A HYBRID CLUSTERING BASED FORMULATION OF SHORT TERM LOAD FORECASTING USING CURVE FITTING

Babita JAIN
Dept. of Electrical and Electronics Engineering, Dadi Institute of Engineering and Technology, Anakapalle, India
Email: jain.babita@gmail.com
Amit JAIN
Joint Director, Central Power Research Institute, Bengaluru, Karnataka, India
Email: dramitjain06@gmail.com

Abstract: Short Term Load Forecasting (STLF) has always been the most critical, sensitive and accuracy demanding factor of the Electricity Markets. An accurate STLF not only improves the system's economic viability but also its safety, stability and reliability. The current research supports the argument of hybrid approach whereby the complementary strengths of different intelligent techniques are combined to offer a better and complete solution to the problem of accurate STLF. This paper presents a very simple Artificial Intelligence (AI) based formulation for STLF combined with statistical techniques. The methodology basically uses the clustering technique of data mining and combines it with the curve fitting technique. The proposed framework is detailed as follows: to forecast the load of a given day, the forecast day and the similar days (all Mondays for Monday and likewise) of the forecast day, selected from the previous two years history data are clustered using the k-medoids clustering algorithm based on their similarity of weather variables i.e. temperature and humidity. Novelty is still introduced in the method by finding the curve fitting constants from the cluster matrix containing the forecast day. The load forecast for the forecast day is then carried out using the curve fitting equation. The zest of the entire work is that clustering amalgamates the very similar days of the forecast day in one cluster and curve fitting technique further encapsulates the total correlation of load and weather variables of the similar days in the cluster, hence enhancing the forecast efficiency. As two years data is good enough to capture the impact of weather variables on the load, the technique has been very successful for all days and all seasons load forecast. The STLF has been carried out using the technique for a real-time data of one year with a history data of two years and the results of all seasons have been found to be very satisfactory. A comparative result analysis has been done for the proposed technique with the Weighted Euclidean Norm based Similar Day with Fuzzy Logic Technique and Weighted Euclidean Norm based Similar Day Evolutionary Particle Swarm Optimization (EPSO) optimized Fuzzy Technique. The proposed technique shows improved performance in comparison to the others.

The results have been quite encouraging with the Mean Absolute Percentage Error (MAPE) for most of the days less than 3.0%.

Key words: *Short Term Load Forecasting, Clustering, k-medoids clustering, Curve Fitting Technique, Fuzzy Systems, MAPE, Evolutionary Particle Swarm Optimization*

1. Introduction

Load forecasting, or demand forecasting, is the process of predicting the amount of electricity demand across a region and/or a transmission network over a specified period of time. Long before the introduction of electricity markets, load forecasting had been exhaustively explored and formulated for short term and medium-term planning. Accurate load forecasting enhances the function of security control such as efficiently schedule spinning reserve allocation. Bunn [1] reported that 1% increase in the forecasting error leads to an increase of £10 million operating cost per year. Thus, to obtain economic, reliable and secure operation of the power system, accurate and timely forecasts are highly needed. With the recent move towards deregulation in the electricity industry, STLF and Very Short-Term Load Forecasting (VSTLF) became more important and are constituent to the spot market. Based on the forecasting period, load forecasting can be categorized into: Long Term Load Forecasting (LTLF) (up to 20-year period), Medium Term Load Forecasting (MTLF) (2-year period), STLF (30-minute period to one day to one-week period) and VSTL forecasting (5-minute period). Each of these classes focuses on different criteria to consider.

The most important factors for STLF include the day of the week, temperature, seasonal effect and humidity [1-3]. MTLF takes into consideration some Macro Economic indicators such as Consumer Price Index and the Average Salary Earning or Currency Exchange Rate [4]. In LTLF, influencing factors include economic aspects, political and industrial development degree [3].

Traditional STLF methods include Classical Multiply Linear Regression, Automatic Regressive Moving Average (ARMA), Data Mining models, Time-Series models and Exponential Smoothing models [5-13]. Similar-Day approach and various Artificial Intelligence (AI), based methods have also been applied [4, 5, 7, 14] to STLF. However, these models take long computational time and have some deficiencies in the case of the load changes taking place abruptly. Evolutionary and Behavioural Random

Search algorithms such as Genetic Algorithm (GA) [15-17], Particle Swarm Optimization (PSO) [18, 19], etc. have been previously implemented for different problems. In spite of its successful implementation, GA does pose some weaknesses such as longer computation time and premature convergence accompanied by a high probability of entrapment into the local optimum [20, 21].

Feed forward Neural Net structures like Multi-Layer Perceptron, Functional Link, Wavelet, Recurrent or Feedback Structures like Hopfield, Elman, Multi Feedback and Hybrid structures using Fuzzy Neural Networks have been widely proposed for non-stationary forecasting applications [22]. But in STLF, actual load data put forth many challenges to design a Predictive Neural Network. Prominent of these challenges are, data pre-processing, input parameter selection, type of neural net structure selection and training algorithm selection. Computational complexity, which is important for real time implementation of algorithms in power systems, is dependent on the structural complexity and the training algorithm. There also exist large forecast errors using ANN method when there are rapid fluctuations in load and temperatures [4, 23]. In such cases, forecasting methods using Fuzzy Logic approach have been employed. Fuzzy Logic allows one to (logically) deduce outputs from fuzzy inputs and in this sense fuzzy logic is one of a technique for mapping inputs to outputs. S. J. Kiartzis et al [24], V. Miranda et al [25], and S. E. Skarman et al [26] described applications of Fuzzy Logic to electric load forecasting as well as many others [27-29].

Off late data mining techniques have been significantly used in STLF. Especially the Clustering Technique of Data Mining has been vastly used [30-33] to group the history data in STLF. Literature survey also brings to light that clustering has been successfully combined with several other variants of AI for accurate STLF. PSO optimized Clustering [34], Clustering combined with Support Vector Machines [35, 36], Hybrid Clustering [37], ANN based Clustering [38], Wavelet Technique Supported Clustering [39] and Fuzzy Clustering [40] are few areas where AI based Clustering was implemented for STLF. The idea of the current research tends to combine these techniques to create a hybrid method, making the most of the strengths of each technique. This paper focuses mainly on four practical methods for load forecasting namely, Similar Day Approach combined with Fuzzy Theory, and Data Mining Approach combined with Regression modeling.

Furthermore, the paper also identifies the important parameters in forecasting hourly load. The discussion serves as a point of departure for a research in developing Data Mining Model for STLF.

