Short Term Load Forecasting using Multiple Linear Regression

N. Amral, C.S. Özveren, D King

University of Abertay Dundee, UK

Abstract

In this paper we present an investigation for the short term (up 24 hours) load forecasting of the demand for the South Sulewesi's (Sulewesi Island – Indonesia) Power System, using a Multiple Linear Regression (MLR) method. After a brief analytical discussion of the technique, the usage of polynomial terms and the steps to compose the MLR model will be explained. Report on implementation of MLR algorithm using commercially available tool such as Microsoft EXCEL™ will also be discussed. As a case study, historical data consisting of hourly load demand and temperatures of South Sulawesi electrical system will be used, to forecast the short term load. The results will be presented and analysed potential for improvement using alternative methods is also discussed.

Key words: Multiple Linear Regression, polynomial terms.

1 INTRODUCTION

Due to its major role in an effective and economic operation of power utilities, electric load forecasting has received increasing attention over the years by academic researchers as well as the power systems engineers in the industry.

An accurate short term load forecasts (STLF) with forecasting horizons up to 24 hours are necessary for scheduling functions such as hydro-thermal power generation coordination (unit commitment) in order to establish the hourly schedules for generation resources that will minimize the system operating cost. Such forecasts are also used for economic dispatch, predictive frequency control, security analysis, systems restoration, and energy trading. Recent global privatisation and deregulation initiatives in the sector has made accurate forecasting vital not only for operational but also for the profitability and sustainability of privatised /deregulated utilities.

Because of its crucial function in the operation of Electrical Power Systems, there are many authors in the literature reporting a wide variety of methods for short term load forecasting. Typically, the models are classified into two basic models[1], these are:

1). Peak load models; in these models, the daily or weekly peak load is modelled, usually as a function of weather. Time does not play a role in such models. Peak load models can be divided into two parts: a base load that independent to the weather and variable load which is weather dependent, this component is then added to the base load [2], [3]. The parameters of the model are estimated through linear or nonlinear regression. These models do not define the time which the peak occurs and contain information about the shape of the load curve. Correlation across the period cannot be forecast as the models are essentially static.

2). Load shape models are observed in particular time intervals, which are usually sampled every one hour or one-half hourly intervals. These models can be classified into:

i). Static models which is described as a combination of explicit time function such as sinusoids, exponential or polynomials, with model parameters are calculated through the usage of a set of historical load data into linear regression or exponential smoothing techniques [4],[5].

ii). Dynamic models which consider that load is also affected by its most recent behaviour as well as weather and random inputs and it is not only a function of the time of the day. This model can be composed with autoregressive moving average (ARMA) and state-space models [1] and Artificial Neural Network (ANN) [6]

The investigation we report in this paper is part of our wider investigations to find an appropriate set of methodologies for the STLF the demand of the South Sulewesi's Electrical Power System. A brief analysis and evaluation of five short term load forecasting techniques, that we will be investigating including Regression based approaches, Stochastic Time Series, General Exponential Smoothing, State Space Method,and Expert Systems, was presented in [7]. In this paper we report on our investigations on the regression based approach to short term system load forecasting. This approach to forecasting has been used by different authors including [8]. In the following we briefly explain the MLR technique and detail our model and methodology.

2 MULTIPLE LINEAR REGRESSIONS

Regression analysis is a modelling technique for analysing the relationship between a continuous (real-valued) dependent variable y and one or more independent variables $x_1, x_2, \cdots x_k$. 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

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \qquad (1)$$

Where y is the load, x_i is the affecting factors,

β_i is regression parameters with respect to x_i, and ε is an error term.

The error term ε has a mean value equal to zero and constant variance.

Since parameters β_i are unknown, they should be estimated from observations of y and x_i. Let $b_i(i = 0,1,2,\ldots k)$ be the estimates in terms of $\beta_i(i = 0,1,2,\ldots k)$.

Hence the predicted value of y is:

$$\hat{y} = b_o + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \qquad (2)$$

The difference between the actual load value of y and the predicted value $\hat{y}$ 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 omited in calculating parameters.

