Energy 49 (2013) 323–328

Contents lists available at SciVerse ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Electricity demand estimation using an adaptive neuro-fuzzy network: A case study from the Ontario province – Canada

Gholamreza Zahedi [a,*], Saeed Azizi [b], Alireza Bahadori [c], Ali Elkamel [d], Sharifah R. Wan Alwi [a]

[a] Process Systems Engineering Centre (PROSPECT), Faculty of Chemical Engineering, Universiti Teknologi Malaysia, UTM Skudai, 81310 Johor Bahru, Johor, Malaysia
[b] Chemical Engineering Department, Razi Un., Kermanshah, Iran
[c] Southern Cross University, School of Environment, Science and Engineering, Lismore, NSW, Australia
[d] Department of Chemical Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

ARTICLE INFO

Article history:
Received 3 May 2012
Received in revised form
8 October 2012
Accepted 11 October 2012
Available online 16 November 2012

Keywords:
Electricity demand
Neuro-fuzzy
Forecasting

ABSTRACT

Electricity is an important asset that influences not only the economy, but political or social security of a country. Reliable and accurate planning and prediction of electricity demand for a country are therefore vital. In this paper, electricity demand in Ontario province of Canada from the year 1976–2005 is modeled by using an (adaptive neuro fuzzy inference system) ANFIS. A neuro fuzzy structure can be defined as an ANN (artificial neural network) which is trained by experimental data to find the parameters of (fuzzy inference system) FIS. Inputs for the model include number of employment, (gross domestic product) GDP, population, dwelling count and two meteorological parameters related to annual weather temperature. The data were collected and screened using statistical methods. Then, based on the data, a neuro-fuzzy model for the electricity demand is built. It was found that electricity demand is most sensitive to employment.

1. Introduction

Several investigations have been carried out to find important parameters affecting electricity demand and also the interaction between these parameters. Designed and calculated model will help us to manage energy consumption and distribution efficiently. Most studies have focused on the relationship between electricity demand and economical parameters such as (gross domestic product) GDP, (gross national product) GNP, national income, and the rate of employment as well as unemployment. Sari and Soytas [1] studied the relationship between different sources of electricity consumption, employment and national income growth in Turkey. Narayan and Smyth [2] carried out the same study in Australia. They evaluated both long and short term relationship between electricity consumption, employment and real income. Relationships between GDP and electricity consumption in ten newly industrialized Asian countries were estimated by Chen et al. [3]. They studied long run relationship in China, Hong Kong, India, Indonesia, Korea, Malaysia, Philippines, Singapore, Taiwan and Thailand. In another attempt, German Institute for Economic Research (DIW) was commissioned by "German Advisory Group on Economic Reform in Ukraine" in 1998 to predict electricity demand in Ukraine until the year 2010. A comparison of the relationship between renewable and non-renewable electricity consumption and real GDP in the US using annual data from 1949 to 2006 was done by Payne [4]. Bowden and Payne [5] used these data in 2008 to check the causal relationship between electricity consumption and real GDP. Studying the time series properties of electricity consumption of G-7 countries was the subject of Soytas and Sari [6]. In Pakistan, Aqeel and Butt [7] found out that economic growth affects the total electricity consumption. They also discovered that economic growth leads to growth in petroleum consumption but however electricity consumption leads to economic growth without feedback. De Vita et al. [8] found the same results for Namibia. Their research for the period between the year 1980–2002 showed that electricity consumption respond positively to changes in GDP and negatively to changes in electricity price and air temperature. Hainoun et al. [9] found that both electricity and electricity demand growth rates are lower than the corresponding GDP growth rates in Syria. In some literature, other parameters which are not economical are also selected. For example Valor et al. [10] tried to analyze the relationship between electricity load and daily air temperature in Spain. More recently many studies have been conducted on short/long term electricity demand/load forecasting [11–25], but application of neuro fuzzy

* Corresponding author. Tel.: +60 7 5535583; fax: +60 7 5581463.
E-mail addresses: grzahedi@cheme.utm.my, grzahedi@yahoo.com (G. Zahedi).

