

Medium-Term Electric Load Forecasting Using Multivariable Linear and Non-Linear Regression

Nazih Abu-Shikhah¹, Fawwaz Elkarmi¹, Osama M. Aloquili²

¹Department of Electronics and Communications, Ahlliya Amman University, Amman, Jordan; ²Department of Electrical Engineering, The Hashemite University, Zarqa, Jordan.

Email: {nabushik, fkarmi}@ammanu.edu.jo, aloquili@hu.edu.jo

Received November 2nd, 2010; revised March 29th, 2011; accepted April 3rd, 2011.

ABSTRACT

Medium-term forecasting is an important category of electric load forecasting that covers a time span of up to one year ahead. It suits outage and maintenance planning, as well as load switching operation. We propose a new methodology that uses hourly daily loads to predict the next year hourly loads, and hence predict the peak loads expected to be reached in the next coming year. The technique is based on implementing multivariable regression on previous year's hourly loads. Three regression models are investigated in this research: the linear, the polynomial, and the exponential power. The proposed models are applied to real loads of the Jordanian power system. Results obtained using the proposed methods showed that their performance is close and they outperform results obtained using the widely used exponential regression technique. Moreover, peak load prediction has about 90% accuracy using the proposed methodology. The methods are generic and simple and can be implemented to hourly loads of any power system. No extra information other than the hourly loads is required.

Keywords: Medium-Term Load Forecasting, Electrical Peak Load, Multivariable Regression, Time Series

1. Introduction

The fact that there are many variables contributing to the electric load makes accurate prediction of electric load a difficult process. These variables involve "uncertainty" and have no direct relation with the final load. Moreover, the load is characterized to be nonlinear and non-stationary process that can undergo rapid changes due to weather, seasonal and macroeconomic variations. So linearization of the load contributes to making many of the classical prediction models inappropriate [1,2].

From the forecasting point of view, the utility/company seeks to operate the power system such that a match is achieved between the electric energy demand and supply. This implies that the more accurate the forecast, the more efficient the operation and management of the power system.

Medium term load forecasting covers a time span of (1 - 12 months) [3]. This type of forecasting depends mainly on growth factors, *i.e.* factors that influence demand such as main events, addition of new loads, seasonal variations, demand patterns of large facilities, and maintenance requirements of large consumers. Moreover, this type of forecast uses hourly loads for prediction of the peak load

of days or for the weeks ahead. With this information it can be decided to whether take certain facilities/plants for maintenance or not during a given period of time. This will also help to plan major tests and commissioning events, and determine outage times of plants and major pieces of equipment. The analysis methods used for this type of forecast are similar to the short term forecast. However, it should be remarked that the sensitivity of medium-term forecasting on power system operations is less than that of the short-term forecasting [4].

Since the electric load varies continuously in time, it is considered to be a time series. This enables applying different time series techniques and methodologies to predicting future loads based on the available historical data of the loads.

Time series techniques are based on the assumption that the data have an internal structure, such as autocorrelation, trend, or seasonal variation. Time series forecasting methods detect and explore such a structure [4,5]. The objective of this paper is to develop and implement a new technique that is based on a non-linear model using multivariable regression. This will work as a filtering process to explore the structure of the load behavior, as

well as, to enable prediction of future loads. In this paper, the time series approach is adopted, however, we propose establishing a fit to an exponential model that takes into account the previous years' hourly loads.

Electric companies/utilities use mostly simple forecasting models like linear regression and simple econometric models of one or two parameters. However, the current trend now is to apply multiple regression forecasting models especially to large systems [6,7]. Multiple regression in addition to ARMAX models showed also good performance [8] and they were applied to both electric load and energy forecasting [9,10].

As a matter of fact the majority of forecasting models use statistical techniques or artificial intelligence algorithms such as regression, neural networks, fuzzy logic, and expert systems. The end-use and econometric approach is broadly used for medium- and long-term forecasting. A variety of methods, which implement the similar day approach, various regression models, time series, neural networks, statistical learning algorithms, fuzzy logic, and expert systems, have been developed and are available for short-term forecasting [4,5,9,10].

The forecasting category belonging to quantitative methods are based on mathematical formulation and include: regression analysis, decomposition methods, exponential smoothing, and the Box-Jenkins methodology [11-13]. The research carried out in this paper belongs to this category, *i.e.* quantitative methods.

This paper is organized as follows: in Section 2, the description of the multivariable regression model is illustrated followed by the introduction of the exponential power model. In Section 3 the developed technique is explained, and in Section 4, results of implementing this technique to the Jordanian power system are discussed and analyzed using several error and accuracy indicators. A comparison of the results obtained using the proposed method with those obtained using the exponential regression method is also demonstrated. Section 5 presents the final outcomes and conclusions of this study.

