An Ensemble Model for Day-ahead Electricity Demand Time Series Forecasting

Wen Shen, Vahan Babushkin, Zeyar Aung, and Wei Lee Woon
Masdar Institute of Science and Technology
PO Box 54224, Abu Dhabi
United Arab Emirates
{wshen, vbabushkin, zaung, wwoon}@masdar.ac.ae

ABSTRACT

In this work, we try to solve the problem of day-ahead prediction of electricity demand using an ensemble forecasting model. Based on the Pattern Sequence Similarity (PSF) algorithm, we implemented five forecasting models using different clustering techniques: K-means model (as in original PSF), Self-Organizing Map model, Hierarchical Clustering model, K-medoids model, and Fuzzy C-means model. By incorporating these five models, we then proposed an ensemble model, named Pattern Forecasting Ensemble Model (PFEM), with iterative prediction procedure. We evaluated its performance on three real-world electricity demand datasets and compared it with those of the five forecasting models individually. Experimental results show that PFEM outperforms all those five individual models in terms of Mean Error Relative and Mean Absolute Error.

Categories and Subject Descriptors

I.5 [**Pattern Recognition**]: Models, Clustering

General Terms

Design, Experimentation

Keywords

Time Series Forecasting, Ensemble Model, Clustering

1. INTRODUCTION

In the electrical power industry, it is essential for decision makers to accurately predict the future values of variables such as electricity demand or price. This process of forecasting or prediction is called *Time Series Forecasting* or *Time Series Prediction*.

To forecast electricity demand and/or price, various forecasting techniques have been studied in the literature. These techniques include the wavelet transform models [1], the ARIMA models [2], the GARCH models [3], the Artificial Neural Network models [4] the hybrid model (basically a combination of Artificial Neural Networks and Fuzzy Logic) [5], the nearest neighbors methodology [6], the Support Vector Machines framework [7] and the Least-Square Support Vector Machine models [8].

Martínez-Álvarez *et al.* [9] proposed a Label-Based Forecasting (LBF) Algorithm using K-means clustering to predict electricity pricing time series. Using the mean squared error as a metric, they demonstrated that the LBF method outperforms several other methods including Naïve Bayes, Neural Networks, ARIMA, Weighted Nearest Neighbors and the Mixed Models. Based on the LBF algorithm, they later developed the Pattern Sequence-based Forecasting (PSF) algorithm [10], which predicts the future evolution of a time series based on the similarity of pattern sequences. They reported that PSF produced a significant improvement in the prediction of energy time series compared to several well-known techniques including its predecessor LBF algorithm.

In this work, we propose a Pattern Forecasting Ensemble Model (PFEM). Five PSF-style forecasting models are deployed, namely the K-means model (PSF itself), Self-Organizing Map model, Hierarchical Clustering model, K-medoids model, and Fuzzy C-means model. Each model is first applied separately, producing its respective forecasted values; we then perform a weighted combination of those values in an iterative manner in order to realize a better energy time series forecasting model.

We evaluated the performance of PFEM on three publicly available electricity demand datasets from the New York Independent System Operator (NYISO) [11], the Australia's National Electricity Market (ANEM) [12], and the Ontario's Independent Electricity System Operator (IESO) [13]. It was observed that the PFEM was able to provide more accurate and reliable forecasts than the five individual models, including PSF.

The rest of the paper is organized as follows. In Section 2, we briefly introduce the general principles of the PSF algorithm and give a general description of the five clustering methods which are the basis for our five individual forecasting models respectively. In Section 3, we present our proposed Pattern Forecasting Ensemble Model (PFEM). In Section 4, we evaluate the performance of PFEM and compare its performance with five individual models based on K-means (i.e., same as original PSF), Self-Organizing Map, Hierarchical Clustering, K-meloids, and Fuzzy C-means. In Section 5, we reach the conclusion and discuss the future work.

e-Energy'13, May 21–24, 2013, Berkeley, California, USA.

2. BACKGROUND

In this section, we present a short description of each of the five clustering methods used in PFEM, namely: K-means, Self-Organizing Map, Hierarchical Clustering, K-medoids, and Fuzzy C-means. Each of these methods can be used to generate a separate forecasting model, as described later in Section 3. However, in the following subsection, we begin by describing the basic concept of the PSF algorithm, upon which the proposed PFEM algorithm is based.

2.1 PSF Algorithm

The Pattern Sequence-based Forecasting (PSF) algorithm [10] is the basis of the proposed prediction algorithm, where the basic set of steps taken are repeated for each of the five individual forecasting models which comprise the PFEM algorithm. The general idea of the PSF algorithm can be described as follows:

1. Firstly, a clustering method is used to divide all of the 24-hour segments in the training dataset into a set of similar categories (K-means was the method of choice for clustering in PSF). This allows us to assign a label for each day in the training set, producing an associated label sequence. This sequence as well as the associated cluster centers are stored for use in the subsequent steps of this process.

2. To predict the demand for a given day, the category label for the days leading up to the day in question are determined based on similarity to the cluster centers previously obtained. This results in a label sequence which forms a kind of "fingerprint".

3. The sequence of labels generated during the training phase is searched for occurrences of this fingerprint, and matching instances are collected.

4. The demand profiles for the days immediately following each of these matches are extracted. The prediction is then generated by taking the average of all of these profiles.

2.2 Clustering Methods

2.2.1 K-means Clustering

K-means clustering [14] is a simple unsupervised learning algorithm which partitions N observations into k disjoint subsets C_j containing N_j data points. The K-means clustering algorithm aims to find the minimum value (or the local minima in most cases) of an objective function. The objective function J is described as follows:

$$J = \sum_{j=1}^{K} \sum_{n \in C_j} d(x_n, \mu_j) \tag{1}$$

where μ_j is the cluster centroid for points in C_j and $d(x_n, \mu_j)$ points from their respective cluster centers, is a chosen distance measure between a data point x_n and the cluster center μ_j. In most cases, the Euclidean distance is used as a metric. The algorithm will terminate when no new partitions occur.

K-means is a greedy algorithm and as such its performance closely relies on the choice of the parameter k and the appropriate selection of the initial cluster centers [15].

2.2.2 Self-Organizing Map

Self-Organizing Map (SOM) [16] is an unsupervised learning artificial neural network which uses a neighborhood function to preserve the topological properties of the input space and maps high-dimensional data onto a 2-dimensional grid.

SOMs can be used to produce low dimensional representations of the data that preserve similarities between points in the data. Due to SOM's ability to preserve topological properties and good visualization features, they perform well for the prediction of non-linear time series [17].

2.2.3 Hierarchical Clustering

Hierarchical Clustering is a widely used data analysis tool which seeks to build a binary tree of the data that successively merges similar groups of points. Among all clustering techniques known in the literature, Hierarchical Clustering offers great versatility since it does not require a prespecified number of clusters [18]. Instead, it only requires a measure of similarity between groups of data points.

In our work, we will use agglomerative Hierarchical Clustering. The algorithm can be described as follows: given a dataset $D = (x_1, x_2, \ldots, x_n)$ of n points, it first calculates a distance matrix M with all of the pairwise distances between points. Then it conducts the following process recursively until D has only a single data point:

1. Choose the two points x_i, x_j from D such that the distance between the two points is the minimum.

   $$i, j = \arg\min_{i,j} d(x_i, x_j) \tag{2}$$

   where $i \neq j$.

2. Cluster x_i, x_j to form a new point c. The new point could be the mean of the two points or some other metric.

3. Remove x_i, x_j from D and insert c into D. Recalculate the pairwise distance matrix M.

2.2.4 K-medoids Clustering

The K-medoids algorithm [19] is an adaptation of the K-means algorithm in which a representative item, or a medoid is chosen for each cluster at each iteration. Medoids for each cluster are calculated by finding object i within the cluster that minimizes $\sum_{j \in C_i} d(i, j)$, where C_i is the cluster containing object i and $d(i, j)$ is the distance between objects i and j.

