, obtaining the amplitude and phase harmonic spectra associated to each load pattern. The harmonic orders are then ranked by using the indicator ς_h introduced in (1). Table 2 shows the resulting harmonic orders listed in descending order of the indicator ς_h. Successively, the various sets Θ_n of harmonic orders and the corresponding feature vectors $\mathbf{f}_n$ are formed for $n \geq 0$. At this point, fixing a value of n, the features included in $\mathbf{f}_n^{(m)}$ are computed from the representative daily load pattern of every customer $m = 1,\ldots,M$.

Table 2
Harmonic order ranking

n	h	ς_h	n	h	ς_h
0	0	0.1595	6	6	0.0091
1	1	0.0580	7	7	0.0062
2	2	0.0278	8	8	0.0060
3	4	0.0147	9	12	0.0057
4	5	0.0118	10	11	0.0053
5	3	0.0117	11	9	0.0053

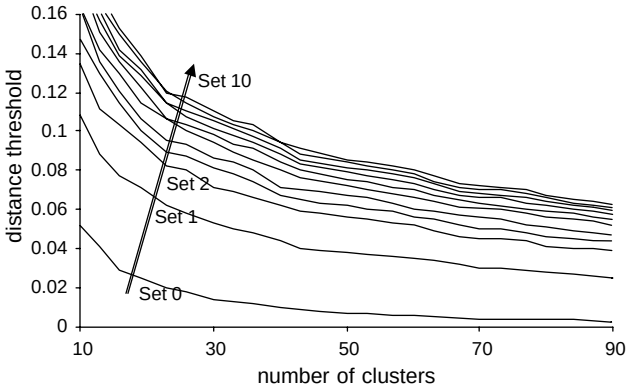

Fig. 2. Distance threshold for the clustering procedure in function of the number of clusters for different sets of harmonics-based features.

In order to investigate the choice of harmonics-based feature sets, the clustering algorithm are run by using the sets $\boldsymbol{F}_n$, for $n = 0$–10. Fig. 2 shows the evolution of the distance thresholds used in the clustering procedure, in function of the number of clusters for the different sets of harmonics-based features. For a specified distance threshold, the number of clusters increases when n increases, showing that using more features is helpful in better discriminating the customers' electrical behavior. In fact, with a larger number of features the distance thresholds are higher for the same number of clusters. This intuitive result is a first confirmation that the proposed approach is viable. In addition, from Fig. 2 it can be noted that adding further ranked harmonic orders to the feature set is increasingly less effective.

The objective is now to establish some motivated limits to the number of ranked harmonic orders to be considered for obtaining significant classification results. For this purpose, on the basis of the clustering results, the effectiveness of the proposed approach is studied by using the adequacy indicators introduced in Section 3.2. The adequacy indicators obtained from the original set of 96 time–domain daily samples forming the *Set T*96 are used as a term of comparison.

Running the clustering procedure with different sets of features results in partitioning the initial data into a different number of clusters, each cluster corresponding to a customer class. The clustering algorithm has been run with the time–domain sample *Set T*96, and with the harmonics-based sets $\boldsymbol{F}_n$, for $n = 0$–10. For each run, the outputs of the clustering procedure are the number of customer classes and the list of the customers belonging to each class. These results are used to compute, for each customer class, the *class representative load patterns*, representing the synthesized behavior of all the customers belonging to the class. The class representative load pattern is built, for each class, by computing the weighted average of the representative load patterns of the customers belonging to the class, using the corresponding reference powers as weights. However, the calculation of the class representative load diagrams is a step independent of the data used for running the clustering procedure. In fact, it is possible to compute the class representative load diagrams from the time–domain load patterns even when the clustering algorithm has been run by using the harmonics-based features, since