Then, the least square estimates method is used to minimize the sum of squared residuals (SSE) to obtain the parameters b_i :

$$\underline{B} = [b_o \quad b_1 \quad b_2 \ldots b_k]^T = \left(\underline{X}^T \underline{X}\right)^{-1} \underline{X}^T \underline{Y} \ (3)$$

Where $\underline{Y}$ and $\underline{X}$ are the following column vector and matrix:

$$\underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \text{ And } \underline{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots x_{1k} \\ 1 & x_{21} & x_{22} & \cdots x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots x_{nk} \end{bmatrix}$$

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

$$s = \sqrt{\frac{SSE}{n-(k+1)}} \quad (4)$$

$$SSE = \sum \left(y_i(t) - \hat{y}_i(t)\right)^2 \quad (5)$$

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

$$R^2 = 1 - \frac{\sum_{t=1}^{n} (y_i(t) - \hat{y}_i(t))^2}{\sum_{t=1}^{n} (y_i(t) - \bar{y}_i(t))^2} \quad (6)$$

$\bar{y}_i(t)$:the average value of $y_i(t)$

Experience about the load to be modelled helps an initial identification of the candidate of explanatory (independent) variables. To determine the significance of regression parameters, the F statistical test is performed and to determine the significance of each of these coefficients; is performed by calculating t ratios. 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.

3 IMPLEMENTATIONS

The MLR technique as has been discussed previously is implemented to predict the hourly load of South Sulawesi Electricity system.

South Sulawesi is one of the provinces of Republic of Indonesia. Indonesia is located on the equator with tropical weather, and have only dry and rainy season throughout the year. The energy sales composition of the consumer base in 2005 was 49% residential, 16% commercial, 26% Industrial and 9% public, with the typical daily load (in MW) and temperature curve as follows:

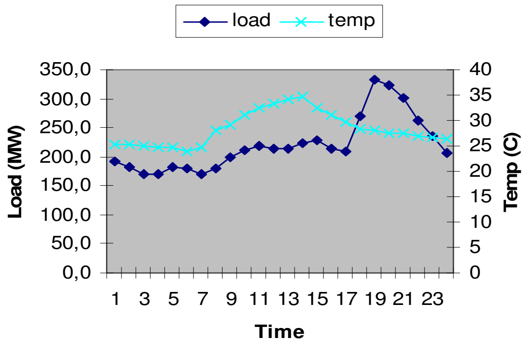

Figure 1 Daily load and temperature curve

For this forecasting study, data during the rainy season and dry season were used separately. In the MLR application, the hourly load is modelled as: (i) Intercept component which is assumed constant for different time intervals of the day, (ii) Time of observation which is represent the load characteristic of the day. (iii) Temperature sensitive component which is function of difference of temperature at time t and the average temperature in time intervals. The relationship between the temperature sensitive component, time of observation and the load fluctuation are not linear but in form of polynomial term.

The model that is used divides the load curve into 3 intervals, 1 to 6 am, 7am to 17pm and 18 to 24 pm.

For the interval 1 to 6 am, $1 \le t \le 6$, the hourly load can be modelled as follows

$$\begin{aligned}\hat{y}_i(t) = b_0 &+ b_1(T_i(t) - T_i(t-1)) \\ &+ b_2(T_i(t-1) - T_i(t-2)) \\ &+ b_3 t + b_4 t^2\end{aligned} \quad (7)$$

For interval 7am to 17 pm, $7 \le t \le 17$, the load curve is modelled by the form:

$$\begin{aligned}\hat{y}_i(t) = b_0 &+ b_1 T_i(t) + b_2(T_i(t) - T_{iav}) \\ &+ b_3(T_i(t) - T_{iav})^2 \\ &+ b_4(T_i(t) - T_{iav})^3 \\ &+ b_5 t + b_6(T_{ava} - T_{avb}) \\ &+ b_7(T_i(t) - T_i(t-1)) \\ &+ b_8(T_i(t-1) - T_i(t-2)) \\ &+ b_9(T_i(t-2) - T_i(t-3))\end{aligned} \tag{8}$$

And for interval 18 to 24 pm, $18 \le t \le 24$, the model is different because the load increase sharply due to the decrease of sky brightness, hence the consumer start to turn on their light. The formula has the form:

$$\begin{aligned}\hat{y}_i(t) = b_0 &+ b_1 T_i(t) + b_2 t + b_3 t^2 \\ &+ b_4 t^3\end{aligned} \tag{9}$$

Where:

$\hat{y}_i(t) =$ predicted load at hour t in the interval i of the day

$b_0 =$ Intercept component (regression constant coefficient)

$b_1, \cdots, b_9 =$ Regression parameters of temperature sensitive component and time of observations

$T_i(t) =$ Temperature (^{0}C) at time t in the interval i of the day

$T_{iav} =$ Temperature average (^{0}C) in the interval i of the day.

$T_{ava} =$ Average temperature of previous 24 hours to the time t (^{0}C)

$T_{avb} = T_{ava}$ lagged 3 hours, (^{0}C)

$t =$ Time of observation.