There are two advantages of using existing objects as the centers of the clusters. Firstly, a medoid object serves to usefully describe the cluster. Secondly, there is no need for repeated calculation of distances at each iteration, since the K-medoids algorithm can simply look up distances from a distance matrix.

The K-medoids algorithm can be described as follows [19]:

1. Choose k data points at random to be the initial cluster medoids.

2. Assign each data point to the cluster associated with the closet medoid.

3. Recalculate the positions of the k medoids.

4. Repeat Step 2 and Step 3 until the medoids become fixed.

2.2.5 *Fuzzy C-means Clustering*

Fuzzy C-means clustering [20] groups data by assigning a membership value linking each data point to each cluster center. This value is a reflection of the distance between the cluster center and the data point. Let $X = \{x_1, x_2, x_3, \ldots, x_n\}$ be the set of data points and $V = \{v_1, v_2, v_3, \ldots, v_c\}$ be the set of cluster centers. The Fuzzy C-means clustering algorithm is described as follows [20]:

1. Randomly select c cluster centers.
2. Calculate the fuzzy membership u_{ij} of i^{th} data point to j^{th} cluster center using:
$$u_{ij} = \frac{1}{\sum_{k=1}^{c}(\frac{d_{ij}}{d_{ik}})^{(\frac{2}{m-1})}} \quad (3)$$
where n is the number of data points, d_{ij} is the distance between i^{th} data and j^{th} center, m is the fuzziness index and greater than or equal to 1.
3. Compute the fuzzy centers v_j using:
$$v_j = \frac{\sum_{i=1}^{n} x_i(u_{ij})^m}{\sum_{i=1}^{n}(u_{ij})^m}, \ \forall j = 1, 2, \ldots, c \quad (4)$$
4. Repeat Step 2 and 3 until the minimum value of objective function J is achieved or $d(U^{(k+1)}, U^{(k)}) < \beta$. β is the termination threshold between [0,1], $U = (u_{ij})_{n\times c}$ is the fuzzy membership matrix and J is defined by the following formula:
$$J(U, V) = \sum_{i=1}^{n}\sum_{j=1}^{c}(u_{ij})^m(x_i, v_j) \quad (5)$$
where $d(x_i, v_j)$ is the distance between i^{th} data point and j^{th} cluster center.

Unlike K-means where data points exclusively belong to one cluster, in the Fuzzy C-means algorithm data points are assigned memberships to each cluster center and as such can belong to more than one cluster at a time. Therefore, Fuzzy C-means clustering produces better results for overlapped dataset compared with K-means clustering.

3. PATTERN FORECASTING ENSEMBLE MODEL (PFEM)

The Pattern Forecasting Ensemble Model (PFEM) consists of four phases: data preprocessing, applying individual clustering methods, building individual forecasting models (based on corresponding clustering results), and iterative ensembling. Figure 1 shows the general idea behind PFEM. Details about the four phases are given below.

3.1 Data Preprocessing

The same feature selection and data normalization methodologies used for the original PSF model [10] were adopted for this study. This will facilitate direct comparison of PSF with the proposed method.

3.2 Applying Individual Clustering Methods

In this phase, five individual clustering methods: K-means, Self-Organizing Map, Hierarchical Clustering, K-medoids and Fuzzy C-means are applied on the preprocessed data. In each clustering exercise, cluster labels (which are discrete

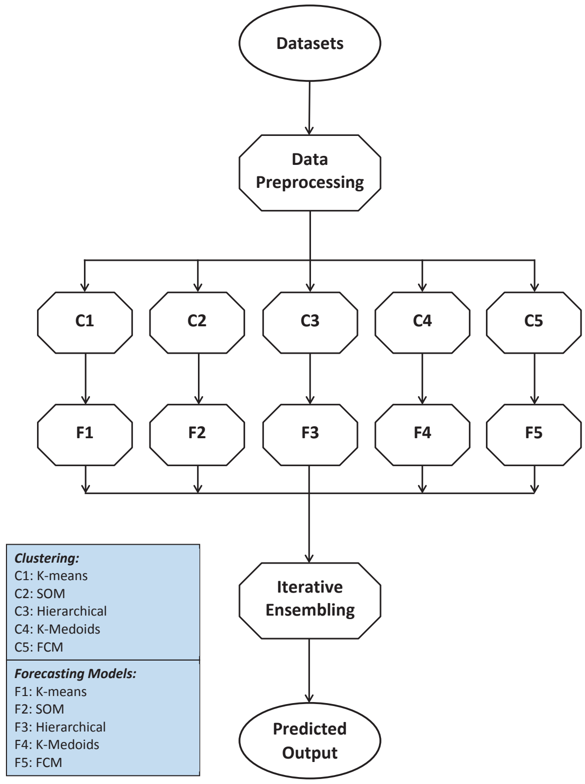

Figure 1: The general framework of Pattern Forecasting Ensemble Model.

values like 1, 2, 3) are assigned to all the real-number values in the preprocessed time-series data. Different clustering methods usually produce different cluster outputs form the same input data. Thus, the sets of labels assigned to the preprocessed data by the five methods will apparently be different from each other.

For each individual clustering method, several parameters need to be configured. The optimal parameter configurations for each method are determined by employing both statistical analyses (using well-defined indexes) and empirical analyses (based on the accuracy of the subsequent forecasting results). These parameters are briefly discussed below.

3.2.1 *Number of Clusters*

For K-means, K-medoids, Fuzzy C-means and SOM, the numbers of clusters are required to be specified in advance. For this reason, statistical analyses, namely the Silhouette index [21], the Dunn index [22] and the Davies-Bouldin index [23] are used to determine in how many groups the original dataset has to be split. For Hierarchical Clustering, it is not necessary to specify how many clusters as a priori. This is because the end result is a dendrogram where the leaves are data points and interior nodes represent a cluster made up of all children of the nodes. With the dendrogram, one can choose various number of clusters to use by cutting the tree at some particular height. However, we also employ the three indexes to determine the optimal number of clusters for Hierarchical Clustering for our convenience.

3.2.2 The Distance Function

In K-means, K-medoids, Fuzzy C-means and SOM, we use the Euclidean distance function to calculate similarities. In Hierarchical Clustering, the Mahalanobis distance [24] is utilized. This is because the respective indexes give more accurate indications than that with Euclidean distance according to our experimental analysis.

3.2.3 Window Size

The window size (the length of the sequence label) is selected through 12 folds cross validation as the K-means (original PSF) model does as described in [10].

3.2.4 Parameters for Fuzzy C-means Clustering

With lower value of the termination threshold β, we can get the better result. But it may require more iterations and thus more time to terminate. Therefore, a compromise between the accuracy and the efficiency is needed. In our experiments, we set the termination threshold $\beta = 10^{-5}$.

The fuzziness parameter m significantly effects the fuzziness of the clustering partition and hence affects the prediction results [25]. As the fuzziness m approaches 1 from above 0, the partition becomes hard ($u_{ik} \in \{0, 1\}$) and v_i are ordinary means of the clustering. As $m \to \infty$, the partition becomes completely fuzzy ($u_{ik} = 1/c$) and the cluster means are all equal to the mean of the dataset X.

3.3 Building Individual Forecasting Models

By changing the underlying clustering algorithm, different label sequences are produced. In this way, five individual predicting models are constructed using the same basic procedure used for the PSF algorithm [10] (see Section 2.1 for details). The mechanisms of all these five models are exactly the same except for the source label sequences they use. In this way, K-means model (same as original PSF), Self-Organizing Map model, Hierarchical Clustering model, K-medoids model, and Fuzzy C-means model are built separately. The five models produce their respective prediction results, which are subsequently used as the inputs in the next Iterative Ensembling phase.

