Arash ASRARI, Dawood Seyed JAVAN, Mohammad Hossein JAVIDI, Mohammad MONFARED

Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

Application of Gray-Fuzzy-Markov Chain Method for Day-Ahead Electric Load Forecasting

***Abstract**. Short-term load forecasting (STLF) plays a decisive role in electric power system operation and planning. Accurate load forecasting not only reduces the generation costs of power systems, but also serves to maximize profit for participants in electricity markets. In recent years, power markets have grown more deregulated and competitive, adding to the complexity and uncertainties of load, and making it more difficult for conventional techniques to accurately forecast the load. To improve the accuracy of load forecasting, this paper suggests a hybrid method, called Gray-Fuzzy-Markov Chain Method (GFMCM), comprising three stages. In the first stage, daily load is forecasted by Gray model, with its training deviations classified, in a second stage, by fuzzy-set theory, and finally, fed into Markov chain model to predict future relative errors that might be supplied by the Gray model. The proposed approach has been verified by the historical data of power consumption in Ontario, PJM and Iranian electricity markets. The obtained forecasts by GFMCM proved to have better prediction properties compared to the other forecasting techniques, such as Gray models, specifically GM(1,1) and GM(1,2), ARIMA time series, wavelet-ARIMA and multi-layer perceptron (MLP) neural network.*

***Streszczenie.** W celu poprawy jakości przewidywania zużycia energii autorzy zaproponowali hybrydową metodę GMMCM (Gray-Fuzzy-Markov Chan Method). W pierwszym etapie prognoza obciążeń jest prowadzona przy wykorzystaniu modelu Gray, następnie stosuje się metody logiki rozmytej. Błąd prognozowania analizowany jest metodą Markova. (**Zastosowanie kombinowanej metody: Gray-Fuzzy-Markov do prognozowania obciążenia sieci elektrycznej**)*

Keywords: Short-term load forecasting, Gray model, Fuzzy clustering, Markov chain model, Transition probability matrix.
Słowa kluczowe: krótkoczasowe prognozy obciażenia, model Gray'a, model Markova, logika rozmyta.

1. Introduction

The short-term load forecasting (STLF) is essential for managing the effective operation of power systems. Consequently, precise load prediction is necessary not only for power system operation, e.g. economic load dispatching (ELD), unit commitment, etc., but also it provides a good insight for participants in the electricity market.

During the last two decades, a wide variety of methods have been proposed due to the importance of short-term load forecasting. In early stages, the bulk of the research focused on time-series methods [1]–[6], which are still widely used due to their stability and efficiency. However, an important shortcoming associated with these methods is that they tend to ignore the impact of relative factors. A solution to this problem can be achieved when the nonlinear relationship between load and relative factors is identified. Hence, much effort has in later stages been directed toward intelligent algorithm procedures which are capable of dealing with multivariable nonlinear regression, such as artificial neural networks (ANN) [7]-[11], support vector machines (SVM) [12]-[13], fuzzy theory [14]-[17], particle swarm optimizer [18], genetic algorithm [19], and wavelets in combination with ANN [20]-[22].

With the development of power industry, load composition and its regularities has become ever more complex, and it has been recognized that one single method may not be capable of load forecasting under such a diverse and complicated condition. Recently, hybrid forecasting methods have become a focus of attention by some researchers [23]–[26]. In these studies, load forecasting procedure was divided into a number of stages, bringing about more flexible forecasting procedures.

Cognitively speaking, information systems can usually be categorized into three classes, such as a white system, a gray system and a black system. A black system is a completely unknown system. By contrast, a white system is fully identified. A gray system is neither black nor white, i.e. in this case, only partial information is known. Deng (1982) was the first one who proposed the application of the Gray system theory (GST or GS) in different fields of technology. The electric load, as a dynamic process, is random and uncertain in nature. In practice, we do not know or cannot determine all effective factors on the load variations. Therefore, the electric load forecast system is a Gray system [27] which may be analyzed by GST. GM(1,1) forecasting model is a basic Gray model that has been successfully employed to predict the electric load variations [28]-[32]. However, this model is most appropriate for prediction of time series with exponential variations, e.g. for prediction of financial ratios of national health insurance [33], development of cell phone market [34], number of Chinese international airlines [35], trend of river water pollution [36] and the global photovoltaic market [37]. Nevertheless, GM(1,1) cannot yield proper results for fluctuating time series, while the GM(1,2) model can fluctuate over the original series (an issue which will be explored in section 2.2). Considering the fluctuating and non-monotonic nature of electric load, we have adopted the GM(1,2) model as a basis for one-day ahead load forecasting, and at a later stage we have used fuzzy clustering and Markov correction to get more accurate results. The proposed hybrid forecasting technique can successfully handle the information of available historical data, compensate for the shortcomings of each of the constitutive models in isolation and increase the forecasting precision.

Markov method predicts the future state of the system by transition probability matrix between different system states. In this way, it can reflect the influence of random factors and the interval law. A number of researchers have worked on this aspect for prediction of power load, e.g. [38], and electricity price, e.g. [39]. In the present study, a Markov chain model has been introduced into the Gray model to reduce the effect of errors. Fuzzy classification has been incorporated into the Markov model to boost its function.

The combined model was called Gray-Fuzzy-Markov Chain Method (GFMCM), which consists of three stages. In the first step, Gray model is developed to fit the historical statistical data. In the next stage, fuzzy classification is used to categorize the fitting errors of the Gray model. According to the fuzzy theory, a certain sample can belong to different clusters or states with various membership degrees. As the variable varies slightly, only its membership degrees in each state vary, resulting in a more accurate classification of variations and thus less severe computation error.

Finally, classified errors output by the fuzzy-set theory are entered into Markov chain model to predict future errors that might be supplied by the Gray model. Application of the method in load forecasting, using historical data of power consumption in Ontario, PJM and Iranian electricity markets confirms the potential of the GFMCM in dealing with electric load time series prediction.

To demonstrate the effectiveness of the proposed technique, the obtained load forecasting results are compared with results produced by other forecasting methods, such as Gray models- GM(1,1) and GM(1,2), ARIMA time series, wavelet-ARIMA and multi-layer perceptron (MLP) neural network. The contributions of this paper elaborated in the following sections are:

A. The procedure through which 24 Gray models are assigned to 24 hours of a day in order to improve accuracy of forecasting;

B. The procedure through which fuzzy approach is utilized to set a link between Gray model and Markov Chain model;

C. The strategy based on which membership vectors of Markov Chain model are computed in order to correct the error of Gray model forecasting.

2. The Gray Model

Gray system theory was proposed by Deng in 1982. He called any random process a gray process and assumed that all gray variables change with certain amplitudes in specified ranges and in a certain time zone [40]. Basically, gray system theory focuses on using a definite amount of available information to build a “Gray Model” (GM) in order to approximate the dynamic behaviour of a system [41]. It is based on GM(n,h), where n is the order of the differential equation, and h is the number of variables. Due to the poor regularity, the accumulated generating operation (AGO) technique is utilized in Gray forecasting to efficiently decrease the uncertainty of raw data.

2.1. Gray Prediction Model GM(1,1)

GM(1,1) refers to a kind of single-variable sequence and a first-order linear gray model [42]. The Gray forecasting method essence is to cumulate the original data sequence without obvious law to generate an obvious exponential law. These data are fitted into a curve to forecast variables. GM(1,1) is composed of a first order differential equation with only a single variable. The procedure to build a GM(1,1) can be summarized as follows [43]:

1) Set up a row matrix $x^{(0)}$ for the non-negative historical time-series data:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n)) \tag{1}$$

2) Build a row matrix $x^{(1)}$ by AGO:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), ..., x^{(1)}(n)) \tag{2}$$

where;

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \qquad (k = 1,2,...,n) \tag{3}$$

As $x^{(1)}(k)$ sequence corresponds to an exponential growth rule, it can be supposed that $x^{(1)}$ sequence satisfies a first order differential equation.

