A hybrid simulation-adaptive network based fuzzy inference system for improvement of electricity consumption estimation

A. Azadeh a,*, M. Saberi b,1, A. Gitiforouz a, Z. Saberi c

a Department of Industrial Engineering and Center of Excellence for Intelligent Based Experimental Mechanics, College of Engineering, University of Tehran, Iran
b Department of Industrial Engineering, University of Tafresh, Iran
c Department of Industrial Engineering, Sharif University of Technology, Iran

ABSTRACT

This paper presents a hybrid adaptive network based fuzzy inference system (ANFIS), computer simulation and time series algorithm to estimate and predict electricity consumption estimation. The difficulty with electricity consumption estimation modeling approach such as time series is the reason for proposing the hybrid approach of this study. The algorithm is ideal for uncertain, ambiguous and complex estimation and forecasting. Computer simulation is developed to generate random variables for monthly electricity consumption. Various structures of ANFIS are examined and the preferred model is selected for estimation by the proposed algorithm. Finally, the preferred ANFIS and time series models are selected by Granger–Newbold test. Monthly electricity consumption in Iran from 1995 to 2005 is considered as the case of this study. The superiority of the proposed algorithm is shown by comparing its results with genetic algorithm (GA) and artificial neural network (ANN). This is the first study that uses a hybrid ANFIS computer simulation for improvement of electricity consumption estimation.

Significance

This is the first study that presents a hybrid simulation-adaptive network fuzzy inference system (ANFIS) for improvement of electricity consumption estimation. The unique features of the proposed algorithm are two fold. First, ANFIS is ideal for complex and uncertain data because it is composed of both ANN and fuzzy systems. Second Monte Carlo simulation is used to generate input variables whereas the conventional methods use deterministic data. The superiority of the proposed algorithm is shown by comparing its results with time series, genetic algorithm (GA) and ANN.

1. Introduction

There have been several studies on ANN and neuro-fuzzy models in different cases (Baylar, Hanbay, & Ozpolat, 2008; Çaydas, Hasçalık, & Ekici, 2009; Dogantekin, Yilmaz, Dogantekin, Avci, & Sengur, 2008; Huang, Kang, Chu, Chien, & Chang, 2009; Khajeh, Modarress, & Rezaee, 2008; Subasi, Serdar Yilmaz, & Binici, 2008; Wang & Chen, 2008).2 ANN is configured for a specific application, such as pattern recognition, function approximation, data classification and so on in different areas of science. Time series modeling is one of the main applications. Many researchers showed ANN’s comparability and superiority to conventional methods for estimating functions (Azadeh, Ghaderi, Tarverdian, & Saberi, 2006; Azadeh, Ghaderi, & Sohrabkhani, 2007; Azadeh, Ghaderi, Tarverdian, & Saberi, 2007; Chiang, Urban, & Baldridge, 1996; Hill, O’Connor, & Remus, 1996; Hwaung, 2001; Indro, Jiang, Patuwo, & Zhang, 1999; Jhee & Lee, 1993; Kohzadi, Boyd, Kermanshahi, & Kaastra, 1996; Stern, 1996; Tang, Almeida, & de Fishwick, 1991; Tang & Fishwick, 1993). Whereas neuro-fuzzy is combination of ANN and fuzzy system, have a benefit of two models and is selected instead of ANN (Werbos, 1974). Some of the neuro-fuzzy systems are well known by their short names. For example, ANFIS (Jang, 1993), DENFIS (Kasabov, 2002), SANFIS (Wang & Lee, 2002) and FLEXNFIS (Rutkowski & Cpakla, 2003), etc. ANFIS is used in present study as one of algorithm tools.

