ISSN (Online): 2230-7893
www.IJCEM.org

Short Term Load Forecasting using Stochastic and Self-Organizing Maps

S. Farook[1] and NMG Kumar[2]

**[1,2] Department of Electrical and Electronics Engineering, Sree Vidyanikethan Engineering College
Tirupathi, Andhra Pradesh, India.**

Abstract

This paper concerns with forecasting short-term electricity load using self-organizing maps. Short term load forecasting has become more important for power producers and consumers to plan their resources and utilities strategically. Misjudgment of load may results in loss of revenue for both consumer and power produces. Knowing the load behavior in advance is an important aspect for power system planning, economic and reliable operation. In this paper, Short Term Load Forecasting (STLF) was carried out using time series, stochastic method (ARIMA) and Self-Organizing Maps (SOM) technique to predict the future demand, where the historical demand and weather factors as inputs. These developed methods give load forecasting for 24 hours in advance. The proposed SOM method is implemented in MATLAB. The performance of the model has been tested with practical data from Ercot load, considering six days historical data with hourly resolution. The Mean Absolute Percentage Error (MAPE) for hourly load demand forecasting is observed as 1.5% where as it is 3% and 6.5% with time series and Auto Regressive Integrated Moving Average (ARIMA). This comparison clearly shows the effectiveness of the SOM model for STLF.

Keywords: *Auto Regressive Integrated Moving Average (ARIMA), Self Organizing Maps (SOM), Short Term Load Forecasting (STLF), Time series.*

1. Introduction

The rapid growth of demand for electrical energy in the past few decades and the extreme fatigue of fossil fuel resources have raised the need to the development of optimal energy planning and forecasting of electrical load demand in power system. Load forecasting is an important component for power system energy management system and has always been the essential part of an efficient power system operation and planning. Precise and load forecasting helps the electric utility to make unit commitment decisions, reduce spinning reserve capacity and schedule device maintenance plan properly. Inaccurate load forecasting can increase the operational costs and maintenance costs. In [1]-[2], infers that a one percent error in forecast could imply in ten million pounds of operational costs. Power system expansion planning and operation starts with a forecasting of future load requirement. Load demand forecasting determines the capacity of generation; transmission and distribution system and energy forecasting determine the type of facilities required. Load forecasting is usually made by constructing models with relative information such as weather and historic load demand data [3]. Load forecasting with time leads, from a few minutes to several months helps the power system operator to efficiently schedule spinning reverse allocation and also provide information which is able to be used for possible electrical energy interchange with other utilities. Various techniques for load forecasting have been proposed in the past few decades. Authors in [10] have addressed the load forecasting problem in the presence of active demand. Authors in [11] have addressed uncertainties issues and improved forecasting accuracy using interval type 2 fuzzy systems. Such forecasting is usually aimed at short-term prediction like one hour or one day ahead prediction since longer load prediction may not be accurate due to error propagation. STLF with time lead of one hour is mainly required for real time control and also as input to security or contingency analysis.

This paper concerns with empirical analysis of short term load forecasting including the influencing factors using stochastic and self-organizing maps. Section II concerns with time series method of load forecasting. Section III concerns with stochastic methods and self-organizing maps were presented in section IV. The SOM models are implemented to predict one hour ahead load demand data. The accuracy of the SOM models are calculated and compared. The paper utilizes the Mean Absolute Percentage Error (MAPE), R-Squared and correlation value as a measure of forecast accuracy. The Section V concerns with the case studies on Ercot data considering different regions. The paper concludes with conclusion in section VI.

2. Time Series

A time series is a time ordered sequence of observations taken at regular intervals over a period of time (e.g., hourly, daily, weekly, monthly, and quarterly, annually). Forecasting techniques based on time series data are made on the assumption that future values of the series can be estimated from historical values. Analysis of time series data requires the analyst to identify the underlying behavior of the series [4]. An important feature of time series is that the successive observations are usually dependent. When successive observations are dependent future values may be predicted from the past observations. Time series model is used to explain the theories and mathematical representations and it establish relation between "cause" and "effects". "Time" is one variable, which is an independent variable and "Data" is another variable, which is a dependent variable. The classical decomposition procedure is used in time series analysis to easily extract patterns from the historical data [5]. Typically, the classical decomposition process takes four components into consideration: the Trend (T), Cycle component(C), Seasonal factor component (S) and Irregular component (I). Those components can be combined into an additive or a multiplicative model. In this analysis of time series model, multiplicative decomposition method is employed. Multiplicative decomposition model of time series can be described as