In this paper, we propose a new approach for the STLF that combines the strengths of Clustering and the Statistical Technique of Curve Fitting. The basic idea of the paper is to cluster the history data containing the hourly load and weather variables using the k-medoids clustering technique and then forecast the hourly load from the cluster containing the forecast day using the curve fitting technique. The history data considered for clustering contains the dataset with all the similar days of the forecast day (past Mondays if the forecast day is a Monday) of past two years.

The paper is organized as follows: Section II deals with the parameter identification for STLF and the overall methodology; Section III explains the Clustering and Curve Fitting based formulation and implementation; Section IV deals with the methods used for comparison: Weighted Similar Day based Fuzzy Logic approach for STLF and EPSO tuned Weighted Similar Day Fuzzy Inference System (FIS) for STLF; Section V presents comparative study of simulation results of the three proposed forecasting methodologies followed by conclusions in Section VI.

2. Parameter Identification and Basic Methodology of STLF

A range of factors affect the electricity demand of a place, locality and a country. The analysis on the annual load and weather data helps in understanding the variables which affect the load. The data analysis is carried out on data containing hourly values of load, temperature, and humidity of 3 years. In the analysis phase, the load curves are drawn and the relationship between the load and weather variables is established [41]. Also, the correlation between the same days (all Mondays, all Tuesdays and so on) load curves has been analyzed.

Good understanding of the impact of weather conditions, like Temperature and Humidity on load demand can significantly improve the forecast accuracy.

A. Temperature

Temperature is the most important weather parameter that affects load consumption behavior. This has been proven by previous studies [42-45]. The inclusion of Temperature as one of the forecast criteria reduces forecast errors, since load consumption change is very sensitive to Temperature changes. Large load demand changes occur at the time with

large Temperature rise and fall. A sudden drop or rise of Temperature is not unusual. Fig. 1 clearly indicates a raise in Average Load with increased Temperature.
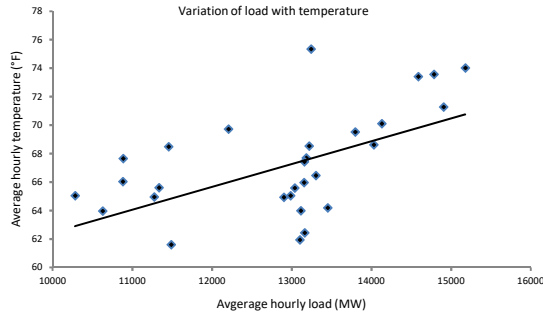

Fig. 1. Variation of Average Load with Average Temperature

B. Humidity

As addressed in [46], Relative Humidity is also important influencing factor to load demand. In summer, for a given range of Temperature, Relative Humidity is significant in affecting the utilization of air conditioning. In winter, rain is very common, which directly affects heating and lighting load because rainfall is related to both Temperature and Sky Cover. Fig. 2 shows how the increased Average Humidity has a positive impact on the Average Load.
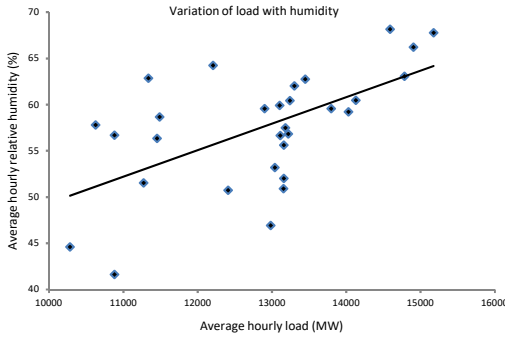

Fig. 2. Variation of Average Load with Average Humidity

C. Day Type

Another prominent factor to determine the shape of the load curve is the Calendar Date. Previous studies have attempted to divide days into different groups and have improved the performance of forecasting models. The patterns of weekdays are typically different from the ones of weekend. Thus, the basic classification falls in two groups: weekdays and weekend days [47, 48]. Further, anomalous load condition periods such as holidays (e.g. Christmas, Easter), special sports events (e.g. Football, Cricket) and long weekends (a weekend followed by the Queen's birthday) should be classified into another group [42]. Sometimes the load patterns of the days preceding the weekends and holidays are

disturbed by the following day and therefore they should be treated with extra attention [45].

Owing to all the above factors it has been understood that load of a Monday is similar to load of previous Mondays and load of a Sunday is similar to load of previous Sundays and the same holds good for all weekdays. Of course, the seasonal changes play a vital role even for the similar day's selection. The data analysis done clearly indicates (Fig. 3 and Fig. 4) that the load curve similarity is more when considering the similar days of forecast day than when considering the previous days to forecast day.
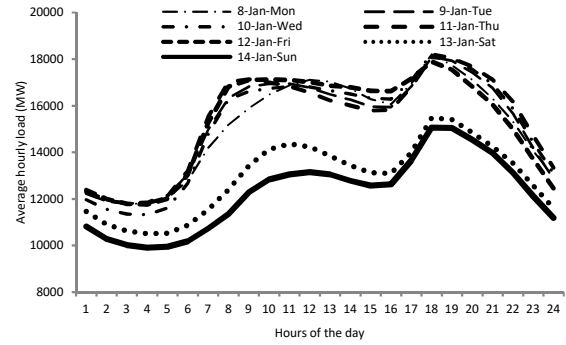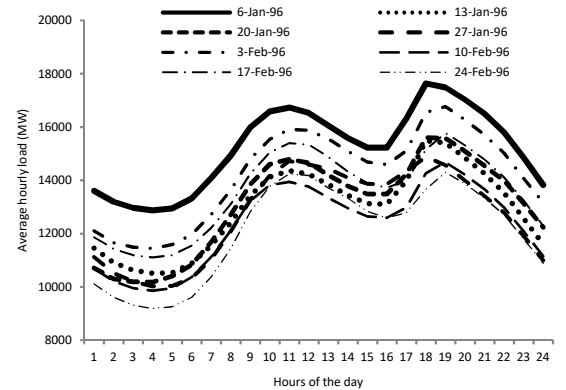

Fig. 3. Weekly Load Curve of Second week of January'96

Fig. 4. Load Curves of all Saturdays of Jan and Feb'96

D. Time of the day

It is common knowledge that electricity demand is usually higher during the day than at night. Peak load demand is generally associated with certain time of the day, when most of the events are going on. In [49], an auto-correlation analysis is applied to the historical load data and result shows that correlation of peak load demand occurs at the multiples of 24-hour lags, which indicates that loads at the same hours have strong correlation with each other. Hence the proposed STLF methodology considers the similar days on hourly basis only.