One of the main objectives of this research is to combine conventional time series concepts with ANFIS. We show these concepts are more useful in improving ANFIS performance. These concepts are preprocessing (for madding process, covariance stationary), post processing (to access main data) and principle component analysis (for input selection). Exploring the literature reveals that combination of traditional concepts with ANFIS to model time series has been rarely done. Although data preprocessing concept is considered in some literature, but the covariance stationary concept in data preprocessing is ignored (Aznarte et al., 2007; Gareta, Romeo, & Gil, 2006; Jain & Kumar, 2007;
Karunasingha & Liong, 2006; Nayak, Sudheer, Rangan, & Ramasastri, 2004; Niskaa, Hiltunen, Karpinnen, Ruuskanen, & Kolehmainen, 2004; Oliveira & Meira, 2006; Tseng, Yu, & Tzeng, 2002). Other aim is to used computer simulation as an overlapping approach with ANFIS. However, this study uses computer simulation to generate random variables to be used in ANFIS, whereas previous studies only use available raw data for ANFIS. Moreover, the integration of ANFIS and simulation is proposed as an alternative forecasting approach in this study and it is compared with ANFIS and time series.

Computer simulation has excellent capabilities such as proper description of system behavior, scenario analysis and forecasting capabilities. Numerous studies have been conducted in domain of ANFIS or computer simulation; however, this is the first study that integrates ANFIS and simulation for forecasting electricity consumption to be used in the proposed algorithm of this paper. In addition, in context of forecasting, simulation is an attractive tool because it allows generation of random variables, which could define the complex behavior of input data for forecasting problems. Simulation could help in modeling of past data to be used for electricity demand process with relatively low cost.

Electricity, as a resource of energy, with its ever growing role in world economy, and its multi-purpose application in production and consumption has gained special attention. Through the development of societies and growth of economical activities, electricity becomes more effective on corporations and their services. Corporations use electricity as a production factor. Also, families directly or indirectly rely on electricity. Thus, energy consumption determines their and society’s economical welfare. A major application is the estimation of electricity consumption, which reveals the consumption growth in the forthcoming years.

This paper is organized as follows. In the next section, ANFIS is described. After introducing one of the most famous conventional models (ARIMA model), the importance of data preprocessing is explained and different data preprocessing methods are proposed. An algorithm for estimation and prediction with ANFIS is also developed in Section 5. The results are shown in Section 6.

2. The hybrid algorithm

The ANFIS model with preprocessed data (ANFISW) and ANFIS model without preprocessed data (ANFISWO) are considered to determine the impact of preprocessing on ANFIS. Moreover, the raw data is simulated by computer simulation to identify its probability distribution and the mean of probability distribution is then used as input data for ANFIS. This is of course repeated for each month. The advantage of simulated-based is to foresee if the stochastic nature of data has any impact on future demand estimation. Except ANFIS, the best time series model for the data set is also identified. This is done to compare the best ANFIS model against the best conventional time series model.

This algorithm has the following general basic steps:

Step 1: Collect data set in all available previous periods. Then, the stationary assumption should be studied for both ANFISW model and time series models. If the models are not covariance stationary, the most suitable preprocessing method should be selected and applied to the model. In addition, simulated data must be generated with computer simulation approach and above process must be considered.

Step 2: Divide data into two sets, one for estimating the models called the train data set and the other one for evaluating the validity of the estimated model called test data set. Usually train data set contains 70–90% of all data and remaining data are used for test data set (Zhang and Hu, 1998).

Step 3: Run and estimate all models. Input variables for ANFIS models can be selected using autocorrelation function (ACF). However, in most heuristic methods, selecting input variables is experimental or based on the trial and error method (Aznarte et al., 2007; Box & Jenkins, 1970; Cybenko, 1989; Garetta et al., 2006; Hwang, 2001; Jain & Kumar, 2007; Karunasingha & Liong, 2006; Nayak et al., 2004; Niskaa et al., 2004; Oliveira & Meira, 2006; Palmer, Montano, & Sese, 2006; Rumelhart & McClelland, 1986; Schifffmann, Joost, & Werner, 1992; Tseng et al., 2002; Zhang & Hu, 1998; Zhang & Qi, 2005). Importance of ACF approach is understood when difficulty and carelessness of trial and error method are considered. Irregular input selection is cause of its lack of preciseness. Even if all the previous lag combinations are used, the trial and error method will be time-consuming. For example, if all the combinations are selected from the recent 12 lag, the number of combination will be:

\[
\sum_{i=1}^{12} \binom{12}{i} = 2^{12} = 4096
\]

