International Journal on "Technical and Physical Problems of Engineering" (IJTPE)
Published by International Organization on TPE (IOTPE)
ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

June 2011	Issue 7	Volume 3	Number 2	Pages 87-91

LONG TERM ELECTRIC PEAK LOAD FORECASTING OF KUTAHYA USING DIFFERENT APPROACHES

Y. Aslan S. Yavasca C. Yasar

Electrical Engineering Department, University of Dumlupinar, Kutahya, Turkey
yaslan@dpu.edu.tr, serkanyavascaoglu@hotmail.com, cyasar@dpu.edu.tr

Abstract- In the system planning, control and management of power distribution companies the long term peak load demand projections has an important role. In general, the load forecasting is performed by studying the past events. Long term forecasting of future peak load demand is very important for the economic and secure operation of power systems. In this study the long term peak load forecasting is performed for the city of Kutahya with the least squares regression based methods and artificial neural networks (ANN) using the load, temperature and population growth data from 2000 to 2008. The results attained are validated with the real data obtained from the Turkish Electricity Distribution Corporation (TEDAS) which represents the monthly peak load electric consumption in Kutahya, Turkey. By comparing the forecasted results with the real data the most suitable method is proposed.

Keywords: Load Forecasting, Least Squares Regression, Artificial Neural Networks.

I. INTRODUCTION

Load forecasting is of the most difficult problems in distribution system planning and analysis. However, not only historical load data of a distribution system play a very important role on peak load forecasting, but also the impacts of meteorological and demographic factors must be taken into consideration [1, 2]. For planning, management and effective operation of electric power systems, load forecasting should be accomplished over a broad spectrum of time intervals [3]. Load forecasting methods are distinguished on the basis of forecasting periods. In general, the required load forecasting can be categorized into short, medium and long term forecasts. Short term forecasting (half hour to one week ahead) represents a great saving potential for economic and secure operation of power systems.

Medium term forecasting (one day to several months) deals with the scheduling of fuel supplies and maintenance operations, and long term forecasting (more than a year ahead) is useful for planning operations [1-3]. Majority of the electric load forecasting methods are dedicated to short term forecasting and relatively less work has been done on long or medium term load forecasting. To date no specific methodology has been developed by which projections regarding anticipated electricity consumption can be made accurately specially over a long-term time frame [4, 5].

Generally, load forecasting methods are mainly classified into two categories: classical approaches and ANN based techniques. Classical approaches are based on statistical methods and forecast future value of a variable by using a mathematical combination of the historic information [6]. For instance with time series model of auto regressive integrated moving average (ARIMA) which incorporates the knowledge of experienced human operators short term load forecasting is carried out by using a linear combination of the past values of the variable [6]. In [5, 7] long-term and mid-term electric load forecasting is presented which incorporates daily and weekly simple linear regression models with annual load growth to predict the future load demand. Meanwhile, for different residential areas by using ANN based techniques short and long term load forecasting is performed in [2, 3, 8-11] and in [12] respectively. In aforementioned studies, the results attained from different approaches are compared and discussed. In these studies, it is pointed out that, while ANN can yield satisfactory results with relatively limited information, in the case of statistical methods longer input data is required.

In this study, by using the peak load data which is recorded from 2000 to 2007 and employing the least squares regression based methods and ANN, peak load demand for 2008 is forecasted. The results are validated by using the real data of 2008.

II. FORECASTING METHODS

In the forecasting process of long term electric peak load, least squares regression based methods and ANN are used.

A. Least Squares Regression Methods

In regression based models, the prediction error is minimised to zero by using the least squares approach as given in Equation (1).

$$S = \sum_{i=1}^{n} (y_{i,real\ value} - y_{i,approximate})^2 \tag{1}$$

In the equation; n is the number of data, $y_{i,real\ value}$ is the existing recorded data, $y_{i,approximate}$ is the type of the function used and S is the sum of the squared prediction errors. In this method, the variable S is equalised to zero after differentiated to each coefficient. By this way the normal equations are attained [13, 14]. In the least squares method, regression models which are explained in detail below and used as $y_{i,approximate}$ in Equation (1).