3. Clustering and Curve Fitting based Formulation of Short Term Load Forecasting

This section presents the architectural details and implementation procedure of the Clustering and Curve Fitting based formulation for the proposed STLF. The initial stages involve building the history data set for clustering. Owing to the impact of temperature and humidity on the load and also similarity in the load curves of the forecast day and its similar days from history as discussed in Section I - C, the history data set is build using the vectors of similar days of past 104 weeks taken from the 2 years of history data. Each vector comprises of the hourly temperature and hourly humidity. The data set to be clustered comprises of 105 vectors i.e. 104 history days and 1 forecast day. Each of these vectors is specified as x_i comprising of f_{it} and f_{ih}, with i ranging from 1 to 105. f_{it} represents the average hourly temperature and f_{ih} represents the average hourly relative humidity of the ith day. This data set is clustered using the k-medoids clustering technique.

A. k-medoids Clustering Algorithm

The k-medoids clustering algorithm derives its base from the famous k-means algorithm. k-means is one of the simplest unsupervised learning algorithms [50] that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k-centroids, one for each cluster. These centroids should be placed in a cunning way because, different locations cause different results. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate k new centroids as barycenter's of the clusters resulting from the previous step. After we have these 'k' new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the 'k' centroids change their location step by step until no more changes are done. In other words, centroids do not move any more.

Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function is given below:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad (1)$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster center c_j, is an indicator of the distance of the n data points from their respective cluster centers.

So far, we have only used Euclidean distance as a distance measure. However, when we have discrete multivariate data, or data that should not be clustered in circles, or data that is on different scales [51,52], as in our case where data is on different scales and is not to be clustered in circles Euclidean distance is not appropriate. Instead, we use the k-medoids algorithm, which does not require us to know the means, only distances between data points.

The k-medoids algorithm is as follows:
1. Initialization
    (a) Data is $x_{1:n}$
    (b) Pick initial cluster identities $m_{1:k}$
2. Repeat
    (a) Assign each data point to its closest center,
    $z_n = \text{argmin}_{i \in 1 \ldots k} d(x_n, m_i)$ (a)
    (b) Find a data point in a cluster that is closest to the other data points in the cluster,

$$i_k = \text{argmin}_{n:z_n=k} \sum_{m:z_m=k} d(x_n, x_m) \quad (b)$$

    (c) New cluster centers are set to the closest data points, $m_k = x_{i_k}$ (c)
3. Until assignments $z_{1:n}$ do not change

B. Implementation of K Medoids Algorithm

In the proposed methodology there are 105 sample feature vectors x_1, x_2, ..., x_{105} all from the same class, with each vector having two dimensions: Hourly Temperature and Hourly Humidity and we know that they fall into k compact clusters, k < n. Let m_p be the medoid of the vectors in cluster p. If the clusters are well separated, we can use a minimum-distance classifier to separate them. That is, we can say that x is in cluster p if $\| x - m_p \|$ is the minimum of all the k distances. Following the k medoids algorithm given above the hourly data set is clustered and finally the cluster 'C_f' comprising of 's' vectors ('s' depends on the clusters formed) which also includes the forecast day is obtained [53,54]. The curve fitting technique is applied to this cluster to obtain the w, x and y constants which are then used to forecast the load of the forecast day.

C. Curve Fitting Technique Algorithm:

The methodology that is developed for the STLF of load using the Curve Fitting method mainly considers the variation of load with the two main parameters we have already mentioned i.e. Temperature and Humidity (TH).

The basic formulation of the curve fitting for the STLF is defined below:

$$P = w + x.T + y.H \qquad (2)$$

Above relation is defining the variation of the load with respect to the T and H (Temperature and Humidity) as main parameters [55, 56].

One of the most important weather factors that influence the load to a great deal be the temperature with its impact ranging at different levels during different seasons of any particular year.

Getting in to the analytical approach of obtaining a well-defined relation between load and temperature, in this part the step by step formulation of short term load forecasting using only temperature (T) is given as:

$$P = w + x\,T \qquad (3)$$

Another vital factor that could never be ignored while considering the load variation with weather conditions is the humidity. The load variation with both (T and H) these factors taken in to account as shown below:

$$P = w + x\,T + y.H \qquad (4)$$

where,   P is the Power in Watts

         T is the Temperature in ^{o}F

         H is the Relative Humidity %

The curve fitting formulation of the load in terms of Temperature and Humidity is done as follows:

$$\sum P = w.N + x.\sum T + y.\sum H \qquad (5)$$

$$\sum P.T = w.\sum T + x.\sum T^2 + y.\sum H.T \qquad (6)$$

$$\sum P.T.H = w.\sum T.H + x.\sum T^2.Hy.\sum H^2.T \qquad (7)$$

where,

$\sum P$ is the sum of hourly load of the considered dataset

N is the total number of vectors of the considered dataset

$\sum T$ is the sum of hourly temperatures of the considered dataset

$\sum H$ is the sum of hourly relative humidity's of the considered dataset

Other terms in the above equations can be understood with a similar ontology.

From (5), (6) and (7) we get:

$$\begin{bmatrix} \sum P \\ \sum PT \\ \sum PTH \end{bmatrix} = \begin{bmatrix} N & \sum T & \sum H \\ \sum T & \sum T^2 & \sum HT \\ \sum TH & \sum T^2 H & \sum H^2 T \end{bmatrix} \begin{bmatrix} w \\ x \\ y \end{bmatrix} \qquad (8)$$

$$\begin{bmatrix} w \\ x \\ y \end{bmatrix} = \begin{bmatrix} N & \sum T & \sum H \\ \sum T & \sum T^2 & \sum HT \\ \sum TH & \sum T^2 H & \sum H^2 T \end{bmatrix}^{-1} \begin{bmatrix} \sum P \\ \sum PT \\ \sum PTH \end{bmatrix} \qquad (9)$$

Hence, $P_{forecasted} = w + x.T + y.H \qquad (10)$

D. Implementation of Curve Fitting Technique for STLF:

For the implementation of the Curve Fitting technique for STLF the cluster 'Cf' consisting of 's-1' vectors, excluding the forecast day vector is considered. For these 's-1' vectors the Hourly Load dimension of each vector is also considered in addition to the Hourly Temperature and Hourly humidity. This cluster of 's-1' vectors form the dataset on which the Curve Fitting technique is implemented. Using the (9) the w, x and y constants are obtained and the Hourly Load of a given forecast day is obtained using (10).