While ACF approach introduces few combinations for model input in comparison with trial and error process. As well, for time series model preferred ARIMA model is selected. Input variables can be selected using autocorrelation function (ACF) and partial autocorrelation function (PACF). By using result of selected model, McLeod–Li test is applied. The result of this test shows that nonlinear time series must be construct or not.

Step 4: Post process the estimated data in the models which the data were preprocessed.

Step 5: The reliability of each model is evaluated in this step. Granger–Newbold test is used to compare the models. First, the four fuzzy models (ANFISW, ANFISWO, ANFISMW & ANFISIMWO) are compared with each other to study the impact of preprocessing on ANFIS. The most suitable ANFIS is called ANFIS*. The preferred ARIMA model is selected from plausible model by aid of Granger–Newbold test and is called Li. The nonlinearity of the process is determined by McLeod–Li test. The result of Li is used for McLeod–Li test. If this test shows the nonlinearity process, a preferred nonlinear model is identified and used (defined as NL). Finally, Granger–Newbold test is used to select either ANFIS or time series. The main elements of the proposed algorithm are described next.

2.1. ANFIS

The fuzzy inference process we have been referring to so far is known as Mamdani’s fuzzy inference method, the most common methodology (Mamdani & Assilian, 1975). The so-called Takagi–Sugeno method of fuzzy inference introduced in 1985 (Sugeno, 1985). It is similar to the Mamdani method in many respects. The first two parts of the fuzzy inference process, fuzzifying the inputs and applying the fuzzy operator, are exactly the same. The main difference between Mamdani and Takagi–Sugeno method is...
that the Takagi–Sugeno output membership functions are either linear or constant. A typical rule in a Takagi–Sugeno fuzzy model has the form:

If Input 1 = x and Input 2 = y then Output is $z = ax + by + c$

For a zero-order Sugeno model, the output level $z$ is a constant $(a = b = 0)$.

A Takagi–Sugeno system is suited for modeling nonlinear systems by interpolating between multiple linear models, are a promising alternative to econometric. ANFIS are configured for specific identification retrieval, database management, and computer vision and data classification, through learning process. e.g. [Jang, 1993].

2.2. Conventional time series models

Time series models are quite well known to predict a variable behavior in the future by knowing its behavior in the past. One of the most famous time series models is Autoregressive Integrated Moving Average (ARIMA) model. The ARIMA model belongs to a family of flexible linear time series models that can be used to model many different types of seasonal as well as non-seasonal time series. In the most popular multiplicative form, the ARIMA model can be expressed as:

$$
\Phi_p(L) \Phi_q(L) \psi_t
$$

with

$$
\Phi_p(L) = 1 - \phi_1 L - \ldots - \phi_p L^p
$$

$$
\Phi_q(L) = 1 - \theta_1 L - \ldots - \theta_q L^q
$$

where $s$ is the seasonal length, $L$ is the back Shift operator defined by $L^s y_t = y_{t-s}$ and $\psi_t$ is a sequence of white noises with zero mean and constant variance. Eq. (1) is often referred to as the ARIMA ($p, q$) model, and $p$ and $q$ are the order of autoregressive and moving average terms, respectively.

