Energy Time Series Forecasting Based on Pattern Sequence Similarity

Francisco Martínez–Álvarez, Alicia Troncoso, José C. Riquelme and Jesús S. Aguilar–Ruiz

Abstract— **This paper presents a new approach to forecast the behavior of time series based on similarity of pattern sequences. First, clustering techniques are used with the aim of grouping and labeling the samples from a dataset. Thus, the prediction of a data point is provided as follows. First, the pattern sequence prior to the day to be predicted is extracted. Then, this sequence is searched in the historical data and the prediction is calculated by averaging all the samples immediately after the matched sequence. The main novelty is that only the labels associated with each pattern are considered to forecast the future behavior of the time series, avoiding the use of real values of the time series until the last step of the prediction process. Results from several energy time series are reported and the performance of the proposed method is compared to that of recently published techniques showing a remarkable improvement in the prediction.**

Keywords— **Time series, forecasting, patterns.**

I. Introduction

The analysis of temporal data and the prediction of future values of time series are among the most important problems that data analysts face in many fields, ranging from finance and economics, to production operations management or telecommunications.

A *forecast* is a prediction of some future event(s). Forecasting problems are often classified as short-term, medium-term and long-term. Short-term forecasting problems involve predicting events only a few time periods (days, weeks, months) into the future. Medium-term forecasts extend from one to two years and long-term forecasting problems can extend beyond that by many years.

Time series data can be defined as a chronological sequence of observations on a variable of interest. Most forecasting problems imply the use of such data whose analysis has traditionally been done by means of classical statistical tools. Nowadays, data mining techniques are acquiring a great relevance due to the large number of samples forming the time series in multiple areas.

A new approach, called Pattern Sequence-based Forecasting (PSF), is here presented in order to forecast time series. This work can be considered a generalization of the algorithm introduced in [33], which is based on nearest neighbors techniques. Nevertheless, the new approach makes predictions using only labels generated by means of clustering techniques. This fact discretizes and simplifies the process of prediction, since during the whole process the PSF algorithm deals with sequences of labels instead of with sets of real values. Despite that a naive use of labels to predict time series was presented in [21], the PSF includes a new methodology to automatize the obtaining of the labels providing rules to assign them to the samples of real values. The sensitivity of the key parameter involved in the selection of the number of underlying patterns is also analyzed in order to study the robustness of the method. The number of labels comprising the pattern sequence, used in each prediction process, is systematically determined in this work.

The PSF algorithm aims to be a general-purpose forecasting procedure. However, electricity-related problems are addressed in this work. To be precise, two major groups of time series are forecasted: electricity prices and electricity demand. These groups belong to three different markets: the Spanish Electricity Market Operator (OMEL), the New York Independent System Operator (NYISO) and the Australia's National Electricity Market (ANEM). Therefore, the overall experimentation consists of six independent time series showing thus the adaptability of the PSF to miscellaneous time series. Moreover, in order to facilitate the comparison of the obtained results, all the data sets analyzed are available on-line [19], [25], [26].

The rest of the paper is organized as follows. Section II presents an exhaustive revision of the state-of-the-art on electricity prices and demand time series forecasting. Section III introduces the proposed methodology and the description of the PSF algorithm, which can be applied to time series of any nature. Section IV shows the results obtained by the PSF approach in electric energy markets of Spain, Australia and New York for the whole year 2006, including measures of the quality of them. In Section V comparisons between the proposed method and other techniques are provided. Finally, Section VI summarizes the main conclusions achieved and gives clues for future work.

II. Related work

The forecasting of energy time series has been widely studied in literature, as it is described in Sections II-A and II-B.

A. Electricity prices time series forecasting

The electric power markets have become competitive markets due to the deregulation carried out in the last years, allowing the participation of all producers, investors, traders or qualified buyers. Thus, the price of the electric-

Francisco Martínez Álvarez, Alicia Troncoso and Jesús S. Aguilar Ruiz are with the Pablo de Olavide University, Seville, Spain (e-mail: fmaralv@upo.es; ali@upo.es; aguilar@upo.es) and José C. Riquelme is with the Department of Computer Science, University of Seville, Seville, Spain (e-mail: riquelme@lsi.us.es). The financial support from the Spanish Ministry of Science and Technology, project TIN2007-68084-C-00, and from the Junta de Andalucía, project P07-TIC-02611, is acknowledged.

ity is determined on the basis of this buying/selling system. Consequently, a will of obtaining optimized bidding strategies has arisen in the electricity-producer companies [29], needing both insight into future electricity prices and assessment of the risk of trusting in predicted prices.

Electricity prices time series presents some peculiarities such as nonconstant mean and variance as well as the presence of outliers that turns the forecasting into a specially difficult task. Due to this fact, the accomplishment of accurate forecasting has motivated research works by many authors nowadays [2], [37].

The authors in [7] used the wavelet transform and autoregressive integrated moving average models (ARIMA) to predict the day-ahead electricity price. Indeed, they first used the wavelet transform to split the available historical data into constitutive series. Then, specific ARIMA models were applied to these series and the forecasts were obtained by applying the inverse wavelet transform to the forecasts of these constitutive series. In [15] ARIMA models, selected by means of Bayesian Information Criteria, were proposed to obtain the forecasts of the prices. In addition, the work analyzed the optimal number of samples used to build the prediction models. Aggarwal et al. [3] divided each day into segments and they applied a multiple linear regression to the original series or the constitutive series obtained by the wavelet transform depending on the segment. Moreover, the regression model used different input variables for each segment.

Equally noticeable was the approach proposed by García et al. [14] in which a forecasting technique based on a generalized autoregressive conditional heteroskedasticity (GARCH) model was presented. Hence, this paper focused on day-ahead forecast of electricity prices with high volatility periods.

Transfer functions models based on past electricity prices and demand were proposed to forecast day-ahead electricity prices by Nogales et al. in [24], but the prices of all 24 hours of the previous day were not known. They used the median as measure due to the presence of outliers and they stated that the model in which the demand was considered presented better forecasts.

Weron et al. [38] presented twelve parametric and semiparametric time series models to predict electricity prices for the next day. Moreover, in this work forecasting intervals were provided and evaluated taking into account the conditional and unconditional coverage. They concluded that the intervals obtained by semiparametric models are better than that of parametric models.

A hybrid model that combined artificial neural networks (ANN) and fuzzy logic was introduced in [4]. As regards the neural network presented, it had a feed-forward architecture and three layers, where the hidden nodes of the proposed fuzzy neural network performed the fuzzification process. Following with this technique, another neural network-based approach was introduced in [5] in which multiple combinations were considered. These combinations consisted of networks with different number of hidden layers, different number of units in each layer and several types of transfer functions. Recently, Pindoriya et al. [28] proposed an artificial neural network in which the output of the hidden layer neurons was based on wavelets that adapted their shape to training data.

A modification of the weighted nearest neighbors (WNN) methodology is proposed in [33]. To be precise, the approach weighted the nearest neighbors in order to improve the prediction accuracy.

The occurrence of outliers (also called spike prices) or prices significantly higher than the expected values is an usual feature found in these time series. With the aim of dealing with this feature, the authors in [41] proposed a data mining framework based on both support vector machines (SVM) and a probability classifier.

Recently, a fuzzy inference system –adopted due to its *transparency and interpretability*–combined with traditional time series methods was proposed for day-ahead electricity price forecasting [18].

B. Electricity demand time series forecasting

The process of forecasting the quantity of electricity required for a specific geographical area during a time period is called load forecasting or demand forecasting. This process is key since current technology allows to store only little amount of electricity in batteries. Therefore, the demand forecasting plays an important role for electricity power suppliers because both excess and insufficient energy production may lead to large costs and significative reduction of benefits.

Load forecasting has been widely studied [31], [37]. The existing procedures are usually divided into two main groups [13]. The first one gathers traditional approaches such as regression, data smoothing techniques or Box and Jenkin's models. Thus, the authors in [27] focussed on the one year-ahead prediction for winter seasons by defining a new Bayesian hierarchical model. They provide the marginal posterior distributions of demand peaks. Also in [8] Bayesian models are used to forecast electricity demand. Moreover, a multiple linear regression model to forecast electricity consumption using some input variables such as the gross domestic product, the price of electricity and the population was proposed in [23].

Taylor et al. [32] compared six univariate time series methods to forecast electricity load for Rio de Janeiro and England and Wales markets. These methods were an ARIMA model and an exponential smoothing (both for double seasonality), an artificial neural network, a regression model with a previous principal component analysis and two naive approaches as reference methods. The best method was the proposed exponential smoothing and the regression model showed a good performance for the England and Wales demand.