$$y(t) = T(t) * S(t) * C(t) * I(t) \tag{1}$$

Where is Y (t) the time series, T(t) is trend component, S(t) is seasonal component, C(t) is cyclic component and I(t) represent the irregular component. The cyclic component is normally in the duration of one year to a few years and is not applicable to STLF. In this irregular component is taken as one. Thus, simplified time series model has only two terms, the above equation is written as;

$$y(t) = T(t) * S(t) \tag{2}$$

3. Stochastic Method

The Auto-Regressive Moving Average (ARMA (p,q)) model is that the value taken by a time series at a given time t, denoted y depends on two additive terms: one is the historical values of the time series (AR component of order p) and another one is the past of the disturbances of the load data generation process (MA component of order q). Hence, the general form of an ARMA model [6] is

$$Y_t = c + \sum_{i=1}^{p}\left(\alpha_i Y_{t-i}\right) + e_t + \sum_{j=1}^{q}\left(\theta_j e_{t-j}\right) \tag{3}$$

Where C, α_i and Θ_j are the model parameters to be estimated and e_t is an error. The Auto-Regressive Integrated Moving Average (ARIMA (p,d,q)) model differs from the ARMA (p,q) model in one significant way ,it contains a d parameter. This parameter defines the level of non-seasonal differencing that is required to convert a non-stationary time series data into stationary data (when differencing is appropriate). Once stationary data is achieved, the series can be represented using an ARMA (p,q) model [7]. If the series is already stationary data and does not need to be differenced, then it can be directly used as ARMA (p,q). The general form of the ARIMA (p,d,q) model is

$$\alpha(B) Y_t = \theta(B) e_t \tag{4}$$

4. Self-Organizing Maps

Group Method of Data Handling (GMDH) works on SOM technique and its algorithm was first introduced G.Ivakhnenko in 1966 (Ivakhnenko, 1971). The SOM algorithm is a multilayered network with a certain structure determined through training [8]. It aims to find relationship between output and a large set of possible inputs. The architecture of GMDH is shown in below Fig.1

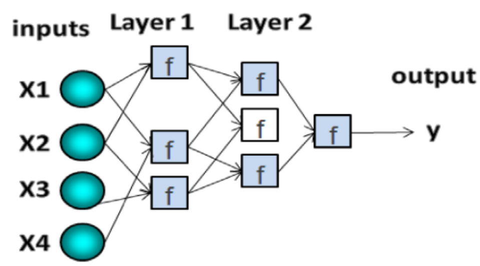

Fig.1: Architecture of GMDH Network.

The general connection between input variables and output is given by volterra series.

$$y = a_0 + \sum_{i=1}^{M}\left(a_i x_i\right) + \sum_{i=1}^{}\sum_{j=1}^{}\left(a_{ij} x_i x_j\right) + \sum_{i=1}^{}\sum_{j=1}^{}\sum_{k=1}^{M}\left(a_{ijk} x_i x_j x_k\right) + \ldots. \tag{5}$$

The proposed algorithm is based on a multilayer network using for each pair of input variables [9]. In each layer, there are neurons with only two input variables; the output of each neuron is a quadratic polynomial of its both input variables.. The quadratic polynomial is in the form:

$$Y = A_0 + A_1 X_i + A_2 X_j + A_3 X_i X_j + A_4 X_i^{\,2} + A_5 X_j^{\,2} \tag{6}$$

Where X_I and X_J are input variables and Y is the corresponding output value. The data values are divided into training and checking sets. The coefficients of the polynomial are found by regression analysis on the training set and its output is evaluated and tested for suitability using the data values in the checking set. A regularity criterion, normally the Mean Squared Error (MSE) is used to select the polynomials that are allowed to proceed for the next layer. The output of the selected polynomials of the previous layer becomes the new input values for the next layer. The whole procedure is repeated until the lowest regularity criteria are no longer smaller than that of

ISSN (Online): 2230-7893
www.IJCEM.org

the previous layer [12]. The flow chart that describes GMDH algorithm is shown in Fig.2.