Box and Jenkins (1970) proposed a set of effective model building strategies for identification, estimation, diagnostic checking, and forecasting of ARIMA models (Werbos, 1974). In the identification stage, the sample autocorrelation function (ACF) is plotted. A slowly decaying autocorrelation function suggests non-stationary behavior. In such circumstances, Box and Jenkins recommend differencing of the data. A common practice is to use a logarithmic transformation, if the variance does not appear to be constant. After preprocessing, if needed, ACF and PACF of preprocessed data are examined to determine all plausible ARIMA models. A well-estimated model should be parsimonious, fits the data well, has residuals that approximate a white noise process and has good out-of-sample forecasts. Box–Pierce Q-statistic can be used to test whether residuals can approximate the white noise process. So, the models which are not white noise process will be eliminated from consideration. Then, parsimony and well-fit of the model are checked using Akaike information criterion (AIC) and Schwartz Bayesian criterion (SBC). Finally, the Granger–Newbold test is applied to compare the forecasting performance of the models (Niskaa et al., 2004). For more information about AIC, SBC, Box–Pierce Q-statistic and Granger–Newbold test, refer to Appendices I–III.

Some nonlinear time series patterns were also developed mainly by Granger and Pristly. One of these nonlinear models is referred to as bilinear of which the first rank model of the bilinear model is as shown in Eq. (3).

$$
X_t = a X_{t-1} + b Z_t + c Z_{t-1} X_{t-1}
$$

In which $Z_t$ is the stochastic procedure and $a$, $b$, and $c$ are the model parameters. It should be noted that only the last part of the above equation is nonlinear. Another type of nonlinear models is the threshold auto regressive (TAR) models in which the parameters are dependent to the past values of the procedure. One example of such models is described by Eq. (4).

$$
X_t = \begin{cases} 
2X_{t-1} + Z_{t-1}^2 & \text{if } X_{t-1} < d \\
Z_{t-1} + Z_{t-1}^2 & \text{if } X_{t-1} \geq d 
\end{cases}
$$

Furthermore, the proposed algorithm fits the best linear or nonlinear model to the data set. This is quite important because most studies assume that linear time series such as ARIMA provide the best fit.

2.3. Data preprocessing

In time series methods creating a covariance stationary process is one of the basic assumptions. Also, using preprocessed data is more useful in most heuristic methods which require the investigation of stationary assumption for the models (Zhang & Hu, 1998). If the models are not covariance stationary, the most suitable preprocessed method should be defined and applied. In forecasting models, a preprocessing method should have the capability of transforming the preprocessed data to its original scale (called post processing). Therefore, in time series forecasting method, appropriate preprocessing method should have two main properties. It should make the process stationary and must have the post processing capability. The most useful preprocessed methods are studied in the sections.

2.3.1. Normalization

There are different normalization algorithms which are Min–Max Normalization, Z-Score Normalization and Sigmoid Normalization. The Min–Max normalization scales the numbers in a data set to improve the accuracy of the subsequent numeric computations. Tseng et al. (2002), Nayak et al. (2004), Niskaa et al. (2004), Karunasinghe and Liong (2006), Oliveira and Meira (2006), Gareta et al. (2006), Aznarte et al. (2007), and Jain and Kumar (2007) used this method in their articles to estimate time series functions using heuristic approach.

If $X_{old}$, $X_{max}$, $X_{min}$, $X_{std}$ are the original, maximum and minimum values of the raw data, respectively and $X_{max}$, $X_{min}$ are the maximum and minimum of the normalized data, respectively, then the normalization of $X_{old}$ called $X_{new}$ can be obtained by the following transformation function:

$$
X_{new} = \frac{(X_{old} - X_{min})}{(X_{max} - X_{min})}(X_{max} - X_{min}) + X_{min}
$$

In Z-score normalization the data are changed so that their mean and variance are 0 and 1, respectively. The transformation function used for this method is as follows where $std$ is the standard deviation of the raw data:

$$
X_{new} = \frac{X_{old} - \text{mean}}{\text{std}}
$$

The Sigmoidal Normalization uses a Sigmoid function to scale the data in the range of $[-1, 1]$. The transformation function is as follows:

$$
X_{new} = \frac{1 - e^{-x}}{1 + e^{-x}}
$$

2.3.2. The first difference method

The first step in the Box–Jenkins method is to transform the data so as to make it stationary. The difference method was proposed by Box and Jenkins (1970) and Werbos (1974). Also