3.4 Iterative Ensembling

In the iterative ensembling phase, the ensemble model is constructed by using the five individual forecasting models produced in the previous phase. The idea of iterative ensembling is to create a new model derived from a linear combination of several models with coefficients (weights) to minimize the forecasting error rates.

In the training stage, the actual values and the predicted values produced by the individual models are employed to select the weights which give the minimum prediction error rates. Then the weights are re-normalized. In the testing phase, the newly produced weights are used to predict the value. After prediction, this sample is incorporated into the training dataset. The weights are re-calculated, and new samples are learned iteratively in the same way. The idea of our iterative ensembling is inspired by AdaBoost [26].

The formal process is described as following:

1. Initialize the vector of observation weights $\mathbf{w}^{(0)} = (w_1^{(0)}, w_2^{(0)}, \ldots, w_n^{(0)})$, where n is the number of participating forecasting models for ensemble learning ($n = 5$ in our case):

$$w_i^{(0)} = 1/n \quad \forall i = 1, 2, \ldots, n. \tag{6}$$

2. For a training dataset with M days, for $m = 1$ to M:

$$\mathbf{P}^{(m)} = \sum_{i=1}^{n} w_i^{(m-1)} \mathbf{P}_i^{(m)} \tag{7}$$

where $\mathbf{P}^{(m)} = (P_1^{(m)}, P_2^{(m)}, \ldots, P_{24}^{(m)})$ is the vector of combined predicted values for 24 hours in day m, and $\mathbf{P}_i^{(m)} = (P_{i1}^{(m)}, P_{i2}^{(m)}, \ldots, P_{i24}^{(m)})$ is the vector of predicted values for 24 hours in day m generated by the individual forecasting model i.

(a) Define the prediction error rate for the iteration m:

$$err^{(m)} = \frac{1}{24} \sum_{h=1}^{24} \frac{\mid P_h^{(m)} - A_h^{(m)} \mid}{\bar{A}^{(m)}} \tag{8}$$

where $\mathbf{A}^{(m)} = (A_1^{(m)}, A_2^{(m)}, \ldots, A_{24}^{(m)})$ is the vector of actual values for 24 hours in day m and $\bar{A}^{(m)}$ is the average of actual values for 24 hours in that day.

(b) Calculate new weights for the iteration m:

$$\mathbf{w}^{(m)} = \underset{w_i^{(m-1)}, i=1,\ldots,n}{\text{argmin}} \quad err^{(m)}, \tag{9}$$

such that $0 \leq w_i^{(m)} \leq 1, \sum_{i=1}^{n} w_i^{(m)} = 1$

3. Calculate the initial weights for testing:

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

Re-normalize $\mathbf{w}$.

4. Produce the predicted values using Equation (7) with the weights obtained in Step 3.

5. Add testing sample to the training dataset, increase M by 1, recalculate the initial weights for testing using Step 2, 3 and 4 for the next testing sample.

4. EXPERIMENTAL ANALYSIS

4.1 Performance Metrics

Several metrics are used to evaluate the performance of the PFEM approach:

- Mean Error Relative(MER):

$$MER = 100 \times \frac{1}{N} \sum_{h=1}^{N} \frac{|\hat{x}_h - x_h|}{\overline{x}} \tag{11}$$

where $\hat{x}_h$ are predicted and actual demand at hour h, respectively, $\overline{x}$ is the mean demand of the day and N is the number of predicted hours.

- Mean Absolute Error(MAE):

$$MAE = \frac{1}{N} \sum_{h=1}^{N} |\hat{x}_h - x_h| \tag{12}$$

- Mean Absolute Percentage Error(MAPE):

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right| \quad (13)$$

where n is the number of the samples. x_i is the actual value and $\hat{x}_i$ is the forecast value.

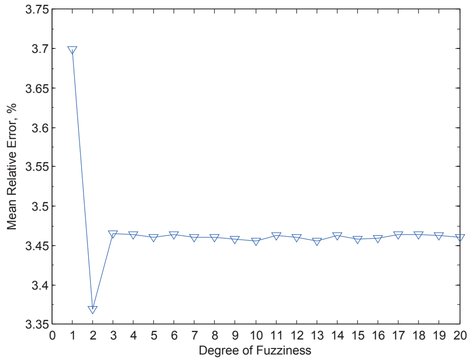

Figure 2: Correlation between mean relative error and degree of fuzziness for NYISO electricity demand data.

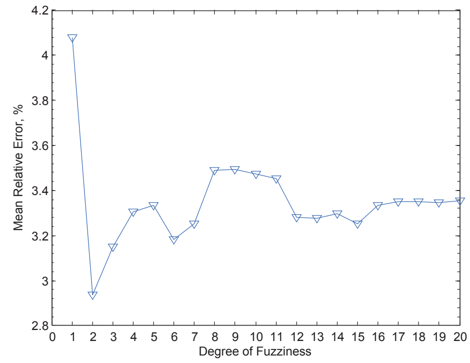

Figure 3: Correlation between mean relative error and degree of fuzziness for ANEM electricity demand data.

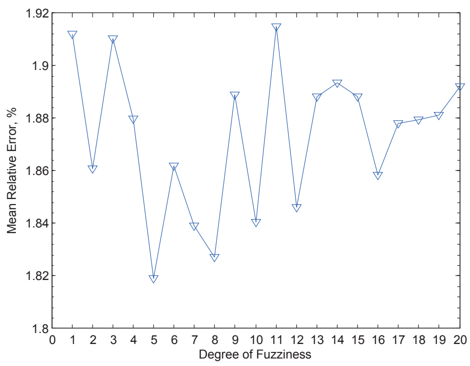

Figure 4: Correlation between mean relative error and degree of fuzziness for IESO electricity demand data.

4.2 Experimental Settings

In order to evaluate the performance of the PFEM, three real-world electricity demand datasets-the New York Independent System Operator (NYISO), the Australian Energy Market Operator (ANEM) and the Ontario's Independent Electricity System Operator (IESO) from 2007-2011 are employed.

For NYISO dataset, historical demand for Capital area was considered because the Capital district is a representative sample in terms of electricity consumption in the New York state. As for AEMO, load data for the New South Wales area was used since New South Wales is the most populated Australian state with high electricity demand.

To construct the Pattern Forecasting Ensemble Model for predicting the electricity demand of a day in 2009, Data from 2007 was used to train the individual clustering classifiers which were subsequently used to generate predictions for the data in 2008 and 2009. Then the predicted values and the actual time series of the data in 2008 were utilized to calculate the initial weights for testing following Step 2 described in section 3.4. This process is categorized as the training phase of the PFEM. With the preliminarily predicted time series in 2009 and the initial weights computed in the previous step, the final prediction of the electricity demand time series in 2009 could be produced by using methods shown in Step 3, 4 and 5 in section 3.4. For comparison, we also employed the five PSF-based individual learning models to forecast the 2009 time series. The 2007 and 2008 datasets were used for training and the 2009 dataset for testing. Likewise, we also performed the forecasting experiments for the 2010 dataset using the 2008 and 2009 datasets for training.

4.2.1 Number of Clusters

In order to find optimal numbers of clusters, we analyze the data for 3 years. For all the three datasets the training set is chosen from Jan 1, 2007 to 31, Dec 2007.

The optimal numbers of clusters are chosen based on the majority voting method of Silhouette, Dunn and Davies-Bouldin indices. The maximum values of Silhouette and Dunn indices indicate the optimal numbers of clusters when the best possible clustering is obtained, that is, the distance among elements in the same cluster is minimized and the distance among elements from different clusters is maximized (for Davies-Bouldin index the optimal number of clusters corresponds the minimum value). If all three indices give different results, the optimal numbers of clusters are chosen based on the sub-optimal numbers of clusters. The numbers of clusters are calculated for 10 times and the most frequent values are selected. Table 1 shows the numbers of clustering for different markets in the year 2007 with the three indices. Similarly, the numbers of clusterings are chosen for the three datasets in the year 2007, 2008, 2007-2008 and 2008-2009, which can be seen in Table 2.