Table 1. Weekday MLR dry season model parameter estimates for different time intervals of the day

	Day time interval		
Parameter	1-6	7-17	18-24
b_0	194,25	203,43	-30050,28
b_1	4,33	0,00	-1,13
b_2	-14,71	7,28	4277,08
b_3	-15,79	-0,02	-198,86
b_4	1,94	-0,04	3,05
b_5	-	0,00	-
b_6	-	38,58	-
b_7	-	-4,04	-
b_8	-	-0,73	-
b_9	-	-0,22	-
R^2	0,9028	0,9935	0,9898

Table 2. Weekday MLR rainy season model parameter estimates for different time intervals of the day

	Day time interval		
Parameter	1-6	7-17	18-24
b_0	228,13	167,24	-29972,81
b_1	8,25	0,00	0,19
b_2	4,07	4,33	4214,20
b_3	-31,69	0,00	-193,82
b_4	4,24	-0,02	2,94
b_5	-	2,77	-
b_6	-	1,50	-
b_7	-	0,67	-
b_8	-	3,73	-
b_9	-	4,12	-
R^2	0,9814	0,9821	0,9850

For this review, the model is derived both for the dry season and the rainy season, with parameters shown in table 1 and 2 respectively. These parameters regression have been calculated using weekday hourly data for every time interval. The division of the day into three unequal interval time zones is based on the characteristics of load compare to the temperature fluctuation

4 RESULT

In order to investigate how effective the implementations of the MLR method will be for STLF in an actual real life network, this technique has been applied to predict the daily load (up to 24 hours) on dry and rainy season days, in the South Sulewesi Network. The error was calculated as the mean average percentage error (MAPE). as follows;

$$\text{MAPE} = 100 \cdot \frac{\sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{n} \qquad (10)$$

-The MAPE in the dry season, is 3,52 %

-The MAPE in the rainy season is 4,34 %

Figure 2 and 3 show the actual and predicted curve in dry season and rainy season respectively. Table 3 describes the actual load, forecast and MAPE for dry and rainy season,

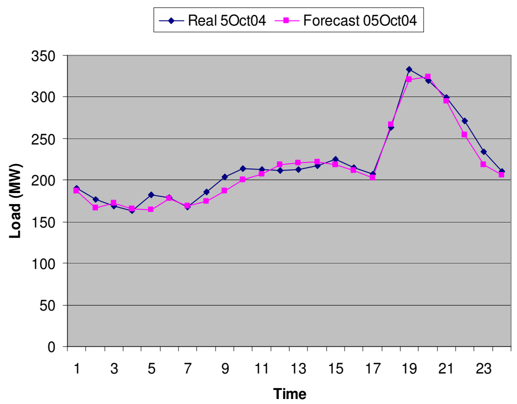

Figure 2. Dry season load curve

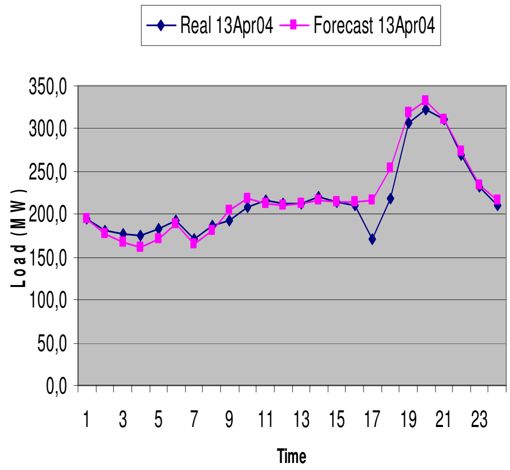

Figure 3. Rainy season load curve

Table 3. Actual, forecast load and mean average percentage error (MAPE) for dry and rainy season

	Dry Season			Rainy Season		
Time	Actual	Forecast	Error	Actual	Forecast	Error
1	190,1	186,7	1,78	194,5	193,7	0,41
2	176,5	166,4	5,75	181,5	177,6	2,15
3	168,3	171,9	2,10	176,4	166,7	5,49
4	163,3	165,7	1,43	174,9	161,1	7,94
5	182,2	164,2	9,92	182,4	170,9	6,32
6	179,1	178,2	0,50	192,7	188,3	2,28
7	167,9	169,3	0,85	171,3	165,3	3,48
8	185,4	174,6	5,85	187,7	181,8	3,17
9	203,4	186,3	8,42	192,8	205,4	6,51
10	214,1	200,6	6,30	207,9	218,5	5,09
11	213,2	207,0	2,92	215,5	212,2	1,52
12	212,0	218,6	3,10	212,2	210,2	0,96
13	213,1	220,8	3,63	212,6	213,0	0,19
14	216,9	222,1	2,41	219,8	215,7	1,88
15	224,7	217,8	3,07	214,6	213,6	0,48
16	215,3	211,9	1,54	211,3	214,8	1,62
17	207,4	202,6	2,31	171,4	216,3	26,19
18	263,4	267,2	1,46	219,0	254,0	16,00
19	332,9	320,9	3,63	307,1	319,2	3,94
20	320,1	324,1	1,26	322,6	332,2	2,97
21	299,1	295,0	1,37	311,5	310,8	0,24
22	271,0	254,9	5,96	270,1	272,5	0,88
23	234,4	218,5	6,80	231,3	235,0	1,58
24	210,4	205,7	2,22	210,0	215,9	2,81
	MAPE		3,52	MAPE		4,34