2. The Models

When time series analysis is in forecasting, past information is used, in conjunction with a forecasting model, to predict future values. This becomes an optimization problem. In general, this process can be expressed as the search for or synthesis of a function f which leads to the prediction accordingly [14]:

$$\hat{Y}_{\lambda+1}(k) = f(X_{\lambda}(k), X_{\lambda-1}(k), \dots, \alpha_1, \alpha_1, \dots, \alpha_p), \quad (1)$$

$$k = 1, 2, \dots, N$$

where,

$\hat{Y}_{\lambda+1}(k)$ is the estimated loads for the next time span,
 N is the signal length = 8760 hours

$X_m(k)$ are the multiple variables ($m = 1, 2, \dots, \lambda, \dots$)

α_i are the model parameters to be computed ($i = 1, 2, \dots, P$)

λ is the number of years span considered in the forecasting process.

Practically speaking, forecasting becomes a problem of approximating a given function as precisely as possible while being able to quantify the performance of forecasting error [15,16].

In this research we propose using the previous year's yearly hourly loads in conjunction with multivariable regression to find the forecast of the next year hourly loads. Three models are investigated: a) Modeling the next year's load as a linear function of multivariable (previous year's hourly loads), b) as a nonlinear function of previous year's hourly loads, and c) as a power exponent of previous year's hourly loads. The mathematical formulation of these models is derived in the subsequent sections.

2.1. Multivariable Regression

Multiple regressions are used in load forecasting when the predictor variable y is set to be a function of more than one variable. A **linear regression** model is given as:

$$\hat{Y}(k) = a_0 + \sum_{i=1}^{\lambda} a_i X_i(k), \quad k = 1, 2, \dots, N \quad (2)$$

where,

$\hat{Y}(k)$ is the estimated loads for the next year,

N is the signal length = 8760 hours

$X_i(k)$ are the multiple variables

α_i are the model parameters to be computed ($i = 0, 1, \dots, \lambda$)

λ is the number of years span considered in the forecasting process.

The linear model for a time span $\lambda = 1$, reduces to:

$$\hat{Y}(k) = a_0 + a_1 X_1(k), \quad k = 1, 2, \dots, N \quad (2a)$$

The non-linear multiple regressions assume that the estimates of loads $\hat{Y}(k)$ have a non-linear relationship with the multiple variables $X_i(k)$. It should be noted that the selection of the multiple variables is open and unlimited to a restricted set. For short-term load forecasting, they can represent the temperature, the wind speed, the cloud density, etc. In this research the selection was directed to the hourly loads of the year. So $X_i(k)$ were selected to represent the previous year's hourly loads.

An m^{th} order multivariable **polynomial regression** may take the following form:

$$\hat{Y}(k) = a_{00} + \sum_{i=1}^{\lambda} a_{1i} X_i(k) + \sum_{i=1}^{\lambda} a_{2i} X_i^2(k) + \dots + \sum_{i=1}^{\lambda} a_{mi} X_i^m(k), \quad k = 1, 2, \dots, N \tag{3}$$

where,

a_{ij} are the unknown parameters to be computed with $[i = 0, 1, \dots, \lambda, \text{ and } j = 0, 1, \dots, m]$

If the estimated hourly load data is assumed to depend on the previous year hourly loads data, then the model of Equation (3) may be written as:

$$\hat{Y}(k) = a_{00} + a_{11} X_1(k) + a_{12} X_1^2(k) + \dots + a_{1m} X_1^m(k) \quad k = 1, 2, \dots, N \tag{3a}$$

The unknown parameters $a_{00}, a_{11}, a_{12}, \dots, a_{1m}$ are found by minimizing the Mean Squared Errors (MSE) between

original and estimated values is given as:

$$E = \sum_{k=1}^N [Y(k) - \hat{Y}(k)]^2 = \sum_{k=1}^N [Y(k) - a_{00} - a_{11} X_1(k) - a_{12} X_1^2(k) - \dots - a_{1m} X_1^m(k)]^2 \tag{4}$$

The unknown parameters are found using the follow-

ing equation:

$$\begin{pmatrix} a_{00} \\ a_{11} \\ a_{12} \\ \vdots \\ a_{1m} \end{pmatrix} = \begin{pmatrix} N & \sum_{k=1}^N X_1(k) & \sum_{k=1}^N X_1^2(k) & \dots & \sum_{k=1}^N X_1^m(k) \\ \sum_{k=1}^N X_1(k) & \sum_{k=1}^N X_1^2(k) & \sum_{k=1}^N X_1^3(k) & \dots & \sum_{k=1}^N X_1^{m+1}(k) \\ \sum_{k=1}^N X_1^2(k) & \sum_{k=1}^N X_1^3(k) & \sum_{k=1}^N X_1^4(k) & \dots & \sum_{k=1}^N X_1^{m+2}(k) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^N X_1^m(k) & \sum_{k=1}^N X_1^{m+1}(k) & \sum_{k=1}^N X_1^{m+2}(k) & \dots & \sum_{k=1}^N X_1^{2m}(k) \end{pmatrix}^{-1} \begin{pmatrix} \sum_{k=1}^N Y(k) \\ \sum_{k=1}^N Y(k) X_1(k) \\ \sum_{k=1}^N Y(k) X_1^2(k) \\ \vdots \\ \sum_{k=1}^N Y(k) X_1^m(k) \end{pmatrix} \tag{4a}$$