0360-5442/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.energy.2012.10.019

logic for forecasting electricity demand is still unexplored. In this paper, an ANFIS network (adaptive neuro fuzzy inference system) was designed to map six parameters as input data (i.e. employment, GDP, dwelling, population, HDD and CDD) to electricity demand as output variable.

1.1. Problem statement

All of the referred literature have applied classical methods and statistical techniques to predict electricity demand, but this paper employs a new approach (neuro-fuzzy) in order to achieve a model which can present more accurate prediction. Some researchers like Abraham and Nath [26] and Beccali et al. [27] have also used new approaches, which have been applied in many branches of science and are becoming more popular than classical methods. Abraham and Nath [26] have applied a neuro fuzzy approach to model the electricity demand in Victoria. They found that even by using auto regressive integrated moving average model and an (artificial neural network) ANN in their paper, neuro fuzzy network provides a more accurate prediction.

In this paper, a neuro fuzzy network has been designed to model the electricity demand of Ontario province in Canada based on six input variables. These variables are employment, GDP, population, dwelling and two other variables to indicate how hot or cold the weather is. All these data were collected from CANSIM II, Statistics Canada's key socio-economic database, and Environment Canada (2005), with statistical analysis. Electricity demand data from 1976 to 2004 is obtained from IESO (Independent Electricity System Operators).

In the literature review, the neuro-fuzzy modeling is described. The section is divided into three sub-sections i.e. description of ANN, fuzzy logic inference systems, and adaptive neuro fuzzy inference system description. In research methodology section, Ontario electricity demand forecasting is presented followed by the designed and validation of ANFIS model in the next section. In the last section, the sensitivity analysis of the designed network is evaluated.

2. Neuro fuzzy system: ANFIS

Fuzzy logic is a system that can be applied to transform linguistic concepts to mathematical and computational structure for many purposes. But fuzzy systems do not have good ability to learn and adapt to changing the conditions [28,29]. Combination of ANN and fuzzy logic methods can overcome this drawback [29,30]. A combined fuzzy logic systems and neural network helps researchers to choose and design parameters of fuzzy logic inferences [31]. Not having a systematic procedure for choosing membership function type and parameters leads us to use an (adaptive neuro fuzzy inference system) ANFIS [32].

ANFIS applies a combination of error back propagation algorithm and least squares method as a hybrid algorithm to adjust the membership functions of a fuzzy logic system optimally [33]. After determining final error (difference between results and targets) of the system, derivative of squared error with respect to each node's output as error signals are fed back to the system to be used to alter membership function parameters [34,35].

There are two fuzzy style inferences which are mostly used, Mamdani-style inference and Sugeno-style inference. Mamdani-style is based on Lotfi Zadeh's 1973 paper [33] while Sugeno-style is based on Takagi-Sugeno-Kang method of fuzzy inference [36,37].

In Fig. 1, a typical structure of an ANFIS based on Sugeno fuzzy modeling is shown. Two inputs x, y and one output z were considered for simplicity. Two membership functions were considered for each input in this network.

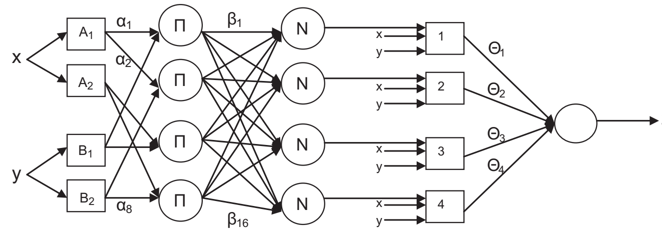

Fig. 1. A typically ANFIS with 2 inputs and 1 output.