\footnote{By definition, an ARIMA model is covariance stationary if it has a finite and time-invariant mean and covariance.}
Tseng et al. used this method in their article to estimation time series functions using heuristic approach (Tseng et al., 2002). The following transformation should be applied for the method:

\[ y_t = x_t - x_{t-1} \]  

(8)

However, for the first difference of the logarithm method the transformation is adjusted as follows:

\[ y_t = \log(x_t) - \log(x_{t-1}) \]  

(9)

2.4. The Mcleod–Li test

McLeod and Li (1983) was proposed to detect nonlinearity in time series data. The McLeod–Li Test seeks to determine if there are significant autocorrelations in the squared residuals from a linear equation (Zhang, 2001). To perform a test, autocorrelation coefficients in the Ljung–Box statistic are replaced by autocorrelation coefficients of the squared residuals. This statistic determines whether the squared residuals exhibit serial correlation. Ljung–Box statistic is discussed in Appendix II.

2.5. The Granger–Newbold test

Granger and Newbold (1974) test was proposed to compare two time series models (Ghiasi, Saidane, & Zimbra, 2005). First, \( x_t \) and \( z_t \) elements are formed as follow:

\[ x_t = er_{1t} + er_{2t} \quad \text{and} \quad z_t = er_{1t} - er_{2t} \]  

(10)

Let \( r_{xz} \) denote the sample correlation coefficient between \( \{x_t\} \) and \( \{z_t\} \). Granger and Newbold show that \( r_{xz}/(1 - r_{xz}^2)/(H - 1)^{0.5} \) has a \( t \)-distribution with \( H - 1 \) degrees of freedom. Thus, if \( r_{xz} \) is statistically different from zero, model 1 has a larger mean square error (MSE) than model 2 if \( r_{xz} \) is positive and model 2 has a larger MSE than model 1 if \( r_{xz} \) is negative.

2.6. Error estimation methods

There are four basic error estimation methods which are listed below:

- Mean absolute error (MAE)
- Mean square error (MSE)
- Root mean square error (RMSE)
- Mean absolute percentage error (MAPE)

They can be calculated by the following equations, respectively:

\[ \text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |x_i - \hat{x}_i| \]

\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \]

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2} \]

\[ \text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right| \]  

(11)

All methods, except MAPE have scaled output. MAPE method is the most suitable method to estimate the relative error because input data used for the model estimation, preprocessed data and raw data have different scales (Azadeh, Ghaderi, & Sohrabkhani, 2007; Azadeh, Ghaderi, Tarverdian et al., 2007). Therefore, MAPE is used as the major reference in this study.

3. The case study

The proposed algorithm is applied to 130 set of data which are the monthly consumption in Iran from April 1994 to February 2005. Detailed information can be obtained from “Electric Power Industry in Iran” (including transmission and distribution) published by the TAVANIR management organization (1992–2005). As simulated data is important for us, the related process discussed at first.

3.1. The simulated data

For this purpose, the distribution function of each month is calculated and then by using the related function for each month, the average value of that month is obtained. By this way, instead of using the deterministic value for each month we have its average value from the probable distribution the real value belongs to. When the distribution of each month is found, the amount of the square error and \( p \)-value of that distribution is also returned.

An example of the selected distribution functions for each month is shown in Table 1. The selected distribution functions are selected from a series of distribution functions according to their \( p \)-values and square errors. Table 2 shows an example of such selection for the period of 3/2001 to 6/2001 (see Fig. 1).

