بسم الله الرحمن الرحيم

Sudan University of Sciences &Technology

Collage of Graduate Studies

Department of Electrical Engineering

Short Term Electrical Load Forecasting Using Time Series

تنبؤات الأحمال الكهربائية قصيرة المدى بإستخدام السلاسل الزمنية

A Thesis Submitted in Partial Fulfillment for the Requirements of the Degree of M.Sc.In Electrical Engineering (Power & Machines)

Prepared by:

Salma Yousif Mohammed

Supervised by:

Dr. Mohamed Osman Hassan

October, 2017

الآية

قال الله تعالى : (قُلْ لَوْ كَانَ الْبَحْرُ مِدَادًا لِكَلِمَاتِ رَبِّي لَنَفِدَ الْبَحْرُ قَبْلَ أَنْ تَنْفَدَ كَلِمَاتُ رَبِّي وَلَوْ جِئْنَا بِمِثْلِهِ مَدَدًا)

سورة الكهف الاية (109)

DEDICATION

This thesis is dedicated to my parents

For their endless love, support and encouragement.

ACKNOWLEDGEMENT

My deepest gratitude goes to Allah who has given me strength to complete this project. Throughout this entire study, He took care of everything that would have stopped me in my tracks and strengthened me even through my most difficult times.

I wish to express my profound sense of deepest gratitude to my best friend and my first teacher my father Yousif thank you for giving me the strength to chase my dreams and ambitions.

I also would like to extend my deepest gratitude to my guide and motivator Dr. Mohammed Osman, Electrical Engineering Department, Sudan university of science and technology for his kind assistance and advice throughout this Master Project.

Last, but not least, I would like to thank the authors of various research articles and book that referred to.

ABSTRACT

Electric load forecasting plays an important role in the planning and operation of the power system. Precise load forecasting helps the electric utility to make unit commitment decisions, reduces spinning reserve capacity and schedule device maintenance plan properly. It also reduces the generation cost and increases reliability of power systems. A short-term electrical peak load forecast for Sudan grid was carried out using time series and the result was compared with actual values of peak load. Historical peak load data was collected from the daily report of operation and control of Sudanese National Grid, the hourly data covers a period from 1^{st} January to 31^{th} January 2014.

Mean absolute percentage error (MAPE) and correlation (R) were used as performance indices to test the accuracy of the forecasted load.

Result obtained from the time series using GMDH gave a correlation (R) range between 0.896 and 0.982 and a mean absolute percentage error (MAPE) range between 1.07 % and 1.72%.Results obtained show the efficacy of the GMDH forecasting.

.

مستخلص

تنبؤ الاحمال هو عنصر هام للمنظومة الكهربائية لادارة الطاقة. دقة تنبؤ الحمل يساعد في إتخاذ القرارات وخفض المفاقيد في القدرة الاحتياطية واستخدام خطة صيانة الاجهزة بشكل صحيح وفقا للجدول الزمني، كما أنه يقلل من تكلفة التوليد ويزيد من موثوقية أنظمة الطاقة .

تم تنفيذ توقعات الأحمال الكهربائية قصيرة المدى لشبكة السودان باستخدام السلاسل الزمنية وتمت مقارنة النتائج مع القيم الفعلية للأحمال الكهربائية. تم جمع بيانات الأحمال الكهربائية من التقرير اليومي للتشغيل والصيانة، وتغطي البيانات الفترة من الأول يناير إلى الواحد والثلاثون من يناير 2014 .

أستخدم متوسط نسبة الخطأ المطلق (MAPE) والارتباط (R) كمؤشرات أداء لاختبار دقة التنبؤ.

النتائج التي تم الحصول عليها من السلاسل الزمنية برنامج (GMDH) أظهرت الارتباط (R) في المدى بين 0.896 و 0.982 و متوسط نسبة الخطأ المطلق في المدى بين % 1.07 و 1.72% . النتائج التي تم الحصول عليها توضح فعالية التنبؤ بإستخدام برنامج (GMDH) .