2.2. The Exponential Power

This model gives an estimate of the current value of a given signal through modeling it as an exponential function of previous year's hourly loads. If we select the previous year dependency, then the model is mathematically described as:

$$\hat{Y}(k) = a X^b(k), \quad k = 1, 2, \dots, N \tag{5}$$

where,

- $\hat{Y}(k)$ is the estimated loads for the next year,
- N is the yearly hours = 8760 hours
- $X(k)$ is the previous year's loads at the k^{th} hour.
- a, b are the model parameters to be computed

It is essential to replace the data by their natural logarithm ($\ln = \log_e$), this will transform Equation (5) to a linear form, given by:

$$\ln(\hat{Y}(k)) = \ln(a) + b \ln(X(k)) = \alpha + b \times \ln(X(k)), \quad k = 1, 2, \dots, N \tag{5a}$$

It can be seen that $\alpha = \ln(a)$ implying that $a = e^\alpha$. The objective is to find optimum values for a , and b .

Hence, by minimizing the (MSE) between original and estimated values is given as:

$$E = \sum_{k=1}^N [\ln(Y(k)) - \ln(\hat{Y}(k))]^2 = \sum_{k=1}^N [\ln(Y(k)) - \alpha - b \times \ln(X(k))]^2$$

This objective is attained by setting the partial differentials of E with respect to a , and b equal to zero. This will

lead to the following:

The unknown coefficients can be found using:

$$\begin{bmatrix} \sum_{k=1}^N \ln(Y(k)) \\ \sum_{k=1}^N \ln(Y(k))\ln(X(k)) \end{bmatrix} = \begin{bmatrix} N & \sum_{k=1}^N \ln(X(k)) \\ \sum_{k=1}^N \ln(X(k)) & \sum_{k=1}^N \ln(X(k))\ln(X(k)) \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} \quad (7)$$

Hence, the optimum values are:

$$\begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} N & \sum_{k=1}^N \ln(X(k)) \\ \sum_{k=1}^N \ln(X(k)) & \sum_{k=1}^N \ln(X(k))\ln(X(k)) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{k=1}^N \ln(Y(k)) \\ \sum_{k=1}^N \ln(Y(k))\ln(X(k)) \end{bmatrix} \quad (7a)$$

Therefore the model parameters a , and b , are assessed such that

$$a = e^{\alpha} \quad (7b)$$

3. The Proposed Technique

3.1. Modeling

The proposed load forecasting technique can be applied to assess the medium term load forecasting in any power system. A simple block diagram that represents the modeling phase of the proposed methodology is illustrated in **Figure 1**.

This figure shows that the input variables are given by previous year's hourly loads $X_m(k)$, ($m = 1, 2, \dots, \lambda$, and $k = 1, 2, \dots, 8760$). The computed model parameters, are used in conjunction with the actual load $Y(k)$ to provide the hourly load forecast $\hat{Y}(k)$. The forecasted loads are then subtracted from the actual recorded hourly loads. This will result in the noise added or the random part of these loads. This analysis is performed for previously known loads, and results are used to define the pattern and behavior of the R component, in addition to the characterization of the computed parameters of different models used.

In fact we can use Signal to Noise ratio as a measuring metric. It is shown in a following part of this paper that the SNR for the proposed models is about 21 dB. The next step is, obviously, to use obtained results to estimate

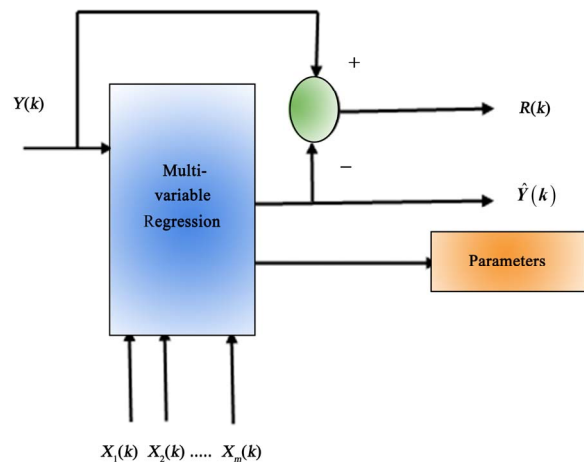


Figure 1. Multi-variable regression modeling.

or forecast the next time span unknown loads.

3.2. Parameters Estimation

The estimation stage is based on three factors:

1) The signal to noise ratio (SNR) between the recorded hourly loads and their corresponding random loads R . This is given as:

$$SNR = 10 \log_{10} \left(\frac{\sum_{k=1}^{8760} Y^2(k)}{\sum_{k=1}^{8760} R^2(k)} \right) \text{ dB} \quad (8)$$

2) The model parameters computed in the previous phase, to compute $\hat{Y}(k)$.

3) The energy growth model which is a polynomial model used to fit the recorded yearly energies. Here, we can also use the sum of hourly loads to represent acceptable energy consumption for certain power systems.

The above factors can then be employed to compute the medium term forecast, for $k = 1, 2, \dots, 8760$, according to the following steps:

1) Knowing the average SNR , and the energy forecast of the recorded load, then the energy of the $R(k)$ component for the next time span is computed.

2) Use the estimated parameters to compute the model-based estimation of loads $\hat{Y}_m(k)$.