3) Form the whitening equation as follows:

$$\frac{dx^{(1)}}{dk} + a \quad x^{(1)} = b \tag{4}$$

where, a is a developing coefficient, and b is a control variable.

4) Solve for a and b by least-square method (LSM):

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1}(B^T Y) \tag{5}$$

where;

$$B = \begin{bmatrix} -0.5(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -0.5(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ . & . \\ . & . \\ . & . \\ -0.5(x^{(1)}(n-1) + x^{(1)}(n)) & 1 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ . \\ . \\ . \\ x^{(0)}(n) \end{bmatrix} \tag{6}$$

5) Solve the differential equation:

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{b}{a}) \quad e^{-a.k} + \frac{b}{a} \tag{7}$$

$$(k = 0,1,...,n-1)$$

6) Take inverse accumulated generating operation (IAGO) for the first-order series. AGO and IAGO are a pair of inverse sequence operators, so the inverse accumulated generating operation returns the data to the original condition as in Eq. (8).

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \tag{8}$$

in which, k=1, 2,…, n-1 and $\hat{x}^{(0)}(1) = x^{(0)}(1)$.

2.2. Gray prediction model GM(1,2)

In GM(1,2), two types of sequences, i.e. the main sequence (MS) and the reference sequence (RS) are defined [44]. Assume that the main raw sequence $P^{(0)}$ and the reference raw sequence $R^{(0)}$ are defined as follows:

$$\begin{cases} P^{(0)} = (P^{(0)}(1), P^{(0)}(2), ..., P^{(0)}(n)) \\ R^{(0)} = (R^{(0)}(1), R^{(0)}(2), ..., R^{(0)}(n+1)) \end{cases} \tag{9}$$

Then, by defining:

$$\begin{cases} P^{(1)}(k) = \sum_{i=1}^{k} P^{(0)}(i) & (k = 1,2,...,n) \\ R^{(1)}(k) = \sum_{i=1}^{k} R^{(0)}(i) & (k = 1,2,...,n+1) \end{cases} \tag{10}$$

We obtain the new series as follows using AGO:

$$\begin{cases} P^{(1)} = (P^{(1)}(1), P^{(1)}(2), ..., P^{(1)}(n)) \\ R^{(1)} = (R^{(1)}(1), R^{(1)}(2), ..., R^{(1)}(n+1)) \end{cases} \tag{11}$$

$P^{(1)}$ is the solution of the following Gray ordinary differential equation (the relevant whitening equation):

$$\frac{dP^{(1)}(k)}{dk} + a \quad P^{(1)}(k) = b \quad R^{(1)}(k) \tag{12}$$

where, a and b are called developing coefficient and Gray input, respectively. Eq. (12) is referred to as GM(1,2). Here, parameters a and b are usually determined by the least squares method as follows:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1}(B^T Y) \tag{13}$$

where,

$$B = \begin{bmatrix} -0.5(P^{(1)}(1)+P^{(1)}(2)) & R^{(1)}(2) \\ -0.5(P^{(1)}(2)+P^{(1)}3)) & R^{(1)}(3) \\ . & . \\ . & . \\ . & . \\ -0.5(P^{(1)}(n-1)+P^{(1)}(n)) & R^{(1)}(n) \end{bmatrix}, Y = \begin{bmatrix} P^{(0)}(2) \\ P^{(0)}(3) \\ . \\ . \\ . \\ P^{(0)}(n) \end{bmatrix} \tag{14}$$

If rank (B)=2, the Eq. (12) has a unique solution:

$$\hat{P}^{(1)}(k+1) = (P^{(0)}(1) - \frac{b}{a} \quad R^{(1)}(k+1)) \quad e^{-a.k} + \frac{b}{a} \quad R^{(1)}(k+1) \tag{15}$$

in which, k=0, 1, 2,…, n-1.

As it can be seen from Eq. (15), unlike GM(1,1) (see Eq. (7)), the calculated series with GM(1,2) model always fluctuate over the original series.

One can compute the main sequence values as:

$$\hat{P}^{(0)}(k+1) = \hat{P}^{(1)}(k+1) - \hat{P}^{(1)}(k) \tag{16}$$

where, k = 1, 2,…, n-1, and $\hat{P}^{(0)}(1) = P^{(0)}(1)$.

As mentioned before, the GM(1,2) model has a unique solution if the rank (B) equals 2; however, in some processes, this condition is not satisfied. Under such circumstances, we need to use other techniques like Genetic Algorithm (GA) or Particle Swarm Optimization (PSO) rather than LSM.

In addition, in the case of short-term load forecasting, it is important to properly choose the reference sequence in the GM(1,2) model. The reference sequence should meet the following conditions:

- The specified reference sequence must have a large correlation with the main sequence.
- The value of the reference sequence should be known at the forecasting point.

The definition of correlation coefficient (r) in Gray system theory is formulated as follows:

$$r = \frac{n\sum_{i=1}^{n} MS(i)RS(i) - (\sum_{i=1}^{n} MS(i))(\sum_{i=1}^{n} MS(i))}{\left(\begin{bmatrix} n\sum_{i=1}^{n} MS^2(i)... \\ ... - (\sum_{i=1}^{n} MS(i))^2 \end{bmatrix} \begin{bmatrix} n\sum_{i=1}^{n} RS^2(i)... \\ ... - (\sum_{i=1}^{n} RS(i))^2 \end{bmatrix} \right)^{\frac{1}{2}}} \tag{17}$$

where, $MS(i)$, $RS(i)$ and n represent the main sequence and reference sequence values at time point i and the length of these sequences, respectively. However, usually the two adjacent samples in the electric load time series have a high correlation. As the time interval between two adjacent electric load samples is assumed one hour, it is reasonable to choose the one hour-before load sequence as the reference sequence.

3. Fuzzy Classification

Fuzzy classification is a method based on the fuzzy-set theory for analysis of uncertain phenomena. The essential benefit of fuzzy logic is that it breaks down knowledge representation into simple “IF-THEN” relations. Situations that cannot be mapped onto a simple and well defined deterministic mathematical model can be more easily handled by the fuzzy-set theory, where simple rules and a set of simple membership functions are used to provide more accurate classifications [45]. However, in classical set theory, the relationship of element x and set A can be determined by a function defined as:

$$f_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \tag{18}$$

The above equation indicates that for a given element and a set, the element either belongs to the set or not. This implies a bi-valued and absolute relationship between the element and the set. It is obvious that such type of sets cannot be used for fuzzy problems with amphibious boundaries. On the other hand, in the fuzzy theory, the value range of the relationship function between an element and a set is modified from a bi-valued set {0, 1} to a closed interval [0, 1] including continuous values. The corresponding relationship function is called membership function. Given a set U, u_A: $U \to [0, 1]$ is a mapping from U to interval [0, 1], which defines a fuzzy subset A on the set U. Here, u_A refers to the membership function of A such that $u_A(x)$ represents the membership degree of x concerning A. Fig. 1 depicts a triangle fuzzy space, which shows that the set U is divided into 5 different fuzzy membership functions A_1, A_2,… and A_5. In this figure, a and b refer to minimum and maximum values of the elements in set U. The fuzzy classification method allows an object to belong to different sets at the same time. This concept is very useful in cases where the boundary between two sets is indistinct and hard to define.