With reference to the second main group, it gathers artificial intelligence techniques among which expert systems, neural networks and fuzzy theory are the most popular [22]. In [11], the authors discussed and presented results by using an ANN to forecast the Jordanian electricity demand, which is trained by a particle swarm optimization

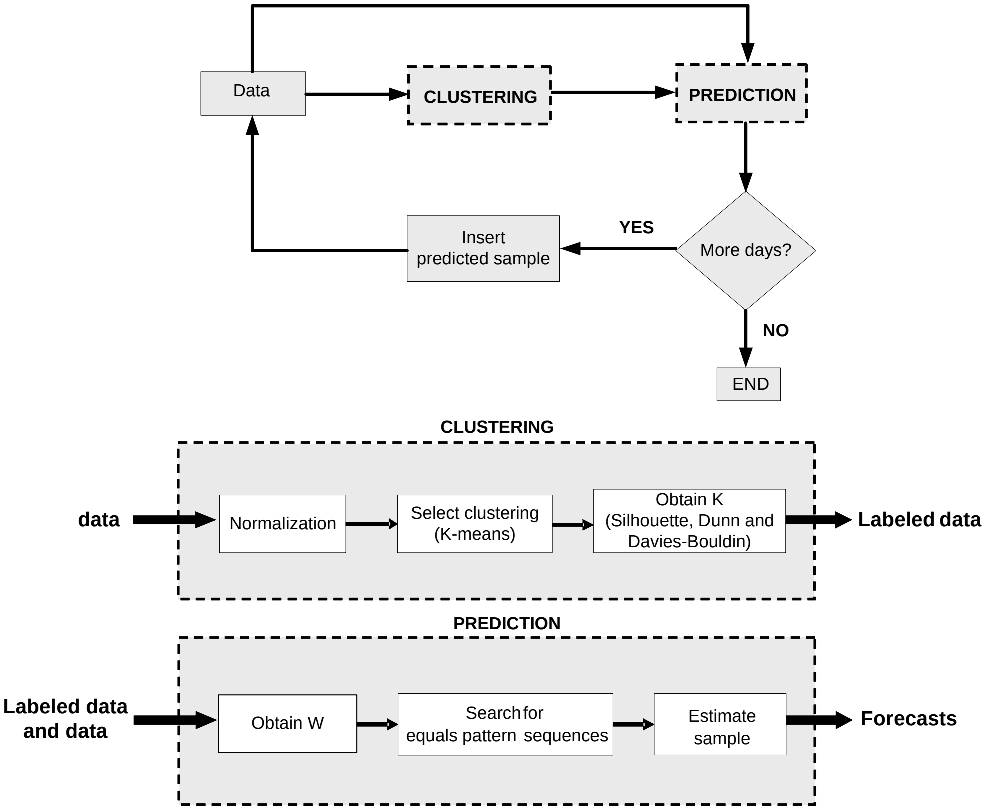


Fig. 1. Illustration of the proposed methodology. The clustering and prediction stages are further detailed.

technique. They also showed the performance obtained by using a back-propagation algorithm and autoregressive moving average (ARMA) models. An ANN-based forecasting technique can also be found in [30]. Another proposal can be found in [36], where a forecasting algorithm based on Grey Models was introduced to predict the load of Shanghai. In the Grey model the original data series was transformed to reduce the noise of the data series and the accuracy was improved by using Markov chains techniques. Fan et al. [12] proposed a hybrid machine learning model based on Bayesian classifiers and SVM. First, Bayesian clustering techniques were used to split the input data into 24 subsets. Then, SVM methods were applied to each subset to obtain the forecasts of the hourly electricity load. In [34], the authors proposed a methodology based on WNN techniques. The proposed approach was applied to the 24-hour load forecasting problem and they built an alternative model by means of a conventional dynamic regression technique to perform a comparative analysis. In [1] the performance of ANN, fuzzy networks and ARIMA models was evaluated to forecast the electricity demand time series in Victoria and the results showed that the fuzzy neural network outperformed the plain ANN and ARIMA models. Finally, [35] proposed a new prediction approach based on SVM techniques with a previous selection of features from data sets by using an evolutionary method. The creation of hybrids methods that highlight most of the strengths of each technique is currently the most popular work among the researchers. And, from all hybrids methods, the combination of ANN and fuzzy set theory has become a new tool to be explored.

III. The proposed methodology

The proposed methodology is divided into two phases clearly differentiated. In a first step, a clustering technique is performed and, secondly, the phase of forecasting is applied by using the information provided by this clustering. The PSF algorithm is focused on predicting samples framed in a time series, either one-dimensional or multidimensional, previously labeled with clustering techniques. As soon as the clustering is applied, the algorithm only processes the number of the cluster –the label associated with each pattern– assigned to the samples, ignoring if they had more than one feature.

With the PSF method, the horizon of prediction can be as long as desired. Hence, more than one sample can be predicted, making predictions of non-restricted length. This fact is possible because it is implemented with a close loop that feeds the prediction of a sample back in the data set in order to predict the following sample. As a consequence, the PSF approach is able to insert the predicted samples in the data set with the aim of forecasting further samples. Therefore, in case the horizon of prediction was longer than one day, every predicted sample would be inserted into the data set and considered to be a regular sample. This feature is specially useful when the prediction has to cover various days or a long-term prediction is required.

Fig. 1 shows the basic idea behind the proposed methodology. All the steps composing this methodology are going to be described in subsequent subsections.

A. Data normalization

The first task to be completed is the normalization of data that is only used for the clustering process. It can be assumed that the prices increase all along the year following a tendency in accordance with the intra-annual inflation. That is, the original trend is smoothed from the initial data. The transformation applied is,

$$x_j \leftarrow \frac{x_j}{\frac{1}{N}\sum_{i=1}^{N} x_i} \tag{1}$$

where x_j is the price/demand of the $j-th$ hour of a day and N is equal to 24 since each value represents one hour of the day.

B. Clustering technique

Given the database of hourly prices/demand, the clustering problem consists of identifying K groups or clusters such that the prices/demand curves of the days belonging to a cluster are similar among them and dissimilar to the curves of those days belonging to other clusters, according to a distance. Clustering is a difficult task due to the great number of possible geometric shapes for the clusters and distances that can be considered.

As a consequence, the dimensionality of the database is drastically reduced from its initial 24 features (equivalent to the 24 hours of the day) to only one dimension (the label of the cluster to which the day belongs).

To achieve this challenge, two questions should be answered: which clustering technique should be chosen? And, if appropriate, how many clusters should be created?

These two topics have widely been discussed in the literature [39]. Nevertheless, it seems that there is not an unique answer because it depends on sensitive factors.

Crisp or fuzzy clustering are the two main branches of non-supervised classification. The discussion of choosing one technique or another can be found in [20], in which the well-known K-means algorithm was the optimal method to classify this kind of data set. For this reason, the K-means algorithm is the clustering technique used in this work during the whole process of prediction.

The K-means algorithm requires that the user provides the number of clusters to be created. However, this number is a priori unknown and its selection and later evaluations of the results obtained by the clustering are crucial for most engineering applications. Thus, the most challenging problem of the clustering realm has been to select the right number of clusters for data sets.

For all these reasons, three well-known validity indices have been applied to data in order to decide how many groups the original data set has to be split into: silhouette index [16], Davies-Bouldin index [9] and the Dunn index [10]. The three of them share a common feature: the new data structure obtained by the clustering algorithm is evaluated to test the validity of the partition.

The procedure to select the number of clusters to be generated, K, is now discussed. From the application of these three indices (see subsections III-B.1, III-B.2, III-B.3), two possible situations can appear: at least two indices –the majority– coincide in selecting the same K (the K eventually chosen) or none of them coincide. When the second situation occurs, the second best values of the three indices are also considered (together with the first best values). The K selected is, then, the one pointed by the majority of all the cases. Further best values will be included and analyzed until one K had more votes than the others.

B.1 The silhouette index

The *silhouette function* provides a quality measure of separation among the clusters obtained by using a clustering technique. The average distance of the object i belonging to the cluster A to all the objects in A is denoted by $a(i)$ and the average distance of i to all objects of the cluster $C \neq A$ is called $d(i,C)$. For every cluster $C \neq A$, $d(i,C)$ is computed and the smallest one is selected as follows,

$$b(i) = \underset{C \neq A}{min} \quad d(i,C) \text{ with } i \in A \tag{2}$$

The value $b(i)$ represents the dissimilarity of the object i to its nearest neighbor cluster. Thus, the silhouette values, $silh(i)$ are given by the following equation,

$$silh(i) = \frac{a(i) - b(i)}{max\{a(i), b(i)\}} \tag{3}$$

The $silh(i)$ can range from -1 to $+1$, where $+1$ and -1 means that the object i belongs to an adequate or inadequate cluster, respectively. If the silhouette value of the object i belonging to the cluster A is close to zero, it means that the object i can also be in the nearest neighbor cluster to A. If cluster A is a set with only one element, the silhouette value of the object i is not defined and in this case, it is concerted be equal to zero. The objective function is the average of $silh(i)$ over the number of objects to be classified, and the best clustering is reached when this function is maximized.