3) The hourly load forecast is found by adding results of 1, and 2 above, resulting in an estimation given by $\hat{Y}(k) = \hat{Y}_m(k) + R(k)$

3.3. Load Forecasting

In this research, the forecasting process using multi-variable method is restricted to the time span of one year. The adopted load forecasting procedure based on this time span, *i.e.* $\lambda = 1$, involves the following steps, which are illustrated in **Figure 2**:

a) Process the hourly loads of previous year using the specific regression model to compute model parameters,

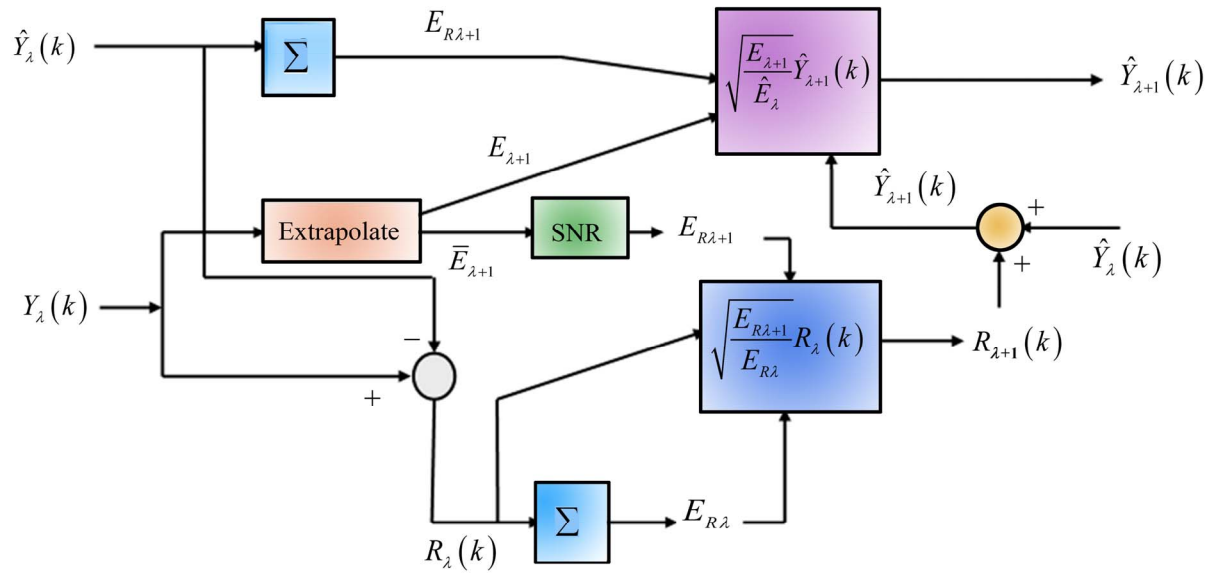


Figure 2. Forecasting process for time span $\lambda = 1$.

estimated load.

b) Compute the energies $E_\lambda = \sum_1^{8760} Y_\lambda(k)$, and

$$\hat{E}_\lambda = \sum_1^{8760} \hat{Y}_\lambda(k).$$

c) Compute the random component

$R_\lambda(k) = Y_\lambda(k) - \hat{Y}_\lambda(k)$, and the energy of this component

$$E_{R\lambda} = \sum_1^{8760} R_\lambda^2(k).$$

d) Compute signal energy $\bar{E}_\lambda = \sum_1^{8760} Y_\lambda^2(k)$, and ex-

trapolate to find the value at the next time span $\bar{E}_{\lambda+1}$.

e) Find an estimate of the noise energy for the next span based on Equation (8), such that

$$E_{R\lambda+1} = \bar{E}_{\lambda+1} \times 10^{-SNR/10}$$

f) Find the estimate of R for the next time span using $R_{\lambda+1}(k) = \sqrt{\frac{E_{R\lambda+1}}{E_{R\lambda}}} R_\lambda(k)$.

g) Find the initial estimate of the load in the next span as: $\hat{Y}_{\lambda+1}(k) = \hat{Y}_\lambda(k) + R_{\lambda+1}(k)$

h) Find the required an estimate of the load $\tilde{Y}_{\lambda+1}(k) = \frac{E_{\lambda+1}}{\hat{E}_\lambda} \hat{Y}_{\lambda+1}(k)$

3.4. Error Performance

Many error measures exists which are defined based on recorded (actual), $Y(k)$, and estimated loads, $\hat{Y}(k)$, $k=1,2, \dots, 8760$. The following were used in this research:

a) The absolute normalized error ($\varepsilon_{\%}$), which is computed based on the following equation:

$$\varepsilon_{\%} = \frac{\sum_{k=1}^{8760} |Y(k) - \hat{Y}(k)|}{\sum_{k=1}^{8760} |Y(k)|} \times 100 (\%) \quad (9a)$$

The absolute percentage error (APE) ($\varepsilon_{APE\%}$), which is computed based on the following equation:

$$\varepsilon_{APE\%} = \sum_{k=1}^{8760} \left| \frac{Y(k) - \hat{Y}(k)}{Y(k)} \right| \times 100 (\%) \quad (9b)$$

b) The average of the absolute error (ε_{av}), which is computed based on the following equation:

$$\varepsilon_{av} = \frac{1}{8760} \sum_{k=1}^{8760} |Y(k) - \hat{Y}(k)| \quad (MW) \quad (9c)$$

4. Implementation to Jordanian System

The hourly loads for the years 1994-2008 were used to explore the characteristics of the Jordanian power system. The three proposed models were applied to the abovementioned loads as explained in the following sections. The first step was to estimate the energy growth of this system.