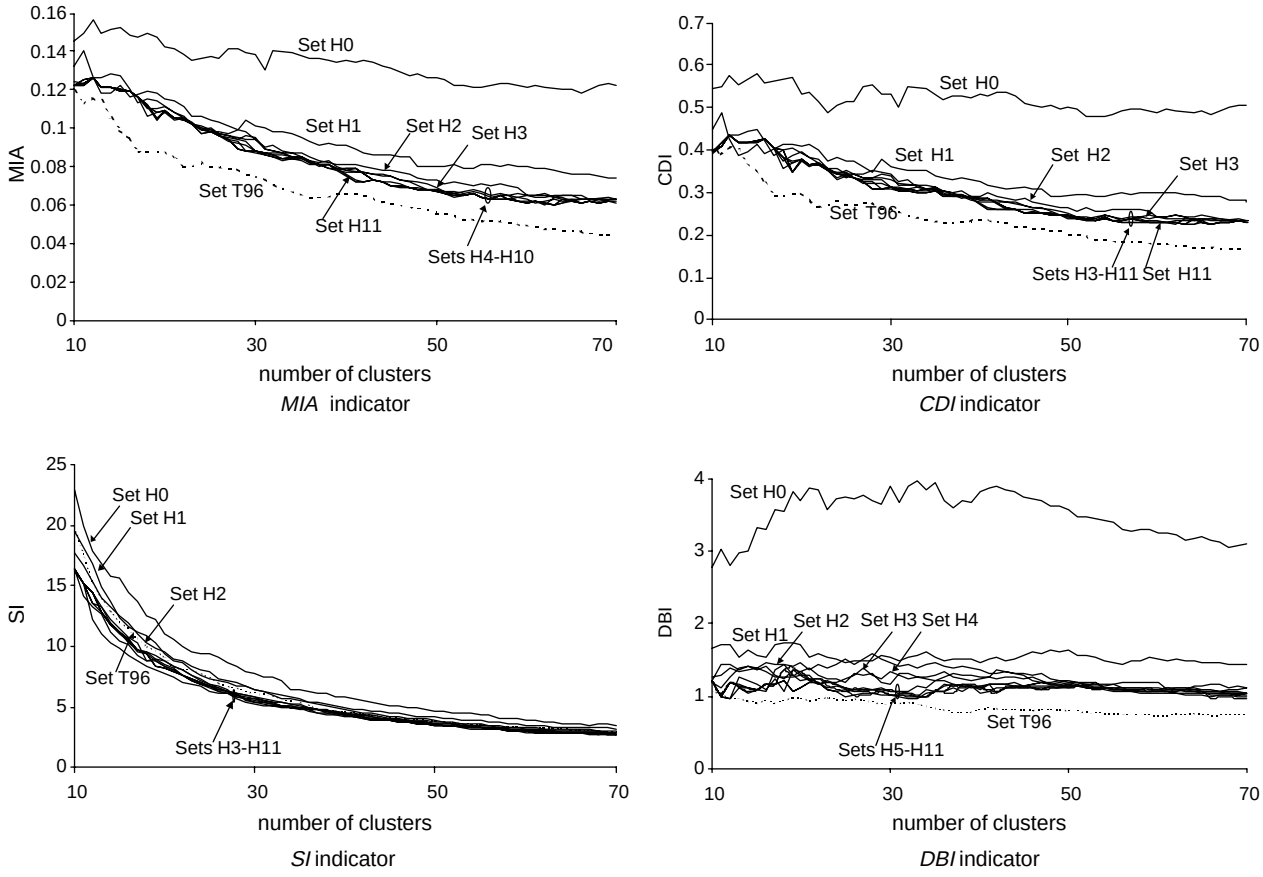

Fig. 3. Adequacy indices computed on the basis of the time–domain class representative load patterns from the modified follow-the-leader clustering performed with 96 time–domain data and various harmonics-based features.