The first layer of this structure contains nodes which generate output of each membership functions for inputs.

$$\alpha_i = \mu_{Aj}(x) \quad i = 1,2 \text{ for } j = 1 \text{ and } 3,4 \text{ for } j = 2 \tag{1}$$

$$\alpha_i = \mu_{Bj}(y) \quad i = 5,6 \text{ for } j = 1 \text{ and } 7,8 \text{ for } j = 2 \tag{2}$$

where μ_{Aj} and μ_{Bj} are membership functions for x and y respectively. In the second layer, nodes which are labeled Π, calculate the firing strength of a rule by using equation (3).

$$\beta_i = \mu_{Aj}(x)\mu_{Bj}(y) \quad i = 1,2,\ldots,6 \tag{3}$$

Nodes labeled N, in third layer calculates the ratio of a rule firing strength to the sum of all rules firing strength:

$$\gamma_i = \frac{\alpha i}{\sum_{i=1}^{16} \alpha i} \quad i = 1,2,3,4 \tag{4}$$

After calculating this ratio, the outputs of fourth layer are obtained by equation (5):

$$\Theta_i = \gamma_i z_i \quad i = 1,2,3,4 \tag{5}$$

Using a Sugeno fuzzy modeling style in this network, z_i is calculated by equation (6):

$$z_i = p_i x + q_i y + r_i \quad i = 1,2,3,4 \tag{6}$$

The last single node in the fifth layer, computes overall output as summation of all incoming signals, which is expressed as below:

$$z = \sum_{i=1}^{4} \Theta_i \tag{7}$$

3. Research methodology

In this paper, an ANFIS has been employed to estimate electricity demand in Ontario, Canada. A statistical study has been carried out on available data to find the affecting factors and prepare data for model building.

3.1. Statistical preprocessing

3.1.1. Available data

Seven parameters in the period from 1976 to 2011 were available. These data included the number of employment, GDP, population, dwelling count, (degree days) DD,[1] number of new housing and bank of Canada interest rate. The first five parameters are obtained from CANSIM II, Statistics Canada's key Socio-economic database, and Environment Canada and two others are obtained from Ref. [38].

[1] For a particular year, DD is summation of temperatures of days during the year.

3.1.2. Data analysis using Pearson coefficient

Francis Galton in the 1880s introduced a factor to show measure of correlation or linear dependence between two variables X and Y [36,39]. This factor, named after Karl Pearson as "Pearson correlation coefficient", is typically denoted by "r" and can be a value between $+1$ and -1. Being more close to $+1$ or -1 shows more correlation between X and Y. Positive sign means that data have direct relationship, but negative measures means indirect relationship between data. 'r' can be calculated by using equation (8).

$$r = \frac{\sum(X - X')(Y - Y')}{\sqrt{\sum(X - X')^2 \sum(Y - Y')^2}} \tag{8}$$

where X' and Y' are average of variables X and Y respectively.

Pearson correlation factor is widely used in sciences as a measure of strength of linear dependence between two variables. Measure of this coefficient between each seven parameters and electricity demand was calculated. The results are listed in Table 1.

It is obviously that employment; GDP, population and dwelling count have high correlation with electricity demand. But the remaining parameters do not have good correlation to be selected as an independent variable for model building.

All these four parameters are social or economic and none of them are climatic variables. Environmental conditions that affect electricity demand is described by Rodgers [36], where the hottest temperature and the coldest temperature of a day were selected as environmental temperature effect (one for showing how cold the weather has been, (HDDd) and the other for showing how hot the weather has been (CDDd)). These two new parameters are defined as below:

$$\mathrm{HDDd} = \max\{0, \mathrm{Tb}_1 - \mathrm{Td}\} \tag{9}$$

$$\mathrm{CDDd} = \max\{0, \mathrm{Td} - \mathrm{Tb}_2\} \tag{10}$$

where Tb_1 is the base temperature, which was set as 10 °C Tb_2 is another base temperature which was chosen to be 20 °C. Td is the mean temperature of a given day and d is the number of days in the year (365 or 366).

In this study, summation of HDDd during a year (HDD) and summation of CDDd during a year (CDD) were selected as new parameters for model development. Measure of correlation between these new variables and electricity demand was calculated. The results are listed in Table 2 which shows that HDD and CDD have higher correlation than DD.