### Table 1

<table>
<thead>
<tr>
<th>Date</th>
<th>Distribution</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/2001</td>
<td>Normal (285000, 20900)</td>
<td></td>
</tr>
<tr>
<td>4/2001</td>
<td>Triangular (297000, 332000, 536000)</td>
<td></td>
</tr>
<tr>
<td>5/2001</td>
<td>Normal (365000, 18300)</td>
<td></td>
</tr>
<tr>
<td>6/2001</td>
<td>Normal (403000, 14200)</td>
<td></td>
</tr>
<tr>
<td>7/2001</td>
<td>Normal (415000, 12300)</td>
<td></td>
</tr>
<tr>
<td>8/2001</td>
<td>Normal (387000, 19000)</td>
<td></td>
</tr>
<tr>
<td>9/2001</td>
<td>Normal (341000, 17200)</td>
<td></td>
</tr>
<tr>
<td>10/2001</td>
<td>Normal (325000, 11400)</td>
<td></td>
</tr>
<tr>
<td>11/2001</td>
<td>Normal (323000, 14700)</td>
<td></td>
</tr>
<tr>
<td>12/2001</td>
<td>Normal (323000, 10000)</td>
<td></td>
</tr>
<tr>
<td>1/2002</td>
<td>Normal (330000, 12000)</td>
<td></td>
</tr>
<tr>
<td>2/2002</td>
<td>Normal (285000, 20900)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Month</th>
<th>Distribution Function</th>
<th>Square Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/2001</td>
<td>Uniform</td>
<td>0.077</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>Erlang</td>
<td>0.153</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>0.153</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>0.100</td>
<td>&gt;0.15</td>
</tr>
<tr>
<td></td>
<td>Triangular</td>
<td>0.099</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>0.160</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Lognormal</td>
<td>0.249</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Lognormal</td>
<td>0.348</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>5/2001</td>
<td>Uniform</td>
<td>0.046</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Erlang</td>
<td>0.144</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>0.144</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>0.025</td>
<td>&gt;0.14</td>
</tr>
<tr>
<td></td>
<td>Triangular</td>
<td>0.002</td>
<td>&gt;0.15</td>
</tr>
<tr>
<td></td>
<td>Gamma</td>
<td>0.150</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Lognormal</td>
<td>0.257</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Lognormal</td>
<td>0.308</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

For the months of interest, the distribution functions were selected from the tables above.
Using these results, the average values for each month is simulated by Visual Slam (Enlin, 1995). The selected distribution functions for each month is then generated 1000 times to obtain steady state. An example of the Visual Slam network can be seen in Fig. 2. The outputs of the simulation are the average and standard deviation of daily consumption values. Then, the upper and lower limits are constructed by \( \mu \pm 3\sigma \). Next, the daily results are multiplied by the number of days in per month and consequently monthly consumption values are obtained.

3.2. Step 1

It can be seen from Fig. 3a that raw data has a trend. As removing the trend is needed for more precise estimation in time series...
methods and also for studying impact of preprocessing on ANFIS, all preprocessing methods are applied on both ANFISW and ARIMA model. Consequently, the best preprocessing method is selected to convert the model to covariance stationary process. The results of applying preprocessing methods for given data set is discussed in the next section.

3.3. Step 2

The 130 rows of data are divided into 118 training data set and 12 test data set. Also, the 129 preprocessed data are divided into 117 training data set and 12 test data set.

3.3.1. Normalization

All three methods of normalization are used to preprocess the data, but as can be seen in Fig. 3b–d, in which the normalized consumption data are shown the trend of data cannot be removed by any of the normalization methods. Therefore normalization is not suitable for preprocessing the data set.

3.3.2. The first difference method

The preprocessed data using the first difference method is shown in Fig. 3e. Although, the first difference of the series seems to have a constant mean, the variance has an increasing pattern over time. Thus, this method is not covariance stationary and cannot be

Fig. 3. Raw and preprocessed data by different methods.
used for data preprocessing as prescribed by the algorithm. It can be seen in Fig. 3f that the first difference of logarithm is the most likely candidate to have covariance stationary process. Moreover, it is the most applicable preprocessing method for the data set.