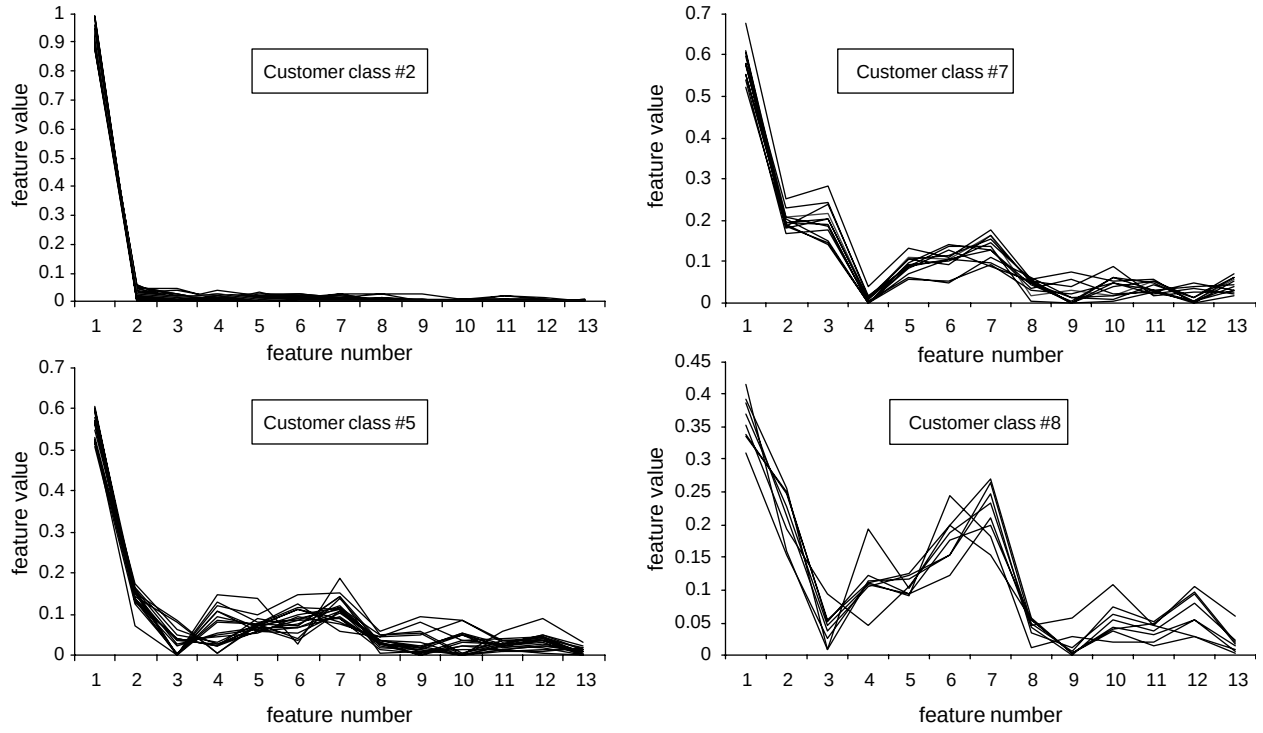

Fig. 4. Class representative feature set for four of the 16 customer classes obtained from clustering by using the *Set H*4 (13 features). Feature #1 is the zero-order harmonic (average value). The successive triplets of features are A_h, ψ'_h and ψ''_h for the harmonic orders ranked in descending order of the indicator (1).

Table 3
Number of customers in the classes obtained from the *Set H*4

Class no.	Number of customers	Class no.	Number of customers	Class no.	Number of customers	Class no.	Number of customers
1	35	5	32	9	1	13	1
2	79	6	15	10	4	14	1
3	26	7	14	11	1	15	1
4	4	8	14	12	4	16	2

the only relevant information in the clustering output concern the number of classes and the class composition.

4.2. Adequacy test

The values of the indicators shown in shown in (7)–(9) and (11) depend on the number K of customer classes, so that a comparison among the clustering results is meaningful only when the number of customer classes obtained from the different clustering techniques is the same. In addition, the adequacy results depend on the load pattern data used to compute the indicators. As such, the adequacy indices should be computed starting from the same set of data. For performing the adequacy test, the class representative load patterns have always been built starting from the *Set T*96, using the customer class compositions resulting from the various sets of initial data. The adequacy indices have been computed by using the class representative load patterns obtained in each case. In general, the number of clusters obtained from different feature sets with the same distance threshold is not the same. In order to get results for comparable numbers of clusters, computations have been repeated for different distance thresholds, building the curves shown in Fig. 3, where the curve labels represent the data set used for running the clustering algorithm.

The results obtained make it possible to perform the classification adequacy ranking, by comparing the solutions resulting in the same number of clusters. Fig. 3 shows that the different indicators give consistent information and that for each indicator the adequacy ranking is similar for different numbers of clusters. In particular, the results obtained by using the top-ranked *Set H*0 (containing information on the average normalized value of the load pattern) are insufficient to justify the use of this feature only. Adding new harmonics-based feature sets corresponding to top-ranked harmonic orders leads to results that, for a given number of clusters, generally correspond to lower values of the adequacy indicators, i.e. to better adequacy. However, it should be noted that using the harmonics-based features the adequacy indices remain higher