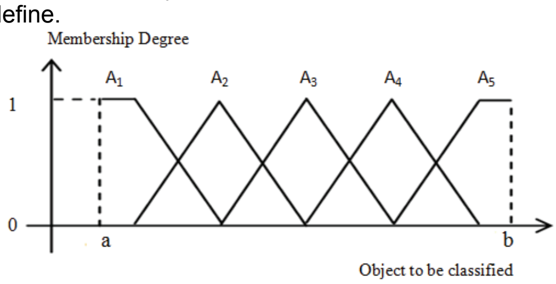


Fig.1. Triangle fuzzy space.

Now, to include fuzzy theory in the Gray model, the relative error of GM should initially be calculated:

$$\varepsilon(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \qquad (k = 1,2,...,n) \tag{19}$$

where, $x^{(0)}(k)$ and $\hat{x}^{(0)}(k)$ represent the actual values and the predicted ones by the Gray model, respectively. Based on the fuzzy classification and according to the relative error of Gray model, the system can be classified into m states. Therefore, the belongingness of each data can be described as a vector: (u_1(x), u_2(x),…, u_m(x)), which is called fuzzy state vector. Based on the greatest membership principle, we can define the state of each variable (i.e. relative error of GM). In other words, for any relative error there is a k (1≤ k ≤m) to satisfy Eq. (20).

$$U(x) = \vee_{k=1}^{m} \{u_{A_K}(x)\} \tag{20}$$

In this equation, we assume that x relatively belongs to A_k, where A_1, A_2,…, A_m refer to m fuzzy sets on u. In Eq. (20), x represents each relative error of GM, and $U(.)$ stands for the function determining the class of x obtained by fuzzy-set theory.

4. Markov Chain Model

When historical load data are used to make forecasts, anomalies such as sudden weather fluctuations or load demand variations can result in a large deviation between the forecast value and the actual value. Under such circumstances, there are two common strategies adopted by researchers to reduce the effect of such uncertainties. The first strategy is selecting the influential variables (e.g. temperature in this case) as the inputs of the forecasting model [2, 5]. But the more difficult and off course the more efficient one is developing hybrid predicting models [23]–[26]. This paper proposes a hybrid model consisting of a homogeneous Markov chain model to detect anomalies and correct the relative errors (see Eq. (19)). In other words, the forecast value by the Gray model should therefore be corrected by introducing yet another forecasting procedure (i.e. Markov Chain model) that is used to predict the relative errors of the initial load forecasting model using the fuzzy approach elaborated in the previous section.

A Markov chain is a special case of a Markov process, which is a special case of a random or stochastic process. As the states of Markov chain are finite and fully connected, transition from one state to any other state is allowed. A random process X_n is called a Markov chain if:

$$P(X_{n+k} = q_{n+k} | X_n = q_n, ..., X_1 = q_1) = P(X_{n+k} = q_{n+k} | X_n = q_n) \tag{21}$$

where, q_1, q_2,…, q_n,…, q_{n+k} take discrete values in a countable state set $\eta_q = \{\Theta_1, \Theta_2, ..., \Theta_N\}$. The index n in X_n is a non-negative integer value which refers to the number of observations; and $P(.)$ stands for the probability function.

Eq. (22) represents a k -step transition probability for the Markov chain of X_n with N states:

$$P_{ij}(n, n+k) = P(q_{n+k} = \Theta_j | q_n = \Theta_i), \quad 1 \le i, j \le N \tag{22}$$

In cases where $P_{ij}(n,n+k)$ is independent of n, X_n is called a homogenous Markov chain. In such a case, $P_{ij}(n,n+k)=P_{ij}(k)$. If $k=1$, $P_{ij}(1)$ is called a one-step transition probability. The transition probability of state is defined as:

$$p_{ij}^{(k)} = \frac{m_{ij}^{(k)}}{M_i} \tag{23}$$

In Eq. (23), $m_{ij}(k)$ represents the number of state u_i transferred into the state u_j by k steps and M_i refers to the number of the appearances of state u_i. The matrix that is made up of elements of k steps transition probability is called k-order state transition probability matrix:

$$P^{(k)} = \begin{bmatrix} p_{11}^{(k)} & p_{12}^{(k)} & . & . & . & p_{1m}^{(k)} \\ p_{21}^{(k)} & p_{22}^{(k)} & . & . & . & p_{2m}^{(k)} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ p_{m1}^{(k)} & p_{m2}^{(k)} & . & . & . & p_{mm}^{(k)} \end{bmatrix} \tag{24}$$

Now, using the transition probability matrix $P^{(k)}$ and the initial state u_i, Markov chain model can be obtained easily.

The vector of membership degrees can be defined as:

$$F(x) = [P_1(x), P_2(x), ..., P_m(x)] \tag{25}$$

where, $\sum_{i=1}^{m} P_i(x) = 1$, x refers to the target variable based on which the Markov Chain model is developed, and $P(.)$ stands for the function determining the probabilities of the state transition probability matrix.

Now, one can follow the procedure to build his/her forecasting Markov Chain model provided that the variables of membership vector as well as the states of the transition probability matrix are defined correctly. Considering that the Markov Chain model in this case is used to correct the load prediction of the Gray model, we suggest defining the relative error of GM (see Eq. (19)) as the variable of the membership vector. Applying the aforementioned fuzzy approach, we consider the classes which such relative errors belong to as the states of the transition probability matrix.

Now, the vector of membership degrees at time step $n+k$ is still a fuzzy vector, which can be calculated by:

$$F(\varepsilon(n+k)) = F(\varepsilon(n)) \quad P^{(k)} = [u_{A1}(\varepsilon(n+k)), ..., u_{Am}(\varepsilon(n+k))] \tag{26}$$

where, $\varepsilon(n)$ refers to the relative error of GM at the time step n and $u_{Ai}(\varepsilon(n))$ represents the membership degree of $\varepsilon(n)$ in the fuzzy state A_i (see section 3) .

Each component of $F(\varepsilon(n+k))$ shows the membership degree of each fuzzy state at the time step $n+k$. Considering the membership degrees in the fuzzy vector as weights, and using the weight sum method, the predicted relative error by the Markov chain model at time step $n+k$ will be calculated as:

$$\varepsilon(n+k) = \frac{1}{2} \sum_{i=1}^{m} U_{Ai}(\varepsilon(n+k))(\varepsilon_{i-1} + \varepsilon_i) \tag{27}$$

where, ε_{i-1} and ε_i are respectively the minimum and maximum values of relative errors of training data in class i.

Finally, the forecasted and corrected value at $n+k$ time step can be calculated by using the following formula:

$$\hat{y}(n+k) = \frac{\hat{x}(n+k)}{1 - \varepsilon(n+k)} \tag{28}$$

where, $\hat{x}(n+k)$ and $\hat{y}(n+k)$ represent the predicted values by Gray and Gray-Fuzzy-Markov models, respectively.

5. Performance Evaluation

5.1. Case Studies and Data Sets

In order to investigate the effectiveness of the proposed strategy, called GFMCM, different benchmark tests have been carried out in this study using real electric load data collected from three electricity markets including Ontario [46] and PJM [47] electricity markets (EMs) for 2004 and Iran EM for 2009-2010 period. The predictions have been calculated for four separate weeks, each of which corresponding to one of the four seasons of the year. For these markets, the hourly measured load data of 20 days prior to the forecasting day have been used to build the forecasting model. Thus, the GFMCM is trained with

$20 \times 24 = 480$ training samples for prediction of the load of the following 24 hours. The training process has been repeated for every day of every test week. Therefore, a new GFMCM is executed for each day of a given set of seven successive days in a given season. Although it sounds simpler to report the forecasting result of just the next day when we aim to predict the day-ahead electric load, the reported results can hardly be taken as strong indicators of the productivity of the method. Since the electric load signal in the next day is likely to have exceptionally soft variations or on the other hand dramatically volatile fluctuations. So, seven days of a week are selected for day-ahead load prediction. This short training time is enough for effective extraction of the data trend [48] and, as a result, for obtaining accurate predictions. In order to measure the performance of the forecasting model, two indices, namely daily mean absolute percentage error (DMAPE) and weekly mean absolute percentage error (WMAPE), are used:

$$DMAPE = \frac{100}{24}\sum_{i=1}^{24}\frac{|x(i)-y(i)|}{x(i)} \tag{29}$$

$$WMAPE = \frac{100}{168}\sum_{i=1}^{168}\frac{|x(i)-y(i)|}{x(i)} \tag{30}$$

In these equations, x and y stand for the actual and predicted load data, respectively. Table 1 shows the historical hourly load data used to build GFMCM, which will be employed to predict the load data of the test week for the above mentioned markets.