4. ANFIS model results

4.1. Designed network structure

As mentioned in the previous section, six parameters that had the most important effects on the actual electricity demand were selected for input variables (i.e. employment, GDP, dwelling, population, HDD and CDD). Based on these data, an ANFIS network with Sugeno-style inference system has been designed which maps six independent variables as input data to electricity demand as output. MATLAB 7.6 was employed for model building. Three Gaussian membership functions have been considered for each input data. Fig. 2 shows the final and best obtained membership functions.

Table 1
Pearson correlation coefficient between electricity demand and input variables.

Parameter	Pearson coefficient	Parameter	Pearson coefficient
Employment	0.965	DD	−0.357
GDP	0.923	Number of housing start	0.225
Population	0.952	Bank of Canada interest rate	−0.695
Dwelling count	0.959		

Table 2
Pearson correlation coefficient between electricity demand and HDD − CDD.

Parameter	Pearson coefficient
HDD	−0.460
CDD	0.466

From 36 available data sets, 30 sets were selected to train the network. Six remaining data sets were applied to validate the trained network. The procedure ensures that the designed network produces good results for any range of data.

After training the network, a (mean square error) MSE of 8.9251×10^{-13} was obtained for training the data. The low training error enabled the trained network to estimate unseen data with high precision. The best obtained network MSE is 0.0016 for test data.

4.2. Model validation

Censuses in Canada are conducted in a five-yearly interval. The latest census run by IESO provides exact data for employment, GDP, population and dwelling count. In order to build a forecasting model to the year 2015, a linear trend was assumed (Table 3).

These equations can be used to find the average values for the next 15 years period from 2012 to 2026. Having the values of all input parameters for the year 2006 the electricity demand for the year can be predicted. The calculated value is 165.46 TWh where the actual data reported by IESO, is 151 TWh. It means a good validation for the model and confirms its power to predict electricity demand in future years.

4.3. Prediction of electricity demand until 2015

After validating the model, it can be used to forecast the electricity demand in future. Linear lines were fitted to the data based on the linear trend of the data. Table 4 provides the linear equation form.

By using these equations, it is possible to predict the values of these parameters in the future (in this paper 2012–2015). The trend of change for HDD and CDD is not liner. In this case, the average values for every 5 years have been calculated to find out whether they have a special trend. To compare the annual changes of parameters, changes of these parameters needs to be performed step by step.

In this case, using the fitted equations, all six independent variables for future (2012–2015) were obtained (Table 5). Using these inputs, ANFIS network can provide estimations. Fig. 3 depicts electricity demand value from 1976 to 2015.

5. Sensitivity analysis

One of the important benefits of having a forecasting model is to find the effects of independent parameters (in this study: employment, GDP, dwelling, population, HDD and CDD) on electricity demand.