4.1. Estimation of Energy

The Jordanian system energy growth over the years and the corresponding 2nd and 5th order polynomial fit are shown in **Figure 3**.

The corresponding equations of the estimated energy ($\hat{E}(k)$) in (MWH) for these fits, with $k = 1, 2, \dots, 8760$,

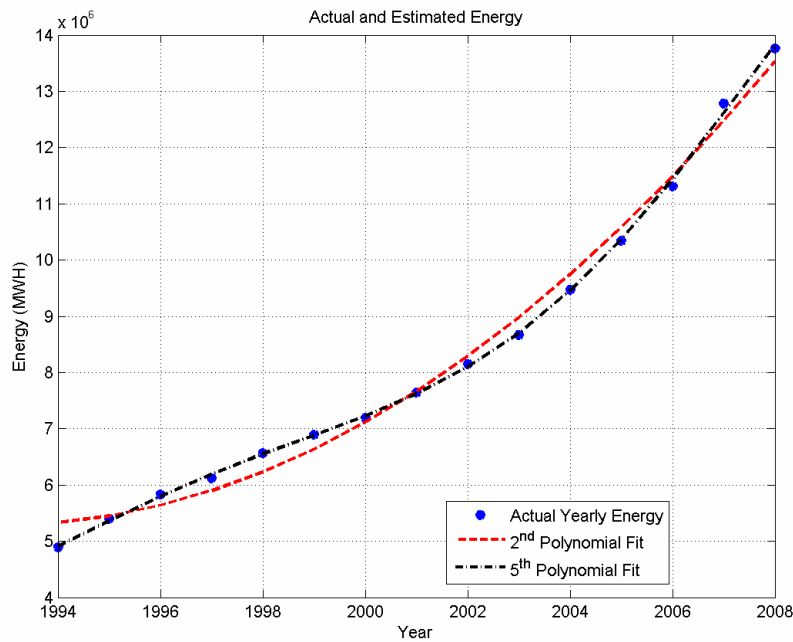


Figure 3. Energy growth and their polynomial fit for the years 1994-2008.

are:

- Second order polynomial:

$$\hat{E}(k) = 0.7205 \times 10^6 k^2 + 2.6192 \times 10^6 k + 7.6632 \times 10^6 \tag{10a}$$

- While, the fifth order polynomial is:

$$\hat{E}(k) = -9.1 \times 10^4 k^5 - 7 \times 10^4 k^4 + 6.1 \times 10^5 k^3 + 8.9 \times 10^5 k^2 + 19.1 \times 10^5 k + 7.6210^6 \tag{10b}$$

It is apparent from **Figure 3** that a fifth order polynomial will result in a very close prediction values of energy over the years of study. However, the second order fit shows acceptable results and can be used if simpler computations are needed. The polynomial fit will be used to find out unknown future energy values, which are used to compute the associated random or noise component for the particular year as explained in the following sections.

4.2. Linear Model Parameters

Here, the model shown in Equation (2) was used where future load is assumed to depend on the loads of several historical years. The SNR was computed for different values of λ between 1-7 years. Results are summarized in **Table 1**:

Table 1 shows that the time span has little effect on the SNR. In fact the overall average of the SNR is 20.2 dB. The coefficients associated with the time span $\lambda = 1$ (Equation (2a)) were computed and results are shown in **Table 2**.

The over all averages for all years are: $SNR = 20.9$, $a_0 = 163.6$ and $a_1 = 0.9$

It is concluded that the average linear model for $\lambda = 1$ can be described by the following formula (for $k = 1, 2, \dots, 8760$):

$$\hat{Y}(k) = 163.6 + 0.9 X_1(k) \tag{11}$$

A typical actual hourly loads $[Y(k)]$, and forecasted hourly loads $[\hat{Y}(k)]$, in addition to the noise hourly values $[R(k)]$ for a selected year (2004) is shown in **Figure 4**.

4.3. Polynomial Model Parameters

Here, the model shown in Equation (3a) was used (*i.e.* $\lambda = 1$). This selection of λ was used to simplify the model and to assure higher load correlation due as the time span λ is reduced. Moreover, this selection will enable comparing results of this model with that of the exponential power model.

The SNR was computed when using the third order polynomial model (*i.e.* $P = 3$), and results showed that the overall average of the SNR is 20.9, while the optimum model using third order polynomial multivariable regression is given (Equation (3a)) on average, for $k = 1, 2, \dots, 8760$, as:

$$\hat{Y}(k) = 253.81 + 0.154 X_1(k) + 0.0013 X_1^2(k) - 0.067 \times 10^{-5} X_1^3(k) \tag{12}$$

Table 1. Effect of time span (λ) on SNR for different years.