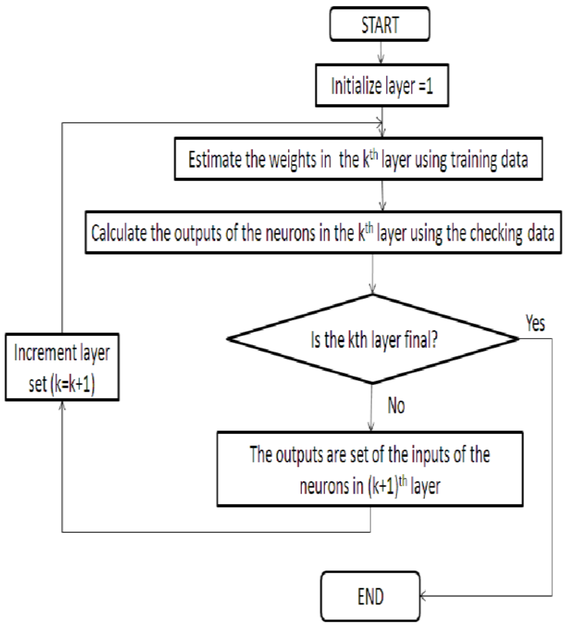

Fig.2: GMDH algorithm

4.1 Performance Measures

To determine the performance of SOM forecasting model for hourly ahead load demand prediction, the error measures namely Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), R-Squared and Correlation are considered and it is given by,

$$\text{Mean Absolute Error } (MAE) = \frac{1}{N}\sum_{i=1}^{N}\left|X_i - Y_i\right| \quad (7)$$

$$\text{Mean Absolute Percentage Error } (MAPE) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{X_i - Y_i}{X_i}\right| \quad (8)$$

$$\text{Root Mean Square Error } (RMSE) = \sqrt{\frac{\sum_{i=1}^{N}(X_i - Y_i)^2}{N}} \quad (9)$$

$$R^2 = 1 - \left(\frac{\text{Sum of squared distancebetween actual and predicted Y values}}{\text{Sum of squared distancebetween actual and predicted Y values and their mean}}\right) \quad (10)$$

$$\text{Correlation } (r) = \frac{N\left(\sum XY\right) - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left(N\sum X^2 - \left(\sum X\right)^2\right)\left(N\sum Y^2 - \left(\sum Y\right)^2\right)}} \quad (11)$$

5. Experimental Results

The results obtained by using proposed model with practical data are discussed and also the comparisons with conventional methods are listed. To assess the performance of the SOM forecasting model a practical data of euro load from www.Ercot.com is considered. The hourly humidity and temperature is collected from the public domain website www.wunderground.com. The demand forecasting for next day using time series, ARIMA and GMDH is shown in Fig. 3 to Fig. 11 respectively. The Table 1 to III shows the error measures with proposed forecasting model for the different regions of Ercot.