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER ONE

INTRODUCTION

1.1 Background

A reliable and continuous supply of electrical energy is necessary for the functioning of today's complex power systems. Due to the increasing consumption and obstruction of various kinds, and the extension of existing electrical transmission networks, power systems operated closer and closer to their limits.

Thus, the chances of overloading occurrences equipment failures and blackout increases. Moreover, another problem to be faced is that the electrical energy cannot be stored whereby the electricity is only generated when needed. As a result from the electricity supply and demand fluctuating, an accurate model for electric power load forecasting is essential to the operation and planning of electricity generation in order to provide an effective and reliable operation.

Load forecasting predict the demand for electricity over the planning period of time for the utility in planning generation schedules in a power system, where it helps on making decision on when and how much electricity need to be generated. Load forecast studies are extremely important for energy suppliers, financial institutions, and other participants in electric energy generation, transmission, distribution, and markets.

Load forecasting can be divided into three major categories:

1. Long-term electric load forecasting, more than a year used to supply electric utility company management with prediction of future needs for expansion, equipment purchases, or staff hiring.

2. Medium-term forecasting, from a week to a year used for the purpose of scheduling fuel supplies and unit maintenance.

3. Short-term forecasting, from one hour to one week used to supply necessary information for the system management of day- to-day operations and unit commitment.

Short-term power load forecasting is used to provide utility company management with future information about electric load demand in order to assist them in running more economical and reliable day-to-day operations. The power load during the year followed the same the daily and weekly periods of electric load, which represents the daily and weekly cycles of human activities and behavior patterns, with some cyclical and random changes. Short term load forecasting is necessary for the control and scheduling operations of a power system and also acts as inputs to the power analysis functions such as load flow and contingency analysis. Owing to this importance, various methods have been reported, that includes time series, Auto-Regressive (AR), Auto-Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving-Average (ARIMA), and Artificial Neural Networks (ANN).

In this thesis, Time series analysis is used to obtain the short term forecast model with the historical load demand data as the input parameters.

1.2 Problem Statement

With the power systems growth and the increase in their complexity, many factors have become influential in electric power generation and consumption. The basic problem for electric utilities is to maximize the short term and long term system operation performance. To support economic growth and meet power requirements continually in the future, load forecasting has become a very important task for electric utilities. Moreover, an accurate load forecast can be helpful in developing a power supply strategy.

Short term load forecasting is essentially a time-varying signal processing problem. The prime duty of any utility is to provide reliable power to customers. Customer load demand in electric distribution systems is subject to change because human activities follow daily, weekly, and monthly cycles .Usually, the future load prediction is carried out by making use of the past load data. Due to these reason time series method of short term load forecasting will used in this thesis.

1.3 Thesis Objectives

This thesis focuses on the following objectives:

- Prediction the peak load for the next 24 hours by using GMDH Shell.
- Prediction the peak load for the next 7days (week) by using GMDH Shell.

1.4 Methodology / Approach

In this thesis, the load demand will be predicted for the next 24 hours to all network of Sudan, and for the next 7 days (week) to Khartoum grid, Aljazeera grid, north grid, East grid and all grid of Sudan. To achieve the research objectives, the data are collected with help Sudanese Electrical distribution Company, and then the output is obtained by using GMDH shell software.

1.5 Thesis Outline

In order to attain the above-stated goals, Thesis contents are broadly grouped into five chapters and they are outlined as below:

Chapter one explains the important in propose a short-term load forecasting model based on the historical data of load demand by using the Time series by using GMDH software. It also outlines objective and scope of this project.

Chapter two presents the literature review of the study, introduction of load forecasting, needs of load forecasts, types of load forecasting and method load forecasting.

Chapter three concentrates on time series for load forecasting, modeling of time series, classification of time series and time series tools.

Chapter four presents the analysis and discusses on the results obtained due to the simulation using GMDH. The results are computed and the performance is measured using MAPE and correlation.

Conclusion and directions for future research are presented in Chapter five.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Electricity is one of the ordinary life necessities, and a major driving force for economic growth and development. The unstorable nature of electricity means that the supply of electricity must be always available to satisfy the growing demand. Therefore, electricity utilities throughout the world have given a remarkable interest for forecasting electricity demand. Decision makers around the world widely use energy demand forecasting as one of the most important policy making tools[2].