**Fig. 4.** The ACF and PACF chart for preprocessed data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of terms for each linguistic variables</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.036</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.025</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.045</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.021</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.015</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0.025</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>0.035</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>0.039</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>0.025</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0.026</td>
</tr>
</tbody>
</table>

**Table 3** Architecture of the 10 models and network’s error.

**Fig. 5.** The comparison of ANFISW output with actual data.

**Fig. 6.** 5th ANFIS model structure.

**Fig. 7.** The comparison of ANFISWO output with actual data.

**Fig. 8.** The comparison of ANFISIMO monthly electricity consumption estimation with actual data.

**Fig. 9.** The comparison of ANFISIMW output with actual data.
3.4. Step 3

For fuzzy regression models, ACF approach is used to select input variables among 12 lags. According to Fig. 4, \( y_t \) is the function of consumption in the 1th lag in preprocessed data. Similarly, \( y_t \) is the function of consumption in the 1th and 12th lags in raw data. Input variables are selected for simulated data with ACF too.

3.4.1. Estimation of electricity consumption by ANFISW

In order to get the best ANFIS for the estimation of electricity consumption, 10 models are tested to find the best architecture. The MAPE values of the mentioned models are shown in Table 3. Gaussian and bell shaped membership functions are becoming increasingly popular for specifying fuzzy sets as they are nonlinear and smooth and their derivatives are continuous. Gradient methods can be used easily for optimizing their design parameters. In present study, Gaussian membership function\(^5\) is considered as term membership function. Therefore, 5th model is selected for estimating the electricity consumption.

The comparison of ANFISW output with test actual data is shown in Fig. 5. Fig. 6 shows structure of fifth model.

3.4.2. Estimation of electricity consumption by ANFISWO

With similar approach, model with 0.05 MAPE value is selected for estimating the electricity consumption. The comparison of ANFISWO with actual data is shown in Fig. 7. Figs. 8 and 9 show the comparison of actual data with ANFISIMW and ANFISIMWO, respectively.

3.4.3. Estimation of electricity consumption by time series model

In order to find the preferred time series model appropriate ARIMA model must be estimated. The preferred ARIMA model is selected by aid of Granger–Newbold test. Then by using the result of this model, McLeod–Li test is applied. The result of this test shows that nonlinear time series must be used or not. The following sections show that ARIMA model is sufficient for the case study. Moreover, various researches show that linear time series is the most ideal for our case study (Zhu, 1998).

3.4.3.1. ARIMA model

To find the best time series model preprocessed approach is used. Fig. 4 shows ACF and PACF charts, respectively. The theoretical ACF of a pure MA(\( q \)) process cuts off to zero at lag \( q \) and theoretical ACF of an AR(1) models decays geometrically. Examination of Fig. 4 suggests that neither of these specifications seems appropriate for the electricity consumption. The ACF does not decay geometrically and it is suggestive of an AR(2) process or a process with both autoregressive and moving average components. A seasonal factor at lag 12 is incorporated due to availability of monthly data. Therefore, 5 models are considered which are AR(1), AR(2), ARIMA(1,1), ARIMA(1,(1,12)) and AR(2, MA(12)) for the training data set. Table 4 shows models’ information including Q-statistics, AIC, SBC and coefficients estimation. Box–Pierce Q-statistic test shows that AR (1) and AR (2) should be eliminated. However, as measured by AIC and SBC, ARIMA(1,(1,12)) and AR(2,MA(12)) do not fit the data as well as the ARIMA (1,(1,12)). Also Granger–Newbold test shows that ARIMA(1,(1,12)) has the better forecasting performance than ARIMA(1,1) and AR(2, MA(12)). The value of r-statistic is 2.52 and 2.32, respectively.

3.4.3.2. The Mcleod–Li test

Residuals of ARIMA(1,(1,12)) are used to run Mcleod–Li test examination of (Table 5) shows that nonlinearity condition is not satisfied. So, ARIMA(1,(1,12)) is called TM.