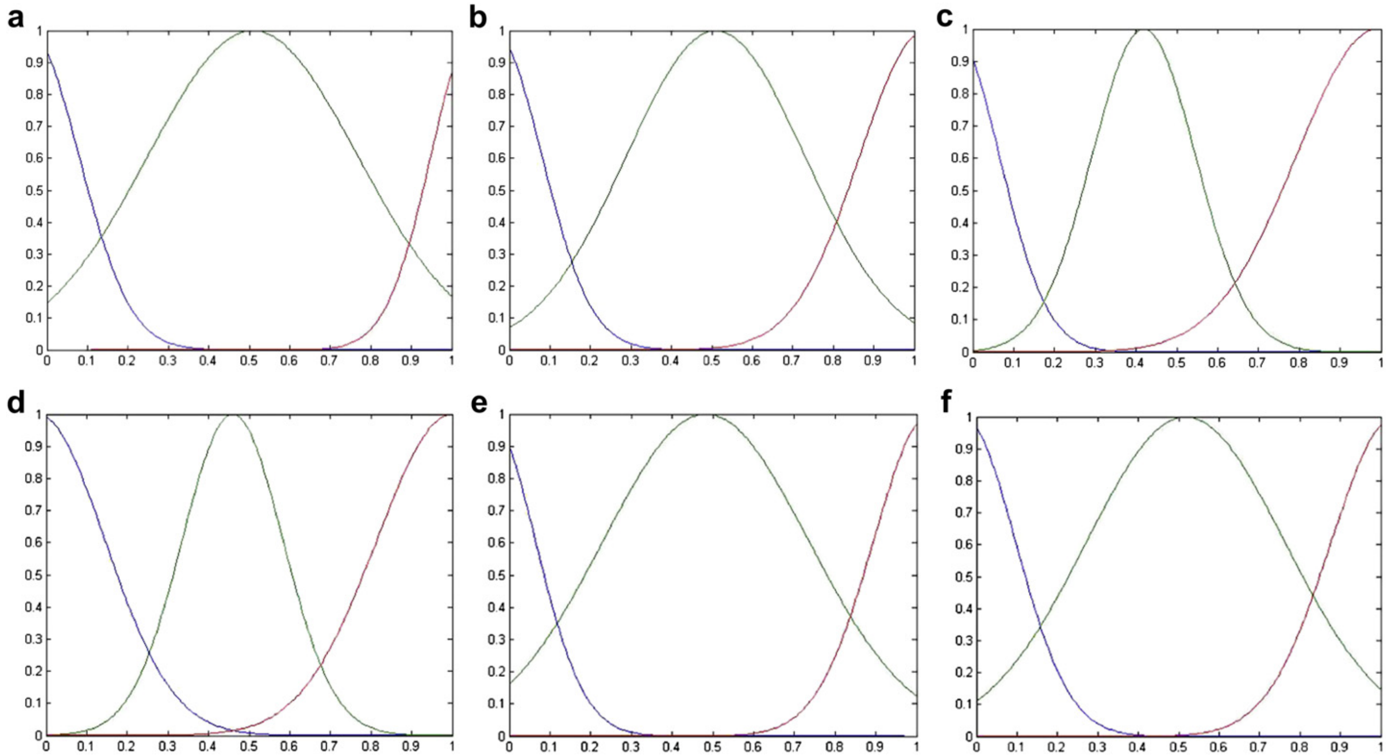

Fig. 2. Membership functions of (a) employment, (b) GDP, (c) dwelling, (d) population, (e) HDD and (f) CDD.

Table 3
Average changing for HDD and CDD for 15 years.

Parameter	Equation of fit curve
HDD	$y = -211.2x + 4544$
CDD	$y = 45.87x + 196.8$

Table 5
Polynomial equations for 5 year average changes for HDD and CDD.

Parameter	Equation of fit curve
HDD	$Y = 16X^2 - 186.5X + 0.4637$
CDD	$Y = 2.994X^2 - 0.752X + 0.2229$

To investigate sensitivity of employment, amount of GDP, dwelling population and also HDD and CDD were fixed in their 2005 values. Amount of employment changed based on trend obtained in previous section (equation of employment in Table 4) and its value for the next 100 years was calculated. Fig. 4 illustrates the results.

The GDP, amount of employment, dwelling, population, HDD and CDD were fixed at 2005 value and GDP was varied using the second equation in Table 4.

Sudden fall and rise in energy demand have been experienced in year 1980 as shown in Fig. 4. So the predictions follow the same pattern. Another reason for observing this sudden fall and rise is that only one parameter is varied and the amount of other variables are constant. In real situation, these input parameters vary and can give us accurate value. In sensitivity analysis, it is obvious that the similar trend is followed in Fig. 5.

Table 4
Input parameters trends.

Parameter	Equation of fitted curve[a]
Employment	$Y = 141.4X + 0.2217$
GDP	$Y = 15{,}628X + 22{,}712$
Dwelling	$Y = 69.10X + 0.2673$
Population	$Y = 150.4X + 0.8090$

[a] Y indicates parameter value and X indicates order of data.

The dwelling, amount of employment GDP, population, HDD and CDD were fixed at 2005 value and the dwelling was varied using the third equation in Table 4.