- **The simple linear regression**

The simple linear regression model is based on the linear relationship between the dependent variable y and independent variable x as shown in Equation (2) [4, 13, 14].

$$y = a + bx \tag{2}$$

This is the equation of a straight line which intercepts y axes at a with a slope of b [14]. In order to attain zero error, by using the least squares errors method, Equations (3) and (4) are obtained.

$$an + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{3}$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \tag{4}$$

In the equations the variables y, x and n represent the peak load, the years and the number of years which the forecasting is based on respectively. The a and b coefficients are calculated from the Equations (3) and (4) and replaced in Equation (2) for the load forecasting [14].

- **Multiple linear regression**

This approach shows a plane in the space with three dimensions which can be expressed as given in Equation (5) [13, 14].

$$y = a + bx_1 + cx_2 \tag{5}$$

In the equation, a, b, and c are regression parameters relating the mean value of y to x_1 and x_2. When the least squares method is applied (to attain zero error), Equation (6) is obtained [15].

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}.x_{2i} \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{1i}.x_{2i} & \sum_{i=1}^{n} x_{2i}^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{1i} y_i \\ \sum_{i=1}^{n} x_{2i} y_i \end{bmatrix} \tag{6}$$

By solving the Equation (6); a, b and c parameters are calculated; where y is the peak load, x_{1i} is the temperature, x_{2i} is the population data and n is the number of years the forecasting algorithm based on. By replacing the regression parameters in Equation (5) the peak load forecasting is performed.

- **The quadratic regression**

In this approach the parabolic function which is given in Equation (7) is used [14].

$$y = a + bx + cx^2 \tag{7}$$

The a, b and c coefficients of the parabolic function can be obtained from Equation (8) which is written in matrix form.

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \end{bmatrix} \tag{8}$$

The load forecasting is performed by replacing the calculated coefficients in Equation (7).

- **The exponential regression**

In this approach, the trend equation is formed by using an exponential function as given in Equation (9) [4, 14].

$$y = ab^x \tag{9}$$

By writing the Equation (9) in logarithmic form and then applying the least squares approach, Equations (10), (11) and (12) are formed.

$$\log y = \log a + x \log b \tag{10}$$

$$\sum_{i=1}^{n} \log y_i = n \log a + \sum_{i=1}^{n} (x_i \log b) \tag{11}$$

$$\sum_{i=1}^{n} (x_i \log y_i) = \sum_{i=1}^{n} (x_i \log a) + \sum_{i=1}^{n} (x_i^2 \log b) \tag{12}$$

Since the Equation (10) is linear, by applying linear trend analysis a and b coefficients are found as shown in Equations (13) and (14).

$$\log a = \frac{1}{n} \sum_{i=1}^{n} \log y_i \tag{13}$$

$$\log b = \sum_{i=1}^{n} x_i \log y_i \Big/ \sum_{i=1}^{n} x_i^2 \tag{14}$$

By replacing a and b coefficients in Equation (9) the peak loads are predicted.

B. ANN

Figure 1 depicts the architecture of typical feed-forward multilayered neural network consist of an input layer, (one or more) hidden layers and an output layer. The number of hidden layers and neurons in layers is subject to problem studied and decided upon trial-error.

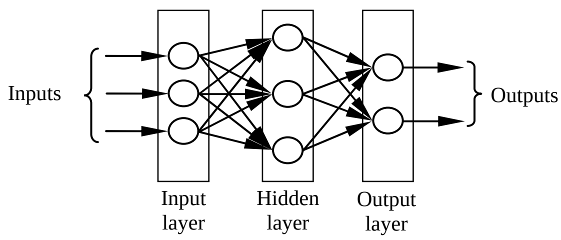


Figure 1. Feed forward multi layer ANN

The input layer receives the signal from outer environment and distributes it to the neurons in the hidden layers. The hidden layers have computational neurons and the number of layers depends on the functions to be used. The network computes the actual outputs of the neurons in the hidden and output layer by using the activation function. The error gradient for the neurons in the output layer is calculated and the weights in the back-propagation network propagating backward the errors associated with output neurons are adjusted. The total error at the output layer is then reduced by redistributing the error backwards through the hidden layers until the hidden layer is reached. The process updating the weights until the desired output is reached defined as training. This process is called as generalised *delta rule* and repeated until the error criterion for all datasets is reached. In general each ANN is trained on a different 80% of the training data and then validated on the remaining 20%. Since each additional layer exponentially increases the computing load, in practice mostly 3-layer ANNs are preferred [3, 12, 15].

