International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 12, December-2014

Factor Affecting Elements and Short term Load forecasting Based on Multiple Linear Regression Method

Girraj Singh[1], D.S. Chauhan[2], Aseem Chandel[3], Deepak Parashar[4], Girijapati Sharma[5]
B S A College of Engineering & Technology, Mathura[1, 3, 4 & 5]
GLA University, Mathura[2]

***Abstract:* Electrical load forecasting plays an important role in planning and operation of power system. The accuracy of this forecasted value is necessary for economically efficient operation and also for effective control. Deregulation in an energy sector and the energy market origin needs an accurate of the Load Forecasting methods. Load Forecast techniques are classified into two group as statistical and intelligence method. In this paper application and factors affecting elements are presented, and short term load forecasting based on Multiple Linear Regression method is discussed.**

***Key words:* Load forecasting, Statistical, intelligence, short medium and long-term load forecasting, multiple linear regression(MLR).**

I. INTRODUCTION

Power systems development and increasing their complexity caused many factors have become much more significant in electric power generation and consumption. In order to meet power systems requirements continually and having sustained economic growth, load forecasting has become a very important task for electric utilities. An accurate load forecast become more imperative in managing utility, developing a power supply strategy, finance planning and electricity market management. Load forecast should be accomplished over time intervals for economical and efficient operation and also control of power systems. In general, the required load forecasts 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.

To supply the load demand over this particular duration of time involves the start up and shutdown of entire generating units, which will be determined by a number of generation control functions such as hydro Scheduling, hydro-thermal coordination, economic dispatch, load management, operation scheduling unit commitment and interchange evaluation [1]. It is a main goal for any utilities to operate at cost as low as possible. One way to achieve this is to minimize the forecast error. Though, it was estimated that an increase of operating cost associated with a 1% increase of forecast error was 10 million pounds per year [2]. Many algorithms have been developed for more accurate load forecasting.

This paper is organized as follows: section II introduces the load forecasting applications and factors affecting elements of load pattern; next, section III for multiple linear regression method and the section IV present the implementation and results of short term load forecasting; section V presents the conclusion.

II. LOAD FORECASTING

Prediction of future events and conditions is called forecasts, and the act of making such predictions is called forecasting. Load forecast is prediction of future load at different time interval. It is play a very vital role in power system planning, operation and control. Load forecasting means estimating active load at various load buses ahead of actual load occurrence. Planning and operational application of load forecasting requires a certain 'lead time' also called forecasting interval.

2.1 Factor Affecting Load Patterns

A large number of factors affect the load demand considerably. The impacts of all these factors which affect the load need to be studied in order to develop a accurate load forecasting model.

Economic factor:

Many economic factors such as the type of customers such as residential, agricultural, commercial and industrial, demographic conditions, population, per capita income, GDP growth, national economic growth and social activities etc. can cause a considerable change in the load pattern. These economic factors mainly affect the long-term load forecasting.

Weather Factors:

Load forecasting is greatly affected by weather conditions such as temperature (dry bulb and wet bulb temperature), humidity, cloud coverage etc. The most important weather factor is the temperature. The changes considerably affect the load requirement for heating in winter and air conditioning in summer. Load forecasting also affected by other factors such as humidity especially in hot and humid areas, precipitation, thunderstorms, wind speed and light intensity of the day.

Time and Seasonal Factors:

Time factors play an important role in accurate load forecasting. It may cause a considerable change in load pattern [3]. There are following factor –:

1. Seasonal variation: change of season (summer, winter, rainy and autumn), change of day light hours, change of average temperature, etc.
2. Daily variation: different day time and night time consumption
3. Weekly cycle: Different weekday and weekend consumption patterns
4. Holidays and special days: Load pattern on holidays will be different from that of weekdays and weekend. Special days such as festive days can affect the load.

Price Factor:

Load forecasting is strongly affected by electricity price. Electrical price which may have a complicated relationship with the system load, it is an important factor in load forecasting. Change in tariff may also change the load pattern [5].

Random Disturbances:

A random disturbance occurs in the power system which may affect the load pattern considerably. The random disturbances include sudden shutdown or start of industries, wide spread strikes, marriages, special functions etc [4].

Other Factors:

In addition to all the factors listed above, the load pattern may also change due to geographical condition (urban or rural areas), type of consumers (rural or urban), home appliances sale data, television program (sports, serial etc.) etc.