5.2. Results and Discussion

5.2.1. Prediction of electric load in Ontario Market

For winter, as shown in Table 1, the load data from January 1, to January 27, 2004 are used. In developing the Gray model, the procedure we adopted builds 24 separate Gray models, corresponding to 24 hours of a day. For instance, to develop the first GM(1,2), we consider the electric load data from hour 0:00 of January 2 to January 20, 2004 as the main sequence, and the electric load data from hour 23:00 of January 1 to January 19, 2004 as the reference sequence. In a similar way, we build 24 separate GM(1,1) models as well, with the difference that there is no reference sequence for GM(1,1). Fig. 2 shows the MAPE values computed for the train data of the 20-day period of each of the 24 GM(1,1) and GM(1,2) models. As it is observed from the figure, the MAPE values for GM(1,2) are constantly lower than those of the GM(1,1), indicating that the process of training with GM(1,2) is more effective than that with GM(1,1).

Table 1. Hourly load data for forecasting model constructions and testing

Markets	Seasons	Historical Hourly Load Data	Test Weeks
Ontario (2004)	Winter	Jan. 1 – Jan. 20	Jan. 21 – Jan. 27
	Spring	April.1 – April. 20	April.21 – April. 27
	Summer	July. 1 – July. 20	July. 21 – July. 27
	Autumn	Oct. 1 – Oct. 20	Oct. 21 – Oct. 27
PJM (2004)	Winter	Jan. 1 – Jan. 20	Jan. 21 –Jan. 27
	Spring	May. 1 – May. 20	May. 21 – May. 27
	Summer	July. 1 – July. 20	July. 21 – July. 27
	Autumn	Oct. 1 – Oct. 20	Oct. 21 – Oct. 27
Iran (2009-2010)	Spring	April. 21–May. 10	May. 11–May. 17
	Summer	July. 23–Aug. 11	Aug. 12 –Aug. 18
	Autumn	Oct. 23–Nov. 11	Nov. 12 –Nov. 18
	Winter	Jan. 21–Feb. 9	Feb. 10 –Feb. 16

Table 2 summarizes the predicted values by GM(1,1) and GM(1,2) for the first test day. The DMAPE values for these two models are 2.9459% and 1.2942%, respectively. This further confirms the more effective performance of GM(1,2) found with the train data.

A question that might arise here is why the MAPE values for the test data (Table 2) are so considerably smaller than those of the train data (Fig. 2). The reason lies in the fact that Gray models are very vulnerable when fitting the first few train samples, which obviously increases the average fitting error of the train data; though these models continue the training process with increasing higher precision. This is clearly observable from column 4 in Table 3. As can be seen in the table, the relative error values of the first GM(1,2) for the rows 2, 3 and 4 are significantly larger than the others. Needless to say, the reason why the relative error of the first row in the table is zero is that the first value fitted by GM is assumed to be equal to the actual value (see Eq. (16)).

Table 2. Comparison of effectiveness of GM(1,1), GM(1,2), GM(1,1)-Fuzzy-Markov Chain and GM(1,2)-Fuzzy-Markov Chain models for test data

Hour	Actual Value	GM(1,1) Forecast	GM(1,2) Forecast	GM(1,1) -Fuzzy-Markov Forecast	GM(1,2) -Fuzzy-Markov Forecast
0	18536	19299	18637	18424	18535
1	18247	18860	18200	17863	18279
2	17972	18692	17952	17759	17994
3	17939	18593	17847	17651	17904
4	18009	18636	17926	17615	17962
5	18570	19152	18375	17916	18686
6	20002	20370	19508	19163	20001
7	21852	22093	21194	21153	21574
8	22262	22590	21822	21699	22031
9	22090	22737	22019	21647	22149
10	21851	22727	22120	22102	22100
11	21682	22494	22036	21887	22064
12	21334	22144	21786	21577	21677
13	21370	21942	21600	21070	21490
14	21302	21759	21475	21119	21394
15	21476	21881	21571	21019	21527
16	22469	22564	22372	21698	22379
17	23660	23815	23669	23092	23645
18	23592	24113	23785	23118	23572
19	23058	23632	23294	22484	23172
20	22520	23181	22823	22304	22598
21	21624	22388	22048	20867	21936
22	20092	21200	20885	19909	20767
23	18959	20019	19616	19007	19595
DMAPE(%)		2.9459	1.2942	1.9465	0.8022

In the next step, 24 Markov Chain models, corresponding to each of the 24 Gray models, are developed. For this purpose, the fuzzy vectors for each 24 Gray models are formed by triangle membership functions (see section 3), based on relative errors of the fitted values of the corresponding model (an example of such membership functions will be provided in Eq. (31)).

Table 3 shows the relevant values computed only for the first GM(1,2) model referring to hour 0:00. In order to obtain the fuzzy vectors, the minimum-maximum error intervals have to be divided into an appropriate number of states (i.e. classes) from the obtained relative errors. A trial and error procedure showed that a division into 3 states can result in lower MAPE values for the proposed model. Therefore, a decision was made to adopt the 3 state division procedure.

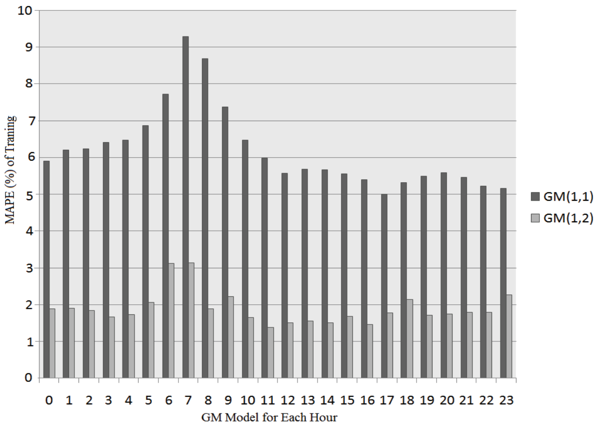


Fig.2. MAPE (%) of GM(1,1) and GM(1,2) for train data.

In section 3, it was explained that the fuzzy membership functions are computed by the use of triangle functions. Applying the elucidated procedure to the relative errors percentage presented in the forth column of Table 3, we can determine the membership functions for the first GM(1,2) as follow:

$$u(k,1)=\begin{cases} 1 & \varepsilon(k)\le -2 \\ -\frac{1}{2}\varepsilon(k) & -2\le\varepsilon(k)\le 0 \\ 0 & otherwise \end{cases}$$

$$u(k,2)=\begin{cases} \frac{1}{2}(\varepsilon(k)+2) & -2\le\varepsilon(k)\le 0 \\ -\frac{1}{2}(\varepsilon(k)-2) & 0\le\varepsilon(k)\le 2 \\ 0 & otherwise \end{cases}$$

$$u(k,3)=\begin{cases} \frac{1}{2}\varepsilon(k) & 0\le\varepsilon(k)\le 2 \\ 1 & \varepsilon(k)\ge 2 \\ 0 & otherwise \end{cases}$$

(31)

In Eq. (31), $u(k,m)$ represents the membership degree of kth relative error for each of the three classes. And $\varepsilon(k)$ represents the relative error corresponding to each train sample (see Eq. (19)).