Finally, the population, amount of employment, GDP, dwelling, HDD and CDD were set at the year 2005 value and the population

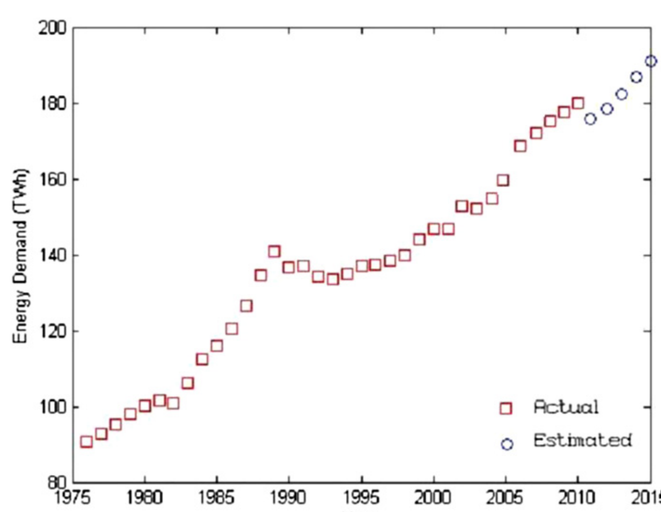

Fig. 3. Estimation of electricity demand from 1976 to 2015.

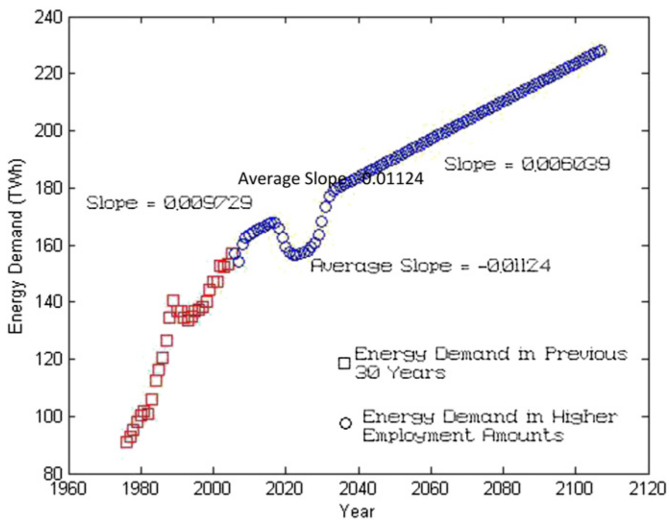

Fig. 4. Effect of employment on electricity demand.

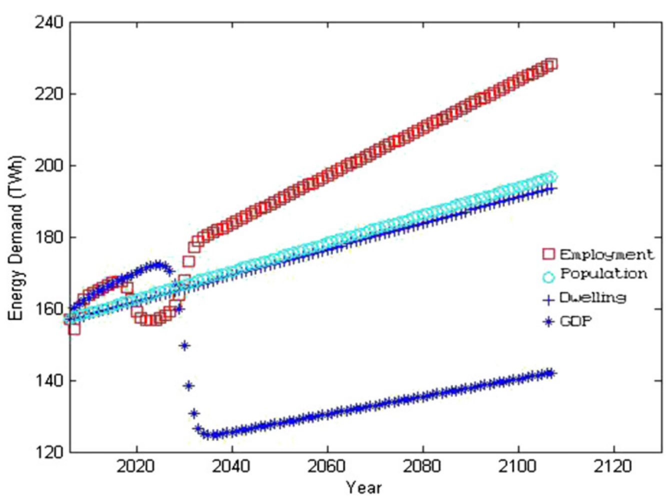

Fig. 5. Comparison of ANFIS outputs for changes in different parameters.

was changed was varied using the last equation in Table 4. Fig. 5 illustrates the sensitivity results.

For the purpose of comparison, all sensitivity analysis results have been plotted in Fig. 5. It can be seen that employment has the most important effect on electricity demand. Comparing the slopes confirm this conclusion. As shown in Fig. 5, the slopes of electricity demand due to changes in dwelling and population are 0.005268 and 0.002596 respectively. But for employment and GDP, there are three regions with three different slopes. If the slope of the decreasing regions i.e. the second region are ignored to compare the effect of parameters on increasing electricity demand, the average slope of increasing regions (first and third regions) is

Table 6
Slope of changing in electricity demand based on different parameters.