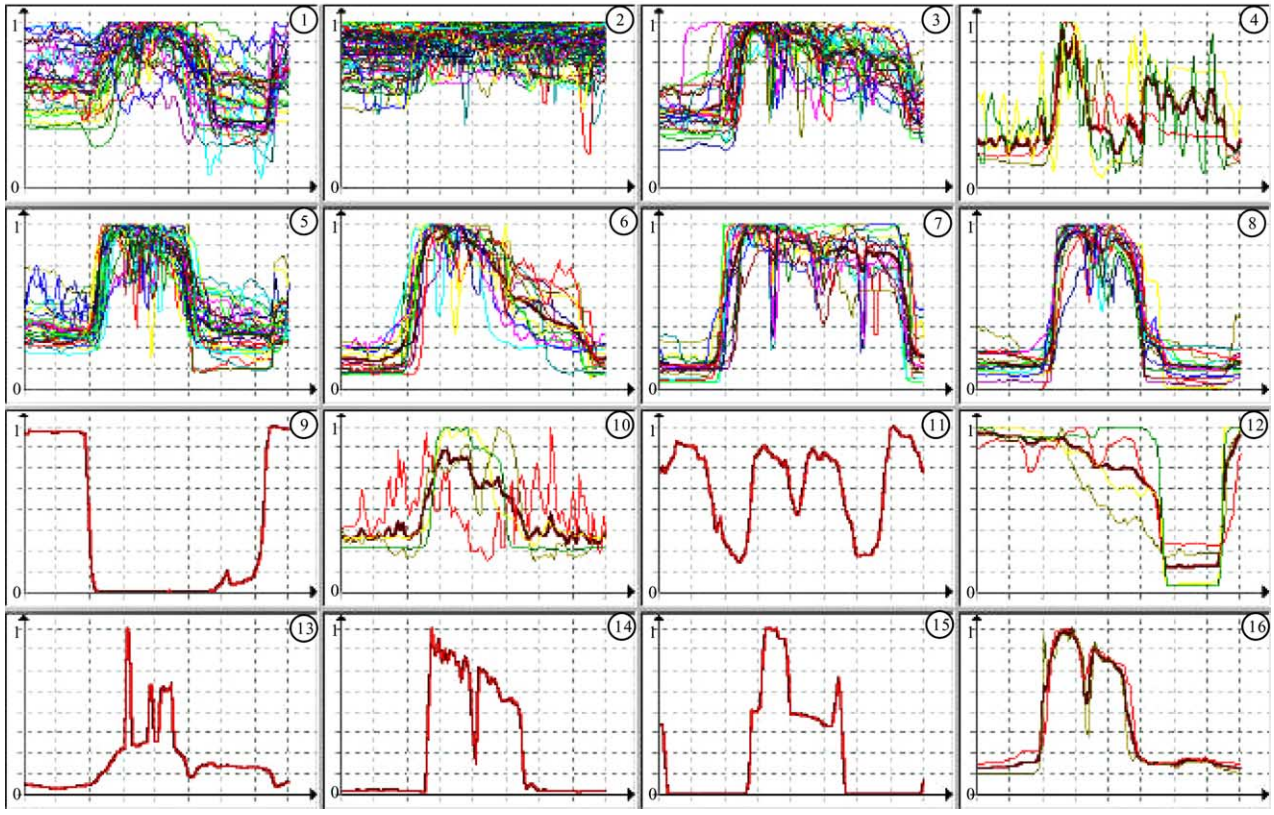

Fig. 5. Clustering results for the 16 customer classes obtained from the *Set* H4, with the load patterns of each customer and the class representative load patterns (thick lines). The horizontal axis contains the 96 time–domain features. The vertical axis contains the normalized power.

than the one obtained by using the time–domain features with the *Set T*96, with the only exception of the indicator SI, for which the results obtained from the harmonics-based features with $n>2$ are generally better than the ones resulting from using the *Set T*96. On the other side, the number of time–domain features (i.e. 96) for the *Set T*96 is significantly larger than the number of features used in every harmonics-based feature sets, so that the clustering procedure run with the *Set T*96 requires a significantly higher computational effort and amount of data stored. As such, the slight worsening in the adequacy indices obtained using the data in the sets *H*4 to *H*11 can be considered to be practically acceptable. The low difference between the results obtained from the *Set H*4 and the successive sets, up to the *Set H*11, suggests that the *Set H*4 can be considered able to ensure an acceptable trade off between classification adequacy and number of data stored, performing the clustering procedure with only 13 features (the amplitude of the zero-order harmonic and three features for each of the other four top-ranked harmonic orders—numbers 1, 2, 4 and 5, as indicated in Table 2). Fig. 4 highlights the discriminatory properties of the proposed method with the clustering algorithm run with the *Set H*4, by showing the frequency–domain features of some classes for the clustering resulting in 16 customer classes. In addition, Table 3 shows the composition of the 16 customer classes and Fig. 5 shows the corresponding groups with the initial load patterns and the class representative load patterns. The latter are built by computing the weighted average of the time–domain load patterns belonging to each customer class formed by the clustering algorithm run with the *Set H*4, by using the corresponding reference powers as weights.