If a power generation system is able to meet consumers demand at both normal and emergency conditions the system is said to be secure. If this is not the case, the system is said to be insecure. The process of specifying whether a system is secure or insecure is called security assessment [8]. The set of actions necessary to restore the secure state of a system is called security enhancement. Both security assessment and security enhancement need load prediction. Load forecasting contributes in the decision-making concerning, among others, the processes of unit commitment and security enhancement. If the forecast is inaccurate the generation will be either above or below the required demand. If the prediction is too high, extra generation units will be put into operation without real need. If the prediction is too low shortfall will take place. Correcting the second situation either by activating standby units or purchasing electric power from neighbor countries. Thus, in both situations the generation utility will pay extra cost. Because of the important role of load forecasting in power system, researchers in the last two decades have been trying and experimenting to develop new techniques to increase the accuracy and efficiency of load forecasting models [9].

2.1.1 Business Needs of Load Forecasts

In today's world, load forecasting is an important process in most utilities with the applications spread across several departments, such as planning department, operations department, trading department, etc. The business needs of the utilities can be summarized, but not limited to, the following:

1) Energy purchasing. Whether a utility purchases its own energy supplies from the market place, or outsources this function to other parties, load forecasts are essential for purchasing energy. The utilities can perform bi-lateral purchases and asset commitment in the long term, e.g., 10 years ahead. They can also do hedging and block purchases one month to 3 years ahead, and adjust (buy or sell) the energy purchase in the day-ahead market.

2) Transmission and distribution (T&D) planning. The utilities need to properly maintain and upgrade the system to satisfy the growth of demand in the service territory and improve the reliability. And sometimes the utilities need to hedge the real estate to place the substations in the future. The planning decisions heavily rely on the forecasts, known as spatial load forecasts, that contain when, where, and how much the load as well as the number of customers will grow.

3) Operations and maintenance. In daily operations, load patterns obtained during the load forecasting process guide the system operators

to make switching and loading decisions, and schedule maintenance outages.

4) Demand side management (DSM). Although lots of DSM activities are belong to daily operations, it is worthwhile to separate DSM from the operations category due to its importance in this smart-grid world. A load forecast can support the decisions in load control and voltage reduction. On the other hand, through the studies performed during load forecasting, utilities can perform long term planning according to the characteristics of the end-use behavior of certain customers.

5) Financial planning. The load forecasts can also help the executives of the utilities project medium and long term revenues, make decisions during acquisitions, approve or disapprove project budgets, plan human resources and technologies, etc[3].

2.1.2 Important factors for forecasts

The time factors include the time of the year, the day of the week, and the hour of the day. There are important differences in load between weekdays and weekends. The load on different weekdays also can behave differently. This is particularly true during the summer time. Holidays are more difficult to forecast than non-holidays because of their relative infrequent occurrence [4].

Weather conditions influence the load. In fact, forecasted weather parameters are the most important factors in short-term load forecasts. Various weather variables could be considered for load forecasting like temperature, humidity and wind speed [5].

Most electric utilities serve customers of different types such as residential, commercial, and industrial. The electric usage pattern is different for customers that belong to different classes but is somewhat alike for customers within each class. Therefore, most utilities distinguish load behavior on a class-by-class basis [4].

Random Events: Like Start or stop of large loads, Sports and TV events. [5].

2.1.3 Classification of Load Forecasts

There is no single forecast that can satisfy all of the needs of utilities. A common practice is to use different forecasts for different purposes. The classification of various forecasts is not only depending upon the business needs of utilities, but also the availability of the crucial elements that affect the energy consumption: weather (or climate in the long periods) and human activities. Weather refers to the present condition of the meteorological elements, such as temperature, humidity, wind and rainfall. . Load forecasting can be divided into three categories:

- Short-term load forecasting (STLF) that covers a period of one hour to one month used to supply necessary information for the system management of day- to-day operations and unit commitment.
- Medium-term load forecasting (MTLF) covers a period of one month up to one year used for the purpose of scheduling fuel supplies and unit maintenance.
- Long–term load forecasting (LTLF) predicts the requirements of energy for more than one year used to supply electric utility company management with prediction of future needs for expansion, equipment purchases.