In this work, in the implementation stage of the ANN Matlab 6.5 software is used. In the program; three layers ANN model including one hidden layer with feed forward and back-propagation algorithm has been trained by using Levenberg Marquardt (LM) algorithm. The network used in this study has 12 neurons in the hidden layer with the logarithmic sigmoid activation function which is non-linear continuous function between 0 and 1 as expressed in Equation (15) where, β is the slope constant and in general assumed equal to 1 [15].

$$f(x) = \frac{1}{1+e^{-\beta x}} \tag{15}$$

For inputs, along the peak load dataset, monthly temperature and population growth are taken into account. In this study, the average monthly temperature values are obtained from the regional meteorological record office. The monthly population growth is calculated from the 1997 and 2000 national population statistics using Equation (16) which gives the population growth on monthly bases [16].

$$P_n = P_0.e^{rn} \tag{16}$$

In this equation; P_0 is the first and P_n is the second of the two consecutive population statistics, n is the time interval between the statistics and r shows population growth rate. For Kutahya, by taking the 1997 and 2000 population statistics which are obtained from Turkish Statistical Institute (TURKSTAT) as 643117 and 656903 respectively; the value of r is calculated as 5.04×10^{-4}. By using these values the population is calculated on monthly bases. All the data which are used for the training and testing of the ANN have been scaled to an interval of [0.25, 0.75].

III. NUMERICAL EXAMPLE

By using least square regression based methods and ANN with the data from TEDAS, electric peak load forecasting of 2008 has been carried out for the city of Kutahya, Turkey. The results attained are summarised in Table 1 and illustrated as graphic form in Figure 2, for simple linear, multiple linear, exponential, quadratic and ANN approaches. The data used are the monthly averages of the peak loads recorded between years 2000 and 2008.

Table 1. Forecasted peak load by different approaches for 2008

Months	Real load (MW)	Simple linear (MW)	Multiple linear (MW)	Exponential (MW)	Quadratic (MW)	ANN (MW)
Jan.	136.7	124.6	121.3	125.6	129.9	123.24
Feb.	137.0	125.2	122.1	127.2	151.6	133.44
March	130.6	122.9	120.2	123.6	129.6	130.70
April	128.9	127.5	125.9	128.9	144.7	133.07
May	127.9	125.5	125.9	126.0	132.4	132.12
June	141.4	131.9	135.3	132.7	141.3	144.37
July	149.8	140.3	136.1	141.5	145.1	143.56
Aug.	134.3	142.5	139.1	145.0	157.1	135.74
Sept.	135.1	123.9	119.7	124.2	136.6	141.75
Oct.	143.0	130.4	127.7	131.0	140.7	137.21
Nov.	138.9	136.5	133.4	137.5	143.4	136.56
Dec.	135.2	142.7	139.6	144.8	144.5	128.00

It can be seen from the results that the results attained with the linear and exponential regression approach are very close to each other. In both approaches, the monthly peak load is increased consistently starting from February and reached the highest value in August of 2008.

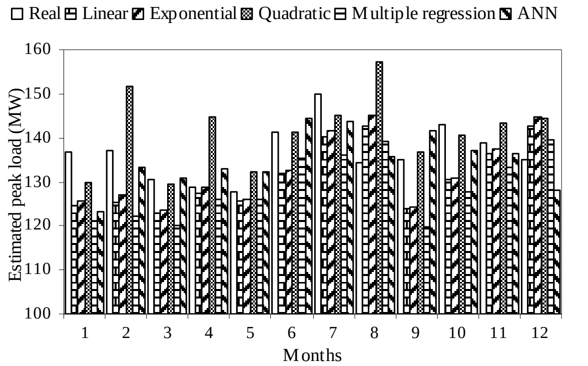


Figure 2. Real and forecasted peak load values for 2008

Figure 2 shows the peak load demand predicted with different forecasting approaches. In multiple regression and ANN approaches, the inputs are supplied with the historic temperature and population growth data along with the peak load data. In the study, since the forecasting is carried out on monthly bases, monthly error analysis performed and the mean errors are given in Table 2 and as graphic form in Figure 3.