After determining the state of each predicted value, based on the greatest membership principle (see Eq. (20)), the 24 transition probability matrices of state, corresponding to the 24 Gray models are developed (see Eq. (23) and (24)). For instance, the transition probability matrix of state, for the first GM(1,2), is as follows:

$$P=\begin{bmatrix} 0.25 & 0.75 & 0 \\ 0.154 & 0.769 & 0.077 \\ 1 & 0 & 0 \end{bmatrix}$$

(32)

Considering that the state fuzzy vector for the 20th day is (0,0.917,0.083) (see Table 3), we calculate the fuzzy vectors of state for the 21st day using Markov chain (see Eq. (26)), and then we use weight accumulation method to compute the predicted relative errors (see Eq. (27)). Finally, the electric load for the 21st day which was forecasted by the first GM(1,2) model is corrected by the first Markov chain model (see Eq. (28)).

Table 3.Fuzzy Classification of relative errors of GM(1,2) model related to hour 0:00

Day	Actual Value	GM(1,2) Forecast	Relative Error (%)	Fuzzy State
2	14093	14093	0	(0,1,0)
3	14682	12851	12.46	(0,0,1)
4	14303	15987	-11.77	(1,0,0)
5	15622	16154	-3.40	(1,0,0)
6	16890	16868	0.13	(0,0.935,0.065)
7	18020	18002	0.10	(0,0.949,0.051)
8	18115	18063	0.28	(0,0.858,0.142)
9	18767	18360	0.72	(0,0.636,0.364)
10	19030	19189	-0.83	(0.417,0.583,0)
11	17854	17805	0.27	(0,0.861,0.139)
12	16885	16812	0.43	(0,0.784,0.216)
13	17227	17404	-1.03	(0.515,0.485,0)
14	19123	19114	0.04	(0,0.976,0.024)
15	19269	19106	0.84	(0,0.577,0.423)
16	19956	19889	0.33	(0,0.832,0.168)
17	18313	18697	-2.09	(1,0,0)
18	17318	17362	-0.25	(0.128,0.872,0)
19	17584	17492	0.52	(0,0.737,0.263)
20	18420	18389	0.16	(0,0.917,0.083)

As predictions for a whole 24 hour day consist of 24 distinct models, the above process is repeated to develop other 23 GM(1,2)-Fuzzy-Markov Chain models as well. In a similar way, this procedure is repeated for 24 GM(1,1)-Fuzzy-Markov Chain models.

The predicted values for test data by GM(1,1)-Fuzzy-Markov Chain and GM(1,2)-Fuzzy-Markov Chain models are summarized in Table 2. As observed from this table, DMAPE values of GM(1,1)-Fuzzy-Markov and GM(1,2)-Fuzzy-Markov are 1.9465% and 0.8022%, respectively. These values are meaningfully smaller than the values obtained for the Gray models. These lower DMAPE values are vitally important as we strive towards more accurate load forecasting to reduce operation costs. As mentioned in [49], a 1% reduction in forecasting error for a 10,000 MW utility in 1999 could save up to $1.6 million annually.

Now, this process is repeated 6 more times to predict the electric load variations for January 22, 2004 through January 27, 2004. The forecasts obtained by GM(1,2) and GM(1,2)-Fuzzy-Markov Chain models along with the actual data during the test week in winter are shown in Fig. 3. It is obvious that the predictions made by GM(1,2)-Fuzzy-Markov better approximate the actual data than those made by GM(1,2).

It is observed from Fig. 3 that there are two peaks on Monday, with the first one for hour 12 and the second (the sharper one) for hour 18. Contrary to what might be expected, the Markov chain model has supplied a considerably more accurate correction for the sharper peak (related to hour 18) than that of the first peak. The reason can be attributed to the existence of a history of comparable sharp variations during the 20-day training period for the hour 18 (see Fig. 4), while the history of training samples for the hour 12 contains less dramatic jumps. Fig. 4 represents the rate of growth of load data between each two consecutive training days for the hours 12 and 18. As can be seen in the figure, there are quite conspicuous variations in the history of the training samples for the hour 18. For instance, notice the growth rate of 34.99% between the days 4 and 5, and that of 27.39% between the days 18 and 19. In contrast, the history of variations related to the hour 12 is comparatively less marked, with a maximum growth rate of 9.82% for the interval of days 12 to 13.

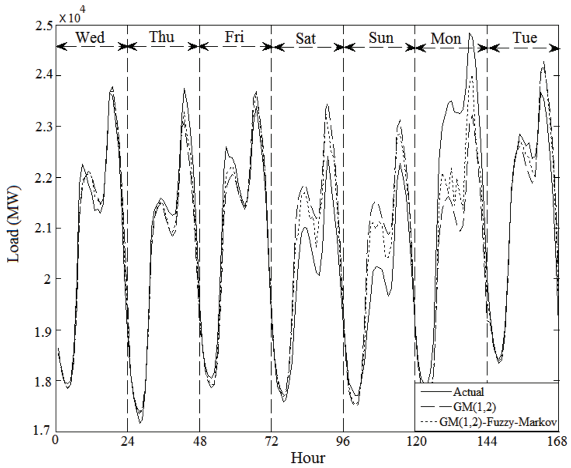


Fig. 3. Forecasted results of winter test week in Ontario market.

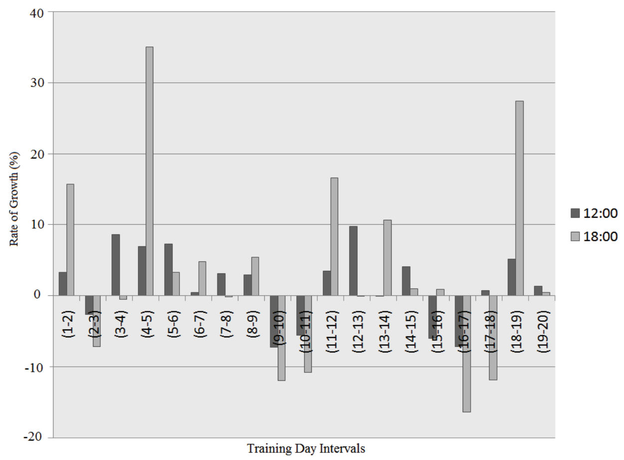


Fig. 4. Growth rate of load data between each two consecutive training days for the hours 12 and 18.

Fig. 5 shows weekly load predictions for spring, summer and autumn, respectively. The WMAPE values of the GFMCM, for all test weeks, are summarized in Table 4. A comparison of the data presented in Table 4 reveals that for seasons with higher load fluctuations, GM(1,2) has a superior prediction performance compared to GM(1,1).

5.2.1.1. Comparison of GFMCM with other techniques

To further illustrate the performance of the proposed strategy, in Table 4, the predictions of GFMCM are compared with the results obtained from Gray models-GM(1,1) and GM(1,2), ARIMA time series, wavelet-ARIMA and multi-layer perceptron neural network which are developed with the same training samples (presented in Table 1) used for GFMCM. The reason we selected these methods to compare their result to that of our proposed model is related to their repetitive utilization for this purpose by other researchers in this scope [6, 48, 55]. It should be emphasised here that we did not use any fuzzy model or any Markov Chain model as a load forecaster persuading us to compare their result with that of the proposed hybrid model. The Markov Chain model was used to forecast the next fuzzy state of the relative errors and the fuzzy approach was used to set a link between the Gray model and the Markov Chain model.