B.2 The Dunn index

One of the most cited indices was proposed in [10]. The Dunn index (DU) aims to identify clusters with high inter-cluster distance and low intra-cluster distance. The Dunn index for K clusters C_i with $i = 1, ..., K$ is defined by,

$$DU_K = \underset{i}{min} \quad \underset{j \neq i}{min} \quad f_{i,j} \tag{4}$$

where

$$f_{i,j} = \frac{d(C_i, C_j)}{\underset{m}{max} \quad diam(C_m)} \tag{5}$$

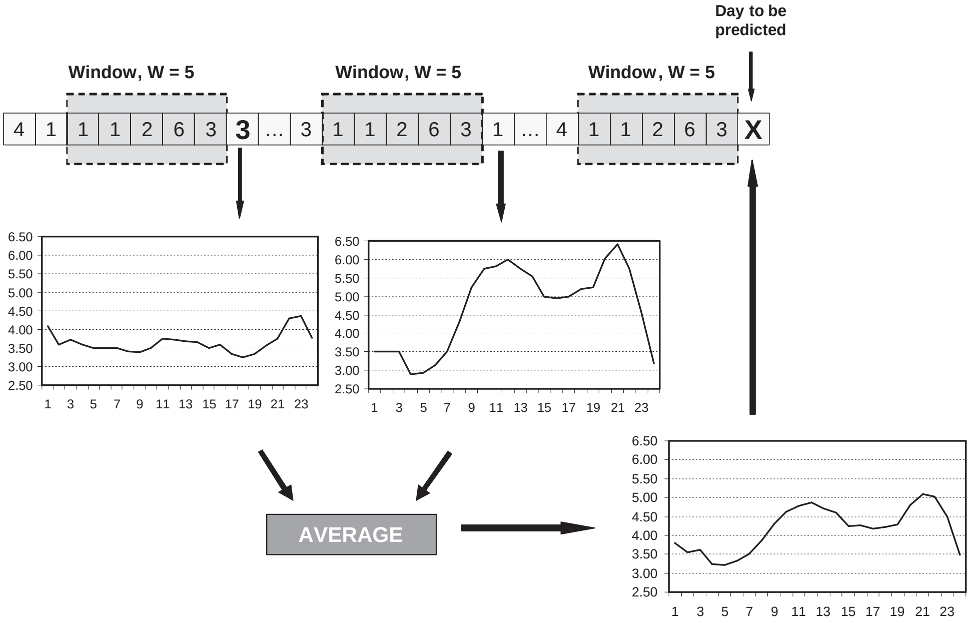


Fig. 2. PSF algorithm.

$d(C_i, C_j)$ is the dissimilarity between clusters C_i and C_j defined by,

$$d(C_i, C_j) = \underset{\substack{x \in C_i \\ y \in C_j}}{min} \|x - y\| \tag{6}$$

and $diam(C)$ is the intra-cluster function or diameter of the cluster defined by this equation,

$$diam(C) = \underset{x, y \in C}{max} \|x - y\| \tag{7}$$

where $\|\cdot\|$ represents a norm.

In short, the existence of compact and well separated clusters is guaranteed if the Dunn index reaches high values. Therefore, the maximum is observed for the most probable number of clusters in the dataset.

B.3 The Davies-Bouldin index

The Davies-Bouldin index identifies as good clusters those compact clusters which are far from each other. Davies-Bouldin index (DB) for K clusters C_i with $i = 1, ..., K$ is defined according to,

$$DB_K = \frac{1}{K} \sum_{i=1}^{K} \underset{j \neq i}{max} \quad f_{i,j} \tag{8}$$

where

$$f_{i,j} = \frac{diam(C_i) + diam(C_j)}{d(C_i, C_j)} \tag{9}$$

and, in this case, the diameter of a cluster is defined as,

$$diam(C_i) = \left(\frac{1}{n_i} \sum_{x \in C_i} \|x - z_i\|^2 \right)^{\frac{1}{2}} \tag{10}$$

with n_i the number of points and z_i the centroid of cluster C_i.

The existence of high–quality clusters is guaranteed if the Davies-Bouldin index reaches small values. Therefore, the optimal number of clusters is found when this index is minimized for the dataset.

C. The PSF algorithm

Given the hourly prices/demand recorded in the past, up to day $d-1$, the forecasting problem aims at predicting the 24 hourly prices/demand corresponding to day d.

Let $X(i) \in \mathbb{R}^{24}$ be a vector composed of the 24 hourly energy prices/demand corresponding to a certain day i

$$X(i) = [x_1, x_2, \ldots, x_{24}]\,. \tag{11}$$

Let $L_i \in \{1, ..., K\}$ be the label of the prices/demand of the day i obtained as a previous step to the forecasting by using a clustering technique, where K is the number of clusters. Let S_W^i be the sequence of labels of the prices/demand of the W consecutive days, from day i backward, as follows,

$$S_W^i = [L_{i-W+1}, L_{i-W+2}, \ldots, L_{i-1}, L_i] \tag{12}$$

where the length of the window, W, is a parameter to be determined (see Section III-D).

The PSF algorithm for the prediction of the hourly prices/demand of the day d first searches for the sequences of labels which are exactly equals to S_W^{d-1} in the database, providing the equal subsequences set, ES_d, defined by this equation,

$$ES_d = \left\{ j \text{ such that } S_W^j = S_W^{d-1} \right\} \tag{13}$$

In case of finding no sequences in database equal to S_W^{d-1}, the procedure searches for the sequences of labels which are exactly equals to S_{W-1}^{d-1} and thus successively. That is, the length of the window composed of the sequence of labels is decreased in one unit. This strategy guaranties that at least some sequences will be found when W is equal to one.

According to the PSF approach, the 24 hourly values of the time series for the day d are predicted by averaging the values of the days following those in ES_d,

$$\widehat{X}(d) = \frac{1}{\text{size}(ES_d)} \cdot \sum_{j \in ES_d} X(j+1) \tag{14}$$

where size(ES_d) is the number of elements that belong to the set ES_d.

The full procedure of the PSF algorithm is detailed in Fig. 2 and a general scheme is presented in Fig. 3. The symbol $\triangleright$ stands for "*append*" (insert at the end).

In case of a medium or long-term prediction, in which the forecasting of more than one sample is required, the following tasks have to be carried out. First of all, the values of the predicted sample are linked to the whole data set. Second, the clustering process is repeated with the enlarged data set and, finally, the prediction step is performed (see Fig. 1).

Input: Dataset D, number of clusters K, labeled dataset $[L_1, L_2, ..., L_{d-2}, L_{d-1}]$, length of the window W and Test Set T
Output: Forecasts $\widehat{X}(d)$ for all days of T

```
PSF()
  $ES_d \leftarrow \{\}$
  $\widehat{X}(d) \leftarrow 0$
  for each day $d \in T$
      $S_W^{d-1} \leftarrow [L_{d-W}, L_{d-W+1}, \dots, L_{d-2}, L_{d-1}]$
      for each $j$ such as $X(j) \in D$
          $S_W^j \leftarrow [L_{j-W+1}, L_{j-W+2}, \dots, L_{j-1}, L_j]$
          if($S_W^j = S_W^{d-1}$)
              $ES_d \leftarrow ES_d \bigcup j$
      for each $j \in ES_d$
          $\widehat{X}(d) \leftarrow \widehat{X}(d) + X(j+1)$
      $\widehat{X}(d) \leftarrow \widehat{X}(d)/size(ES_d)$
      $D \leftarrow D \triangleright \widehat{X}(d)$
      $[L_1, L_2, ..., L_{d-1}, L_d] \leftarrow$ clustering(D,K)
      $d \leftarrow d+1$
  return $\widehat{X}(d)$ for all days of $T$
```

Fig. 3. A general scheme of the algorithm PSF.

D. Determining the size of the window

The previous clustering generates a sequence of labels associated with every day (in Fig. 2 the sequence of numbers are these labels). Now, a sequence of labels is taken into consideration for further steps; concretely, if the day d has to be predicted, the sequence of labels $S_W^{d-1} = [L_{d-W}, L_{d-W+1}, \dots, L_{d-2}, L_{d-1}]$ is extracted from the data set and is used as a pattern of search, where W is the length of this sequence (or window).

The selection of W depends on the case under study but it can be systematically tuned. Thus, it is compulsory to perform a training phase to find an adequate value of W before applying the PSF approach.

The optimal number of labels comprising the window (parameter W) –that will be used as a pattern of search to find all equal sequences of labels in dataset– is determined by minimizing the forecasting error when the PSF method is applied to a training set.

Mathematically, that means to find the value of W that minimizes the following function:

$$\sum_{d \in TS} ||\widehat{X}(d) - X(d)|| \tag{15}$$

where $\widehat{X}(d)$ are forecasted prices/demand for day d, according to the PSF method, $X(d)$ are actual recorded prices/demand and TS refers to the training set. Notice that, according to (14), $\widehat{X}(d)$ is an implicit function of the discrete variable W. Hence, the application of standard mathematical programming methods is not possible when searching for W.

In practice, W is calculated by means of cross-validation. The cross-validation was originally defined as: "the statistical practice of partitioning a sample of data into subsets such that the analysis is initially performed on a single subset, while the other subsets are retained for subsequent use in confirming and validating the initial analysis" [17].

In this work, the $n-$fold cross-validation is used to obtain the optimal value of W. In $n-$fold cross-validation, the original dataset is split into n subsets. From all the n subsets, one subset is used to validate the model, which is generated by the remaining $n-1$ subsets. Thus, this process is repeated n times, using each of the n subsets exactly once to validate. The n results are then combined – usually averaged– in order to generate the final estimation. The advantage of this method lies on the use of all samples for both training and validation.

For the training phase, twelve folds have been used in this work ($n = 12$), where each fold represents a month of the year under study. The $12-$fold cross-validation is then evaluated. The forecasting errors are calculated in every fold by varying the length of W. These monthly errors are denoted by $e_{month}\{W = j\}$ for $j = 1 \dots W_{max}$, where $W_{max} = 10$ –as empirically is shown in Section IV-B.2. Then, the average errors are calculated for each window size as follows,

$$\mathbf{e}_j = \frac{1}{n} \sum_{i=1}^{n} e_{month}\{W = j\} \tag{16}$$

where $n = 12$ and $month = \{Jan, \dots, Dec\}$.