4. Bird's eye view of the proposed methodology for STLF:

The history dataset consists of the hourly load, hourly temperature and hourly humidity data of the years 1995 and 1996. The hourly load forecasting is done for all the days of 1997 using the proposed methodology. Basically, for forecasting a particular day's load, its hourly load is forecasted. Suppose the load forecast for first hour of Jan 1'97 is to be done: it being Wednesday the 104 previous Wednesdays' first hour temperatures and humidity data from the previous two years 1995 and 1996 form the input dataset for the clustering section of the proposed methodology. Hence the dataset for clustering consists of 105 vectors (including the forecast day since the forecast day hourly temperature and humidity are known). Clustering of the dataset is performed using the k-medoids clustering algorithm to obtain the cluster 'C$_f$' containing the forecast day with total of 's' vectors. This dataset excluding the forecast day vector, consisting of 's-1' vectors form the input for the curve fitting technique to obtain the w, x and y constants and finally the hourly load forecast of first hour of Jan 1'97 is done using (2).

The same procedure is repeated for all 24 hours of Jan 1'97 and for all days of Jan'97 and then for all months of the year 1997. The objective function considered for testing the accuracy of the proposed methodology is the Mean Absolute Percentage Error (MAPE). MAPE is defined as the forecast results

deviation from the actual values which is defined as in the (11).

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \frac{|P_A^i - P_F^i|}{P_A^i} \times 100 \qquad (11)$$

P_A, P_f are the actual and forecast values of the load. N is the number of the hours of the day i.e. 24 and i = 1, 2,….,.24.

5. Methods Used for Comparison
A. Similar Day based FIS for STLF

One of the methods used for the result comparison is the similar day based Fuzzy Inference System for Short Term Load Forecasting and it is discussed in this section. In this method initially, a set of five similar days to the forecast previous day are selected from the history data based on the Weighted Euclidean Distance norm given by (12).

$$EN = \sqrt{W_1(\Delta T_{max})^2 + W_2(\Delta H_{avg})^2 + W_3(\Delta D)^2} \qquad (12)$$

where

ΔT_{max} is given by $(T_{max} - T_{max}^p)$
ΔH_{avg} is given by $(H_{avg} - H_{avg}^p)$
ΔD is given by $(D - D^p)$

where, T_{max} and H_{avg} are the forecast day maximum temperature and average humidity respectively. Also, T_{max}^p and H_{avg}^p are the maximum temperature and average humidity of the searched previous days, D and D^p are the day type values of the forecast day and the searched previous days and w_1, w_2, w_3 are the weight factors determined by least squares method based on the regression model constructed using historical data given (13).

$$L_{peak}^t = w_0 + L_{peak}^{t-1} + w_1 T_{max}^t + w_2 H_{avg}^t + w_3 D^t \qquad (13)$$

Once the five similar days are obtained, they are used to generate five correction factors using the Fuzzy Inference System developed as follows:

The formulation of the developed Fuzzy Inference System comprises of three input membership functions with membership limits a1…a6:ΔE_L, ΔE_T, ΔE_H, average load difference, average temperature difference and average humidity difference respectively of the forecast previous day and its selected similar days and one output membership function i.e. the correction factor as shown in the Fig. 5. The limits of all these membership functions are initially fixed using the history data.

The membership function of an inferred fuzzy output variable using a fuzzy centroid defuzzification scheme translates fuzzy output statements into a crisp output value, W_k using the firing strength μ_i and the defuzzification coefficient α_i of the fuzzy rule applicable to the input data.

$$W_k = \sum_{i=1}^{27} \alpha_i \mu_i^k \Big/ \sum_{i=1}^{27} \mu_i^k \qquad (14)$$

The output value is expressed by W_k which is the correction factor for the load curve on the kth similar day to the shape on the forecast day. Now the five similar days of the forecast day are found using the same methodology explained above. W_k is applied to each similar day and corrects the hourly load curve on similar days. The forecast day load curve L (t) is then given by averaging the corrected loads on similar days.

$$L(t) = \frac{1}{N} \left[\sum_{k=1}^{N} (1 + W_k) L_s^k(t) \right] \qquad (15)$$

where $L_s^k(t)$, is the load at t'o clock on the kth corrected similar day, N is the number of similar days and t is hourly time from 1 to 24. Few example FIS rules are given in Table I.

TABLE I
FEW FUZZY RULES OF THE INFERENCE SYSTEM

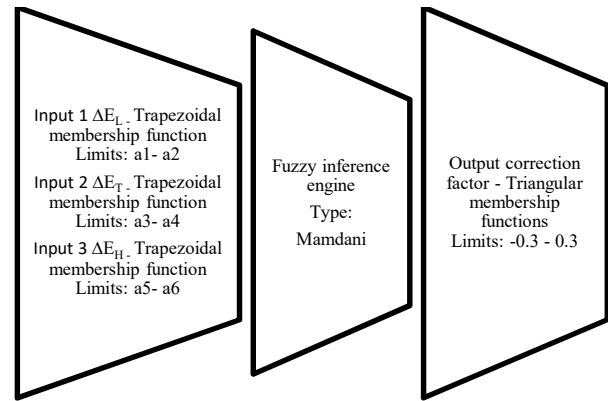| Rule No	EL	ET	EH	Output Value
R1	H	H	H	PVB (Positive Very Big)
R7	M	M	H	PB2 (Positive Big 2)
R14	M	M	M	ZE (Zero Error)
R23	L	L	H	NB1 (Negative Big 1)

Fig. 5. Overview of the fuzzy inference system

B. Similar Day based EPSO tuned FIS for STLF

The second method used for the comparison of results is the extension of the previous method. In this case the Evolutionary Particle Swarm Optimization technique is used to optimize the input membership function limits a1…..a6, of the Fuzzy Inference System (FIS) using the history data.

EPSO is a general-purpose algorithm, whose roots are in Evolutions Strategies (ES) [57-59] and in Particle Swarm Optimization (PSO) [60] concepts. The EPSO technique, a new variant in the meta-heuristic set of tools, is capable of dealing with complex, dynamic and poorly defined problems that AI has problem with, has an advantage of dealing with the nonlinear parts of the forecasted load curves, and also has the ability to deal with the abrupt change in the weather variables such as temperature, humidity and also including the impact of the day type. According to a thorough literature survey performed by authors, any application of EPSO to STLF has not been reported in literature as of today.