4.2.2 Window Size

The dataset for 1 year (2007) was normalized (in order to eliminate the noise introduced by the presence of global trend) and divided into 12 parts, each part consisting of a one-month long subset of given data for one year (2007). Then, starting from the first month and rotating through all twelve months in the year, one of the months was used for testing and the remaining 11 months for training. For each day of the testing set, given the known optimal number

Market	K-means				SOM				Hierarchical				k-medoid				Fuzzy C-means			
	Silh.	DU	DB	**sel.**	Silh.	DU	DB	**sel.**	Silh.	DU	DB	**sel.**	Silh.	DU	DB	**sel.**	Silh.	DU	DB	**sel.**
NYISO	3	3	4	**3**	4	4	4	**4**	2	6	5	**5**	6	8	6	**6**	3	3	4	**3**
ANEM	5	5	4	**5**	2	4	4	**4**	4	4	4	**4**	4	6	5	**6**	2	4	3	**3**
IESO	3	4	3	**3**	3	4	3	**3**	3	3	3	**3**	4	4	3	**4**	3	4	4	**4**

Table 1: Numbers of clusters for different markets evaluated with Silhouette (denoted as Silh.), Dunn (DU) and Davies-Bouldin (DB) indexes for five clustering methods(2007). "sel." indicates the selected optimal numbers of clusters.

Method	2007			2008			2007-2008			2008-2009		
	NYISO	ANEM	IESO	NYISO	ANEM	IESO	NYISO	ANEM	IESO	NYISO	ANEM	IESO
K-means	3	5	3	3	4	3	5	5	3	6	5	3
SOM	4	4	3	6	3	8	7	6	4	4	5	6
Hierarchical	4	4	3	5	2	4	5	4	4	5	5	5
K-medoids	6	5	4	6	5	3	6	5	5	7	6	4
Fuzzy c-means	3	3	4	6	5	4	6	3	4	3	4	4

Table 2: Optimal numbers of clusters, when the 5 clustering methods achieve the best clustering, for NYISO, ANEM and IESO markets (2007, 2008, 2007-2008 and 2008-2009)

Method	W=1	W=2	W=3	W=4	W=5	W=6	W=7	W=8	W=9	W=10	W=11	W=12	W=13	W=14	Selected
K-means	3.0940	2.9106	**2.7445**	2.7570	2.8066	2.8485	2.8835	2.9108	2.9733	2.8843	2.9120	2.8740	3.0363	2.8942	3
SOM	3.4371	**3.4217**	3.5281	3.5693	3.7037	3.7310	3.8234	3.7682	3.8117	3.7612	3.8120	3.8012	3.8411	3.8697	2
Hierarchical	3.5894	3.5067	3.4721	**3.4541**	3.6008	3.6149	3.6478	3.6845	3.7076	3.7423	3.8077	3.8259	3.8264	3.8273	4
K-medoids	3.3078	3.2503	**3.2092**	3.3177	3.3370	3.3669	3.5249	3.5572	3.4905	3.4773	3.5413	3.4736	3.5832	3.4611	3
Fuzzy C-means	3.4852	3.4788	**3.4526**	3.4602	3.5494	3.5669	3.6486	3.6482	3.6210	3.6671	3.6289	3.6653	3.6792	3.6947	3

Table 3: MER-based selection of optimal window sizes for NYISO Demand Dataset(2007) through 12 folds cross-validation

Method	W=1	W=2	W=3	W=4	W=5	W=6	W=7	W=8	W=9	W=10	W=11	W=12	W=13	W=14	Selected
K-means	3.0940	2.9106	**2.7445**	2.7570	2.8066	2.8485	2.8835	2.9108	2.9733	2.8843	2.9120	2.8740	3.0363	2.8942	3
SOM	3.0176	2.9963	**2.8940**	2.9266	2.9685	2.8958	2.8962	2.9515	2.9874	3.0121	3.0818	3.1324	3.1456	3.0827	3
Hierarchical	3.1001	2.8099	2.7063	2.6131	**2.4989**	2.5378	2.5636	2.5896	2.6228	2.6762	2.7008	2.7121	2.7157	2.7223	5
K-medoids	3.0831	2.8597	**2.7627**	2.8017	2.8297	2.9381	2.8089	3.0244	2.8452	2.9224	2.9302	2.9858	2.9351	3.0995	3
Fuzzy C-means	3.1135	3.0842	2.9290	2.8923	2.8718	2.8538	**2.8304**	2.8512	2.8827	2.8912	2.9262	2.9769	3.0288	3.0519	7

Table 4: MER-based selection of optimal window sizes for ANEM Demand Dataset(2007) through 12 folds cross-validation.

Method	W=1	W=2	W=3	W=4	W=5	W=6	W=7	W=8	W=9	W=10	W=11	W=12	W=13	W=14	Selected
K-means	2.6414	2.5821	2.5561	2.5948	2.5621	**2.5560**	2.5716	2.5964	2.5946	2.5935	2.5687	2.5887	2.5710	2.5966	6
SOM	2.6650	2.4498	**2.4414**	2.4440	2.4705	2.4976	2.5375	2.5569	2.5904	2.6308	2.6372	2.6392	2.6562	2.6322	3
Hierarchical	2.5956	2.4754	**2.4404**	2.4580	2.4588	2.4755	2.4734	2.4769	2.4624	2.4597	2.4608	2.4700	2.4811	2.4902	3
K-medoids	2.6218	2.4190	2.5062	**2.3822**	2.4567	2.4788	2.5751	2.4740	2.5218	2.5317	2.5806	2.5735	2.5629	2.6450	4
Fuzzy C-means	2.5261	2.3559	**2.3413**	2.3476	2.3746	2.3938	2.3932	2.4416	2.4662	2.5103	2.5100	2.5031	2.5102	2.5003	3

Table 5: MER-based selection of optimal window sizes for IESO Demand Dataset(2007) through 12-fold cross-validation.

Method	2007			2008			2007-2008			2008-2009		
	NYISO	ANEM	IESO	NYISO	ANEM	IESO	NYISO	ANEM	IESO	NYISO	ANEM	IESO
K-means	3	3	6	5	5	5	4	5	8	5	5	7
SOM	2	3	3	2	9	7	3	5	4	4	8	4
Hierarchical	4	5	3	5	6	4	4	5	3	5	5	5
K-medoids	3	3	4	2	9	4	2	12	5	3	6	5
Fuzzy c-means	3	7	3	2	9	7	2	9	4	5	6	6

Table 6: The selected window sizes of NYISO, ANEM and IESO Demand Datasets (2007, 2008, 2007-2008, 2008-2009) through 12 folds cross-validation

of clusters and proposed sizes of window, predictions were made and the respective MER calculated. The MER for given month is obtained by averaging across all days in the testing set. Iterating through the different window sizes, the correlation and dependence between size of window and the MER is evident. Finally, from obtained MER and size of

Market	Error	K-means		SOM		Hierarchical		K-medoids		Fuzzy C-means		PFEM	
		Err	σ	Err	σ	Err	σ	Err	σ	Err	σ	Err	σ
NYISO	MER	3.11	0.41	3.06	0.44	2.92	0.31	2.97	0.37	3.18	0.43	2.76	0.35
	MAE	39.16	6.88	38.5	7.38	36.79	5.92	37.27	6.21	39.89	6.61	34.78	6.31
	MAPE	3.18	0.42	3.12	0.44	2.99	0.32	3.03	0.38	3.26	0.43	2.82	0.36
ANEM	MER	2.98	0.862	3.18	0.76	2.76	0.91	2.79	0.73	2.58	0.67	2.55	0.80
	MAE	259.66	85.92	283.23	76.98	244.83	85.89	249.25	71.88	229.28	65.22	228.35	79.00
	MAPE	2.96	0.902	3.25	0.81	2.78	0.95	2.86	0.77	2.63	0.71	2.61	0.84
IESO	MER	2.42	0.36	2.67	0.36	2.5	0.41	2.34	0.29	2.31	0.27	2.23	0.25
	MAE	384.02	59.27	422.61	49.22	394.26	58.62	371.21	45.23	364.71	39.31	354.99	46.92
	MAPE	2.49	0.38	2.74	0.36	2.58	0.43	2.41	0.31	2.37	0.30	2.30	0.27

Table 7: Summary performance results of models tested on demand data of NYISO, ANEM and IESO markets for 2009.

window values, the optimal size of window that corresponds to minimal mean relative error is selected.