Table 2. Forecasting errors by different approaches for 2008

Months	Simple linear (%) error	Multiple linear (%) error	Exponential (%) error	Quadratic (%) error	ANN (%) error
Jan.	-8.85	-11.27	-8.12	-4.97	-9.85
Feb.	-8.61	-10.88	-7.15	10.66	-2.60
March	-5.90	-7.96	-5.36	-0.77	0.08
April	-1.09	-2.33	0.00	12.26	3.24
May	-1.88	-1.56	-1.49	3.52	3.30
June	-6.72	-4.31	-6.15	-0.07	2.10
July	-6.34	-9.15	-5.54	-3.14	-4.17
Aug.	6.11	3.57	7.97	16.98	1.07
Sept.	-8.29	-11.40	-8.07	1.11	4.92
Oct.	-8.81	-10.70	-8.39	-1.61	-4.05
Nov.	-1.73	-3.96	-1.01	3.24	-1.68
Dec.	5.55	3.25	7.10	6.88	-5.33
MAPE(%)	5.83	6.71	5.54	5.45	3.53

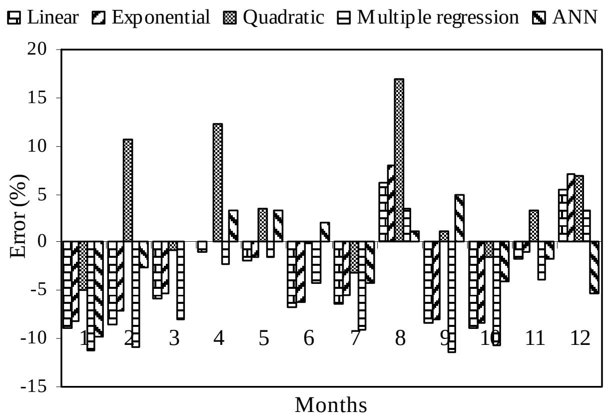


Figure 3. Forecasting errors for 2008

From the Figure 3 it is seen that, for the peak load forecasting for 2008, with the ANN for ten months, with the quadratic regression for eight months, with the multiple linear regression algorithm for six months, with the simple linear and exponential regression approaches for three months the mean error remained less than 5%. When the approaches are investigated for the prediction errors between 5 to 10%, for nine months with the simple linear and exponential regression, for two months with the multiple linear regression and ANN approaches and only for one month for the quadratic regression the error remained in these limits. The error remained between 10 to 15%; only for two months with the quadratic regression and for four months with the multiple linear regression approaches. The error has been found between 15 to 20% for only one month with the quadratic regression. When the approaches are compared according to their highest prediction error; the highest error for quadratic and multiple linear regression approaches is in August and September with an error of 16.98 and -11.40 % respectively. For exponential and simple linear regression approaches the highest prediction errors are found in October with a forecasting error of -8.39 and -8.85 % in January respectively. As for the ANN the highest forecasting error is in December corresponding to -9.85 %. From these results it is evident that with the lowest mean absolute percentage error (MAPE) of 3.53 %, the ANN based algorithm, in general has relatively lower forecasting error and superior to the classical approaches studied.

IV. CONCLUSIONS

In order to provide a high quality and reliable service to the consumers, it is essential for electric power distribution companies to carry out projections regarding anticipated electric peak load demand in the future. In the planning and management of power generation, transmission and distribution systems, load forecasting is the key factor. The errors in the parameters which are used in the load forecasting, the sudden changes in the input parameters or unexpected changes in the peak load data may negatively affect the peak load forecasting process and contribute to further errors. For the city of Kutahya with 200,000 inhabitants which has an increasing demand in electric power consumption, both in residential and industrial areas, the factors which contribute to the uncertainty in peak load forecasting are very high. Hence forecasting methods based on different methodologies may yield more realistic results. By the comparison of the forecasting for 2008 results attained, it is concluded that with the use of longer input data, the forecasting error is decreased. In this study, when the different load forecasting techniques are compared for Kutahya, it is seen that the ANN approach has produced better results.