5.1 Coastal region

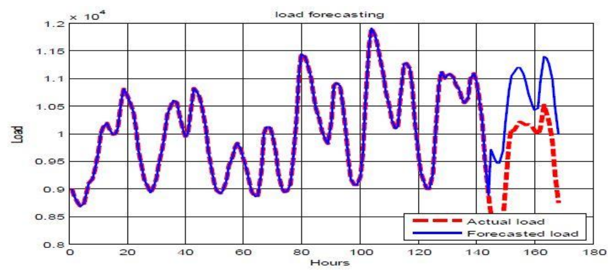

Fig. 3: Hourly ahead Actual vs. Forecasted power using time series

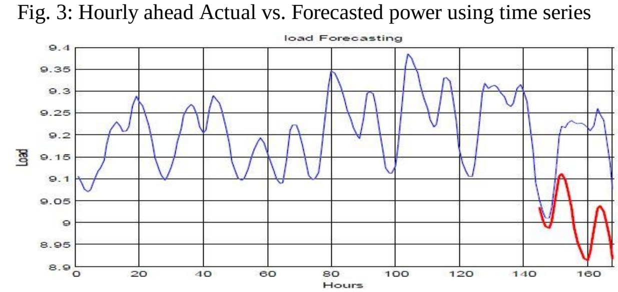

Fig. 4: Hourly ahead Actual vs. Forecasted power using ARIMA

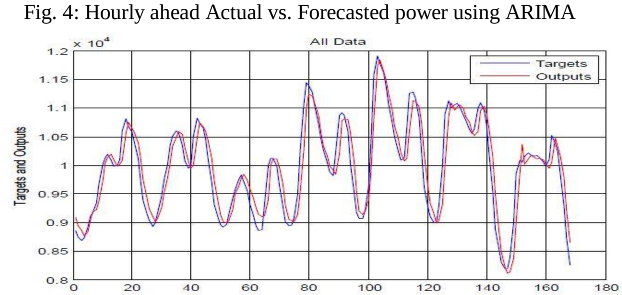

Fig. 5: Hourly ahead Actual vs. Forecasted power using GMDH

Table 1: Error measures of three methods for coastal region

Method	MAPE (%)	Max. MAPE (%)	RMSE (%)	R^2	Correlation(r)
Time series	3	9	5.6	0.998	0.9253
Stochastic	8.9	20	1.4	0.986	0.8979
SOM technique	**2.4**	**8**	**3**	**0.998**	**0.9305**

5.2 East region

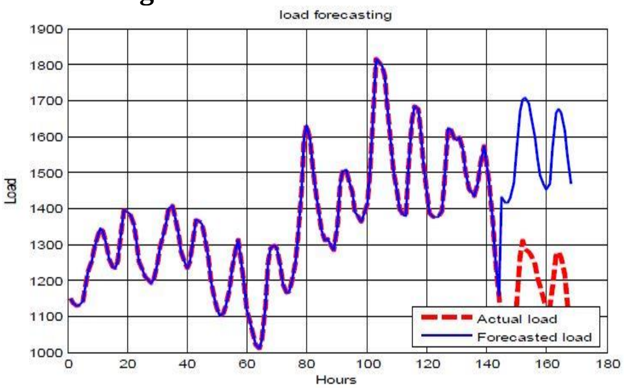

Fig. 6: Hourly ahead Actual power vs. Forecasted power using time series

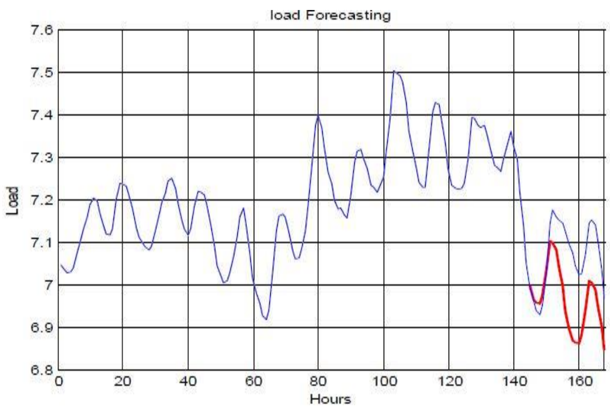

Fig. 7: Hourly ahead Actual power vs. Forecasted power using ARIMA

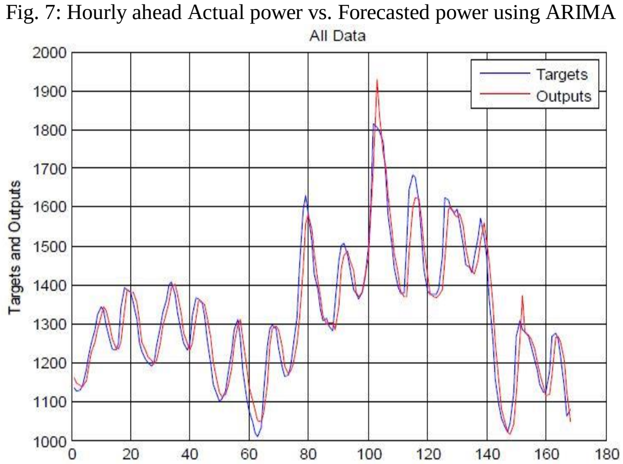

Fig. 8: Hourly ahead Actual power vs. Forecasted power using GMDH

Table 2 Error measures of three methods for east region

Method	MAPE (%)	Max. MAPE (%)	RMSE (%)	R^2	Correlation(r)
Time series	4	8	4.5	0.995	0.9392
Stochastic	9.7	18	10.4	0.986	0.8169
SOM technique	**3.3**	**6**	**4**	**0.997**	**0.954**