2.2 Application of Short term load forecasting

STLF is aimed at predicting system load over an interval of one day to one week. Basic operating functions such as unit commitment, economic dispatch, fuel scheduling, hydro-thermal co-ordination, interchange evaluation, unit maintenance and security assessments require a reliable short term load forecast. It plays a vital role in optimum unit commitment, start up and shut down of thermal plants, control of spinning reserve and buying and selling of power inter connected systems. A precise short term electrical load forecasting, results in cost saving and secure operational allowing utilities to commit their production resources to optimize energy prices and exchange with vender and clients.

III. MULTIPLE LINEAR REGRESSION

A large variety of statistical and intelligence techniques have been developed for short-term load forecasting. The statistical techniques are such as regression, time-series, expert system, state-space, exponential Smoothing, similar day approach, support vector machine and knowledge based. The intelligence techniques are such as artificial neural network (ANN), Fuzzy method, Genetic algorithm and Particle Swarm optimization (PSO), and combination of these techniques have also been developed for achieving more accurate short term load forecasting.

Regression method is one of the most widely used statistical techniques for short term and long-medium load forecasting. Regression models are capable of characterizing the relationship between load demand and other important factors such as holidays, average load and weather changes in short term load forecasts but complicated modeling technique and enormous computational efforts are required to produce acceptable result. In short-term load forecasting, the regression methods are generally used to model the relationship between load consumption and factors affecting load consumption such as weather, day type, time of the day and customer class etc. The main drawback of this method is the relationship between the determined or reference control variation is unclear, a great forecasting error value is produced. Regression analysis is a modeling technique for analyzing the relationship between a continuous (real-valued) independent variable *y* and one or more independent variables x_1 , $x_{2.....}$,x_k . The goal in regression analysis is to identify a function that describes, as closely as possible, the relationship between these variables so that the value of the dependent variables can be predicted using a Range of independent variables values. In the multiple linear regression method, the load is found in terms of explanatory (independent) variable such as weather and other variables which influence the electrical load. The load model using this method is expressed in the form as

$$Y = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \ldots\ldots + \beta_k x_{ki} + \varepsilon \qquad (1)$$

where, Y is the actual load, x_{1i} ,$x_{2i,}$ ...x_{ki} of the k[ith] independent variables, β_k is regression parameters with respect to x_k , *and* ε is an error term. The error term ε has a mean value equal to zero and constant variance. Since parameters β_k are unknown, they should be estimated from observations of y_i and x_{ki}. Let b_k be the estimates in terms of β_k. Hence the predicted value of *y* is:

$$\hat{Y} = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + \ldots\ldots + b_k x_{ki} \qquad (2)$$

The difference between the actual load value of *y* and the predicted value ŷ would, on average, tend toward 0, for this reason it can be assumed that the error term in equation (1) has an average, or expected, value of 0 if the probability distributions for the dependent variable *y* at the various level of the independent variable are normally distributed (bell shaped). The error term can therefore be omitted in calculating parameters. Then, the least square estimates method is used to minimize the sum of squared residuals (SSE) to obtain the parameters b_k:

Suppose that the experimenter has *k* independent variables $x_1 , x_2 , \ldots, x_k$ and n observations $y_1 , y_2 , \ldots, y_n$, each of which can be expressed by the equation

$$Y_i = \beta_o + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_k x_{ki} + \varepsilon_I \qquad (3)$$

This model essentially represents n equations describing how the response values are generated in the scientific process. Using matrix notation, we can write the equation

$$\mathbf{Y} = \mathrm{X}\,\boldsymbol{\beta} + \varepsilon$$

Where

$$\mathbf{Y}=\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}, \mathbf{X}=\begin{bmatrix} 1 & x_{11} & x_{21} & . & . & . & x_{k1} \\ 1 & x_{12} & x_{22} & . & . & . & x_{k2} \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ 1 & x_{1n} & x_{2n} & . & . & . & x_{kn} \end{bmatrix}, \boldsymbol{\beta}=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix}$$

Then the least squares solution for estimation of **β**, we minimize the expression of sum of squares error

$$SSE = (y - X\,b)^T (y - X\,b)$$

Differentiating SSE in turn with respect to b, and equating to zero

$$\frac{\partial}{\partial b}(\mathrm{SSE}) = 0,$$

The result reduces to the solution of b in

$$(X^T X)\,b = X^T y,$$

or $$Ab = g,$$

where $$A = X\,X^T, \quad g = X^T y$$

If the matrix A is nonsingular, we can write the solution for the regression coefficient as

$$b = A^{-1} g = (X^T X)^{-1} X^T y \qquad (4)$$

Thus we can obtain the prediction equation or regression equation by solving a set of equation in a like number of unknowns. So prediction equation

$$\hat{Y} = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + \ldots + b_k x_{ki} \qquad (5)$$