The W selected is the one that minimizes the average error corresponding to the 12 folds (months) evaluated.

$$W = \arg\min\{\mathbf{e}_j\} \text{ with } j = 1, ..., W_{max} \tag{17}$$

IV. Results

The above described methodology has been applied to the electricity prices and demand of Spanish [25], Australian [19] and New York [26] markets. These six data sets have been selected due to the great amount of forecasting results published in the literature. These results will be used to establish a comparison with that of the proposed method in this work.

This section is structured as follows. First, the accuracy of the predictions is validated. Thus, the usual quality parameters are presented. Second, the PSF approach is trained in order to produce accurate predictions and, for this reason, the election of both W and K is discussed here. Third, the prediction of the year 2006 is provided. Finally, a sensitivity analysis of the proposed method with regard to the number of clusters is presented.

A. Parameters of quality.

In order to assess the performance of the PSF approach, several measures have been considered:

- Mean error relative to $\bar{x}$ (MER).

$$MER = 100 \cdot \frac{1}{N} \sum_{h=1}^{N} \frac{\mid \hat{x}_h - x_h \mid}{\bar{x}} \tag{18}$$

where $\hat{x}_h$ and x_h are the predicted and current prices/demand at hour h respectively, $\bar{x}$ is the mean price/demand for the period of interest (a day or a week in this work) and N is the number of predicted hours. Note that, the mean price/demand is used in the denominator of (18) to avoid the effect of prices close to zero.

- Mean absolute error (MAE)

$$MAE = \frac{1}{N} \sum_{h=1}^{N} \mid \hat{x}_h - x_h \mid \tag{19}$$

- Standard deviation of MER/MAE (σ).

$$\sigma = \sqrt{\frac{1}{N} \sum_{h=1}^{N} (e_h - \bar{e})^2} \tag{20}$$

where

$$e_h = \frac{\hat{x}_h - x_h}{\bar{x}} \tag{21}$$

and $\bar{e}$ is the mean of the hourly errors

B. Training the PSF algorithm

In this subsection the number of clusters to be generated, as well as the length of the window comprising the sequence of labels that has to be searched along the time series, are presented.

B.1 Selecting the number of clusters

First of all, the number of clusters K has to be chosen and, for this purpose, the twelve months of the year 2005 are considered for training the algorithm.

In order to validate the quality of the clusters produced by K-means algorithm, the silhouette, Dunn and Davies-Bouldinthree indices have been used in the experiments. Thus, the optimal value of K is selected from a system based on majority votes.

TABLE I

Number of clusters selected for all the markets.

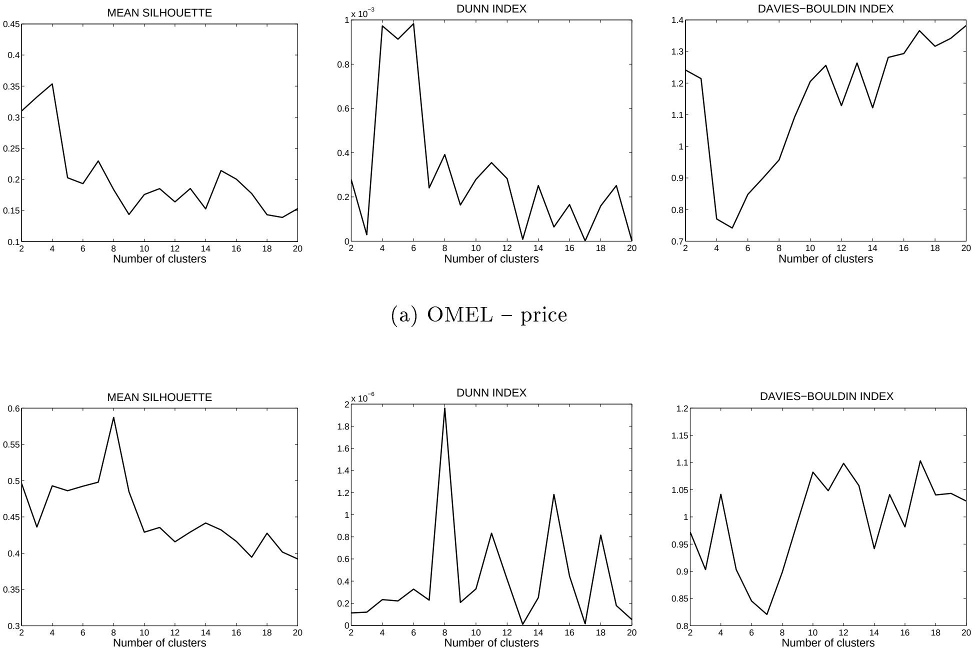
	Market	Silhouette	DU	DB	Selection
prices	OMEL	4 (3)	6 (4)	5 (4)	4
	NYISO	5	5	5	5
	ANEM	3	4	3	3
demand	OMEL	8	8	7	8
	NYISO	4	4	3	4
	ANEM	5	5	5	5

Table I summarizes the results obtained by the three indices of cluster validity, in which the numbers in brackets represent the second best values, as described in Section III-B. A further study about the influence of the number of clusters on the results of the prediction is made in the section IV-D.

The values of the indices when K varies from 2 to 20 are depicted in Figs. 4, 5, and 6. Fig. 4(a) shows the results of all the three indices when varying K in the Spanish electricity price market. Apparently, all of them have different optimum values since the silhouette and the Dunn indices reach the maximum values in $K = 4$ and $K = 6$ (0.3536 and 0.0010 respectively), while the Davies-Bouldin index reaches its optimum value when $K = 5$ (0.7417). Nevertheless, a thorough analysis of all the values reveals that for $K = 4$ both Dunn and Davies-Bouldin indices have the second best result (0.0010 and 0.7703), with values really close to the optimum values. For these reasons, the number of clusters selected for this time series is $K = 4$. On the other hand, Fig. 4(b) illustrates the results of the Spanish demand market. As it can be appreciated, both silhouette and Dunn indices select $K = 8$ –reaching the maximum values in 0.5872 and 1.964E-06, respectively. On the contrary, the Davies-Bouldin index reaches its minimum and, consequently, its optimum value when $K = 7$. However, when $K = 8$ this index also presents a low value close to its global minimum (0.8983 versus 0.8208, respectively).

With reference to the New York electricity prices, the situation shown in Fig. 5(a) reveals an easy selection of the number of clusters. Actually, the three methods select $K = 5$ since both silhouette and Dunn indices reach the maximum values (0.5324 and 0.0004, respectively) while the Davies-Bouldin index reaches its minimum value (0.6602). The selection of K for the New York electricity demand time series is shown in Fig. 5(b). It can be noticed, both silhouette and Dunn indices reach the optimum values in $K = 4$ (0.7895 and 0.0002, respectively), but the Davies-Bouldin index reaches its minimum when $K = 3$ (0.8193). Even if the value in $K = 4$ is not specially low (0.9739), it can be considered a globally low value since this index reaches values such as 1.3125 in $K = 6$.

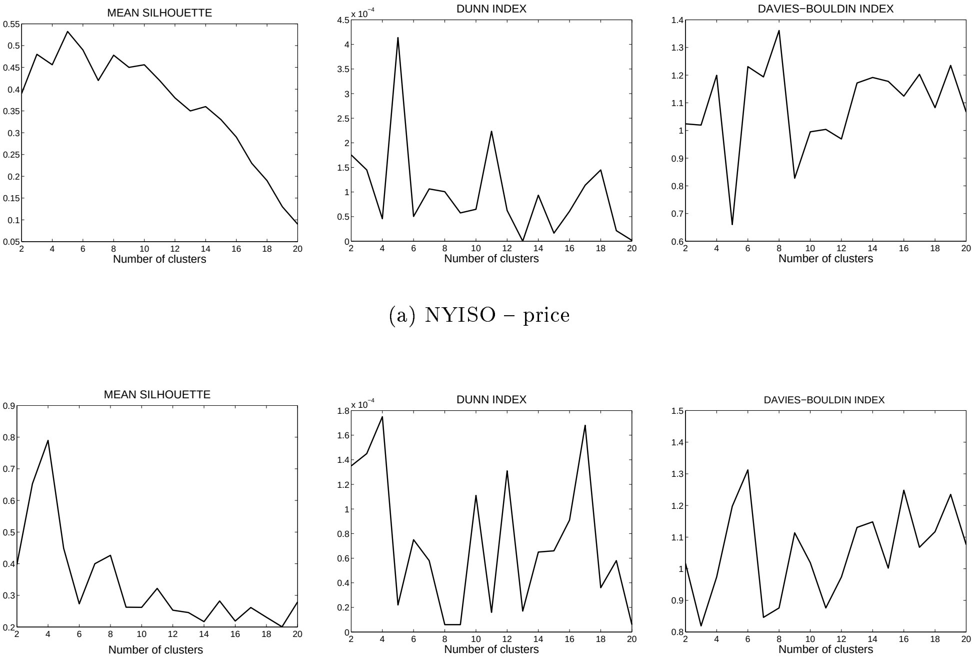
Fig. 6(a) shows the results obtained in the Australian electricity price market. As it can be noticed, both silhouette and Davies-Bouldin indices reach the maximum and minimum values (0.4858 and 0.6931 respectively) in $K = 3$. Oppositely, the Dunn index reaches its maximum and, consequently, its optimum value when $K = 4$. However, when $K = 3$ this index also presents a high value verging



(a) OMEL – price

(b) OMEL – demand

Fig. 4. Selecting the optimal number of clusters in OMEL time series.