2.1.4 FORECASTING ERRORS

Unfortunately, all forecasting situations involve some degree of uncertainty which makes the errors unavoidable [18].

The forecast error for a particular forecast $\widehat{X}_t$ with respect to actual value X_t

$$e_t = X_t - \widehat{X}_t \tag{2.1}$$

To avoid the offset of positive with negative errors, we need to use the absolute deviations.

$$|e_t| = \left|X_t - \widehat{X}_t\right| \tag{2.2}$$

Hence, we can define a measure known as the mean absolute deviation (MAD) as follows:

$$\text{MAD} = \frac{\sum_{t=1}^{n} |e_t|}{n} = \frac{\sum_{t=1}^{n} \left|X_t - \widehat{X}_t\right|}{n} \tag{2.3}$$

Another method is to use the mean-squared error (MSE) defined as follows:

$$\text{MSE} = \frac{\sum_{t=1}^{n} e_t^2}{n} = \frac{\sum_{t=1}^{n} (X_t - \widehat{X}_t)^2}{n} \tag{2.4}$$

2.2 Methods of Load Forecasting

2.2.1 Medium- and long-term load forecasting methods

The end-use modeling, econometric modeling, and their combinations are the most often used methods for medium- and long-term load forecasting. Descriptions of appliances used by customers, the sizes of the houses, the age of equipment, technology changes, customer behavior, and population dynamics are usually included in the statistical and simulation models based on the so-called end-use approach. In addition, economic factors such as per capita incomes, employment levels, and electricity prices are included in econometric models. These models are often used in combination with the end-use approach. Long-term forecasts include the forecasts on the population changes, economic development, industrial construction, and technology development.

End-use models

The end-use approach directly estimates energy consumption by using extensive information on end use and end users, such as appliances, the customer use, their age, sizes of houses, and so on. Statistical information about customers along with dynamics of change is the basis for the forecast.

End-use models focus on the various uses of electricity in the residential, commercial, and industrial sector. These models are based on the principle that electricity demand is derived from customer’s demand for light, cooling, heating, refrigeration, etc. Thus end-use models explain energy demand as a function of the number of appliances in the market [7]. Ideally this approach is very accurate. However, it is sensitive to the amount and quality of end-use data. For example, in this method the distribution of equipment age is important for particular types of appliances. End-use forecast requires less historical data but more information about customers and their equipment.

Econometric models

The econometric approach combines economic theory and statistical techniques for forecasting electricity demand. The approach estimates the relationships between energy consumption (dependent variables) and factors influencing consumption. The relationships are estimated by the least-squares method or time series methods.

One of the options in this framework is to aggregate the econometric approach, when consumption in different sectors (residential, comer coal, industrial, etc.) is calculated as a function of weather, economic and other variables, and then estimates are assembled using recent historical data. Integration of the econometric approach into the end-use approach introduces behavioral components into the end-use equations.

Statistical model-based learning

The end-use and econometric methods require a large amount of information relevant to appliances, customers, economics, etc. Their application is complicated and requires human participation. In addition such information is often not available regarding particular customers and a utility keeps and supports a profile of an "average" customer or average customers for different type of customers. The problem arises if the utility wants to conduct next-year forecasts for sub-areas, which are often called load pockets. In order to simplify the medium-term forecasts, make them more accurate and avoid the use of the unavailable information. The focus of the study was the summer data. We compared several load models and came to the conclusion that the following multiplicative model is the most accurate

$$L\,(t) = F\,(d\,(t), h\,(t)) \cdot f\,(w(t)) + R(t) \tag{2.5}$$

Where

L(t) is the actual load at time t,

d (t) is the day of the week,

h (t) is the hour of the day,

f(d, h) is the daily and hourly component,

w(t) is the weather data that include the temperature and humidity,

f(w) is the weather factor, and R(t) is a random error.