REFERENCES

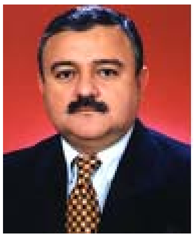
[1] M., Beccali, M. Cellura, V. Lo Brano, A. Marvuglia, "Forecasting Daily Urban Electric Load Profiles using Artificial Neural Networks", Energy Conversion and Management, Vol. 45, pp. 2879-2900, 2004.
[2] M. Djukanovic, B. Babic, D.J. Sobajic, Y.H. Pao, "Unsupervised/Supervised Learning Concept for 24-Hour Load Forecasting", IEE Proceedings-C, Vol. 140, No. 4, pp. 311-318, 1993.
[3] T. Yalcinoz, U. Eminoglu, "Short Term and Medium Term Power Distribution Load Forecasting by Neural Networks", Energy Conversion and Management, Vol. 44, pp. 1393-1405, 2005.
[4] S.C. Tripathy, "Demand Forecasting in a Power System", Energy Conversion Management, Vol. 38, No. 14, pp. 1475-1481, 1997.
[5] H.M. Al-Hamadi, S.A. Soliman, "Long Term/Mid-Term Electric Load Forecasting Based on Short-Term Correlation and Annual Growth", Electric Power Systems Research, Vol. 74, pp. 353-361, 2005.
[6] N. Amjady, "Short-Term Hourly Load Forecasting Using Time Series Modeling with Peak Load Estimation Capability", IEEE Transactions on Power Systems, Vol. 16, No. 3, pp. 498-505, August 2001.
[7] M.S. Kandil, S.M. El-Debeiky, N.E. Hasanien, "Overview and Comparison of Long-Term Forecasting Techniques for a Fast Developing Utility: Part I", Electric Power Systems Research, Vol. 58, pp. 11-17, 2001.
[8] S. Krunic, I. Kircmar, N. Rajakovic, "An Improved Neural Network Application for Short-Term Load Forecasting in Power Systems", Electric Machines and Power Systems, Vol. 28, pp. 703-721, 2000.
[9] D. Singh, S.P. Singh, "A Self Selecting Neural Network for Short-Term Load Forecasting", Electric Power Components and Systems, Vol. 29, pp. 117-130, 2001.
[10] T. Senjyu, H. Sakihara, Y. Tamaki, K. Uezato, "Next-Day Load Curve Forecasting using Neural Network based on Similarity", Electric Power Components and Systems, Vol. 29, pp. 939-948, 2001.
[11] C.C. Hsu, C.Y. Chen, "Regional Load Forecasting in Taiwan Applications of Artificial Neural Network", Energy Conversion and Management, Vol. 44, pp. 1941-1949, 2003.
[12] S.A. Kalogirou, "Applications of Artificial Neural Networks in Energy Systems a Review", Energy Conversion and Management, Vol. 40, pp. 1073-1087, 1999.

[13] S.C. Chapra, R.P. Canale, “Numerical Methods for Engineers”, McGraw-Hill, 2006.
[14] B. Bowerman, R.T. O’connell, “Applied Statistics”, McGraw-Hill, 1997.
[15] M. Ilunga, D. Stephenson, “Infilling Stream Flow Data using Feed-Forward Back-Propagation (BP) Artificial Neural NETWORKS: Application of Standard BP and Pseudo Mac Laurin Power Series BP Techniques”, Water SA, Vol. 31, No. 2, pp. 171-176, 2005.
[16] E. Kreyszig, “Advanced Engineering Mathematics”, 6th Edition, John Wiley and Sons, Canada, 10-11, 1988.

BIOGRAPHIES

Yilmaz Aslan was born in Trabzon, Turkey, 1965. He received the B.Sc. in Electrical Engineering in 1986 and the Ph.D. degree in Electrical Engineering from University of Bath, UK, in 1997. Currently, he is an Assistant Professor at electric and electronic engineering department of Dumlupinar University, Kutahya, Turkey. His research interests are analysis of shunt faults in power transmission and distribution systems, load forecasting, renewable energy.

Serkan Yavasca was born in Bursa, Turkey on January 25, 1977. He received the B.Sc. degree in 1999 from Marmara University, Istanbul, Turkey. In 2007 he started his master degree in the Department of Electrical Engineering of Dumlupinar University, Kutahya, Turkey. Currently, he works as an instructor at Yalova Vocational High School. His research interests include load forecasting and distribution automation.



Celal Yasar was born in Kutahya, Turkey, in 1958. He received his B.Sc. degree from Yildiz Technical University, M.Sc. degree from Anadolu University and Ph.D. degree from Eskisehir Osmangazi University, Turkey, all in Electrical Engineering in 1980, 1988 and 1999, respectively. He is currently working as an Assistant Professor at the Electrical and Electronics Engineering Department of Dumlupinar University, Kutahya, Turkey. His research interests include analysis of power systems, economic operation of power systems, power distribution systems and renewable energy systems.