Tables 3, 4 and 5 present the prediction errors with respect to MER in relation to the size of window and the clustering technique used for the NYISO, ANEM and IESO electricity demand datasets respectively. The window sizes selected for each of the three datasets are shown in the last column of each table. Correspondingly, we selected the optimal sizes of window for 2007, 2008, 2007-2008, 2008-2009 datasets using the same method. The results are shown in Table 6.

4.2.3 Determining Fuzziness Parameter

In order to find the optimal values of the fuzziness parameters that minimize the mean relative error we employed the same cross validation technique that was used to determine the optimal size of window.

In the case of NYISO and ANEM datasets, as fuzziness increases, the error rates first drop to the bottom and then increase to a certain level. Then there will be fluctuations when the fuzziness is greater that 2. The is because the degree of data overlap in this real-world dataset is relatively low. The optimal fuzziness parameter is 2 with corresponding MER values 3.37% and 2.93% as shown in Figure 2 and Figure 3.

For IESO dataset, the optimal fuzziness parameter is 5, when MRE of 1.82% is achieved as shown in Figure 4.

4.3 Experimental Results

Table 7 and Table 11 show the annual mean and the annual standard deviation of MER, MAE and MAPE for each learning model based on the testing results of three datasets in 2009 and 2010, respectively.

4.3.1 NYISO Dataset

Table 8 shows the Mean Error Relative (MER) and Mean Absolute Error (MAE) obtained for all five individual models based on the PSF algorithms and the proposed PFEM in the year 2009. Among all the five individual models, Hierarchical Clustering gives the forecasting results with the lowest MER (2.92%) and MAE(36.79MW). The PFEM outperforms all other models in terms of both MER(2.76%) and MAE(34.78MW).

Figure 5 demonstrates the best results for a day's electricity demand forecasting with respect to MER in the year 2009. Figure 6 gives the worst prediction results for a day in the same year. Table 12 summarizes the performance of all 5 PSF-style algorithms and the PFEM algorithm for 2010. In this case, the lowest mean annual MER (2.77%) and MEA (36.57MW) are achieved with K-medoids model. Again, PFEM outperforms all 5 individuals models with MER (0.74%) and MAE (36.27MW). Figure 11 illustrates the best prediction electricity demand and Figure 12 gives the worst prediction results for a day in the same year.

4.3.2 ANEM Dataset

As for the Australian electricity demand time series, our ensemble model also produces the best prediction results with the MER of 2.55% and a MAE of 228.35MW.

The Fuzzy C-means defeats all other models except the PFEM. Table 9 presents the details about the performance of all models for the year 2009 dataset. Figure 7 and 8 illustrate the best and the worst forecasting results for the ANEM dataset respectively. The lowest MER it can be obtained is 0.744% while the highest error rate is 7.051%.

For 2010 the Hierarchical Clustering based model beats all other individual models with mean annual MER 2.45% and MAE 214.61MW, only PFEM achieves MER and MAE of 2.39% and 210.91MW. The best and the worst day-ahead demand forecasting results for 2010 are represented on Figure 13 and Figure 14 correspondingly.

4.3.3 IESO Dataset

With respect to the IESO dataset, Fuzzy C-means based PSF outperforms all other individual models with a MER of 2.31% and a MAE of 364.71MW. The PFEM gives a slightly better results than the Fuzzy C-means based model. Figure 9 and 10 present the best and worst forecasting of the PFEM in a day of the year 2009. Figure 15 and Figure 16 illustrate the best and worst prediction curves respectively. It shows that the PFEM is able to predict the day-ahead electricity load time series very accurately in the best case.

Considering the results for 2010, the Hierarchical Clustering based model outperforms all other 4 individual models in terms of MER (2.18%) and MEA (354.15MW). The PFEM outperforms all models with MER = 2.10% and MAE = 345.30MW. The best and worst day-ahead electricity demand prediction results for 2010 are represented on Figure 15 and Figure 16. Noticeably, that the best and worst forecasting results for 2009 and 2010 are achieved with MER of 0.81% and 0.59% for best and 4.05% and 6.26% for worst predictions correspondingly.

5. CONCLUSION AND FUTURE WORKS

In this work, a novel forecasting model for electricity demand time series is proposed. Named the "Pattern Forecasting Ensemble Model" (PFEM), the new method is based on the pre-existing PSF algorithm, but uses a combination of five separate clustering models: the K-means model (PSF itself), the SOM model, the Hierarchical Clustering based model, the K-means model and the Fuzzy C-means model. The optimal values of parameters such as k, c and the win-

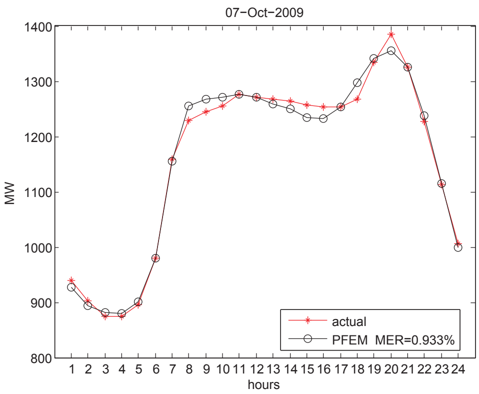

Figure 5: Best prediction of PFEM for NYISO dataset(2009).

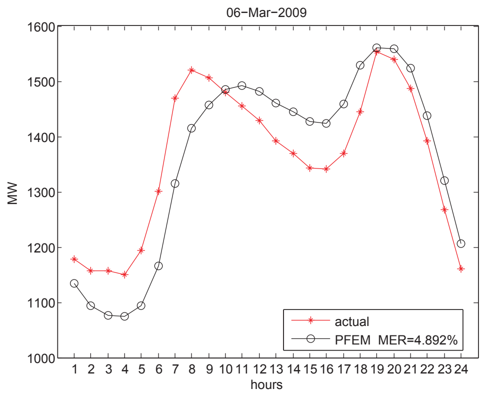

Figure 6: Worst prediction of PFEM for NYISO dataset(2009).

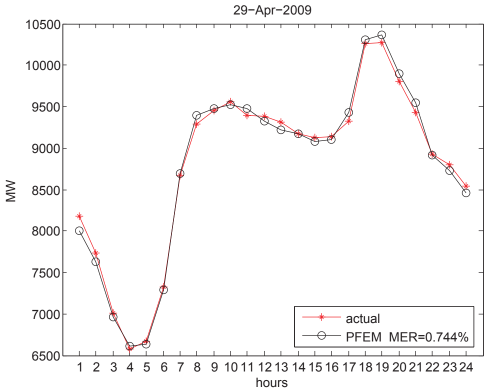

Figure 7: Best prediction of PFEM for ANEM dataset(2009).