In fact, w (t) is a vector that consists of the current and lagged weather variables. This reflects the fact that electric load depends not only on the current weather conditions but also on the weather during the previous hours and days. In particular, the well-known effect of the so-called heat waves is that the use of air conditioners increases when the hot weather continues for several days.

The described methods can be applied to both medium- and long term forecasting. However, the long-term forecasts should incorporate economic and population dynamic forecasts as input parameters [6].

2.2.2 Short term load forecasting methods

In Short Term Load Forecasting there are several target values which can be the forecasting goal. Many tasks, in power generation industry, such as unit commitment, security assessment and enhancement of security depend on the near future load prediction including daily peak load[10]. Peak load forecasting inaccuracy has a negative impact on the economics of these utilities. For these reasons, many researchers in the last 20 years have tackled this area to devise more accurate and efficient techniques of load prediction [11].

The main resource of any power generation utility is the generation units. Efficient management of these units means running them in minimum cost to satisfy the requirements of consumers. To achieve this purpose starting–up and shutting down the generating units should be performed according to schedule. The scheduling process is also known as unit commitment. As load demand varies from hour to hour and from day to day, and starting-up a unit needs time, it is necessary to have a demand prediction on hourly basis. This prediction should be provided one hour, one day, or one month ahead [10].

There is variety of statistical and artificial intelligence techniques have been developed for short-term load forecasting for example similar-day approach, regression methods, time series, expert system and fuzzy logic. Each of the methods will be discuss in detail in the next session.

Similar-day approach

This approach is based on searching historical data for days within one, two, or three years with similar characteristics to the forecast day. Similar characteristics include weather, day of the week, and the date. The load of a similar day is considered as a forecast.

Instead of a single similar day load, the forecast can be a linear combination or regression procedure that can include several similar days. The trend coefficients can be used for similar days in the previous years.

Fuzzy Logic

Fuzzy logic is a logic having many values, approximate reasoning and have a vague boundary. The variables in fuzzy logic system may have any value in between 0 and 1 and hence this type of logic system is able to address the values of the variables (called linguistic variables) those lie between completely truths and completely false. Each linguistic variable is described by a membership function which has a certain degree of membership at a particular instance. The human knowledge is incorporated in fuzzy rules. The fuzzy inference system formulates suitable rules and based on these rules the decisions are made. This whole process of decision making is mainly the combination of concepts of fuzzy set theory, fuzzy IF-THEN rules and fuzzy reasoning. The fuzzy inference system makes use of the IF-THEN statements and with the help of connectors present (such as OR and AND), necessary decision rules are constructed. The fuzzy rule base is the part responsible for storing all the rules of the system and hence it can also be called as the knowledge base of the fuzzy system. Fuzzy inference system is responsible for necessary decision making for producing a required output [6].

The fuzzy control systems are rule-based systems in which a set of fuzzy rules represent a control decision mechanism for adjusting the effects of certain system stimuli. The rule base reflects the human expert knowledge, expressed as linguistic variables, while the membership functions represent expert interpretation of those variables.

Fuzzy logic is a generalization of the usual Boolean logic used for digital circuit design. An input under Boolean logic takes on a truth value of “0” or “1”. Under fuzzy logic an input has associated with it a certain qualitative ranges. For instance a transformer load may be “low”, “medium” and “high”. Fuzzy logic allows one to (logically) deduce outputs from fuzzy inputs. In this sense fuzzy logic is one of a number of techniques for mapping inputs to outputs (i.e. curve fitting). Among the advantages of fuzzy logic are the absence of a need for a mathematical model mapping inputs to outputs and the absence of a need for precise (or even noise free) inputs. With such generic conditioning rules, properly designed fuzzy logic systems can be very robust when used for forecasting. Of course in many situations an exact output is needed. After the logical processing of fuzzy inputs, a "defuzzification" process can be used to produce such precise outputs describe applications of fuzzy logic to electric load forecasting [13].