Year	Time Span (λ)						
	1	2	3	4	5	6	7
1995	21.4						
1996	21.6	21.2					
1997	20.9	20.8	20.9				
1998	21.1	20.6	20.6	20.9			
1999	20.5	20.3	20.2	20.1	20.4		
2000	20.2	20.1	20.1	20	20	20.2	
2001	20.2	19.7	19.8	19.9	20.3	20.7	19.7
2002	20.8	19.9	19	19.6	20	20	19.7
2003	20.9	19.9	20.2	19.5	20.2	23.4	20
2004	21	20.4	20.1	20.3	22.1	21	20.3
2005	21.4	20.5	20	20.1	20.5	20.5	20.1
2006	21.2	19.9	19.8	19.4	19.6	19.6	18.9
2007	21	20.2	19.9	19.8	20.1	21.9	19.6
2008	19.9	19.2	19.1	18.6	18.7	18.6	18.5
Mean	20.9	20.2	20	19.8	20.2	20.7	19.6

Table 2. Linear model coefficients for time span $\lambda = 1$.

Year	SNR	a_0	a_1	Year	SNR	a_0	a_1
1995	21.4	115.6	0.9	2002	20.8	113.8	0.9
1996	21.6	125.7	0.9	2003	20.9	163.8	0.9
1997	20.9	129.3	0.9	2004	21	179.8	0.9
1998	21.1	141.7	0.9	2005	21.4	177.9	0.9
1999	20.5	150.7	0.8	2006	21.2	128.3	1
2000	20.2	112	0.9	2007	21	285.6	0.9
2001	20.2	185.6	0.8	2008	19.9	281	0.9

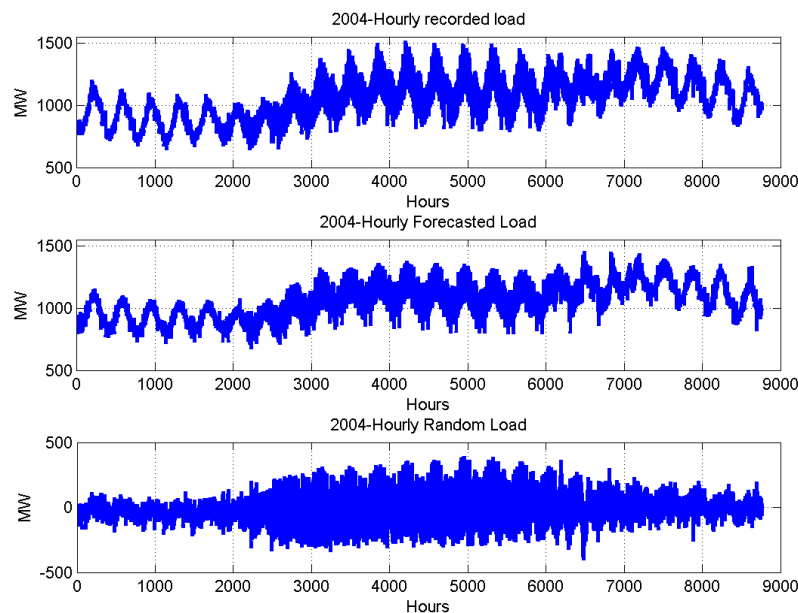


Figure 4. Load components related to linear model analysis for 2004.

4.4. Exponential Power Model Parameters

Here, the model shown in Equation (5) was used with one year span. The SNR values, and the model parameters were computed using the matrix equation given by Equation (7). Results showed that the SNR for the individual years are close, and the average SNR values is about 21 dB. Similarly, the computed model parameters (found by using Equation (7)) have also close values for different years, and it can be concluded that a good model, for $k = 1, 2, \dots, 8760$, would be given by:

$$\hat{Y}(k) = 3.4 X^{0.83}(k) \tag{13}$$

It should be emphasized that other time span value can be used, and hence the procedure must be modified accordingly. However, this is out of the scope of this paper and shall be investigated in future research. We shall illustrate using the above procedure to forecast the load of the year 2008 for the three models adopted.

4.5. Forecasting Results

4.5.1. Linear Model

The results of application of the linear model are summarized in **Table 3**. The table shows the errors incurred when applying the linear model and the corresponding yearly peak loads errors. It can be seen that the average error in estimation reaches 9.8% while the absolute error averages to 6.7%. The mean error per hour is about 96 MW. On the other hand the percentage error in peak forecasting is about 5.3% which corresponds to 76.7 MW.

4.5.2. Polynomial Model

The results of application of the third degree polynomial model are summarized in **Table 4** which indicates that the errors incurred of the polynomial model are very similar to those obtained for the linear model. It can be seen that the average error in estimation reaches 9.7% while the absolute error averages to 6.4%. The mean error per hour is about 97 MW. On the other hand the percentage error in peak forecasting is about 4.7% which corresponds to 72.2 MW.