After regression parameters are calculated, this model can be used for prediction. Assuming that all the independent variables have been correctly identified and therefore the standard error will be small. The standard error is obtained by the equation below:

$$S = \sqrt{\frac{SSE}{n-(k+1)}} \qquad (6)$$

$$SSE = \sum_{t=1}^{n} (y_i(t) - \hat{y}(t))^2 \qquad (7)$$

$y_i\,(t)$ is observed and $\hat{y}_i\,(t)$ is estimated

Goodness of fit measurement is represent by

$$R^2 = 1 - \frac{\sum_{t=1}^{n} (y_i(t) - \hat{y}_i(t))^2}{\sum_{t=1}^{n} (y_i(t) - \bar{y}_i(t))^2} \qquad (8)$$

A goodness of fit measurement is represented by the R^2 statistic which ranges from 0 to 1 and indicates the proportion of the total variation in the dependent variable Y around its average that is counted for by the independent variable in the estimated regression function. The closer the R^2 statistic to the value 1, the better the estimated regression function fits the data.

IV. IMPEMENTATION & RESULTS

The MLR technique as has been discussed previously is implemented to predict the hourly load of power utility system. For this forecasting study, data during the dry season is used. In this method dry bulb, dew bulb temperature, Energy PR, and TMSR PR independent variable are used. The effectiveness of MLR method for STLF in an actual real life network, the method has been applied to predict the daily load (up to 24 hours) and seven days ahead load, in the US based power utility. The error was calculated as the mean average percentage error (MAPE). As follows

$$\mathrm{MAPE} = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \qquad (9)$$

where,

y_t = Actual load at time t

$\hat{y}_t$ = Forecasted load at time t

n= 24

By Equation (4), we can calculate the regression parameters such as b_0, b_1, b_2, b_3 and b_4. After calculating the regression parameters, then we can calculate the forecasted value at particular time by using equation (5).

Case-1:

In this case we have used data of 1/01/2000, by using this data we find regression parameter such as,

bo = 5537.7, b_1 = 54.9, b_2 = 71, b_3 = 175.5, b_4 = -202.3.

$$\hat{Y} = 5537.7 + 54.9x_1 + 71x_2 + 175.5x_3 - 202.3\,x_4 \qquad (10)$$

where,

x_1 = dry bulb temperature

x_1 = dew bulb temperature

x_3 = energy PR(Public Relation)

x_4 = Ten-Minute Spinning Reserve PR

Now using this equation (10), we can forecast one day head load (24 hour), i.e. 2/01/2000.

Results shown in table -1

Case-2:

In this case, we include some polynomial term, such as

$$\hat{Y}(t) = b_0 + b_1\,x_1 + b_2\,x_2 + b_3\,x_3 + b_4\,x_4 \qquad (11)$$

where

$x_1 = x$, $x_2 = x^2$, $x_3 = x^3$, $x_4 = (T_p(t) - T_{pi})$,

$x = (T_d(t) - T_{ci})$

Where

$T_d(t)$ = dry bulb temperature at time t

$T_p(t)$ = dew point temperature at time t

T_{pi} = cut-off of dew temperature for the interval t (i=24)

T_{ci} = cut-off of dry temperature for the interval t (i=24)

b_0 = base load component (regression constant coefficient)

$b_1 - b_4$ = regression coefficient

By using the data of 1/01/2000, we find the value of regression coefficient such as

b_0=11103, b_1 = -89, b_2 = 38, b_3 = -2, b_4 = 65,

Now using equation (6) we can forecast one day ahead load, result shown in table-2

Table 4.1 Actual forecast load and mean average percentage error (MAPE) of 24 hours one day ahead load forecasting (2/01/2000 by 1/10/2000)

Time(hr)	Actual load (y_t)	Forecasted load($\hat{y}_t$)	MAPE (%)
1	10637	12163.48	14.3
2	10116	12437.78	22.95
3	9848	12395.76	25.8
4	9731	12376.00	27.18
5	9746	12118.00	24.3
6	9987	12336.89	23.5
7	10476	12361.46	17.99
8	11031	10006.80	9.28
9	11990	12419.00	3.5
10	13000	14463.30	11.25
11	13592	15085.78	10.9
12	13853	15007.23	8.3
13	13722	15524.57	13.13
14	13417	15582.29	16.1
15	13113	14997.30	14.3
16	13136	14653.00	11.5
17	14354	15543.46	8.2
18	15526	16136.55	3.9
19	15409	15940.39	3.4
20	14991	15419.80	2.8
21	14396	15367.64	6.7
22	13367	14809.90	10.7
23	12076	14177.60	17.4
24	10843	13975.90	28.8
	Average MAPE = 13.59		

The maximum error is 28.8% and lowest error 3.4%, average MAPE of one day ahead load forecast, is 13.59%. From above results, we can see forecasting error for time 1:00 to 7:00 am is higher than time from 16:00 to 22:00 pm.