One of the most widely used models among neural networks is MLP, for which the back-propagation learning rule is used as a training procedure. This rule tries to minimize the error between desired and calculated output by adjusting the weights within the network. In addition, MLP networks are able to model autoregressive non-linear processes with exogenous variables.

Noteworthy, however, is that MLP networks have the potential to contain high degrees of freedom, resulting in existence of numerous weights [50]. This in turn causes the system to become overfitted during training, i.e. to fit itself to the received data so much so that it loses the ability to generalize. Overfitting causes the system to get entangled in local optimums rather than global ones. Furthermore, another drawback that arises from high degrees of freedom is that a large number of training samples is required to obtain reliable results.

In [51], can be found a thorough account of how MLP can be simulated to predict short term load. Based on a trial and error procedure, we utilized a 3-layer MLPNN which has 3 inputs in the input layer, and also has 8 and 2 neurons in the first and second hidden layers.

Another widely used technique for load prediction is the ARIMA model family proposed by Box et al. [52]. A key advantage of ARIMA approaches over neural networks is their better understanding of the studied phenomenon [50]. However, ARIMA models are not without their drawbacks: they assume a linear relation between independent and dependent variables. Researchers of the present study used the sources in [5] to simulate ARIMA, while [53] can be consulted for an understanding of how ARIMA parameters are selected. While in the ARIMA method, one time series is assigned to the whole load series, in wavelet-ARIMA model, one ARIMA model is allocated to each of four constitutive series of the model. The reason why the constitutive series function more accurately is the filtering effect of the wavelet transform. Detailed explanation of how the wavelet-ARIMA model is simulated can be found in [54]. The data presented in Table 4 show that the proposed model (GM(1,2)-Fuzzy-Markov Chain) considerably outperforms the other techniques in all cases by generating the smallest WMAPE values.

5.2.2. Prediction of electric load in Iranian Electricity Market

The load forecasting results for the test weeks given in Table 1 in Iranian electricity market for 2009-2010 are shown in Fig. 6. Since each year in Iranian calendar begins on March 20 (the beginning of spring), it contains parts of two consecutive years in Christian calendar. The WMAPE values are reported in Table 4. As it can be observed from the table, WMAPE values for the proposed method are smaller than those of the other methods presented here.

5.2.3. Prediction of electric load in PJM Market

The proposed model was applied to recorded load data of PECO control zone of PJM market in 2004. The calculated WMAPE values for all models are presented in Table 4. Once again, the results confirm the superiority of the proposed model with regard to WMAPE values.

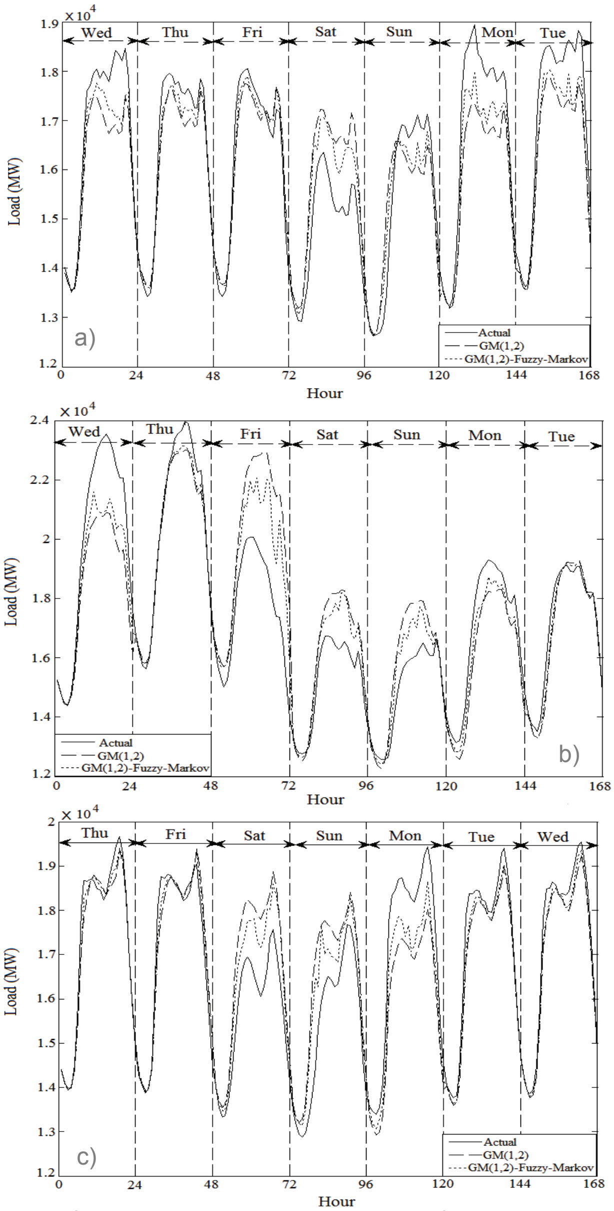
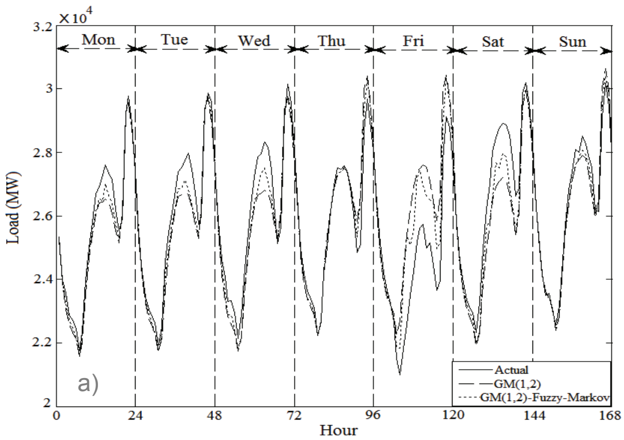


Fig. 5. Ontario market. a) Forecasted results of spring test week. b) Forecasted results of summer test week. c) Forecasted results of autumn test week.