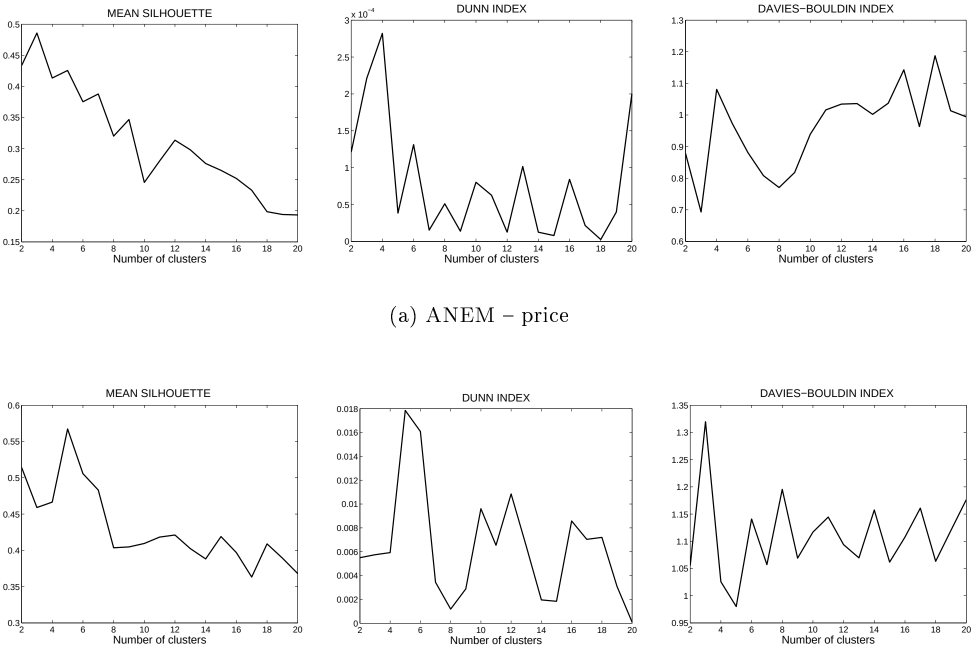


(a) NYISO – price

(b) NYISO – demand

Fig. 5. Selecting the optimal number of clusters in NYISO time series.

on its global maximum (0.0003 and 0.0002 respectively). With regard to the Australian electricity demand, the situation shown in Fig. 6(b) is conclusive. Hence, the three methods agree in selecting $K = 5$ since both silhouette and Dunn index reach the maximum values (0.5673 and 0.0178 respectively) while the Davies-Bouldin index reaches its minimum value (0.9800).



(a) ANEM – price

(b) ANEM – demand

Fig. 6. Selecting the optimal number of clusters in ANEM time series.

B.2 Selecting the length of the window

Once the number of clusters is already decided, the next step is to select the optimal length of the window W. Thus, this step is focused on finding the W that obtains the minimum prediction error in the training set.

Therefore, it is required to evaluate the performance of the PSF algorithm when W varies according to the methodology presented in Section III-D.

Table II shows how the prediction error varies in accordance with the number of patterns considered in the window. Note that the symbol '–' means that similar sequences of length W were not found when K clusters were considered in the training set. Finally, the W that allows a lower prediction error is the value chosen for further forecasting on real data. It can be concluded that the optimal lengths of the windows that have to be used are $W = 5$, $W = 3$ and $W = 6$ for the OMEL, NYISO and ANEM price time series since they reach the lower prediction errors (2.23%, 3.27% and 5.81%, respectively) and $W = 2$, $W = 5$ and $W = 3$ for the OMEL, NYISO and ANEM demand time series (2.87%, 4.99% and 3.43%, respectively).

C. Forecasting results

In this subsection the results obtained from the three different markets are provided. Precisely, Tables III, IV and V show the MER and the MAE (and the standard deviations σ in brackets) in the Spanish, Australian and New York electricity time series –both for prices and demand– for the whole year 2006. In spite of the average of the MER for the year 2006 in the OMEL prices time series is greater than that corresponding to NYISO (6.15% versus 5.53%), it can be noticed that the mean of standard deviation of MER in the Spanish market is lower than that of the New York market (0.27% versus 1.94%). This fact means that the maximum errors in the OMEL prices time series are closer to the average errors than the maximum errors obtained when the prices are predicted in the NYISO market. The standard deviation for the Australian market is the highest (4.40%) due to the many peak prices –considered as outliers– that occur in this prices time series.

TABLE III

Performance of the PSF algorithm for the year 2006 in OMEL time series.

	PRICES		DEMAND	
Month	MER (σ)	MAE (σ)	MER (σ)	MAE (σ) in MW
Jan.	7.26% (0.25)	0.53 (0.07)	3.12% (1.86)	744.32 (82.12)
Feb.	4.93% (0.19)	0.36 (0.04)	4.21% (2.26)	1033.97 (109.21)
Mar.	5.88% (0.22)	0.43 (0.05)	5.07% (4.17)	1001.71 (98.85)
Apr.	3.62% (0.18)	0.28 (0.03)	4.18% (1.28)	1006.30 (107.58)
May	8.11% (0.21)	0.64 (0.05)	5.90% (2.33)	1129.76 (96.37)
Jun.	3.76% (0.24)	0.29 (0.05)	2.89% (1.81)	693.60 (84.60)
Jul.	4.30% (0.23)	0.33 (0.04)	2.34% (1.19)	585.88 (63.17)
Aug.	5.37% (0.34)	0.42 (0.06)	3.61% (2.17)	792.21 (86.94)
Sep.	6.41% (0.31)	0.50 (0.06)	3.15% (1.55)	757.02 (75.39)
Oct.	7.89% (0.29)	0.58 (0.08)	2.89% (3.40)	1121.43 (149.75)
Nov.	8.30% (0.40)	0.64 (0.05)	4.72% (2.39)	982.19 (120.52)
Dec.	8.02% (0.36)	0.59 (0.07)	6.21% (3.82)	1503.44 (198.49)
Mean	6.15% (0.27)	0.47 (0.05)	4.02% (2.35)	945.99 (106.08)

Fig. 7 illustrates several prediction curves obtained for the Spanish market for the year 2006. Concretely, Fig. 7(a) shows the best and worst predictions generated by the PSF algorithm when electricity prices curves were considered. With regard to the prices, the best prediction occurred on June 23rd in which the MER was 3.10% and the MAE 0.12cE/KWHr, while the worst took place on May 8th in which the MER was 9.39% and the MAE 0.80cE/KWHr. Note that these curves are expressed in cents of Euro per

TABLE II

MER OBTAINED WITH THE PSF ALGORITHM ON THE ALL THE MARKETS.

Market	W=1	W=2	W=3	W=4	W=5	W=6	W=7	W=8	W=9	W=10	Selected W
OMEL – price ($K = 4$)	10.32%	8.44%	8.21%	4.39%	**2.23%**	2.89%	–	–	–	–	5
OMEL – demand ($K = 8$)	3.11%	**2.87%**	–	–	–	–	–	–	–	–	2
NYISO – price ($K = 5$)	7.09%	5.98%	**3.27%**	6.98%	4.45%	13.20%	10.31%	–	–	–	3
NYISO – demand ($K = 4$)	5.16%	6.21%	5.68%	5.02%	**4.99%**	6.23%	7.14%	6.90%	8.91%	–	5
ANEM – price ($K = 3$)	9.58%	7.91%	6.26%	6.17%	7.33%	**5.81%**	6.04%	9.12%	–	–	6
ANEM – demand ($K = 5$)	3.45%	4.17%	**3.43%**	6.10%	5.89%	4.02%	7.11%	–	–	–	3

TABLE IV

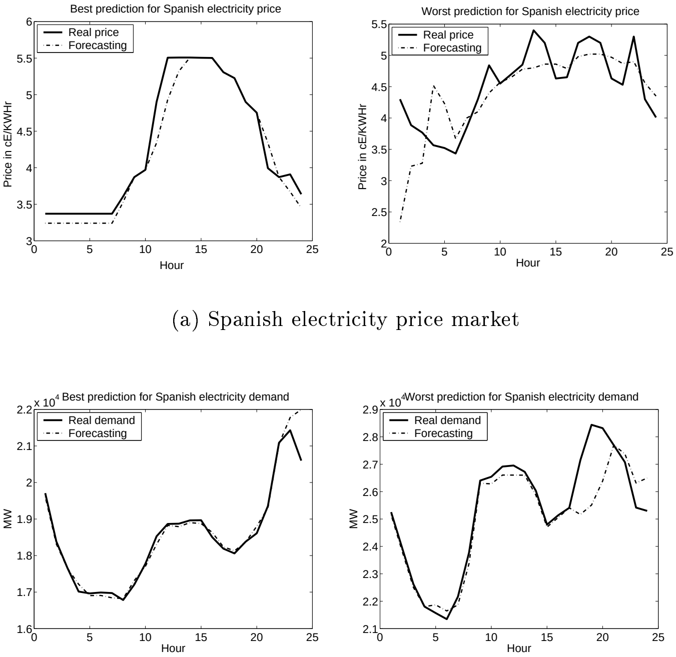
PERFORMANCE OF THE PSF ALGORITHM FOR THE YEAR 2006 IN NYISO TIME SERIES.