5.3 North Region

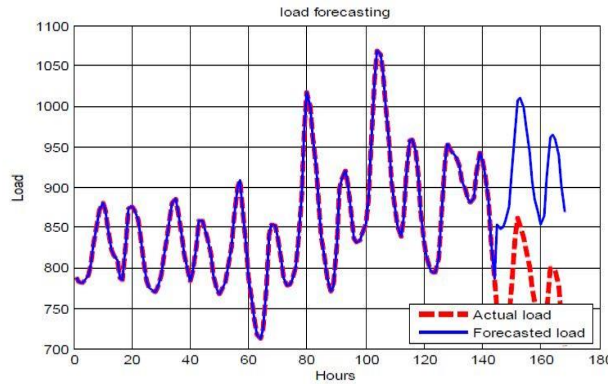

Fig. 9: Hourly ahead Actual power vs. Forecasted power using time series

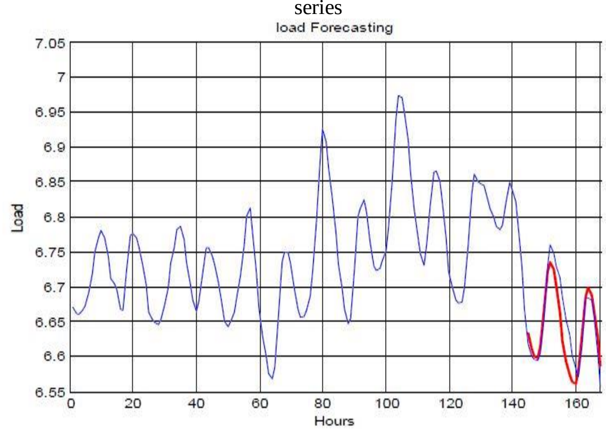

Fig. 10: Hourly ahead Actual power vs. Forecasted power using ARIMA

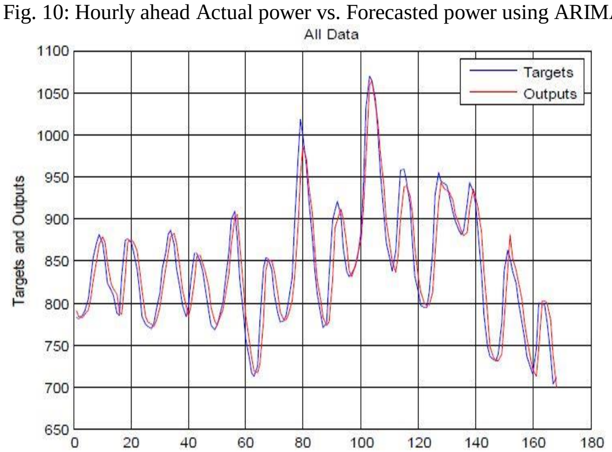

Fig. 11: Hourly ahead Actual power vs. Forecasted power using GMDH

Table 3 Error measures of three methods for north region

Method	MAPE (%)	Max. MAPE (%)	RMSE (%)	R^2	Correlation(r)
Time series	3.8	9	4.05	0.996	0.9263
Stochastic	6	14	6	0.995	0.8818
SOM technique	**2.6**	**7**	**3.4**	**0.998**	**0.935**

The results of the SOM forecasting model are compared with those obtained with time series and ARIMA model in terms of MAPE and R-squared value. The following Table 4 shows comparisons.

Table 4 Comparison of various methods

S. No	Error	Time series	ARIMA	SOM
1	**MAPE(%)**	3	6.5	**1.5**
2	**Max. MAPE(%)**	6	11	**3**
3	**RMSE(%)**	3.1	6.3	**1.6**
4	**R^2**	0.9988	0.9948	**0.9996**
5	**Correlation**	0.8985	0.8621	**0.9065**

From the Table 4 it is observed that SOM Technique has less MAPE (1.5%), maximum MAPE (3%) and R-squared (0.9996) as compared with remaining models for hourly ahead load forecasting.