The idea behind EPSO [59] is to grant a PSO scheme with an explicit selection procedure and with self-adapting properties for its parameters. The variables in an EPSO formulation are divided, according to the vocabulary used in the Evolution Strategies community, composed of *object parameters* (the X variables) and *strategic parameters* (the weights w). At a given iteration, consider a set of solutions or alternatives that we will keep calling particles. A particle is a set of object and strategic parameters [X, w]. The general scheme of EPSO is the following:

Replication - each particle is replicated r times
Mutation - each particle has its weights w mutated
Reproduction - each mutated particle generates an off spring according to the *particle movement* rule
Evaluation - each offspring has its fitness evaluated
Selection - by stochastic tournament or elitist selection, the best particles survive to form a new generation

The *particle movement* rule for EPSO is that given a particle x_i, a new particle x_i^{new} results from

$$x_i^{new} = x_i + v_i^{new} \tag{16}$$

$$v_i^{new} = w_{i0} * v_i + w_{i1} * (b_i - x_i) + w_{i2} * (b_g^* - x_g) \tag{17}$$

This formulation is very similar to classical PSO – the movement rule keeps its terms of inertia, memory and cooperation. However, the weights, taken as object parameters, undergo mutation which is not the case with PSO:

$$w_{ik}^* = w_{ik} + \mu N(0,1) \tag{18}$$

where N (0, 1) is a random variable with Gaussian distribution, 0 mean and variance 1.

The global best b_g is randomly disturbed to give:

$$b_g^* = b_g + \mu' N(0,1) \tag{19}$$

The logic behind this modification from PSO is the following: a) if the current global best is already the global optimum, this is irrelevant; but b) if the optimum hasn't yet been found, it may nevertheless be in the neighbourhood and it makes all sense not to aim *exactly* at the current global best – especially when the search is already focused in a certain region, at the latter stages of the process.

The μ, μ' are learning parameters (either fixed or treated also as strategic parameters and therefore subject to mutation-fixed in the present case). This scheme benefits from two "pushes" in the right direction: the Darwinist process of selection and the particle movement rule; therefore, it is natural to expect that it may display advantageous convergence properties when compared to ES or PSO alone.

A. EPSO implementation for FIS optimization

Optimization of the fuzzy parameters a1......a6 is done using the EPSO. For the data set considered the fuzzy inference system has been optimized for six parameters (maxima and minima of each of the input fuzzy variable E_L, E_T, E_H), considering 49 particles. Hence each particle is a six dimensional one. The initial values of the fuzzy inference system are obtained by using the history data set. These values are incorporated into the FIS to obtain the forecast errors of forecast previous month (in the example case forecast previous month is June '97).

The EPSO tuner function accepts the training data i.e. 120 days, and the objective is to reduce the RMS MAPE error of the 30 forecast days (June '97) using the 90 days history data (May 95, 96 and 97). The MAPE is taken as the fitness function and the EPSO tuner is run for 50 iterations (by then the RMS MAPE is more or less fixed and comes less than 3%). After every iteration, the EPSO tuner updates the latest particle position using the optimizer equations based on the PBest and Gbest of the previous iteration if the fitness function value is better than the previous one. The parameters thus obtained after the EPSO optimization are the final input parameters of the

designed FIS. These fuzzy parameter values are set as the input parameter limits of the FIS and this FIS is used to forecast the load of the current forecast month (in the example case current forecast month is July '97).

B. Forecast of current forecast month load

The data of the current forecast month (Example: July1 to July30) is taken as the testing dataset for the problem at hand. The short-term load forecasting for the month of July is now done using the FIS optimized by the EPSO i.e. EPSO-FIS. The five similar days are selected from the history 90 days (for July1'97 the history 90 days are June'97, June'96, June'95) of the forecast day and the hourly correction factors to these similar days are obtained by the five similar days of the forecast previous day (for June 30 the previous 90 days are May31 to June 29 of 97, 96 and 95) and the EPSO-FIS. These five correction factors are then applied to the five similar days of the forecast day and the average of the corrected five values is considered as the load forecasting for each hour. The same procedure is done for all the 24 hours of the day. The same procedure is followed for all days of July i.e. Jul 1 to Jul 30.

The MAPE is calculated for each day of the 30 days of forecast of the Jul data (using the actual hourly values and the forecast hourly values). The MAPE is less than 0.03 for maximum days of forecast of the Jul month. The FIS is formulated and optimized for every month of 1997 using the same methodology and is then implemented for the load forecasting of all the months of 1997 year. The results obtained for the STLF using EPSO-FIS have been quite satisfactory.

6. Comparative Simulation Results

This section of the paper deals with detailed analysis and possible outcomes of the results of the carried research work. The performance of the proposed clustering and curve fitting based STLF is compared with the earlier works of the authors, Similar Day Based FIS and Similar Day based EPSO tuned FIS which is tested by using the 38 months data, Nov'94 to Dec'97 of a real data set. The clustering, curve fitting and EPSO implementation has been done using the MATLAB coding and the Fuzzy Inference System has been developed using fuzzy logic toolbox available in MATLAB and load forecasting is done for the all days of all months of the year 1997.

The Table II gives the various fuzzy parameter limits obtained for the Similar Day FIS and EPSO tuned FIS used for the STLF. Table III gives the values of various parameters used in the EPSO algorithm.

TABLE II
PARAMETERS OF MEMBERSHIP FUNCTIONS OF FUZZY VARIABLES

Parameters of Membership	(a1, a2)	(a3, a4)	(a5, a6)
Values for Fuzzy Stand Alone	(-1000,1000)	(-20,20)	(-20,20)
Values for EPSO Optimized FIS	(-1959,6439)	(-7, 17)	(-33,44)
Parameters of membership functions of output variables	(b1, b2, b3)	(-0.3, -0.25, -0.2)	
	(b3, b4, b5)	(-0.2, -0.15, -0.10)	
	(b4, b5, b6)	(-0.15, -0.10, -0.05)	
	(b5, b6, b7)	(-0.10, -0.05,0)	
	(b6, b7, b8)	(-0.05,0,0.05)	
	(b7, b8, b9)	(0,0.05,0.1)	
	(b8, b9, b10)	(0.05,0.1,0.15)	
	(b9, b10, b11)	(0.1,0.15,0.2)	
	(b10, b11, b12)	(0.15,0.2,0.25)	
	(b11, b12, b13)	(0.2,0.25,0.3)	

TABLE III
PARAMETERS OF THE EPSO ALGORITHM

Parameters	EPSO
Population Size	49
Number of Iterations	20
w_{i0}^{*} (initial)	0.6
w_{i1}^{*} (initial)	0. 1
w_{i2}^{*} (initial)	0.3
$\mu=\mu`$	1.5

Table IV, V and VI presents the forecast error in Mean Absolute Percentage Error for a week each in winter, spring and summer seasons respectively. This reflects the behavior of the developed techniques during seasonal changes. The index used for testing the performance of forecasters is the MAPE. The designed technique is used to forecast the week ahead forecast on an hourly basis.