	PRICES		DEMAND	
Month	MER (σ)	MAE (σ)	MER (σ)	MAE (σ) in MW
Jan.	4.45% (2.07)	2.25 (0.34)	5.05% (1.95)	53.21 (6.03)
Feb.	5.53% (1.52)	3.02 (0.28)	6.88% (2.62)	83.76 (9.19)
Mar.	6.30% (2.52)	3.97 (0.43)	5.31% (2.42)	59.63 (6.13)
Apr.	4.94% (1.47)	3.51 (0.61)	4.97% (2.22)	52.18 (7.21)
May	7.59% (2.13)	4.63 (0.43)	6.18% (2.39)	61.12 (5.74)
Jun.	3.34% (1.92)	2.31 (0.29)	3.75% (2.06)	44.17 (4.86)
Jul.	3.93% (1.68)	2.28 (0.20)	3.41% (1.78)	37.54 (4.01)
Aug.	5.37% (1.87)	3.49 (0.41)	3.99% (2.13)	39.86 (5.52)
Sep.	6.24% (1.74)	4.49 (0.53)	4.83% (2.16)	54.14 (6.87)
Oct.	7.43% (2.33)	4.23 (0.49)	5.37% (2.25)	65.08 (8.01)
Nov.	5.19% (2.09)	3.53 (0.30)	4.86% (1.99)	50.25 (5.11)
Dec.	6.04% (1.99)	3.08 (0.33)	6.80% (2.40)	82.55 (9.97)
Mean	5.53% (1.94)	3.40 (0.39)	5.97% (2.20)	56.96 (6.55)

TABLE V

PERFORMANCE OF THE PSF ALGORITHM FOR THE YEAR 2006 IN ANEM TIME SERIES.

	PRICES		DEMAND	
Month	MER (σ)	MAE (σ)	MER (σ)	MAE (σ) in MW
Jan.	5.58% (1.34)	1.51 (0.52)	4.74% (3.54)	412.38 (58.02)
Feb.	8.59% (3.24)	5.15 (2.21)	4.98% (2.98)	445.12 (43.29)
Mar.	7.84% (2.98)	1.73 (0.34)	5.02% (5.27)	430.00 (38.90)
Apr.	9.92% (3.90)	1.98 (0.63)	6.03% (7.46)	519.73 (57.71)
May	12.85% (4.03)	3.21 (1.02)	4.17% (2.72)	373.22 (43.28)
Jun.	22.04% (12.34)	6.81 (2.89)	5.67% (3.84)	561.44 (60.32)
Jul.	17.11% (10.58)	8.16 (3.42)	4.91% (5.84)	481.18 (53.67)
Aug.	11.71% (5.08)	3.32 (0.40)	5.88% (6.01)	555.07 (65.05)
Sep.	8.23% (2.45)	2.34 (0.23)	3.99% (2.74)	350.88 (32.41)
Oct.	7.66% (2.89)	1.92 (0.11)	4.04% (3.34)	340.73 (48.91)
Nov.	6.76% (1.94)	2.09 (0.34)	6.12% (5.90)	504.16 (70.42)
Dec.	6.42% (2.01)	1.41 (0.28)	3.91% (3.22)	329.26 (33.81)
Mean	10.39% (4.40)	3.30 (1.03)	4.96% (4.41)	441.93 (50.48)



(a) Spanish electricity price market

(b) Spanish electricity demand

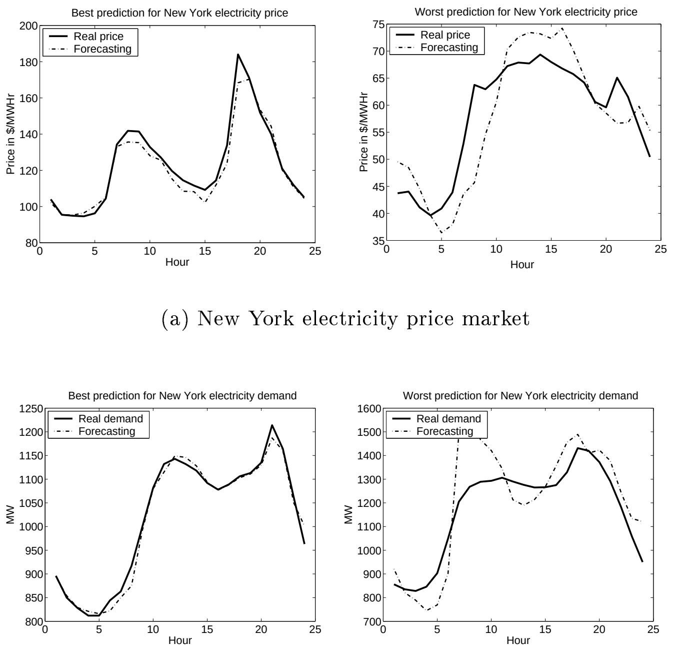
Fig. 7. Best and worst predictions for the Spanish electricity market in 2006.

kilowatts per hour (cE/KWHr). Fig. 7(b) presents the best and worst predictions when the electricity demand was analyzed. The best one took place on May 16^{th} in which the MER was 1.16% and the MAE 253.49MW. On the other side, the worst prediction had a MER of 8.67% and a MAE equal to 1759.03MW and it took place on December 12^{th}. Note that these curves are expressed in megawatts (MW).

The results obtained for the New York market are illustrated in Fig. 8. Fig. 8(a) presents the best and worst prediction curves obtained for the New York electricity prices, which took place on July 8^{th} (with a MER of 2.76% and a MAE of 1.41$/MWHr) and on May 12^{th} (its MER was 8.89% and the MAE was equal to 6.89$/MWHr), respectively. Note that these curves are expressed in dollars per MWHr ($/MWHr). Alternatively, Fig. 8(b) references to the best –December 10^{th} with a MER and MAE equals to 2.67% and 28.47MW, respectively– and the worst –February 15^{th} with a MER and MAE equals to 10.56% and 97.89MW, respectively– predictions in the demand time series.

With respect to the Australian market, it is important to remark that it shows the information structured in different areas. Thus, the National Electricity Market in Australia is comprised of five jurisdictions: Queensland, New South Wales, Victoria, Tasmania and South Australia. The results in Table V refers to the Queensland market.

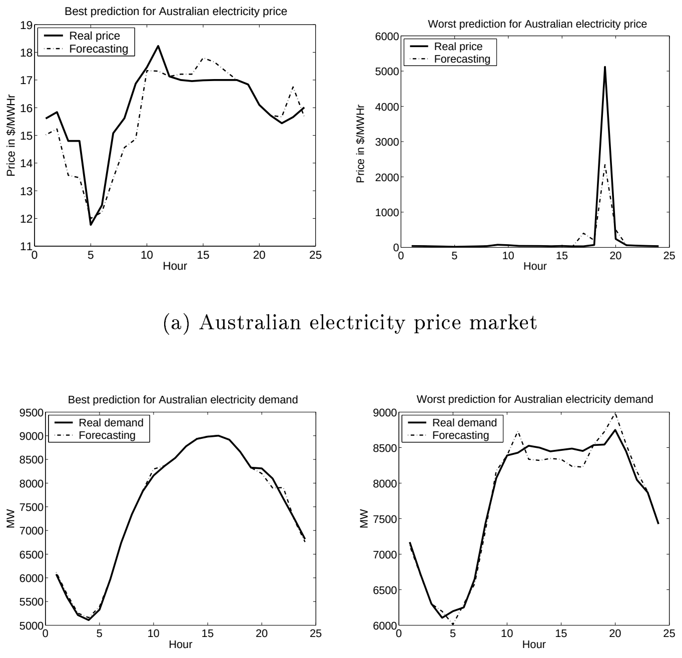
Fig. 9 shows the best and worst prediction curves obtained for the Australian market in the year 2006 for both electricity prices and demand markets. Fig. 9(a) illustrates the best and worst prediction curves obtained for the electricity prices, which took place on May 12^{th} (with associated MER and MAE of 3.66% and 0.98$/MWHr, respectively) and on July 20^{th} (with associated MER and MAE of 65.60% and 28.39$/MWHr, respectively). The Australian electricity price market is characterized by the existence of many spike prices during the year. Indeed, many authors have studied how to perform accurate predictions in that market [41]. The PSF algorithm, even if it is not able to find the real magnitude of such peaks, it is able to forecast the existence of them. This fact justifies the higher value of the MER obtained for that day. It can be observed how the proposed algorithm captures the trend of the prices



(a) New York electricity price market

(b) New York electricity demand

Fig. 8. Best and worst predictions for the New York electricity market in 2006.



(a) Australian electricity price market

(b) Australian electricity demand

Fig. 9. Best and worst predictions for the Australian electricity market in 2006.

time series in the Australian market detecting a peak –an outlier– at 7:00pm. Note that all the curves are expressed in Australian dollars per MWHr ($/MWHr).

Finally, Fig. 9(b) references to the best –December 8^{th} with a MER and MAE equals to 2.67% and 322.82MW, respectively– and the worst –November 19^{th} with a MER and MAE equals to 10.56% and 638.92MW, respectively– predictions in the demand time series.

D. Sensitivity to the parameter K

In this subsection a posteriori analysis of sensitivity to the parameter K is carried out in order to show the good performance of the three indices of cluster validity presented for these six time series and the robustness of the proposed method with regard to this parameter.

Fig. 10 shows the MER provided by the PSF algorithm in 2006 for the prices and demand time series when the number of clusters K ranges from 2 to 15.