Conclusion

In this project, three forecasting models has been developed to predict an hourly load demand, in first method, Time Series is used to predict the next day hourly load demand. In second method, stochastic method (ARIMA) is used to predict the next day hourly load demand. In third method, Self Organizing Maps (SOM) Technique is used to predict the next day hourly load demand based on Neural Networks. The proposed model is implemented in MATLAB to forecast an hourly ahead load demand. Finally, the results are compared with those obtained by Time series and Stochastic method in terms of MAPE, R-Squared and Correlation value. The actual, maximum MAPE, R-Squared and Correlation value of SOM model for an hourly ahead load demand prediction is 1.5%, 3%,0.9996 and 0.9065 respectively, where as in case of Time series these are 3%, 6%, 0.9988 and 0.8985 respectively and in case of Stochastic method these are 6.5%, 11%, 0.9948 and 0.8621 respectively. From the above results it has been observed that Self Organizing Maps method has better accuracy.

References

[1] A. D. Papalexopoulos and T. Hesterberg, “A Regression-based approach to short-term system load forecasting,” IEEE Trransactions on Power Systems, vol. 5, pp. 1535-1550, Nov. 1990.

[2] P.F.Pai,W.C. Hong, “Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms,” ElectricPower Systems Research, vol. 74, pp. 417-425, Apr. 2005.

[3] Indraraj B. Sindhav, KaushikKhatua, “ Short-term Load Forecasting using traditional demand forecasting”, IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) , vol. 10. No. 2, pp. 01-06, Feb.2015.

[4] Javier Contreras, Jesús Riquelme Santos, “Short-term Demand and Energy Price Forecasting”,IEEE Melecon,May 16-19,2006.

[5] Girraj Singhet al., “Factor Affecting Elements and Short term Load forecasting Based on Multiple Linear Regression Method”,International Journal of Engineering Research & Technology (IJERT),Vol. 3 Issue 12, December-2014.

[6] Tarjei Kristiansen, “Forecasting Nord Pool Day-Ahead Prices With An Autoregressive Model”, Electrical Power and Energy Systems, Elsevier, 2012.

[7] J.F. Chen, W.M. Wang et. al., “Analysis of Adaptive Time-Series Autoregressive Moving Average Model For Short Term Load Forecasting”, Electrical power systems research, Elsevier, 1995.

[8] Ehab. E. Elattar, et al., “Generalized Locally Weighted GMDH for Short Term Load Forecasting”IEEE Transactions On Systems,Vol. 42, No. 3, May2012.

[9] Ruchika Malhotra, Anuradha Chug, “Application of Group Method of Data Handling (GMDH) model for software maintainability prediction using object oriented Systems”.Assu.rEng Manag,Spinger,2014.

[10] S. Paoletti, M. Casini, A. Giannitrapani, A. Facchini, A. Garulli, and A. Vicino, “Load forecasting for active distribution networks,” in Inno- vative Smart Grid Technologies (ISGT Europe), 2011 2nd IEEE PES International Conference and Exhibition on, pp. 1–6, 2011.

[11] A. Khosravi, S. Nahavandi, D. Creighton, and D. Srinivasan, “Interval type-2 fuzzy logic systems for load forecasting: A comparative study,” Power Systems, IEEE Transactions on, vol. 27, no. 3, pp. 1274–1282, 2012.

[12] Nadira Mohamed Isa, ShabriAni and SamsudinRuhaidah, “The Group Method Of Data Handling (GMDH) and Artificial Neural Networks (ANN) In Time-Series Forecasting of Rice Yield”, Journal of Materials Science and Engineering,2011.

[13] A. Setiawan, I. Koprinska, and V. Agelidis, “Very short-term electricity load demand forecasting using support vector regression,” in Neural Net- works, 2009. IJCNN 2009. International Joint Conference on, pp. 2888–2894, 2009.

[14] Bautu E, Barbulescu A. Forecasting Meteorological Time Series using Soft Computing Methods: An Empirical Study, Appl. Math. 2013; 7(4):1297−306.

[15] Khashei M, Bijari M. An Artificial Neural Network (p,d,q) Model for Timeseries Forecasting, Expert Syst. Appl.2010;37(1):479−89. Doi: 0.1016/j.eswa.2009.05.04.