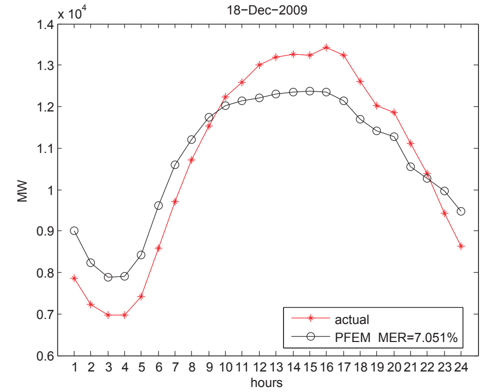

Figure 8: Worst prediction of PFEM for ANEM dataset(2009).

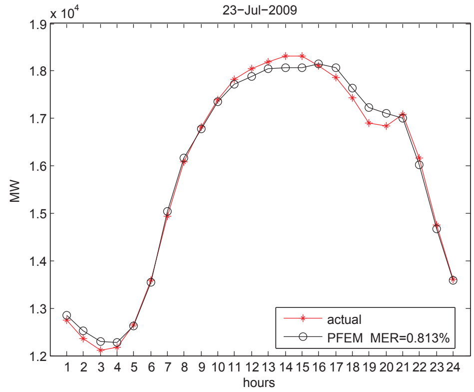

Figure 9: Best prediction of PFEM for IESO dataset(2009).

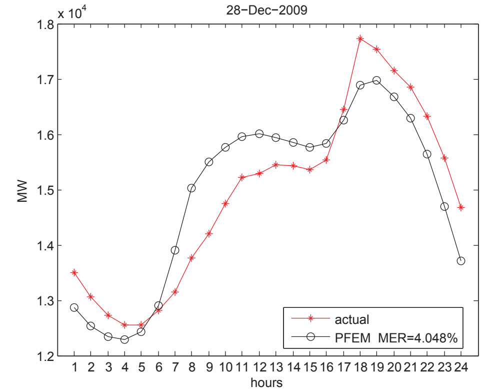

Figure 10: Worst prediction of PFEM for IESO dataset(2009).

dow size were also determined through an empirical approach. We then evaluate the performance of the PFEM and that of the other five individual models on three real-world electricity demand datasets. Experimental results indicate that our proposed approach gives superior results compared with all the other five individual models in terms of both MER and MAE. For future work, we intend to explore non-linear combinations of these individual models and conduct a quantitative evaluation of this approach along with a comparison to the related linearly combined ensemble models.

6. REFERENCES

[1] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1035–1042, 2005.

[2] G. Janacek, "Time series analysis forecasting and control," *Journal of Time Series Analysis*, vol. 31, no. 4, p. 303, 2010.

[3] R. Garcia, J. Contreras, M. Van Akkeren, and J. Garcia, "A GARCH forecasting model to predict day-ahead electricity prices," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 867–874, 2005.

[4] B. Neupane, K. S. Perera, Z. Aung, and W. L. Woon, "Artificial neural network-based electricity price forecasting for smart grid deployment," in *Proceedings of the 2012 IEEE International Conference on Computer Systems and Industrial Informatics (ICCSII'12)*, pp. 1–6, IEEE, 2012.

[5] N. Amjady, "Day-ahead price forecasting of electricity markets by a new fuzzy neural network," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 887–896, 2006.

[6] A. T. Lora, J. M. R. Santos, A. G. Expósito, J. L. M. Ramos, and J. C. R. Santos, "Electricity market price forecasting based on weighted nearest neighbors techniques," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1294–1301, 2007.

[7] 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*, vol. 22, no. 1, pp. 376–385, 2007.

[8] Z. Aung, M. Toukhy, J. Williams, A. Sanchez, and S. Herrero, "Towards accurate electricity load forecasting in smart grids," in *Proceedings of the 4th International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA'12)*, pp. 51–57, IARIA, 2012.

[9] 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 *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM'08)*, pp. 453–461, IEEE, 2008.

[10] F. Martínez-Álvarez, A. Troncoso, J. C. Riquelme, and J. S. Aguilar Ruiz, "Energy time series forecasting based on pattern sequence similarity," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 8, pp. 1230–1243, 2011.

[11] "New York Independent System Operator (NYISO)." `http://www.nyiso.com/public/index.jsp`.

[12] "Australia's National Electricity Market (ANEM)." `http://www.aemo.com.au/Electricity`.

[13] "Independent Electricity System Operator (IESO)." `http://www.ieso.ca/default.asp`.

[14] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, University of California, Berkeley, 1967.

[15] K. Hammouda and F. Karray, "A comparative study of data clustering techniques," tech. rep., Pattern Analysis and Machine Intelligence Research Group, University of Waterloo, 2000.

[16] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 586–600, 2000.

[17] A. Chitra and S. Uma, "An ensemble model of multiple classifiers for time series prediction," *International Journal of Computer Theory and Engineering*, vol. 2, no. 3, pp. 454–458, 2010.

[18] P. Rodrigues, J. Gama, and J. P. Pedroso, "LBF: Hierarchical time-series clustering for data streams," in *Proceedings of the 1st International Workshop on Knowledge Discovery in Data Streams (IWKDDS'04)*, pp. 22–31, ACM, 2004.

[19] A. Reynolds, G. Richards, and V. Rayward-Smith, "The application of k-medoids and PAM to the clustering of rules," pp. 173–178, Springer, 2004.

[20] R. L. Cannon, J. V. Dave, and J. C. Bezdek, "Efficient implementation of the fuzzy c-means clustering algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 2, pp. 248–255, 1986.

[21] P. J. Rousseeuw and L. Kaufman, *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, 1990.

[22] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," *Journal of Cybernetics*, vol. 4, no. 1, pp. 95–104, 1974.

[23] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 2, pp. 224–227, 1979.

[24] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The Mahalanobis distance," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, no. 1, pp. 1–18, 2000.

[25] D. Dembele and P. Kastner, "Fuzzy C-means method for clustering microarray data," *Bioinformatics*, vol. 19, no. 8, pp. 973–980, 2003.

[26] J. Zhu, S. Rosset, H. Zou, and T. Hastie, "Multi-class AdaBoost," *Statistics and Its Interface*, vol. 2, no. 3, pp. 349–360, 2009.

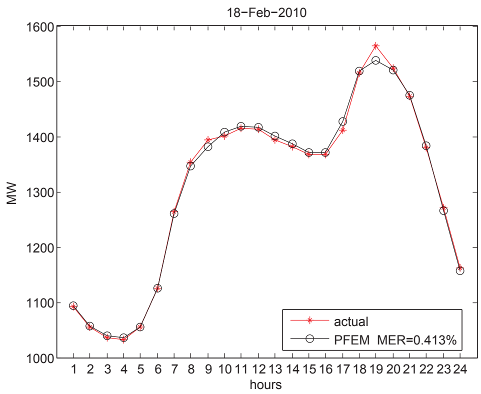

Figure 11: Best prediction of PFEM for NYISO dataset(2010).

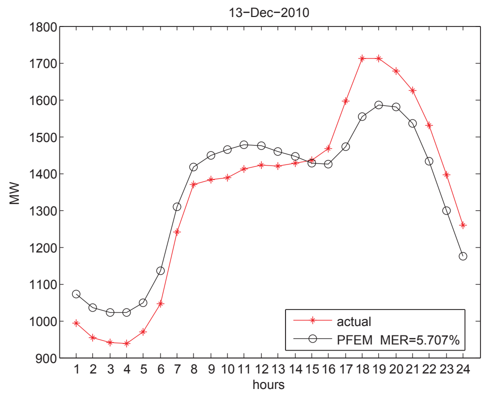

Figure 12: Worst prediction of PFEM for NYISO dataset(2010).