TABLE IV
COMPARATIVE PERFORMANCE (WINTER, Feb. 06-12)

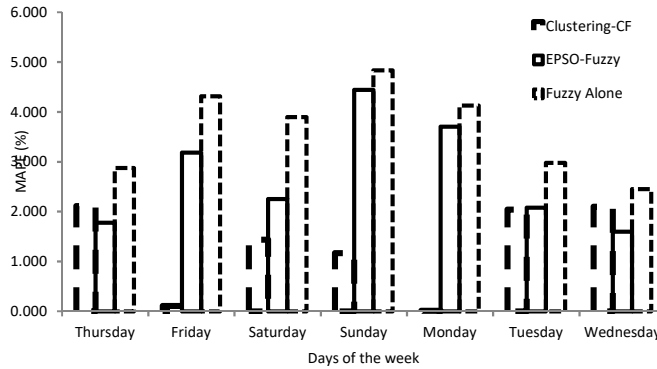| Day	Average Load (MW)	MAPE (%) Clustering-CF	MAPE (%) EPSO-Fuzzy	MAPE (%) Fuzzy Alone
Thursday	13977	2.111	1.775	2.876
Friday	14193	0.105	3.183	4.317
Saturday	13358	1.434	2.250	3.896
Sunday	12744	1.169	4.443	4.831
Monday	14408	0.012	3.702	4.126
Tuesday	14621	2.040	2.078	2.978
Wednesday	14524	2.100	1.599	2.453
Average		1.281	2.719	3.640

Fig. 6. Comparative chart showing accuracy of different approaches for winter

TABLE V
COMPARATIVE PERFORMANCE (SPRING, Apr 28 - MAY 04)

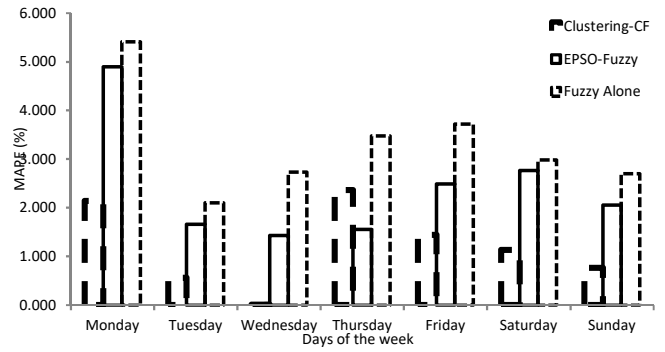| Day	Average Load (MW)	MAPE (%) Clustering-CF	MAPE (%) EPSO-Fuzzy	MAPE (%) Fuzzy Alone
Monday	12748	2.141	4.899	5.413
Tuesday	12406	0.564	1.661	2.098
Wednesday	12300	0.001	1.431	2.732
Thursday	12306	2.361	1.558	3.475
Friday	12034	1.445	2.486	3.722
Saturday	11090	1.137	2.764	2.980
Sunday	10199	0.768	2.056	2.701
Average		1.203	2.408	3.303

Fig. 7. Comparative chart showing accuracy of different approaches for Spring/Fall

TABLE VI
COMPARATIVE PERFORMANCE (SUMMER, August 20-26)

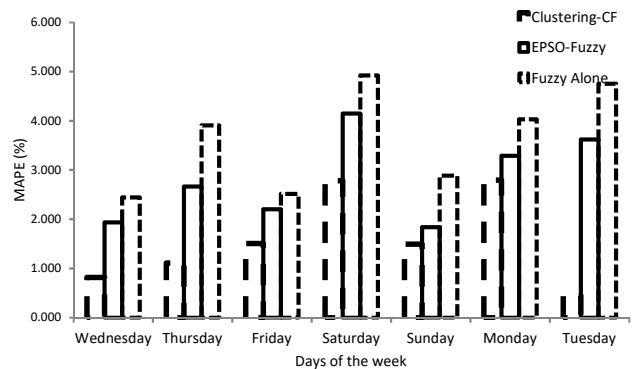| Day	Average Load (MW)	MAPE (%) Clustering-CF	MAPE (%) EPSO-Fuzzy	MAPE (%) Fuzzy Alone
Wednesday	13375	0.816	1.937	2.443
Thursday	13387	1.109	2.664	3.907
Friday	13304	1.505	2.200	2.512
Saturday	11825	2.780	4.149	4.920
Sunday	11054	1.490	1.836	2.887
Monday	13470	2.795	3.287	4.029
Tuesday	13813	0.448	3.624	4.756
Average		1.563	2.814	3.636

Fig. 8. Comparative chart showing accuracy of different approaches for summer

Forecasting has been done on one-year load data of ISO New England. Load varies from 13000 MW to 17000 MW. It is observed (Figs. 6-8) that the proposed Clustering and CF based formulation for

STLF works well in all the seasons irrespective of the variation in temperature. However, the EPSO tuned FIS is also good for all the seasons. For having a comparative study, the proposed clustering CF method is compared with other two methods, EPSO-FIS (EPSO tuned Fuzzy Inference System) and Similar Day FIS (Similar Day Approach corrected by Fuzzy Inference System). Comparison has been done for the same set of data and for the same period of time. In winters there is wide variation in temperatures and therefore in the loads also. It is observed that the forecaster captures the load shape quite accurately and the forecasting errors on most of the days including weekends are very low. The accuracy of the proposed Clustering Curve Fitting (CF) based model is better than the EPSO tuned FIS and Similar Day based FIS for all seasons of the year. With Clustering CF based formulation, MAPE is the minimum and this demonstrates the superiority of the Clustering Curve Fitting method for STLF.

7. Conclusions

A Clustering and Curve Fitting based formulation has been developed using MAPE as an objective function. STLF method, proposed above is feasible and effective. Comparative study shows that the proposed Clustering CF based approach for STLF is better and is giving more accurate results than the other two methods, EPSO-FIS and Similar Day FIS for the same period of time and same set of data. The error depends on many factors such as homogeneity in data, network parameters, choice of model and the type of solution. This approach requires fewer amounts of data and also does not require training. As the history data considered for the proposed methodologies includes two years all seasons data the Clustering CF load forecaster captures the load shape quite accurately and hence forecasting errors on most off the days is much less than 3.0%. Forecasting result shows that Clustering CF is very good for week ahead forecasting and it can also forecast future loads with very good accuracy, whereas the training time required for the method zero. The proposed method is computationally simpler than a Swarm Intelligence and Fuzzy System and also aids in selecting the best set of independent variables due to its speed and also giving better accuracy than the other two methods.