Neural Network

Artificial Neural networks ANN is not something new; they have been used at a large scale in the electric power industry mainly for load forecasting. The goal in the power industry is to predict the consumers demand for short-term. The concept is based on computing systems that are able to learn through experience by recognizing patterns existing within a data set. ANN experience is acquired by a process called training. The feed-forward network consists of a collection of processing units called neurons (nods). The neurons are arranged in layers. These layers are a single input layer figure 2.3, one or more hidden layers and a single output layer. The input layer contains a Number of neurons equal to the number of input variables; the same applies to the output layer out with the output variable. Both the number of hidden layers and the number of neurons in each layer are experimentally determined [12]. Each layer can then have several neurons; these neurons are interconnected to the neurons in the next layer by means of information channels with different strengths called weights. Each neuron can have multiple inputs. The inputs to a neuron could be from external stimuli or could be from the output of other neurons.

The output from a neuron could be an input to many other neurons in a network. Signals flow into the input layer, pass though the hidden layers, and arrive at the output layer. With the exception of the input layer, each neuron receives signals from the neurons of the previous layer linearly weighted by the interconnect values between neurons. The neuron then produces its output signal by passing the summed signal through a non-linear function. Training a network is an iterative process. It means adjusting the weights using some learning algorithm. Most ANN is trained using the back propagation algorithm. Initially, training data is fed to the network via the input layer. This data is in the form of input/output pairs of vectors. The network computes an output for each input vector. The sum squared error between the calculated and actual outputs of neural network over the training data is propagated backward to the input layer[15]. This process takes place by the end of each training iteration. A gradient descent algorithm is used to modify the weights at the end of each epoch and the process is repeated until sum squared error is less than a preset value or a specified number of epochs are reached. Once the neural network is trained, it produces very fast output for a given set of input data [16].

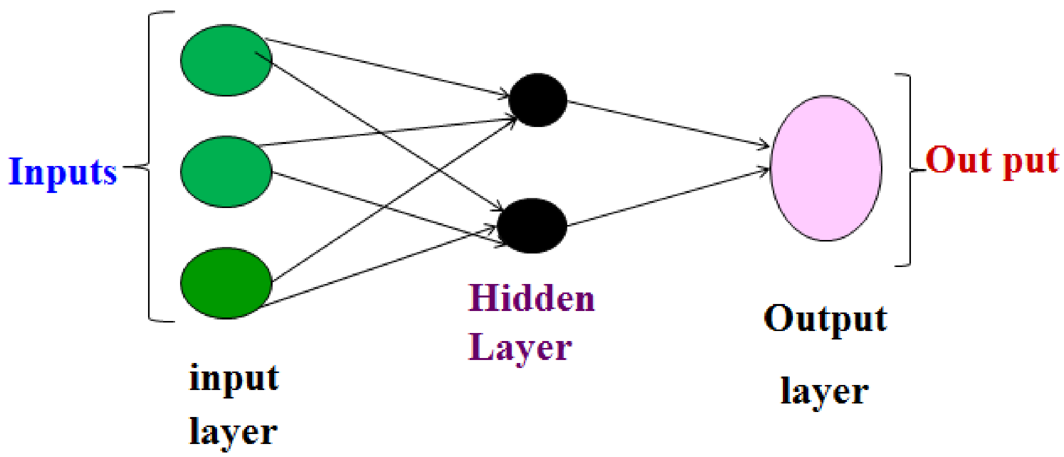

Figure (2.1): Feed forward multi-layer ANN

Expert system

Rule based forecasting makes use of rules, which are often heuristic in nature, to do accurate forecasting. Expert systems, incorporates rules and procedures used by human experts in the field of interest into software that is then able to automatically make forecasts without human assistance.

Expert system use began in the 1960’s for such applications as geological prospecting and computer design. Expert systems work best when a human expert is available to work with software developers for a considerable amount of time in imparting the expert’s knowledge to the expert system software. Also, an expert’s knowledge must be appropriate for codification into software rules (i.e. the expert must be able to explain his/her decision process to programmers). An expert system may codify up to hundreds or thousands of production rules.