From Table I and Fig. 10, it can be stated the following.

Remark 1. The MER is minimum when five clusters ($K = 5$) are considered for both prices and demand in New York and Australia, respectively. All indices –silhouette, DU and DB– coincided in the optimal selection of this parameter for the aforementioned markets.

Remark 2. The indices silhouette and DU selected the same number of clusters for the demand time series of the Spanish and New York markets ($K = 8$ and $K = 4$, respectively) and for the prices time series of the Australian market ($K = 4$). Indeed, the global minima of the MER for these three time series are reached in these values.

Remark 3. The three indices of cluster validity provided different values of the parameter K for the OMEL prices time series ($K = 4$, $K = 6$, and $K = 5$ for silhouette, DU and DB indices, respectively). However, the MER obtained does not present significative differences when the number of clusters is equal to some of these three values. Nevertheless, for this time series, the global minimum is reached in the selected number of clusters from the proposed system based on majority votes.

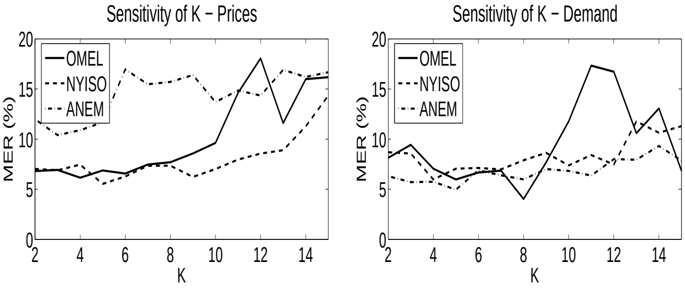


Fig. 10. Sensitivity of the PSF approach to the K parameter.

This analysis highlights the validity of the methodology followed in order to select K, since it reveals that the MER is minimized when the three indices agreed, that the optimality is guaranteed when two of them –a majority– agreed and, finally, that the MER does not vary significantly when all of them are different and K is selected as described in subsection III-B.

TABLE VI
MER for some weeks of the year 2002 (OMEL – price).

Week	Naive	ANN	ARIMA	Mixed models	WNN	**PSF**
18^{th}–24^{th} Feb 2002	7.68%	5.23%	6.32%	6.15%	5.15%	5.98%
20^{th}–26^{th} May 2002	7.27%	6.36%	6.36%	4.46%	4.34%	4.51%
19^{th}–25^{th} Aug 2002	27.30%	11.40%	13.39%	14.90%	10.89%	9.11%
18^{th}–24^{th} Nov 2002	19.98%	13.65%	13.78%	11.68%	11.83%	10.07%
Average	15.56%	9.16%	9.96%	9.30%	8.05%	7.42%

V. Comparative analysis of PSF

A comparison between the results obtained by the PSF method and the most representative approaches reported in the literature is provided in this section, showing that the proposed approach improves the aforementioned techniques. Thus, in order to validate the accuracy of the proposed algorithm, it has been applied to specific periods of time in which other authors evaluated their own approaches.

Furthermore, this section is divided into two subsections. The first one gathers the forecasting results related to the electricity price markets, while the second one points out the enhancements achieved in the electricity demand forecasting with the proposed methodology.

A. Electricity prices time series

A.1 The Spanish electricity prices market

The *Spanish electricity prices market* has been widely analyzed. Many authors have evaluated their own approaches over the time series for the year 2002 and, as a consequence, the literature offers multiple results for this year. The PSF algorithm is compared to four published approaches: ARIMA [7], ANN [5], mixed models [15] and WNN [33]. Finally, it is also compared to the naive Bayes classifier as a reference method.

As it can be observed in Table VI, the proposed method has improved most of the MER rates. However, there are some exceptions, such as for the week of February 18^{th}–24^{th}, in which the ANN obtained an error of 5.23% and 5.15% for the WNN versus the 5.98% provided by the PSF method. The mixed models and the WNN method also obtained lower errors in the week of May 20^{th}–26^{th} (4.46% and 4.34% versus 4.51%, respectively). Apart from these two weeks, the PSF algorithm was much more efficient than the others. The mean errors improved by more than 0.5% the best method compared to (7.42% for the PSF versus 8.05% for the WNN).

The authors in [15] also forecasted a week of the year 2000. The comparative of the MER rates is shown in Table VII. The average of the MER for this week is 5.46% when the PSF method was applied, whereas the mixed models and ARIMA models yield an average of 7.04% and 8.17%, respectively. For this week, the average results are 1.5% better than those obtained by the others methods. Therefore, the improvement reached by the proposed algorithm is considered successful.

TABLE VII
MER for one week of the year 2000 (OMEL – price).

Day	ARIMA	Mixed Models	**PSF**
Day 1	4.30%	4.80%	3.74%
Day 2	7.99%	7.30%	6.91%
Day 3	4.57%	5.40%	3.45%
Day 4	10.81%	4.60%	5.21%
Day 5	6.12%	5.10%	4.48%
Day 6	17.34%	14.90%	9.63%
Day 7	6.05%	7.20%	4.81%
Average	8.17%	7.04%	5.46%

A.2 The New York electricity prices market

As for the *New York electricity prices time series,* the authors in [6] compared some forecasting algorithms with their own approach called STR. They applied manifold-based dimensionality reduction to electricity prices curve modeling. Hence, they showed that it exists a low-dimensional manifold representation for the price curve in the New York electricity market. They compared with an ARIMA model and a naive Bayes as reference method.

Table VIII presents the MER obtained for the one week-ahead electricity price forecasting for each second week for every month of the year 2005. The last row shows the average errors when the horizon of prediction is 24 hours. It can be noticed that the PSF approach provides better predictions in most months. There are just two cases in which the STR overcomes the PSF algorithm: February 2005 and May 2005 (7.65% and 7.53% for STR versus 7.89% and 7.58% for PSF, respectively). Note that in these two cases the MER obtained by the PSF is not significantly high. In addition, when the average error is evaluated, all the approaches obtained worse results than those of the PSF algorithm, which improves 1.5% the result of STR. An increment of 2% approximately can be observed when the horizon of prediction is one week instead of one day.

A.3 The Australian electricity prices market

The prices in the *Australia's National Electricity Market* have also been predicted in [41]. It is remarkable that this market presents an especial behavior since many spike prices are observed. Despite the authors in [41] have developed techniques based on SVM in order to deal with this particular days, the PSF algorithm does not make any assumption about the nature of the days to be predicted, insofar it uses unsupervised learning and, consequently, no a priori information is known about data.

Table IX shows the MER obtained for, precisely, these days of the year 2004 with peak prices. It can be observed

TABLE VIII
MER for the year 2005 (NYISO – price).

Month	ARIMA	Naive	STR	**PSF**
Feb 2005	14.57%	12.19%	7.65%	7.89%
Mar 2005	13.28%	12.33%	10.19%	9.12%
Apr 2005	10.68%	14.59%	10.53%	8.23%
May 2005	14.21%	6.71%	7.53%	7.58%
Jun 2005	21.64%	26.68%	13.88%	8.85%
Jul 2005	14.63%	14.44%	10.41%	9.21%
Aug 2005	9.49%	10.28%	6.42%	5.56%
Sep 2005	10.36%	13.17%	7.31%	6.59%
Oct 2005	11.84%	11.57%	10.11%	8.03%
Nov 2005	11.24%	15.18%	8.99%	7.89%
Dec 2005	21.78%	23.94%	13.30%	11.21%
Jan 2006	26.01%	11.52%	13.28%	10.77%
Average	14.98%	14.38%	9.95%	8.41%
Average (one-day-ahead)	7.39%	16.07%	7.10%	6.11%

TABLE IX
MER for some days of the year 2004 (ANEM – price).

Day (2004)	ARIMA	SVM	**PSF**
5^{th} June	32.31%	18.09%	16.72%
17^{th} June	29.09%	13.31%	8.31%
20^{th} June	33.73%	17.11%	14.23%
21^{st} June	24.18%	19.20%	18.93%
Average	29.82%	16.93%	14.55%

that the proposed method outperforms all the predictions produced by both ARIMA and SVM approaches.

According to [40] four weeks were predicted with different methods: a discrete wavelet transform (DWT), a multi-layer perceptron (MLP) and a SMV approach. Table X presents the MER provided by the PSF method and the aforementioned techniques when the horizon of prediction is one week. The PSF algorithm outperforms the average MER provided by all these methods.

B. Electricity demand time series

B.1 The Spanish electricity demand

In order to compare the performance of the proposed approach in the Spanish electricity demand time series, the results provided in [34] are analyzed.

Table XI shows the comparison between a dynamic regression (DR), a method based on nearest neighbors techniques (kNN) and the PSF algorithm for the period from June to November of the year 2001. As it can be noticed, the proposed algorithm obtained better predictions –not only for MER but also for MAE– when it was compared to the other methods considered in the literature.

Although the kNN had a good performance, the PSF was able to reduce from 2.30% to 1.89%.

TABLE X
MER for some weeks of the year 2004 (ANEM – price).

Week	DWT	MLP	SVM	**PSF**
Second of January	12.94%	25.81%	23.37%	15.62%
First of July	12.23%	8.36%	15.03%	9.12%
First of August	16.17%	15.85%	36.18%	13.98%
Third of December	10.01%	47.41%	33.74%	10.23%
Average	12.84%	24.36%	27.08%	12.23%

TABLE XI
MER, MAE and σ_{MER} for several months of the year 2001 (OMEL – demand).