Table 4.2 Actual load, Forecasted load, and MAPE (%) of one day ahead load forecast (2/01/2000 by 01/01/2000)

Time(hr)	Actual load (y_t)	Forecasted load($\hat{y}_t$)	MAPE (%)
1	10637	11086.70	4.2
2	10116	11152.20	10.2
3	9848	11151.20	13.2
4	9732	11151.70	13.4
5	9746	11038.00	13.4
6	9987	11057.50	10.8
7	10476	11070.50	5.7
8	11031	11079.80	0.4
9	11990	11081.00	7.1
10	13000	11132.91	12.9
11	13592	11314.34	14.3
12	13853	11634.90	13.1
13	13722	12035.56	8.4
14	13417	12568.36	5.0
15	13113	12739.10	7.1
16	13136	12171.00	10.5
17	14354	11751.10	15.1
18	15526	12178.50	13.96
19	15409	13357.58	13.18
20	14991	13377.00	9.1
21	14396	12357.58	13.1
22	13367	12504.58	6.9
23	12078	12442.00	1.5
24	10843	12259.90	9.3
	Average MAPE = 9.70		

From above results, average MAPE is 9.7%, which is relatively less than previously forecast error (13.59%). The error is reduced by including polynomial term with cut-off dry and wet bulb temperature. The maximum and minimum errors are 14.3% and 0.4% respectively.

This model is very sensitive to the fluctuation of temperature. It needs a very accurate temperature forecast, as a small change of temperature will cause a significant change in load prediction. This forecasting model uses the next day temperature forecast as an input which will introduce further errors, as there was no temperature forecast data available, temperature forecast was not included in this regression analysis.

V. CONCLUSION

The principle driving element for all daily and weekly operation scheduling of electric power system is accurate load forecasting. The multi linear regression models for short term load forecasting are relatively easy to develop. The drawback of this model is its dependency on the accuracy of previously recorded load and temperature data which will greatly affect the forecast yield. The accuracy of predictions made using regression models depends on how well the regression function fits the data. There should be regular checks to see how well a regression function fits a given data set.

In This paper presents our preliminary investigations of the application of MLR for STLF of the US based power utility, and in this study only weekdays are modeled. Investigations are underway to model weekend and holidays.

REFERENCES

1. Gross. G., Galiana, F.D, “Short-Term Load Forecasting,” Proceeding of IEEE, Vol. 75, No. 12, pp. 1558-1573, December 1987.
2. D.W. Bunn and E.D. Farmer, Comparative Models for Electrical Load Forecasting, John Wiley & Son, New York, 1985.
3. G. Gross and F. D. Galiana, “Short-term load forecasting”, Proceedings of IEEE. Vol. 75, No. 12, Dec. 1987, pp. 1558-1572.
4. K. Y. Lee, Y.T. Cha and C.C. Ku, “A study of neural networks for Short-term load forecasting”, Application of Neural Network to power Systems, Proceedings of the IEEE, 23-26 July 1991 pages(s): 26-30.
5. H. Chen , C.A. Canizares, A. Singh, “ ANN-based short-term load forecasting in electricity markets”, Power Engineering Society Winter Meeting, IEEE Volume 2, 28 jan-1 Feb. 2001 pages(s): 411-415
6. Moghram,I., Rahman, S., Analysis and Evaluation of Five Short-Term Load Forecasting Techniques, IEEE Transactions on Power Systems, Vol.4, No.4, pp. 1484-1491, October 1989.
7. Papalexopoulos, A.D., Hesterberg, T.C., A Regression-Based Approach to Short-Term Load Forecasting, IEEE Transactions on Power Systems, Vol.5, No. 4, pp 1535-1547, November 1990.
8. Amral, N., Ozveren C.S., and King D., “Short Term Load Forecasting using MLR”,UPEC-1192, 2007.
9. Shahidehpour M., Yamin H., and Li Zuyi “ Market Operation in Electric Power Systems: Forecasting, Scheduling, and Risk Management”, Wiley-IEEE press, ISBN:-0-471-44337-7. April 2002
10. Kang Chongqing, Xia Qing and Zhang Boming. “Review of power system load forecasting and its development”. Automation of Electric Power Systems, Vol.28, No. 17, pp.1-11, 2004.