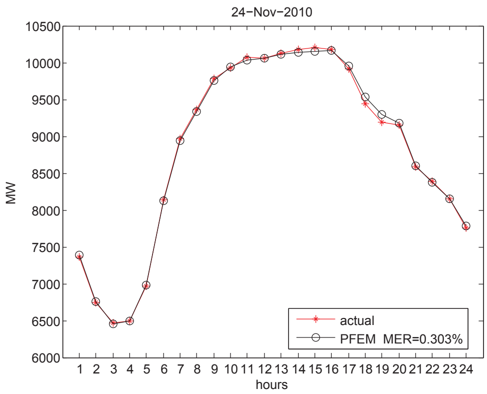

Figure 13: Best prediction of PFEM for ANEM dataset(2010).

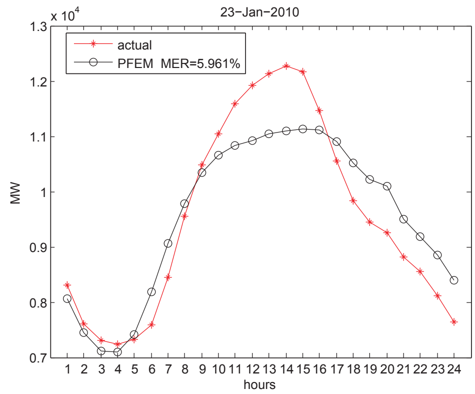

Figure 14: Worst prediction of PFEM for ANEM dataset(2010).

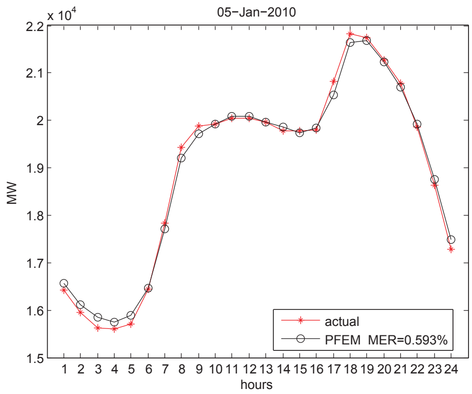

Figure 15: Best prediction of PFEM for IESO dataset(2010).

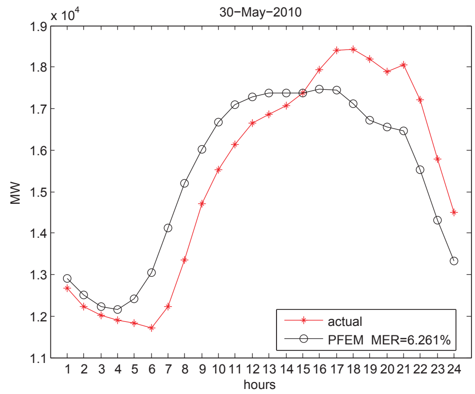

Figure 16: Worst prediction of PFEM for IESO dataset(2010).

Month	K-means		SOM		Hierarchical		K-medoids		Fuzzy C-means		PFEM	
	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)
Jan.	3.09	43.45	2.89	40.57	2.61	36.72	2.76	38.83	2.85	40.17	2.55	35.85
Feb.	2.89	39.02	2.93	39.72	2.62	35.39	2.75	37.37	2.83	38.32	2.46	33.33
Mar.	3.38	41.43	3.26	39.94	3.24	39.79	3.19	39.39	3.69	45.22	3.07	37.65
Apr.	3.41	38.09	3.38	37.84	3.05	34.37	3.65	41.54	3.64	40.75	3.13	35.02
May.	3.44	37.94	3.03	33.36	2.76	30.32	2.69	28.36	3.32	36.53	2.56	28.62
Jun.	3.15	39.09	3.12	38.69	3.09	38.15	2.86	35.36	2.99	37.07	2.72	33.84
Jul.	3.04	40.89	2.85	37.57	3.02	39.81	2.98	39.08	2.98	39.17	2.89	38.27
Aug.	3.59	51.99	3.61	52.29	3.41	49.65	3.55	51.36	3.64	52.76	3.31	48.19
Sept.	2.47	29.98	2.37	28.75	3.02	36.81	2.37	28.84	2.57	31.32	2.29	27.83
Oct.	2.26	26.17	2.27	26.24	2.33	27.78	2.87	33.11	2.57	29.89	2.22	25.52
Nov.	3.06	35.02	3.24	37.15	2.74	31.17	2.91	33.16	3.36	38.38	2.74	31.41
Dec.	3.53	46.89	3.75	49.91	3.14	41.59	3.08	40.84	3.72	49.17	3.16	41.89
Mean	3.11	39.16	3.06	38.50	2.92	36.79	2.97	37.27	3.18	39.89	2.76	34.78

Table 8: Performance of all models tested on NYISO demand data for 2009.

Month	K-means		SOM		Hierarchial		K-medoids		Fuzzy C-means		PFEM	
	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)
Jan.	5.92	468.94	5.15	482.18	4.97	457.47	4.42	410.97	4.09	376.24	4.52	420.75
Feb.	3.93	369.69	3.68	347.89	3.29	310.16	3.27	309.89	2.94	277.27	3.13	295.26
Mar.	1.88	162.09	2.67	230.52	2.37	196.76	2.12	181.62	2.19	170.89	1.86	160.05
Apr.	2.62	210.94	2.76	223.38	2.09	169.46	2.18	177.11	2.07	169.42	1.97	159.63
May.	2.48	208.27	2.37	205.69	1.86	162.91	2.32	200.77	2.04	177.88	1.87	163.81
Jun.	2.32	220.06	2.68	246.29	2.02	192.58	2.25	212.23	1.94	184.12	1.96	186.31
Jul.	2.38	224.77	2.61	246.67	2.08	198.16	2.24	212.59	2.14	202.69	2.21	208.52
Aug.	2.61	231.65	2.97	264.59	2.27	201.07	2.58	230.34	2.28	201.73	2.18	193.78
Sept.	2.24	187.95	2.83	238.16	2.29	185.64	2.28	192.23	2.28	191.59	2.03	170.49
Oct.	2.96	248.99	3.17	266.39	3.19	261.27	2.83	236.04	2.54	214.06	2.52	211.71
Nov.	3.23	296.75	3.72	338.18	3.29	300.62	3.52	320.97	3.25	297.08	3.24	295.43
Dec.	3.25	285.82	3.59	308.92	3.44	301.88	3.57	306.34	3.28	288.41	3.11	274.51
Mean	2.98	259.66	3.18	283.23	2.76	244.83	2.79	249.25	2.58	229.28	2.55	228.35

Table 9: Performance of all models tested on ANEM demand data for 2009.

Month	K-means		SOM		Hierarchial		K-medoids		Fuzzy C-means		PFEM	
	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)
Jan.	1.94	347.33	1.96	357.19	1.63	295.57	1.95	355.46	1.95	346.89	1.81	329.19
Feb.	2.31	399.73	2.36	406.96	1.91	331.03	2.04	354.55	2.68	347.96	2.18	376.81
Mar.	2.27	369.71	2.75	449.66	3.05	495.62	2.68	439.05	2.29	374.83	2.27	370.83
Apr.	2.54	376.76	3.78	450.44	2.46	367.02	2.43	362.47	2.29	341.58	2.27	337.68
May.	2.24	310.57	2.74	383.07	2.83	400.38	2.38	334.71	2.68	381.81	2.09	294.88
Jun.	2.89	444.41	2.62	406.57	2.74	420.28	2.46	376.94	2.45	376.13	2.49	382.84
Jul.	2.17	326.39	2.55	387.35	2.43	366.49	2.25	340.94	2.07	313.55	2.03	301.91
Aug.	2.93	482.93	2.96	494.02	2.85	473.09	2.75	454.78	2.79	461.18	2.75	456.64
Sept.	2.63	400.62	2.32	351.71	2.49	364.35	2.23	337.52	2.19	331.82	2.21	336.37
Oct.	1.96	285.77	2.96	445.78	2.34	353.89	2.16	325.12	2.22	334.91	2.01	302.12
Nov.	2.84	438.99	3.28	504.81	2.82	434.97	2.83	433.89	2.63	402.67	2.51	387.04
Dec.	2.53	425.14	2.58	433.88	2.57	428.49	2.02	340.67	2.17	363.35	2.28	383.68
Mean	2.42	384.02	2.67	422.61	2.50	394.26	2.34	371.21	2.31	364.71	2.23	354.99

Table 10: Performance of all models tested on IESO demand data for 2009.