Proposed knowledge based expert system for the short term load forecasting of the Taiwan power system. Operator’s knowledge and the hourly observations of system load over the past five years were employed to establish eleven day types. Weather parameters were also considered. The developed algorithm performed better compared to the conventional Box-Jenkins method developed a site-independent technique for short-term load forecasting. Knowledge about the load and the factors affecting it are extracted and represented in a parameterized rule base. This rule base is complemented by a parameter database that varies from site to site. The technique was tested in several sites in the United States with low forecasting errors [14].

Time series

It has been observed that unique patterns of energy and demand pertaining to fast-growing areas are difficult to analyses and predict by direct application of time series methods. However, these methods appear to be among the most popular approaches that have been applied and are still being applied to ST L F. Using the time series approach, a model is a first developed based on the previous data, then future load is predicted based on this model. The remainder of this section discusses some of the time series models used for load forecasting.

Auto-Regressive (AR):

In this model, the current value X_t of the time series is expressed linearly in terms of its previous valuesX_{t-1},X_{t-2}, … and a white noise series {Ɛt} with zero mean and variance σ2

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} \qquad (2.6)$$

Moving Average (MA):

In the MA process, the current value of the time series X_t is expressed linearly in terms of current and previous values of a white noise series ($\mathcal{E}_t$). This noise series is constructed from the forecast errors or residuals when demand observations become available. The order of this process depends on the oldest noise value at which X_t is regressed on. For an MA of order q (i.e., MA (q)), this model can be written as:

$$X_t = \eta_t - \theta_1 \eta_{t-1} - \theta_2 \eta_{t-2} - \cdots - \theta_q \eta_{t-q} \qquad (2.7)$$

Auto-Regressive Moving Average (ARMA): Box-Jenkins

If we combine the MA and AR models together, we can present a broader class of model, that is, ARMA model as follows:

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \eta_t + \theta_1 \eta_{t-1} + \theta_2 \eta_{t-2} + \cdots + \theta_q \eta_{t-q} \quad (2.8)$$

Where φ_i and $\theta_j$are called the ARMA parameters, respectively. It is an ARMA (p,q) model.

A methodology for ARMA models was developed largely by Box and Jenkins (1976), and so the models are often called Box-Jenkins models. Recall that an MA process is equivalent to an infinite AR process. Similarly, one can show that an ARMA process can be expressed as an infinite AR process or an infinite MA process.

Auto-Regressive Integrated Moving-Average (ARIMA): Box-Jenkins

The time series defined previously as an AR, MA, or as an ARMA process is called a stationary process. This means that the mean of the series of any of these processes and the covariance among its observations do not change with time. Unfortunately, this is not often true in power demand. But previous knowledge is definitely useful in that the non-stationary series can be transformed into a stationary one with some tricks. This can be achieved, for the time series that are non-stationary in the mean, by a differencing process. By introducing the ∇ operator, a differenced time series of order 1 can be written as:

$$\nabla X_t = X_t - X_{t-1} = (1 - B)x_t \quad (2.9)$$

Consequently, an order d differenced time series is written as:

$$\nabla^d X_t = (1 - B)^d X_t \quad (2.10)$$

Autoregressive integrated moving-average (AR IMA) model:

If the process is non-stationary, then transformation of the series to the stationary form has to be done first. This transformation can be performed by the differencing process.

For a series that needs to be differenced d times and has orders p and q for the AR and MA components, i .e. ARIMA (p,d,q), the model is written as:

$$\varphi(B)\nabla^{d}X_{t} = \theta(B)\eta_{t} \tag{2.11}$$

Others (ARMAX, ARIMAX, SARMAX, NARMA)

Auto-Regressive Moving Average and ARIMA use only the time and demand as input parameters. Since demand generally depends on the weather and the time of day, exogenous variables (X) sometimes can be included to give an ARMAX and ARIMAX model. There are other versions of ARIMA, such as Piecewise ARMA (PARMA), Seasonal ARMA (SARMAX), etc. The concept is to identify the time series model as a single expression instead of preprocessing and post-processing.