Authors hope that the proposed methodology will further propagate research for STLF using new optimization techniques to get even more improvement in forecasting results.

9. References

1. Bunn, D. W.: *Forecasting Loads and Prices in Competitive Power Markets.* In: Proceedings of the IEEEE, 2000, vol. 88, pp.163-169.
2. Bunn, D. W.: *Short Term Forecasting: A review of procedures in the electricity supply industry.* In: Journal of the Operational Research Society, 1982, Vol. 33, pp.533-545.
3. Lambert Torres, G., Traore C. O.: *Short -term load forecasting using a fuzzy engineering tool.* In: Proceedings of the First International Forum on Applications of Neural Networks to Power Systems, 1991.
4. Gavrilas, M., Ciutea, I.: *Medium-term load forecasting with artificial neural network models.* In: 16th Int. Conf. on Electricity Distribution, 2001.
5. Papalexopoulos, A. D. and Hesterberg, T. C.: *A regression-based approach to short-term load forecasting.* In: IEEE Transactions on Power Systems, vol. 5, pp. 1535-1550, 1990.
6. Haida, T. and Muto, S.: *Regression based peak load forecasting using a transformation technique.* In: IEEE Transactions on Power Systems, vol. 9, pp. 1788-1794, 1994.
7. Rahman, S. and Hazim, O.: *A generalized knowledge-based short-term load-forecasting technique.* In: IEEE Transactions on Power Systems, vol. 8, pp. 508-514, 1993.
8. Huang, S. J. and Shih, K. R.: *Short-term load forecasting via ARMA model identification including nongaussian process consideration.* In: IEEE Transactions on Power Systems, vol. 18, pp. 673-679, 2003.
9. Wu, H. and Lu, C.: *A data mining approach for spatial modeling in small area load forecast.* In: IEEE Transactions on Power Systems, vol. 17, pp. 516-521, 2003.
10. Rahman, S, and Shrestha, G.: *A priority vector based technique for load forecasting.* In: IEEE Transactions on Power Systems, vol. 6, pp. 1459-1464, 1993.
11. Moghram, I. and Rahman, S.: *Analysis Evaluation of Five Short Term Load Forecasting Techniques.* In: IEEE Transactions on Power Systems, vol. 4, 1989.
12. Lru, K., Subbarayan, S., Shoults, R.R., Manry, M. T., Kwan, C., Lewis, F. W. and Naccarino, J.: *Comparison of very short-term load forecasting techniques.* In: IEEE Transactions on Power Systems, vol. 11, 1996.

13. Eugene, A. F. and Genethliou, D.: *Applied Mathematics for Power Systems: Load Forecasting.*

14. Kun-Long Ho, Yuan-Yih Hsu, Chih-Chien Liang and Tsau-Shin Lai: *Short-Term Load Forecasting of Taiwan Power System Using a Knowledge-Based Expert Systems.* In: IEEE Transactions on Power Systems, vol. 5, Nov. 1990.

15. Holland, J. H.: *Adaptation in Natural and Artificial Systems* Ann Arbor, MI: Univ. Michigan Press, 1975.

16. Pham, D. T. and Karaboga, D.: *Intelligent Optimization Techniques, Genetic Algorithms, Tabu Search, Simulated Annealing and Neural Networks.* In: New York: Springer-Verlag, 2000.

17. Davis, L.: *Handbook of Genetic Algorithm.* New York: Van Nostrand Reinhold, 1991.

18. Eberhert, R. C. and Kennedy, J.: *A new optimizer using particle swarm theory.* In: Proceeding of the Sixth International Symposium on Micro Machine and Human Science, pp. 39-43, Nagoya, Japan, 1995.

19. Park, J. B., Lee, K. S., Shin, J. R. and Lee, K. Y.: *A particle swarm optimization for economic dispatch with non-smooth cost function.* In: IEEE Transactions on Power Systems, vol. 20, pp. 34–42, 2005.

20. Fogel, D. B.: *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence.* In: Second ed., IEEE Press, Piscataway, NJ, 2000.

21. Eberhart, R.C. and Shi, Y.: *Comparison between genetic algorithms and particle swarm optimization.* In: Proc. IEEE International Conf. Evolutionary Computing, May, 1998, pp. 611–616.

22. Savran, A.: *Multifeeback-layer Neural Network.* In: IEEE Transactions on Neural Networks, vol. 18, pp. 373-384, 2007.

23. Lu, C. N. and Vemuri, S.: *Neural network based short term load forecasting.* In: IEEE Transactions on Power Systems, vol. 8, pp. 336-342, 1993.

24. Kiartzis, S. J. and Bakirtzis, A. G.: *A Fuzzy Expert System for Peak Load Forecasting: Application to the Greek Power System.* In: Proceedings of the 10th Mediterranean Electro Technical Conf., 2000, pp. 1097-1100.

25. Miranda, V. and Monteiro, C.: *Fuzzy Inference in Spatial Load Forecasting.* In: Proceedings of IEEE Power Engineering Winter Meeting, 2, 2000, pp. 1063-1068.

26. Skarman, S. E. and Georgiopoulous, M.: *Short-Term Electrical Load Forecasting using a Fuzzy ARTMAP Neural Network.* In: Proceedings of SPIE, 1998, pp. 181-191.

27. Senjyu, T., Mandal, P., Uezato, K., and Funabashi, T.: *Next Day Load Curve Forecasting using Hybrid Correction Method.* In: IEEE Transactions on Power Systems, vol. 20, pp. 102-109, 2005.

28. Mastorocostas, P. A., Theocharis, J. B. and Bakirtzis, A. G.: *Fuzzy modeling for short term load forecasting using the orthogonal least squares method.* In: IEEE Transactions on Power Systems, vol. 14, pp. 29-36, 1999.

29. Chow, M. and Tram, H.: *Application of fuzzy logic technology for spatial load forecasting.* In: IEEE Transactions on Power Systems, vol. 12, pp. 1360-1366, 1997.

30. Ajay Shekhar Pandey, Singh, D. and Sinha, S. K.: *Clustering based formulation for Short Term Load Forecasting.* In: World Academy of Science, Engineering and Technology, 2008.

31. Michel Misiti, de Math´ematique and Ecole Centrale: *Optimized clusters for disaggregated electricity load forecasting.* In: Revstat– Statistical Journal Volume 8, Number 2, November 2010, 105–124.