4.5.3. Exponential Power Model

The results of application of the exponential power are summarized in **Table 5**. The table shows that the errors incurred of the exponential power vary to a limited extent from those obtained for the linear model. It can be seen that the average error in estimation reaches 10.3% (worse than linear) while the absolute error averages to 4.6% (better than linear). The mean error per hour is about 111 MW. On the other hand the percentage error in peak forecasting is about 9.2% which corresponds to 120 MW (worse than linear).

4.5.4. Comparison with Exponential Regression

A comparison between the proposed methods and the exponential regression method, which is widely used by many electric utilities, was performed. Hourly load data for the period 1994 - 2007 were used to predict the hourly loads of next time span using exponential regression. Results show that the average error incurred is 299 MW corresponding to 20% in the peak load estimation using the exponential regression. For the year 2008, in particu-

Table 3. Linear model: overall incurred errors and peak load errors.

Year	Incurred errors			Recorded Peak (MW)	Forecasted Peak (MW)	Peak load forecasting Errors	
	$\epsilon_{\%}$	$\epsilon_{APE\%}$	$\epsilon_m(\text{MW})$			$ \epsilon (\text{MW})$	$ \epsilon (\%)$
1996	11.5	9.7	76.3	902	983	81	9
1997	13.2	11.4	92.4	971	1047	76	7.8
1998	10.6	8.4	79.7	1020	1073	53	5.2
1999	11.5	9.3	90.6	1099	1161	62	5.6
2000	11.9	9.7	97.7	1206	1200	6	0.5
2001	10.7	7.7	93.6	1225	1305	80	6.5
2002	9.1	5.8	84.8	1370	1392	22	1.6
2003	8.8	5	86.8	1387	1451	64	4.6
2004	6.8	1.4	73.2	1515	1578	63	4.2
2005	6.4	1	76	1710	1639	71	4.2
2006	7.5	3.2	96.3	1860	1742	118	6.3
2007	9.9	7.7	144.3	2130	1979	151	7.1
2008	9.7	6.9	151.8	2230	2081	149	6.7
Mean	9.8	6.7	95.7			76.6	5.3

Table 4. Polynomial model: overall incurred errors and peak load errors.

Year	Incurred errors			Recorded Peak (MW)	Forecasted Peak (MW)	Peak load forecasting Errors	
	$\varepsilon\%$	$\varepsilon_{APE}\%$	ε_{av} (MW)			$ \varepsilon $ (MW)	$ \varepsilon $ (%)
1996	11.5	9.7	76.3	902	961	59	6.5
1997	13.2	11.4	92.4	971	1015	44	4.5
1998	10.6	8.4	79.7	1020	1069	49	4.8
1999	11.5	9.3	90.6	1099	1180	81	7.5
2000	11.9	9.7	97.7	1206	1208	2	0.2
2001	10.7	7.7	93.6	1225	1255	30	2.4
2002	9.1	5.8	84.8	1370	1371	1	0.1
2003	8.8	5	86.8	1387	1454	67	4.8
2004	6.8	1.4	73.1	1515	1549	34	2.2
2005	6.4	1	76.1	1710	1610	100	5.8
2006	7.5	3.2	96.4	1860	1740	120	6.5
2007	9.9	7.6	144.2	2130	1972	158	7.4
2008	9.6	6.9	151.2	2230	2037	193	8.7
Mean	9.7	6.4	97.2			72.2	4.7

Table 5. Exponential power model: overall incurred errors and peak load errors.

Year	Incurred errors			Recorded Peak (MW)	Forecasted Peak (MW)	Peak load forecasting Errors	
	$\varepsilon\%$	$\varepsilon_{APE}\%$	ε_{av} (MW)			$ \varepsilon $ (MW)	$ \varepsilon $ (%)
1996	6.8	1.4	45.3	902	950	48	5.3
1997	10	3.5	70.2	971	1121	150	15.4
1998	9.2	1.2	68.6	1020	1194	174	17.1
1999	9.7	2.4	76.4	1099	1259	160	14.6
2000	9	3	73.5	1206	1277	71	5.9
2001	9.4	1.4	82.2	1225	1389	164	13.4
2002	8.1	0.3	75.2	1370	1516	146	10.7
2003	9.2	0.8	90.6	1387	1568	181	13
2004	9.5	4.1	103.2	1515	1644	129	8.5
2005	10.3	6.3	121.7	1710	1695	15	0.9
2006	11.3	8.1	145.3	1860	1792	68	3.7
2007	14.3	12.3	208.9	2130	2000	130	6.1
2008	13.6	11.4	213.8	2230	2112	118	5.3
Mean	10.3	4.6	110.8			120	9.2

lar, the forecasted peak load was 1 632 MW compared to the actual peak load of 2 230 MW. This corresponds to an absolute error of about 27%. On the other hand, the estimated peak loads using the proposed techniques results in errors in the range of 4.7% - 9.2% (see **Tables 3-5**). This means that the proposed methods outperform, to a large extent, the exponential regression. It is worth nothing that the large error observed in the forecast of the regression method can be attributed to the fact that this technique performs better when applied to monthly or yearly peak loads rather than hourly loads.