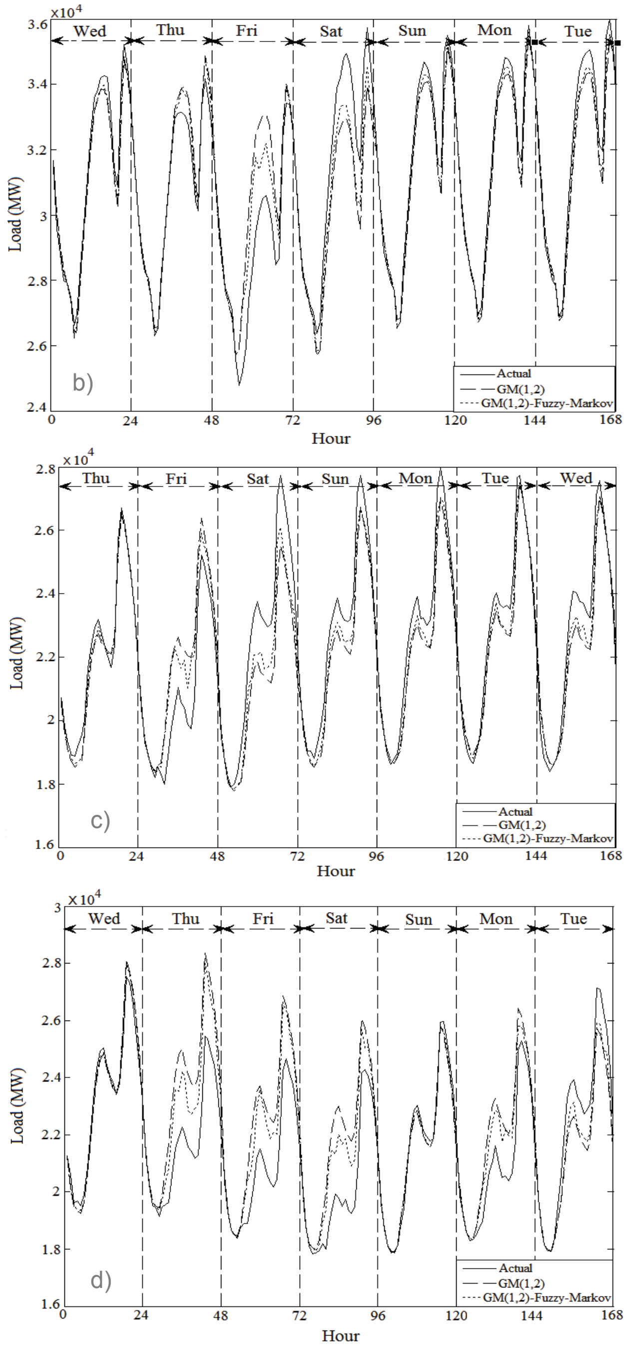


Fig. 6. Iranian electricity market. a) Forecasted results of spring test week. b) Forecasted results of summer test

5.3. Computation Burden

One of the main shortcomings attributed to artificial neural networks is their longer training process. This is basically because of the iterative nature of training of such networks. For example, the training process of MLP neural network for load prediction in Ontario market takes 6 seconds on a PC with Intel(R) core2 Due CPU E7500, 2.93 GHz, and 2 GB RAM. On the other hand, the training process of the proposed method (GFMCM), for the same data and on the same computer, takes 0.4 seconds. The main reason is the non-iterative nature of the GFMCM, which also brings more stability over the neural networks method.

Table 4. WMAPE in Ontario, PJM and Iranian electricity markets for the four weeks of study

Markets	Test Weeks	MLPNN	ARIMA	wavelet-ARIMA	GM(1,1)	GM(1,1) -Fuzzy-Markov	GM(1,2)	GM(1,2) -Fuzzy-Markov
Ontario (2004)	Winter	3.423	4.162	2.428	3.616	2.721	2.802	1.948
	Spring	3.706	4.906	3.654	4.820	3.616	3.908	2.835
	Summer	7.576	8.114	6.980	7.808	6.078	5.902	4.298
	Autumn	5.618	5.774	3.904	5.050	4.237	3.679	2.579
	Average	5.081	5.739	4.241	5.323	4.163	4.073	2.915
PJM (2004)	Winter	3.362	3.728	2.220	3.872	2.846	3.677	2.004
	Spring	4.550	5.014	4.656	5.285	3.730	3.951	3.061
	Summer	7.945	7.802	6.991	8.9375	7.278	7.481	6.025
	Autumn	6.025	6.743	4.097	4.524	3.433	4.265	3.255
	Average	5.470	5.822	4.491	5.655	4.322	4.843	3.586
Iran (2009-2010)	Spring	3.429	4.317	2.655	3.826	2.105	2.645	1.916
	Summer	3.285	3.950	2.969	3.747	2.072	2.091	1.510
	Autumn	4.626	5.078	3.441	3.602	2.740	3.335	2.537
	Winter	5.446	5.913	4.263	5.536	4.111	4.882	3.598
	Average	4.196	4.814	3.332	4.178	2.757	3.238	2.390

6. Conclusion

A new method based on fuzzy-set theory has been proposed for forecasting the electricity load variations. The method uses Gray-Fuzzy-Markov Chain Model for short-term load forecasting.

A comparison between the two Gray models – GM(1,1) and GM(1,2) shows the better performance of GM(1,2) model for prediction of fluctuating load data. A classification based on a fuzzy approach to develop a Markov chain considerably improves the prediction accuracy of the hybrid Gray-Fuzzy-Markov Chain model.

The experimental results, for various load variation patterns, demonstrate that the proposed strategy provides more accurate predictions compared to Gray models-GM(1,1) and GM(1,2), ARIMA time series, wavelet-ARIMA and multi-layer perceptron neural network.

Acknowledgment

The authors would like to thank KREC (Khorasan Regional Electric Company) for providing the load data of Iran.

REFERENCES

[1] Haida T., Muto S., Regression based peak load forecasting using a transformation technique, IEEE Trans. Power Syst., 9 (1994), No. 4, 1788–1794

[2] Papalexopoulos A. D., Hesterberg T. C., A regression-based approach to short-term system load forecasting, IEEE Trans. Power Syst., 4 (1990), No. 4, 1535–1547

[3] Huang S. J., Shih K. R., Short-term load forecasting via ARMA model identification including non-Gaussian process considerations, IEEE Trans. Power Syst., 18 (2003), No. 2, 673–679

[4] Song K. B., Baek Y. S., Hong D. H., Jang G. S., Short-term load forecasting for the holidays using fuzzy linear regression method, IEEE Trans. Power Syst., 20 (2005), No. 1, 96–101

[5] Cho M. Y., Hwang J. C., Chen C. S., Customer short term load forecasting by using ARIMA transfer function model, in Proc. Int. Conf. EMPD, 1 (1995), 317–322.

[6] Amjady N., Short-term hourly load forecasting using time-series modeling with peak load estimation capability, IEEE Trans. Power Syst., 16 (2001), No. 3, 498–505

[7] Lu C. N., Vemuri S., Neural network based short-term load forecasting, IEEE Trans. Power Syst., 8 (1993), No. 1, 336–342

[8] Srinivasan D., Tan S. S., Chang C. S., Chan E. K., Parallel neural network-fuzzy expert system strategy for short-term load forecasting: System implementation and performance evaluation, IEEE Trans. Power Syst., 14 (1999), No. 3, 1100–1105

[9] Papadakis S. E., Theocharis J. B., Kiartzis S. J., Bakirtzis A. G., A novel approach to short-term load forecasting using fuzzy neural networks, IEEE Trans. Power Syst., 13 (1998), No. 2, 480 489

[10] Fan S., Chen L., Lee W. J., Short-term load forecasting using comprehensive combination based on multi-meteorological information, in Proc. IEEE Ind. Commercial Power Syst. Tech. Conf., 2008, 1–7.

[11] Hippert H. S., Pedreira C. E., Souza R. C., Neural networks for short-term load forecasting: A review and evaluation, IEEE Trans. Power Syst., 16 (2001), 44–55

[12] Chen B. J., Chang M. W., Lin C. J., Load forecasting using support vector machines: A study on EUNITE competition 2001, IEEE Trans. Power Syst., 19 (2004), No. 4, 1821–1830

[13] Fan S., Chen L., Short-term load forecasting based on an adaptive hybrid method, IEEE Trans. Power Syst., 21 (2006), No. 1, 392–401

[14] Song K. B., Baek Y. S., Hong D. H., Jang G., Short-term load forecasting for the holidays using fuzzy linear regression method, IEEE Trans. Power Syst., 20 (2005), No. 1, 96–101

[15] Khotanzad A., Enwang Z., Elragal H., A neuro-fuzzy approach to short-term load forecasting in a price-sensitive environment, IEEE Trans. Power Syst., 17 (2002), No. 4, 1273–1282

[16] Gwo-Ching L., Ta-Peng T., Application of a fuzzy neural network combined with a chaos genetic algorithm and simulated annealing to short-term load forecasting, IEEE Trans. Evol. Comput., 10 (2006), No. 3, 330–340

[17] Wu H. C., Lu C., Automatic fuzzy model identification for short term load forecast, Proc. Inst. Elect. Eng., Gen., Transm., Distrib., 146 (1999), No. 5, 477–482