Parameter	Slope
Employment[a]	0.0643242
GDP[a]	4.39×10^{-5}
Dwelling	0.005268
Population	0.002596

[a] Slope of decreasing region is ignored.

calculated to be 0.0643242 for Employment and 4.39×10^{-5} for GDP. These slopes are listed in Table 6.

It is clear that employment yields the biggest slope and it confirms our conclusion that it is the most important parameter affecting electricity demand.

6. Conclusion and remarks

In this paper, an ANFIS network (adaptive neuro fuzzy inference system) was designed to map six parameters as input data (i.e. employment, GDP, dwelling, population, HDD and CDD) to electricity demand as output variable. To reduce the number of independent variables, input parameters were selected using statistical analysis in order to determine the parameter that has the highest impact on electricity demand. The network had excellent forecasting capacity with MSE of 0.0016.

Electricity demand until 2015 was predicted. By analyzing sensitivity of electricity demand based on changes of independent parameters, it was found out that employment affects electricity demand the most.

In term of econometric systems, euro fuzzy systems are more accurate than regression models. Compared to neural network models the neuro fuzzy models are robust in future energy estimations while ANN models fail in such extrapolations. Also neuro fuzzy models require less data compared to ANN models. Developing neuro fuzzy models are time consuming which is a drawback of this method compared to regression methods or Fourier series. Also for operators and technicians it is easy and understandable to work with regressions which are feasible. If the system is not nonlinear it is recommended the regression models be employed in demand predictions. For non-linear systems, neuro fuzzy, Fourier or semi empirical models can be used for forecasting.

In this paper, all inputs were considered as independent variables. The inputs can be further improved by using hybrid models such as the use of fuzzy inference systems or neural networks (or even statistical methods like regression or time series) to find values of these parameters before entering them into the ANFIS network. It is suggested that this case to be studied by using hybrid methods in future researches.

References

[1] Sari R, Soytas U. Disaggregate electricity consumption, employment and income in Turkey. Energy Economics 2004;26:335–44.
[2] Narayan PK, Smyth R. Electricity consumption, employment and real income in Australia evidence from multivariate Granger causality tests. Energy Policy 2005;33:1109–16.
[3] Chen ST, Kuo HI, Chen CC. The relationship between GDP and electricity consumption in 10 Asian countries. Energy Policy 2007;35:2611–21.
[4] Payne JE. On the dynamics of electricity consumption and output in the US. Applied Electricity 2009;86:575–7.
[5] Bowden N, Payne JE. The causal relationship between U.S. electricity consumption and real output: a disaggregated analysis. Journal of Policy Modeling 2008;31:180–8.
[6] Soytas U, Sari R. Electricity consumption and GDP: causality relationship in G-7 countries and emerging markets. Energy Economics 2003;25:33–7.
[7] Aqeel A, Butt MS. The relationship between electricity consumption and economic growth in Pakistan. Asia-Pacific Development Journal 2001;8:101–10.
[8] De-Vita G, Endresen K, Hunt LC. An empirical analysis of electricity demand in Namibia. Energy Policy 2006;34:3447–63.
[9] Hainoun A, Seif-Eldin MK, Almoustafa S. Analysis of the Syrian long-term electricity and electricity demand projection using the end-use methodology. Electricity Policy 2006;34:1958–70.
[10] Valor E, Meneu V, Caselles V. Daily air temperature and electricity load in Spain. Journal of Applied Meteorology and Climatology 2001;40:1413–21.
[11] Deihimi A, Showkati H. Application of echo state networks in short-term electric load forecasting. Energy 2012;39(1):327–40.
[12] Behrang MA, Assareh E, Ghalambaz M, Assari MR, Noghrehabadi AR. Forecasting future oil demand in Iran using GSA (gravitational search algorithm). Energy 2011;36(9):5649–54.