32. Marcelo Espinoza, Caroline Joye, Ronnie Belmans: *Short- Term Load Forecasting, Profile Identification and Customer Segmentation: A Methodology based on Periodic Time Series.*

33. Yuhang Yang, Yao Meng, Yingju Xia, Yingliang Lu, and Hao Yu: *An Efficient Approach for Short Term Load Forecasting.* In: Proceedings of the International Multi Conference of Engineers and Computer Scientists 2011 Vol I, IMECS 2011, March 16-18,2011, Hong Kong.

34. Song Yufei and Jiang Chuanwen: *Power system short-term load forecasting based on PSO clustering analysis and Elman neural network.* In: Proceedings of the 5th WSEAS Int. Conf. on Simulation, Modeling and Optimization, Corfu, Greece, August 17-19, 2005, pp268-273.

35. Jain, A. and Satish, B.: *Short term load forecasting by clustering technique based on daily average and peak loads.* In: Power and Energy Society General Meeting, 2009, PES '09, IEEE.

36. Jain, A. and Satish, B.: *Clustering based Short Term Load Forecasting using Support Vector Machines.* In: Power Tech, 2009 IEEE Bucharest.

37. Sfetsos, A.: *Short-term load forecasting with a hybrid clustering algorithm.* In: Generation, Transmission and Distribution, IEE Proceedings, 13 May 2003.

38. Liu Jin, Yu Feng and Yu Jilai: *Peak Load Forecasting Using Hierarchical Clustering and RPROP Neural Network.* In: Power Systems Conference and Exposition, 2006 IEEE.

39. Pandey, A. S., Singh, D. and Sinha, S. K.: *Intelligent Hybrid Wavelet Models for Short-Term Load Forecasting Power Systems.* In: IEEE Transactions on Power Systems, Aug. 2010, Volume: 25.

40. Hinojosa, V. H. and Hoese, A.: *Short-Term Load Forecasting Using Fuzzy Inductive Reasoning and Evolutionary Algorithms Power Systems.* In: IEEE Transactions on Power Systems, 2010, Volume: 25, Issue:1

41. Fogel, D.B.: *Evolving Artificial Intelligence* Ph.D. Thesis, University of California, San Diego, 1992.

42. Charytoniuk, W., Chen, M. S.: *Nonparametric Regression Based Short-Term Load Forecasting.* In:

IEEE Transaction on Power Systems Vol.13, 3, pp.725-730.

43. Chow, T. W. S. and C. T. Leung: *Nonlinear autoregressive integrated neural network model for short-term load forecasting.* In: IEE Proceedings on Generation, Transmission and Distribution Vol.143, 5, pp.500 -506.

44. Nakirtzis, A. G., Petridis, V., Kiartzis, S. J., Alexiadis, M.C. and Malssis, A. H.: *A Neural Network Short Term Load Forecasting Model for The Greek Power System.* In: IEEE Transactions on Power Systems, Vol. 11, No. 2, pp. 858-863.

45. Alfuhaid, A. S., El-Sayed, M. A., and Mahnoud, M. S.: *Cascaded Artificial Neural Networks for Short-term Load Forecasting.* In: IEEE Transactions on Power Systems, Vol. 12, No. 4, pp.1524-1529.

46. Lamedica, R., Prudenzi, A., Sforna, M., Caciotta, M., and Orsolini M.: *A Neural Network Based Technique for Short -Term Forecasting of Anomalous Load Periods.* In: IEEE Transactions on Power Systems, Vol. 11, No. 4, pp. 1749-1756.

47. Drezga, I. and Rahman, S.: *Short-Term Load Forecasting with Local ANN Predictors.* In: IEEE Trans on Power Systems, Vol. 14, pp. 844-850, No. 3.

48. Lee, K. Y.: *Short-Term Load Forecasting Using an Artificial Neural Networks.* In: IEEE Transactions on Power Systems, Vol. 8, No. 1, pp. 124-131, Feb. 1992.

49. National Institute of Economic and Industry Research, *Factors Affecting Electricity Demand in the NEM.* http://www.nemmco.com/publications/soo on 19 May 2001.

50. Xie, J. Y. and Gao, R.: *K-medoids Clustering Algorithms with Optimized Initial Seeds by Variance.* In: Journal of Frontiers of Computer Science and Technology, No. 8, pp. 973984, 2015.

51. Park, H. S. and Jun, C. H.: *A Simple and Fast Algorithm for K-medoids Clustering.* Expert Systems with Applications, Vol. 36, No. 2, pp. 33363341, 2009.

52. Reynolds, A. P., Richards, G. and Rayward Smith, V. J.: *The application of K-Medoids and PAM to the clustering of rules.* Lecture Notes in Computer Science, vol. 3177, pp. 173-178, 2004.

53. Zongyun Song, Dongxiao Niu, Xinli Xiao and Han Wu: *Daily Peak Load Forecasting Based on Fast K-medoids Clustering, GARCH Error Correction and SVM Model.* In: Journal of Applied Science and Engineering, Vol. 19, No. 3, pp. 249258, 2016.

54. Fintan McLoughlin, Aidan Duffy and Michael Conlon: *A clustering approach to domestic electricity load profile for characterization using smart metering.* In: http://dx.doi.org/10.1016, Published by Elsevier Ltd.

55. Farahat, M. A. and Talaat, M.: *A New Approch for short term load forecasting using curve fitting prediction optimized by genetic algorithm.* In: MEPSC, Cario University, Egypt, December 19-21,2010.

56. Hiroyuki Mori and Noriyuki Kosemura: *Optimal regression Tree based rule Discovery for short term load forecasting.* In: IEEE, 2001.

57. Fogel, D.B.: *Evolving Artificial Intelligence* Ph.D. Thesis, University of California, San Diego, 1992.

58. Kennedy, J. and Eberhart, R. C.: *Particle Swarm Optimization.* In: IEEE International Conference on Neural Networks, Pert, Australia, IEEE Service Center, Piscataway, NJ., 1995.

59. Fukuyama, Y. and Yoshida, H.: *A particle swarm optimization for reactive power and voltage control in electric power systems.* In: IEEE Proc. of Evolutionary Computation 2001, vol.1, pp. 87 -93, 2001.

60. Yoshida, H., Fukuyama, Y., Takayama, S. and Nakanishi, Y.: *A particle swarm optimization for reactive power and voltage control in electric power systems considering voltage security assessment.* In: IEEE Proc. of SMC '99, vol. 6, pp.497 -502, 1999.