5. Conclusions

A novel frequency–domain technique for grouping the customers' representative load patterns into a number of distinct classes has been presented. The results obtained show that the proposed technique is viable and effective for performing a customer classification by characterizing the features on the basis of the information extracted from harmonic analysis of the input load patterns. A key characteristic of the proposed method is the possibility of obtaining satisfactory customer classification by using a number of harmonics-based features lower significantly than the number of features required by using time-domain samples. This allows for a sensible reduction of the number of information that need to be stored, for each customer, in the electricity company's database. In addition, the reduced number of features has the positive effect of speeding up calculations when running the clustering algorithm. The adequacy of the classification, measured by means of suitable indicators, is almost comparable to the one obtained by using a larger number of time–domain features as inputs of the clustering procedure.

References

[1] Chen CS, Hwang JC, Huang CW. Application of load survey to proper tariff design. IEEE Trans Power Syst 1997;12(4):1746–51.
[2] Stephenson, P, Lungu, I, Paun, M, Silvas, I, Tupu, G. Tariff development for consumer groups in internal European electricity markets. Proceedings of CIRED 2001, Amsterdam, The Netherlands, June 18–21, 2001. Paper 5.3.
[3] Chicco G, Napoli R, Postolache P, Scutariu M, Toader C. Customer characterisation options for improving the tariff offer. IEEE Trans Power Syst 2003;18(1):381–7.
[4] Chen CS, Hwang JC, Tzeng YM, Huang CW, Cho MY. Determination of customer load characteristic by load survey system at Taipower. IEEE Trans Power Deliv 1996;11(3):1430–6.
[5] Neyman J. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. J R Stat Soc, Part IV 1934;558–606.
[6] Bompard, E, Carpaneto, E, Chicco, G, Napoli, R, Piglione, F, Postolache, P, Scutariu, M. Stratified sampling of the electricity customers for setting up a load profile survey. Proceedings of RIMAPS 2000, Funchal, Madeira, Portugal, September 25–28, 2000. Paper RUR-017.
[7] Walker CF, Pokoski JL. Residential load shape modeling based on customer behavior. IEEE Trans Power App Syst 1985;PAS-104(7): 1703–11.
[8] Capasso A, Grattieri W, Lamedica R, Prudenzi A. A bottom-up approach to residential load modeling. IEEE Trans Power Syst 1994;9(2):957–64.
[9] Cagni A, Carpaneto E, Chicco G, Napoli R. Characterisation of the aggregated load patterns for extra-urban residential customer groups. Proceedings of the IEEE Melecon 2004, Dubrovnik, Croatia, 12–15 May, vol. 3; 2004. p. 951–954.
[10] Petrescu M, Scutariu M. Load diagram characterisation by means of wavelet packet transform. Proceedings of the second Balkan power conference, Belgrade, Yugoslavia, 19–21 June 2002. p. 15–19.
[11] Carpaneto, E, Chicco, G, Napoli, R, Scutariu, M. Customer classification by means of harmonic representation of distinguishing features. Proceedings of the IEEE Bologna Power Tech 2003, Bologna, Italy, June 23–26, 2003. Paper no. 136.
[12] Chen CS, Kang MS, Hwang JC, Huang CW. Synthesis of power system load profiles by class load study. Elect Power Energy Syst 2000;22(5): 325–30.
[13] Pitt BD, Kirschen DS. Application of data mining techniques to load profiling. Proceedings of the IEEE PICA'99, Santa Clara, CA, May 16–21, 1999. p. 131–136.
[14] Chicco G, Napoli R, Piglione F, Scutariu M, Postolache P, Toader C. Load pattern-based classification of electricity customers. IEEE Trans Power Syst 2004;19(2):1232–9.
[15] Chicco, G, Napoli, R, Piglione, F. Application of clustering algorithms and self organising maps to classify electricity customers. Proceedings of the IEEE Bologna Power Tech 2003, Bologna, Italy, June 23–26, 2003. Paper no. 333.
[16] Pao Y-H, Sobajic DJ. Combined use of unsupervised and supervised learning for dynamic security assessment. IEEE Trans Power Syst 1992; 7(2):878–84.
[17] Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1979;PAM-1(2):224–7.