Parameter	DR	kNN-based	**PSF**
MER	2.82%	2.30%	1.89%
σ_{MER}	0.019	0.015	0.014
MAE (MW)	572	471	454

TABLE XII
MER and MAE for January of the year 2004 (NYISO – demand).

Parameter	NYISO	SVM	MLF	**PSF**
MAE (MW)	214.40	226.87	178.21	176.03
MER	3.16%	3.27%	2.51%	2.39%

B.2 The New York electricity demand

The authors in [12] used a model of machine learning called MLF to predict the electricity demand for the next day. The forecasted period was January of the year 2004. Moreover, two methods were used in order to validate the forecasting: a SVM-based model and the prediction itself provided by the New York Independent System Operator (NYISO).

Table XII shows the results of comparing PSF and the aforementioned methods in the same period. Given the difficulty of predicting demand time series, an improvement of about 5% (2.39% for PSF versus 2.51% for MLF) it can be considered a remarkable enhancement.

B.3 The Australian electricity demand

To compare the results provided by the proposed method for the Australian market the work in [1] was considered. Two days were predicted –October 1^{st} and 2^{nd} of the year 1998– and three methods were used: fuzzy and plain neural networks and an ARIMA model.

As it can be observed in Table XIII, the PSF algorithm improves the prediction error with respect to the other methods, including the very accurate fuzzy neural network.

TABLE XIII
MER for some days of the year 1998 (ANEM – demand).

Parameter	ARIMA	ANN	Fuzzy-ANN	**PSF**
MER	4.23%	3.23%	0.92%	0.90%

VI. Conclusions

In this paper, a new forecasting algorithm has been proposed to predict real-world time series. As previous step to the prediction, a clustering technique to label 24-dimensional time series has been used and the main novelty lies on the exclusive use of the labels obtained by the clustering to forecast the future behavior of the time series, avoiding the use of the real values of the time series until the last step of the prediction process. Moreover, an automatization of the selection of the critical parameters –K and W– has been proposed.

The algorithm has been successfully applied in electricity prices and demand time series of Spanish, Australian and New York markets providing very competitive results. The performance was accurate in all of them, showing thus the robustness and adaptability of the proposed approach for time series of different nature. This fact is specially remarkable since the approaches found in literature are usually focussed on only one specific time series.

Future work is focussed on adjusting the model with dynamical lengths of window and on smoothing the matching sequence criterion.

References

[1] A. Abraham and B. Nath. A neuro-fuzzy approach for forecasting electricity demand in victoria. *Applied Soft Computing Journal*, 1(2):127–138, 2001.

[2] S. K. Aggarwal, L. M. Saini, and A. Kumar. Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power and Energy Systems*, 31(1):13–22, 2009.

[3] S. K. Aggarwal, L. M. Saini, and Ashwani Kumar. Price forecasting using wavelet transform and lse based mixed model in australian electricity market. *International Journal of Energy Sector Management*, 2(4):521–546, 2008.

[4] N. Amjady. Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *IEEE Transactions on Power Systems*, 21(2):887–896, 2006.

[5] J. P. S. Catalao, S. J. P. S. Mariano, V. M. F. Mendes, and L. A. F. M. Ferreira. Short-term electricity prices forecasting in a competitive market: a neural network approach. *Electric Power Systems Research*, 77:1297–1304, 2007.

[6] J. Chen, S. J. Deng, and X. Huo. Electricity price curve modeling by manifold learning. *IEEE Transactions on Power Systems*, 15:723–736, 2007.

[7] A. J. Conejo, M. A. Plazas, R. Espínola, and B. Molina. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Transactions on Power Systems*, 20(2):1035–1042, 2005.

[8] R. Cottet and M. Smith. Bayesian modeling and forecasting of intraday electricity load. *Journal of the American Statistical Association*, 98(464):839–849, 2003.

[9] D. L. Davies and D. W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(4):224–227, 2000.

[10] J. Dunn. Well separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4:95–104, 1974.

[11] M. El-Telbany and F. El-Karmi. Short-term forecasting of jordanian electricity demand using particle swarm optimization. *Electric power systems research*, 78:425–433, 2008.

[12] S. Fan, C. Mao, J. Zhang, and L. Chen. Forecasting electricity demand by hybrid machine learning model. *Lecture Notes in Computer Science*, 4233:952–963, 2006.

[13] E. A. Feinberg and D. Genethliou. *Applied Mathematics for Restructured Electric Power Systems, Chapter 12.* Springer, 2005.

[14] R. C. García, J. Contreras, M. van Akkeren, and J. B. García. A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems*, 20(2):867–874, 2005.

[15] C. García-Martos, J. Rodríguez, and M. J. Sánchez. Mixed models for short-run forecasting of electricity prices: Application for the spanish market. *IEEE Transactions on Power Systems*, 22(2):544–552, 2007.

[16] L. Kaufman and P. J. Rousseeuw. *Finding groups in Data: an Introduction to Cluster Analysis.* Wiley, 1990.

[17] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 1137–1143, 1995.

[18] G. Li, C. C. Liu, C. Mattson, and J. Lawarrée. Day-ahead electricity price forecasting in a grid environment. *IEEE Transactions on Power Systems*, 22(1):266–274, 2007.

[19] Australia's National Electricity Market.
http://www.nemmco.com.au.

[20] F. Martínez-Álvarez, A. Troncoso, J. C. Riquelme, and J. M. Riquelme. Partitioning-clustering techniques applied to the electricity price time series. *Lecture Notes in Computer Science*, 4881:990–999, 2007.

[21] F. Martínez-Álvarez, A. Troncoso, J. C. Riquelme, and J. S. Aguilar Ruiz. LBF: A labeled-based forecasting algorithm and its application to electricity price time series. In *Prooceedings of the eighth IEEE International Conference on Data Mining*, pages 453–461, 2008.

[22] K. Metaxiotis, A. Kagiannas, D. Askounis, and J. Psarras. Artificial intelligence in short term electric load forecasting: A state-of-the-art survey for the researcher. *Energy Conversion and Management*, 44:1525–1534, 2003.

[23] Z. Mohamed and P. Bodger. Forecasting electricity consumption in new zealand using economic and demographic variables. *Energy*, 30:1833–1843, 2005.

[24] F. J. Nogales and A. J. Conejo. Electricity price forecasting through transfer function models. *Journal of the Operational Research Society*, 57:350–356, 2006.

[25] Spanish Electricity Price Market Operator. *http://www.omel.es.*

[26] The New York Independent System Operator.
http://www.nyiso.com.

[27] S. Pezzulli, P. Frederic, S. Majithia, S. Sabbagh, E. Black, R. Sutton, and D. Stephenson. The seasonal forecast of electricity demand: a hierchical bayesian model with climatological weather generator. *Applied Stochastic Models in Business and Industry*, 22:113–125, 2006.

[28] N. M. Pindoria, S. N. Singh, and S. K. Singh. An adaptative wavelet neural network-based energy price forecasting in electricity markets. *IEEE Transactions on Power Systems*, 23(3):1423–1432, 2008.

[29] M. A. Plazas, A. J. Conejo, and F. J. Prieto. Multimarket optimal bidding for a power producer. *IEEE Transactions on Power Systems*, 20(4):2041–2050, 2005.

[30] J. M. Riquelme, J. L. Martínez, A. Gómez, and D. Cros. Load pattern recognition and load forecasting by artificial neural networks. *International Journal of Power and Energy Systems*, 22(1):74–79, 2002.

[31] L. F. Sugianto and X. B. Lu. Demand forecasting in the deregulated market: a bibliography survey. In *Proceedings of the Australasian Universities Power Engineering Conference*, pages 1–6, 2002.

[32] J. W. Taylor, L. M. de Menezes, and P. E. McSharry. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 22:1–16, 2006.

[33] A. Troncoso, J. C. Riquelme, J. M. Riquelme, J. L. Martínez, and A. Gómez. Electricity market price forecasting based on weighted nearest neighbours techniques. *IEEE Transactions on Power Systems*, 22(3):1294–1301, 2007.

[34] A. Troncoso, J. M. Riquelme, J. C. Riquelme, A. Gómez, and J. L. Martínez. Time-series prediction: Application to the short term electric energy demand. *Lecture Notes in Artificial Intelligence*, 3040:577–586, 2004.

[35] J. Wang and L. Wang. A new method for short-term electricity load forecasting. *Transactions of the Institute of Measurement and Control*, 30(3):331–344, 2008.

[36] X. Wang and M. Meng. Forecasting electricity demand using grey-markov model. In *Proceedings of the Seventh International Conference on Machine Learning and Cybernetics*, pages 1244–1248, 2008.

[37] R. Weron. *Modeling and Forecasting Electricity Loads and Prices.* Wiley, 2006.

[38] R. Weron and A. Misiorek. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International Journal of Forecasting*, 24:744–763, 2008.

[39] R. Xu and D. C. Wunsch II. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678, 2005.

[40] Z. Xu, Z. Y. Dong, and W. Liu. *Neural Networks Applications in Information Technology and Web Engineering, Chapter 22.* Borneo Publishing, 2005.

[41] J. H. Zhao, Z. Y. Dong, X. Li, and K. P. Wong. A framework for electricity price spike analysis with advanced data mining methods. *IEEE Transactions on Power Systems*, 22(1):376–385, 2007.