Market	Error	K-means		SOM		Hierarchal		K-medoids		Fuzzy C-means		PFEM	
		Err	σ	Err	σ	Err	σ	Err	σ	Err	σ	Err	σ
NYISO	MER	3.09	0.61	3.11	0.55	2.97	0.54	2.77	0.54	3.07	0.56	2.74	0.57
	MAE	40.72	11.45	40.87	10.86	40.87	10.86	36.57	10.38	40.35	10.84	36.27	11.11
	MAPE	3.16	0.49	3.12	0.55	2.99	0.58	2.79	0.57	3.26	0.63	2.78	0.58
ANEM	MER	2.78	0.56	2.89	0.39	2.45	0.71	2.64	0.52	2.73	0.51	2.39	0.46
	MAE	243.88	52.78	254.17	38.63	214.61	60.99	232.11	47.29	239.93	46.91	210.92	44.11
	MAPE	2.83	0.58	2.96	0.41	2.48	0.74	2.69	0.55	2.78	0.52	2.44	0.48
IESO	MER	2.29	0.48	2.52	0.47	2.19	0.45	2.28	0.46	2.59	0.53	2.10	0.44
	MAE	372.48	91.21	410.51	89.67	354.15	83.38	368.67	86.72	422.13	102.99	345.30	87.04
	MAPE	2.35	0.52	2.59	0.51	2.24	0.48	2.34	0.49	2.66	0.56	2.18	0.48

Table 11: Summary performance results of models tested on demand data of NYISO, ANEM and IESO markets for 2010.

Month	K-means		SOM		Hierarchical		K-medoids		Fuzzy C-means		PFEM	
	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)
Jan.	2.68	36.77	2.89	39.63	2.61	35.63	2.61	35.83	2.73	37.32	2.46	33.95
Feb.	2.53	33.14	2.33	30.87	2.04	27.18	2.03	26.89	2.28	30.28	1.96	25.89
Mar.	2.52	29.63	2.57	30.14	2.61	30.65	2.52	29.76	2.64	31.06	2.42	28.40
Apr.	2.93	31.87	3.01	32.72	2.63	28.56	2.28	24.93	3.12	33.88	2.42	26.26
May.	3.35	42.02	3.41	42.64	3.23	40.45	2.99	37.83	3.43	42.85	2.94	37.30
Jun.	3.06	41.34	2.97	40.06	3.26	44.17	2.97	40.39	2.92	39.55	2.89	39.10
Jul.	3.98	62.22	3.95	62.41	3.93	54.28	3.31	47.94	4.05	63.71	3.44	50.46
Sept.	3.83	50.44	3.68	48.68	3.71	48.93	3.49	45.96	3.67	52.98	3.27	42.65
Oct.	2.32	26.04	2.45	27.88	2.33	26.42	2.27	25.62	3.81	50.29	2.08	23.51
Nov.	2.71	32.35	2.83	33.97	2.82	33.71	2.67	32.02	2.48	27.84	2.57	30.88
Dec.	3.27	44.62	3.28	44.87	2.73	37.26	2.38	32.37	3.16	43.07	2.59	35.32
Mean	3.09	40.72	3.11	40.87	2.97	39.06	2.77	36.57	3.01	40.35	2.74	36.27

Table 12: Performance of all models tested on NYISO demand data for 2010.

Month	K-means		SOM		Hierarchical		K-medoids		Fuzzy C-means		PFEM	
	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)
Jan.	4.41	400.34	3.84	351.47	4.15	378.93	3.91	356.08	3.93	359.73	3.63	335.61
Feb.	2.73	254.43	3.32	309.16	2.55	240.69	3.14	290.54	2.62	244.52	2.64	246.29
Mar.	2.54	221.64	2.77	239.93	2.32	200.71	2.43	211.67	2.13	185.98	2.15	186.79
Apr.	2.55	204.82	2.84	229.27	2.68	216.33	2.58	207.24	2.62	211.01	2.35	188.32
May.	2.48	217.79	2.74	239.92	1.96	171.71	2.35	205.07	2.93	258.05	2.06	181.64
Jun.	2.42	227.29	2.34	221.69	1.74	166.11	2.14	203.51	2.35	222.64	2.02	192.97
Jul.	2.28	217.94	2.51	240.11	1.54	150.04	1.97	190.69	2.47	236.19	1.98	188.91
Aug.	2.71	247.68	2.81	257.91	1.91	176.93	2.37	218.62	2.72	249.59	2.27	209.07
Sep.	2.75	228.26	2.58	217.47	2.00	170.51	2.39	202.92	2.64	222.79	2.02	171.29
Oct.	3.31	271.44	3.06	251.00	2.86	234.09	2.95	240.93	3.48	285.75	2.67	218.72
Nov.	2.63	218.76	3.05	254.64	2.72	226.68	2.58	214.96	2.41	200.01	2.36	197.03
Dec.	2.67	216.29	2.89	237.65	2.98	242.67	2.97	243.05	2.52	202.94	2.64	214.38
Mean	2.78	243.88	2.89	254.17	2.45	214.61	2.64	232.10	2.73	239.93	2.39	210.91

Table 13: Performance of all models tested on ANEM demand data for 2010.

Month	K-means		SOM		Hierarchical		K-medoids		Fuzzy C-means		PFEM	
	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)	MER(%)	MAE(MW)
Jan.	2.09	364.46	2.16	376.11	1.61	281.01	1.85	321.80	2.13	370.98	1.53	304.54
Feb.	1.85	318.37	1.79	309.31	1.68	289.51	1.66	287.60	1.81	312.67	1.61	275.35
Mar.	2.06	323.03	2.49	391.73	2.28	357.84	2.26	352.95	2.47	387.77	2.12	330.70
Apr.	1.80	256.00	2.79	404.61	2.02	291.81	1.95	279.72	2.75	396.48	1.82	259.58
May.	2.57	397.83	2.92	456.09	2.59	397.54	2.48	383.41	2.82	443.48	2.37	370.40
Jun.	2.76	439.81	2.46	398.17	2.38	367.08	2.41	386.29	2.79	451.16	2.25	358.47
Jul.	3.09	548.15	3.42	612.35	2.96	525.82	3.23	564.81	3.74	671.03	3.09	549.91
Aug.	2.77	478.41	3.02	533.66	2.76	480.63	2.81	484.82	3.16	562.05	2.63	462.74
Sep.	2.88	441.52	2.75	420.97	2.58	395.93	2.64	403.06	2.71	415.73	2.38	361.03
Oct.	1.67	242.83	1.98	292.05	1.85	270.93	1.83	266.62	2.25	329.57	1.66	243.50
Nov.	2.09	326.84	2.42	378.94	1.85	290.25	2.33	365.46	2.51	391.24	2.06	322.79
Dec.	1.97	332.63	2.08	352.31	1.79	301.48	1.94	327.58	1.98	333.53	1.85	304.69
Mean	2.32	372.48	2.52	410.51	2.18	354.15	2.28	368.670	2.59	422.14	2.10	345.30

Table 14: Performance of all models tested on IESO demand data for 2010.