[18] Bashir Z. A., El-Hawary M. E., Applying wavelets to short-term load forecasting using PSO-based neural networks, IEEE Trans. Power Syst., 24 (2009), No. 1, 20–27

[19] Ling S. H., Leung F. H., Lam H. K., Yim-Shu L., Tam P. K., A novel genetic-algorithm-based neural network for short-term load forecasting, IEEE Trans. Ind. Electron., 50 (2003), No. 4, 793–799

[20] Oonsivilai A., El-Hawary M. E., Wavelet neural network based short- term load forecasting of electric power system commercial load, in Proc. IEEE Can. Conf. Electrical and Computer Engineering, (1999), 1223–1228

[21] Huang C. M., Yang H.T., Evolving wavelet-based networks for short term load forecasting, Proc. Inst. Elect. Eng., Gen., Transm., Distrib., 148 (2001), No. 3, 222–228

[22] Agnaldo R. R., Alexandre P. A., Feature extraction via multiresolution analysis for short-term load forecasting, IEEE Trans. Power Syst., 20 (2005), No. 1, 189–198

[23] Senjyu T., Mandal P., Uezato K., Funabashi T., Next day load curve forecasting using hybrid correction method, IEEE Trans. Power Syst., 20 (2005), No. 1, 102–109

[24] Srinivasan D., Liow A. C., Chang C. S., Forecasting daily load curves using a hybrid fuzzy_neural aproach, Proc. Inst. Elect. Eng., Gen., Transm., Distrib., 141 (1994), No. 6, 561–567

[25] Kim K. H., Park J. K., Hwang K. J., Kim S. H., Implementation of hybrid short-term load forecasting system using artificial neural networks and fuzzy expert systems, IEEE Trans. Power Syst., 10 (1995), No. 3, 1534–1539
[26] Kim K. H., Youn H.S., Kang Y.C., Short-term load forecasting for special days in anomalous load conditions using neural network and fuzzy inference method, IEEE Trans. Power Syst., 15 (2000), No. 2, 559–565
[27] Cui H-R., A long-term electrical power load forecasting model based on grey feed-back modification, 2008 International Conference on Machine Learning and Cybernetics, 4 (2008), 2198-2201
[28] Li W., Han Z-h., Application of improved grey prediction model for power load forecasting, 12th International Conference on Computer Supported Cooperative Work in Design, CSCWD, (2008), 1116-1121
[29] Peirong J., Juan C., Wenchen Z., Theory of grey systems and its application in electric load forecasting, 2008 IEEE Conference on Cybernetics and Intelligent Systems, (2008), 1374-1378
[30] Zhang D., Ren Z., Bi Y., Zhou D., Bi Y., Power load forecasting based on grey neural network, IEEE International Symposium on Industrial Electronics, (2008), 1885-1889
[31] Zhao W., Niu D., A mid-long term load forecasting model based on improved grey theory, The 2nd IEEE International Conference on Information Management and Engineering (ICIME), (2010), 633-635
[32] Yang S., Zhang Y., Short-term load forecast based on decomposition of daily load curve, Computer, Mechatronics, 2010 International Conference on Control and Electronic Engineering (CMCE), 3 (2010), 65-68
[33] Hsu K-T., Yan T-M., Hsu T-H., Applying grey models to forecast financial ratios of national health insurance in Taiwan, IEEE World Congress on Computational Intelligence, (2008) ,475-480
[34] Kung C-Y., Chen H-S., Huang C-C., Wen K-L., Using GM(1,1) method to forecast the development of cell phone market in Taiwan, International Conference on on Computational Intelligence, (2008), 539-543.
[35] Xie J., Zhang Y., Jiang W., An Improved Grey-Markov Chain Method with an Application to Predict the Number of Chinese International Airlines", International Symposium on Information Science and Engineering, (2008), 716 – 720
[36] Zheng-wen X., Kai-yu S., Improved Grey Model Base on Exponential Smoothing for River Water Pollution Prediction, International Conference on Bioinformatics and Biomedical Engineering (iCBBE), (2010), 1-4
[37] Huang C-Y., Tzeng W-C., Liu Y-W., Wang P-Y., Forecasting the Global Photovoltaic Market by Using the GM(1,1) Grey Forecasting Method, Green Technologies Conference (IEEE-Green), (2011), 1 - 5
[38] Wei L., Zhen-Gang Z., Ning Y., Jia-Liang L., Improved Markov Residual Error to Long-Medium Power Load Forecast Based on SVM Method, Education Technology and Computer Science. ETCS '09, First International Workshop on, (2009), 128 – 132
[39] Gonzalez A.M., Roque A.M.S. Garcia-Gonzalez J, Modeling and forecasting electricity prices with input/output hidden Markov models, IEEE Transactions on Power Systems, (2005), 13 – 24
[40] Deng J. L., Introduction to gray system theory, Journal of gray System, 1, (1989), No.1, 1–24
[41] Deng J. L., Control problems of gray systems, Systems & Control Letters, 5 (1982.), No.1, 228–294
[42] Liu S. F., Lin Y., an Introduction to Gray Systems, Slippery Rock, PA: IIGSS Academic, (1998)
[43] El-Fouly T.H.M., El-Saadany E.F., Salama M.M.A., Grey predictor for wind energy conversion systems output power prediction, IEEE Transactions on Power Systems, (2006), 1450 – 1452
[44] Lei M., Feng Z., Song Q., A novel grey model to forecast short-term electricity price for Nordpool using particle swarm optimization and correlation hours method, 2010 International Conference on Networking, Sensing and Control (ICNSC), (2010), 631-636
[45] Youssef O.A.S., Combined fuzzy-logic wavelet-based fault classification technique for power system relaying, IEEE Transactions on Power Delivery, (2004), 582 – 589
[46] Ontario Power Market Website [Online]. Available: http://www.ieso.ca/.
[47] PJM Power Market Website [Online]. Available: http://www.pjm.com/.
[48] Amjady N., “Day-ahead price forecasting of electricity markets by a new fuzzy neural network,” IEEE Trans. Power Syst., 21 (2006), No. 2, 887–896
[49] Hobbs B.F., Jitprapaikulsarn S., Konda, S., Chankong, V., Loparo, K.A., Maratukulam, D.J., Analysis of the value for unit commitment of improved load forecasts, Transactions on Power Systems, IEEE, 14 (1999), No. 4, 1342-1348
[50] Aburto L., Weber. R., Improved supply chain management based on hybrid demand forecasts, Journal Applied Soft Computing archive, 7 (2007), No. 1, 136-144
[51] Haykin S., Neural Networks, A Comprehensive Foundation, second edition, Pearson education
[52] Box G.E.P., Jenkins G.M., Reinsel G., Time Series Analysis: Forecasting and Control, third ed., Prentice Hall, (1994)
[53] Contreras J., Espinola R., Nogales, F.J., Conejo, A.J., ARIMA models to predict next-day electricity prices, IEEE Transactions on Power Systems, 18 (2003), No. 3, 1014- 1020
[54] Conejo A.J., Plazas M.A., Espinola R., Molina A.B., Day-ahead electricity price forecasting using the wavelet transform and ARIMA models, IEEE Transactions on Power Systems, 20 (2005), No. 2, 1035 - 1042
[55] Pindoriya N.M., Singh S.N., Singh S.K., Application of adaptive wavelet neural network to forecast operating reserve requirements in forward ancillary services market, Applied Soft Computing, 11 (2011), No. 2, 1035 - 1042

***Authors**: A. Asrari, D. S.Javan, M. H.Javidi and M. Monfared are with the Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran .*
(E-mail: arash.asrari@ieee.org; dawoodjavan@ieee.org; h-javidi@ferdowsi.um.ac.ir; m.monfared@um.ac.ir).