[13] Wang J, Zhu S, Zhang W, Lu H. Combined modeling for electric load forecasting with adaptive particle swarm optimization. Energy 2010; 35(4):1671–8.
[14] Nguyen HT, Nabney IT. Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. Energy 2010;35(9):3674–85.
[15] Ekonomou L. Greek long-term energy consumption prediction using artificial neural networks. Energy 2010;35(2):512–7.
[16] Azadeh A, Saberi M, Seraj O. An integrated fuzzy regression algorithm for energy consumption estimation with non-stationary data: a case study of Iran. Energy 2010;35(6):2351–66.
[17] Pao HT. Forecasting energy consumption in Taiwan using hybrid nonlinear models. Energy 2009;34(10):1438–46.
[18] Amjady N, Keynia F. Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. Energy 2009;34(1):46–57.
[19] Pappas SS, Ekonomou L, Karamousantas DC, Chatzarakis GE, Katsikas SK, Liatsis P. Electricity demand loads modeling using auto regressive moving average (ARMA) models. Energy 2008;33(9):1353–60.
[20] Yao AWL, Chi SC, Chen CK. Development of an integrated Grey-fuzzy-based electricity management system for enterprises. Energy 2005;30(15):2759–71.
[21] Tan Z, Zhang J, Wang J, Xu J. Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. Appl Energy 2010;87(11):3606–10.
[22] García-Ascanio C, Maté C. Electric power demand forecasting using interval time series: a comparison between VAR and iMLP. Energy Policy 2010;38(2): 715–25.
[23] Mamlook R, Badran O, Abdulhadi E. A fuzzy inference model for short-term load forecasting. Energy Policy 2009;37(4):1239–48.
[24] Barros CP, Managi S. Productivity assessment of Angola's oil blocks. Energy 2009;34(11):2009–15.
[25] Barros CP. Efficiency analysis of hydroelectric generating plants: a case study for Portugal. Energy Economics 2008;30:59–75.
[26] Abraham A, Nath B. A neuro-fuzzy approach for modeling electricity demand in Victoria. Applied Soft Computing 2001;1:127–38.
[27] Beccali M, Cellura M, LoBrano V, Marvuglia A. Short-term prediction of household electricity consumption: assessing weather sensitivity in a Mediterranean area. Renewable and Sustainable Electricity Reviews 2008;12: 2040–65.
[28] Negnevitsky M. Artificial intelligence: a guide to intelligent systems. Addison-Wesley; 2005.
[29] Goonatilake S, Sukhdev S. Intelligent hybrid systems. Toronto: Wiley Publishing; 1995.
[30] Nauck D, Klawonn F, Kruse R. Foundations of neuro-fuzzy systems. New York: John Wiley; 1997.
[31] Ahmed T, Hamza A, Abdel GA. Adaptive neuro-fuzzy controller of switched reluctance motor. Serbian Journal of Electrical Engineering 2007;4:23–34.
[32] Rojer-Jang JS. ANFIS: adaptive network, based fuzzy inference system. IEEE Transactions on Systems, Man and Cybernetics 1993;23:665–84.
[33] Konstantaras A, Varley MR, Vallianatos F, Collins G, Holifield P. A neuro-fuzzy approach to the reliable recognition of electric earthquake precursors. Natural Hazards and Earth System Sciences 2004;4:641–6.
[34] Panda G, Panda S, Ardil C. Hybrid neuro fuzzy approach for automatic generation control of two – area interconnected power system. International Journal of Computational Intelligence 2009;5:80–4.
[35] Zadeh LA. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man, and Cybernetics 1973; 3(1):28–44.
[36] Rodgers JL, Nicewander WA. Thirteen ways to look at the correlation coefficient. The American Statistician 1988;42:59–66.
[37] Chinnam RB, Baruah P. A neuro-fuzzy approach for estimating mean residual life in condition-based maintenance systems. International Journal of Materials and Product Technology 2004;20:166–79.
[38] Chui F, Elkamel A, Surit R, Crosiet E, Douglas PL. Long term electricity demand forecasting for power system planning using economic, demographic, and climatic variables. European Journal of Industrial Engineering 2009;3:277–304.
[39] Stigler SM. Francis Galton's account of the invention of correlation. Statistical Science 1989;4:73–86.