3.4.3.3. Nonlinear time series model

As mentioned, ARIMA model is sufficient for our data and there is no need to identify an appropriate nonlinear time series model. The comparison of ARIMA (1,(1,12)) output with test actual data is shown in Fig. 10.

<table>
<thead>
<tr>
<th>( \text{Ljung–Box Q-statistic} )</th>
<th>( Q(4) )</th>
<th>( Q(8) )</th>
<th>( Q(12) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>[(1.4(0.846), 7.3(0.5), 11.2(0.51)]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^5\) Gauss \( \phi(x; \sigma, c) = e^{-\frac{x^2}{2\sigma^2}} \).
3.5. Step 4

Since data are preprocessed for ANFISW, ANFISIMW and ARIMA models, the estimated data obtained by these models should be post processed. Let ANFISXi (i: 1…12) be the ANFISW output for preprocessed test data. ANFISXi is postprocessed by this formula:

\[(10^{\text{ANFISXi}}) / c3 \cdot x(i / c01)^{24} \]

ARIMA output postprocessed is similar to above mentioned case.

3.6. Step 5

ANFISWO (model 1) and ANFISW (model 2) are compared by Granger–Newbold test. The value of t-statistic is statistically different from zero (2.602). Since \( r_{xz} \) (which is explained in Appendix I) is positive, then ANFISW has better forecasting performance than ANFISWO model. The same procedure is also performed to compare FRSIMW with FRSIMW. At last, the test results show that ANFIS* (ANFISW) has better forecasting performance than TM(ARIMA (1,(1,12))). The test results show that ANFIS* has better forecasting performance than TM. Also, it can be seen from Table 6 that ANFIS* has the least MAPE which shows the efficiency of ANFISW among other models.

4. Comparison with other intelligent methods

With the aid of ANFISW model, electricity for the next 12 month is forecasted (Fig. 11). Table 7 shows the MAPE estimation for genetic algorithm (GA), artificial neural network (ANN) versus the proposed algorithm (Azadeh, Ghaderi, & Sohrabkhani, 2007; Azadeh, Ghaderi, Tarverdian et al., 2007). Examination of this table shows that the proposed algorithm provides good estimation with respect to GA and ANN.

5. Conclusion

This paper presented a hybrid adaptive network fuzzy inference system (ANFIS), computer simulation and time series techniques to estimate and predict energy consumption estimation. The difficulty with electricity consumption estimation modeling approach such as time series is the reason for proposing the hybrid approach of this study. The algorithm is ideal for uncertain, ambiguous and complex estimation and forecasting. Computer simulation was developed to generate random variables for monthly electricity consumption. Various structures of ANFIS were examined and the preferred model is selected for estimation by the proposed algorithm. Finally, the preferred ANFIS and time series models were selected by Granger–Newbold test. Monthly electricity consumption in Iran from 1995 to 2005 was considered as the case of this study. The superiority of the proposed algorithm is shown by comparing its results with genetic algorithm (GA) and artificial neural network (ANN).

This is the first study that presented a hybrid simulation-adaptive network fuzzy inference system (ANFIS) for improvement of electricity consumption estimation. The unique features of the proposed algorithm are two fold. First, ANFIS is ideal for complex and uncertain data because it is composed of both ANN and fuzzy systems. Second Monte Carlo simulation is used to generate input variables whereas the conventional methods use deterministic data. The superiority of the proposed algorithm is shown by comparing its results with time series, genetic algorithm (GA) and ANN. Furthermore, the MAPE estimation of GA and ANN versus the proposed algorithm showed the appropriateness of the proposed algorithm.

### Appendix I. Akaike information criterion (AIC) and Schwartz Bayesian Criterion (SBC)

The two most commonly used model selection criteria are the AIC and the SBC. These criteria are used to select the most appropriate model. They have the following formulas:
Appendix II. Box–Pierce Q-statistic

If AIC (or SBC) for A is smaller than for B, significant autocorrelations can be rejected.

References