5. Conclusions and Recommendations

We have proposed three models to perform load forecasting based on multi-variable regression (linear, poly-

nomial, and exponential power). These models are generic and can be used in medium-term load forecasting for any power system. Results showed that the performance of the linear and polynomial models perform was close, when applied to the hourly loads of the Jordanian power system for different years. The exponential power model performs close to the linear model, however, due to its more complex nature; it is only applied to a time span of one year. The incurred forecasting errors for the investigated three models is about 10% while the absolute error (*APE*) ranges between 4.6% (exponential)-to 6.7% (polynomial).

Peak load forecasting results showed that the exponential model performance is far behind the performance of the linear and polynomial models. In fact, the average

error in peak load forecasting using the exponential model reaches 9.2% which is almost double the error of the other models. The average incurred error in peak load forecasting using linear model was 5.3%, which is a reasonable percentage. Results also showed that all three methods perform much better than the exponential regression method when hourly loads are used to forecast peak loads.

It was also concluded that the linear model is good and simple and suits the needs of the National Electric Power Company (NEPCO) of Jordan. The application of the linear model can also be extended for different time spans, λ , which will provide deeper insight of the load growth pattern of the Jordanian power system. It should be emphasized that the information needed by the proposed methodologies is only the hourly loads of the year. This means that weather, demographic, socio/economic and other exogenous data will not be required in the load forecasting process. This is an advantage since various pieces of information within the power utility may not be available, or may have high degree of inaccuracy.

Finally, the authors recommend that when the proposed models are adopted, they need to be tested on a collectively different time spans of hourly loads.

REFERENCES

- [1] D. Bassi and O. Olivare, "Medium Term Electric Load Forecasting Using TLFN Neural Networks," *International Journal of Computers, Communications & Control*, Vol. I, No. 2, 2006, pp. 23-32.
- [2] B. Bowerman, R. O'Connell and A. Koehler, "Forecasting, Time Series and Regression: An Applied Approach," Thomson Brooks/Cole, California, 2005.
- [3] E.A. Feinberg, D. Genethliou, "Chapter 12 Load forecasting", *Applied Mathematics for Power Systems*, pp.269-282.
<http://www.ams.sunysb.edu/~feinberg/public/lf.pdf>
- [4] D. Montgomery, L. Johnson and J. Gardiner, "Forecasting & Time Series Analysis," McGraw-Hill, New York, 1990.
- [5] J. Rothe, A. Wadhvani and S. Wadhvani, "Short Term Load Forecasting Using Multi Parameter Regression," *International Journal of Computer Science and Information Security*, Vol. 6, No. 2, 2009, pp. 303-306.
- [6] H. Hippert, C. Pedreira and R. Souza, "Neural Networks for Short-Term Load Forecasting: A Review and Evaluation," *IEEE Transactions on Power Systems*, Vol. 16, No. 1, February 2001, pp. 44-55.
[doi:10.1109/59.910780](https://doi.org/10.1109/59.910780)
- [7] H. Willis and J. Green, "Comparison Tests of Fourteen Distribution Load Forecasting Methods," *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-103, No. 6, June 1984, pp. 1190-1197.
[doi:10.1109/TPAS.1984.318448](https://doi.org/10.1109/TPAS.1984.318448)
- [8] H. Yang, C. Ming and C. Huang, "Identification of AR-MAX Model for Short Term Load Forecasting: An Evolutionary Programming Approach," *IEEE Transactions on Power Systems*, Vol. 11, No. 1, February 1996, pp. 403-408. [doi:10.1109/59.486125](https://doi.org/10.1109/59.486125)
- [9] M. Kandil, S. El-Debeiky and N. Hasanien, "Long-Term Load Forecasting for Fast Developing Utility Using a Knowledge-Based Expert System," *IEEE Transactions on Power Systems*, Vol. 17, No. 2, May 2002, pp. 491-496.
- [10] J. Rothe, A. Wadhvani and S. Wadhvani, "Short Term Load Forecasting Using Multi Parameter Regression," *International Journal of Computer Science and Information Security*, Vol. 6, No. 2, 2009
- [11] G. Janacek and L. Swift, "Time Series: Forecasting, Simulation, Applications," West Sussex: Ellis Horwood Limited, 1993.
- [12] N. Draper and H. Smith, "Applied Regression Analysis," 3rd Edition, Wiley, New York, 1998.
- [13] K. Schittkowski, "Numerical Data Fitting in Dynamical Systems - A Practical Introduction with Applications and Software," Kluwer Academic Publishers, Dordrecht, 2002.
- [14] D. Bassi and O. Olivares, "Medium Term Electric Load Forecasting Using TLFN Neural Networks," *International Journal of Computers, Communications & Control*, Vol. 1, No. 2, 2006, pp. 23-32.
- [15] S. Bengio, F. Fessant and D. Collobert, "A Connectionist System for Medium - Term Horizon Time Series Prediction," *International Workshop on Applications of Neural Networks to Telecommunications*, Stockholm, Sweden, 1995.
- [16] A. Khotanzad and A. Abaye, "ANNSTLF - A Neural Network Based Electric Load Forecasting System," *IEEE Transactions on Neural Networks*, Vol. 8, No. 4, 1997, pp. 835-846. [doi:10.1109/72.595881](https://doi.org/10.1109/72.595881)