5 DISCUSSION

The result of load forecasting contains several component of errors, namely, modelling error (error introduced to regression), error caused by system disturbances such as load shedding and irregular events and also errors of temperature forecast.

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 were no temperature forecast data available, temperature forecast was not included in this regression analysis. Further studies should include this in real life applications.

The load characteristic is not only affected by the temperature, but also affected by the other weather factors such as humidity, cloud cover and brightness of the day. Knowledge based on pre-analysis of the load, identified that in South Sulawesi system, the dominant portion of the load is residential (lighting) and this load is dependent mostly on the levels and duration of the daylight. It can be seen from load curve that the load rises sharply at 18 pm. Therefore the choice of time observation (t) as an independent variable in equation (7), (8) and (9) of the model, increased the accuracy of the forecasting.

6 CONCLUSIONS

To support the operation of electric power system securely and economically, needs an accurate load forecasting method as the principal driving element for all daily and weekly operations scheduling.

The multi linear regression models for short term load forecasting are relatively easy to develop and update regularly with widely available commercial computational software such as Microsoft Excel^{TM} as used in this investigations. The drawback of this model is its dependency on the accuracy of previously recorded load (this will be a common drawback for all STLF techniques) 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. This can be done through regular updates or monitoring to ensure that the error values are always below a pre-spesified error threshold

A pre analysis of the load is necessary for successful MLR forecast as the result of the analysis showed, because in South Sulawesi system, the load is not only affected by the fluctuation of temperature but also by other weather factors such as humidity and cloud cover. It is the daylight change that changes the lighting levels in residential and commercial spaces and demand of consumers according to the brightness of the day.

In this paper we have reported our preliminary investigations the use of MLR for STLF of the South Sulawesi Electrical System, and in our studies to date we have only modelled weekdays in wet and dry seasons. Investigations are underway to model weekend and holidays.

7 ACKNOWLEDGEMENTS

The author would like to thank PT. PLN (persero), the Indonesian state electricity company for financing the study in University of Abertay Dundee

8 REFERENCES

1. Gross,G., Galiana, F.D, Short-Term Load Forecasting, Proceeding of IEEE, December 1987.
2. Goh, T.N., Long, H.L., and Lee, Y.O., A New Approach to statistical Forecasting of daily Peak power Demand, Elect. Power Syst. Res., Vol.10, No.2, pp 145-148, March 1986.
3. Gupta, P.C., A Stochastic Approach to Peak Power Demand Forecasting in Electric Utility Systems, IEEE Trans. Power App. Syst., Vol. PAS-90, March/April 1971.

4. Thomson, R.P., Weather Sensitive Electric Demand and Energy Analysis on a large geographically Diverse Power System – Application to Short Term Hourly Electric Demand Forecasting, IEEE Trans. Power App. Syst., Vol. Pas-90, No 1, pp. 385-393, Jan./Feb. 1976.
5. Christiaanse, W.R., Short Term Load ForecastingUsing General Exponential Smoothing, IEEE Trans. Power App. Syst., Vol. PAS-90, No. 2, pp 900-911, March/April 1971.
6. Park, D.C, El-Sharkawi, M.A, Marks II, R.J, Atlas, L.E., Damborg, M,J., Electric Load Forecasting Using An Artificial Neural Network, IEEE Transaction on Power Systems, Vol. 6, No. 2, pp 442-449, May 1991.
7. 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.
8. 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.
9. Ragsdale, C.T., Spreadsheet Modeling and Decision Analysis a Practical Introduction to Management Science, Course Technology Inc, 1995.
10. Bowerman, B.L., O'Connel, R.T., Forecasting and Time Series an Applied Approach, Duxbury, 3rd Ed, 1993.

AUTHOR ADDRESS

The first author can be contacted at

Mphil Research Programme

Computing & Creative Technologies

University of Abertay Dundee

Bell Street, Dundee DD1 1HG

Email: 0604719@abertay.ac.uk