CHAPTER THREE

TIME SERIES FOR LOAD FORECASTING

3.1 Introduction

Amount of digital information in the world today is unimaginably enormous and is growing even more rapidly every day. Also, large volumes of data are being collected in almost every (scientific) field, usually in the form of time-stamped data and time series. Time series are collections of events or observations, predominantly numeric in nature, sequentially recorded on a defined regular or irregular time basis. Time series data are being generated at an unprecedented speed and volume in a wide range of applications in almost every domain. For example, daily fluctuations of the stock market, traces produced by a computer cluster, medical and biological experimental observations, readings obtained from sensor networks, position updates of moving objects in location-based services, etc. Are all represented in time series .Time series are thus becoming increasingly important in nearly every organization and industry, including banking, finance, telecommunication, medical sciences and transportation. Banking institutions, for instance, rely on the analysis of time series for forecasting economic indices, elaborating financial market models, and registering international trade operations. In medical sciences, analysis of time series nowadays covers a wide range of real-life problems including gene expression analysis and medical surveillance. Notably, the availability of these enormous time series data has brought about huge advantages in many areas. There is no doubt that proper management and analysis of time series has the potential to become a driving force for innovation and decision making. If managed well, time series can be used to unlock new sources of economic value, provide fresh insights into science and hold governments to account. However, the easy management and analysis of these data still pose a great challenge to scientists worldwide. Despite, the abundance of tools to capture and process all this information, issues like: ensuring data security, easy management of data, analysis of huge datasets, optimizing response time for data retrieval, protecting privacy etc. are still being faced as information is shared widely around the world.

3.2 Theoretical Background on Time Series

Today's organizations in both the public and the private sectors are collecting large volume of data. Data is what businesses have been demanding for years in order to perform better analysis, make better decisions, and consequently become more competitive. Evidently, the possibility of such analysis and decision making depends most times on how data is collected and measured. Apparently, time scales provide the simplest and most sensible way of measuring collected data; this explains why time series are common in the world of business today. Currently, time series underlie countless business activities. Businesses are thus often interested in analyzing and forecasting time series variables. Time series are used in statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and largely in any domain of applied science and engineering which involves temporal measurements .

3.3 What is a Time series?

A time series is a sequence of data points, typically consisting of successive measurements or observations on quantifiable variable(s), made over a time interval . Usually the observations are chronological and taken at regular intervals (hours, days, months, years), but the sampling could also be irregular.

Typical examples of time series also include historical data on sales, inventory, customer counts, interest rates, costs, etc. Time series data are also often seen naturally in many application areas including:

• Economics - e.g. monthly data for unemployment, hospital admissions, etc.

• Finance - e.g. daily exchange rate, share prices, etc.

• Environmental - e.g. daily rainfall, air quality readings.

• Medicine - e.g. ECG brain wave activity every 2−8 secs.

Time series can be categorized into two major classes namely: univariate or multivariate. A univariate time series is a sequence of measurements of the same variable collected over time. Most often, the measurements are sequence of events made at regular time intervals. However, when a time series involves more than one variable, it is said to be multivariate. Most economic and financial information is structured in the form of multivariate time series. Multivariate time series can be further categorized into homogenous and heterogeneous multivariate time series, based on the relationships between the measured variables. If a variable X is useful to predict future values of another variable Y, the multivariate time series is said to be homogeneous, else it is heterogeneous. In homogenous multivariate time series, changes in one element in the observations vector of one variable imply corresponding changes in other variables that belong to the phenomenon under study.

Time series can be further classified based on their regularity and seasonality. Classifying on the basis of regularity, time series can be: regular or irregular. These are defined by the duration between the timestamps of their elements. All regular time series have a predictable number of units between them, whereas irregular time series do not.

When classified based on seasonality, a time series can be seasonal or non-seasonal. When a repetitive pattern is observed over some time horizon, the series is said to have seasonal behavior. Seasonal effects are usually associated with calendar or climatic changes. Seasonal variation is frequently tied to yearly cycles. Figure3.1 shows a classification of time series.

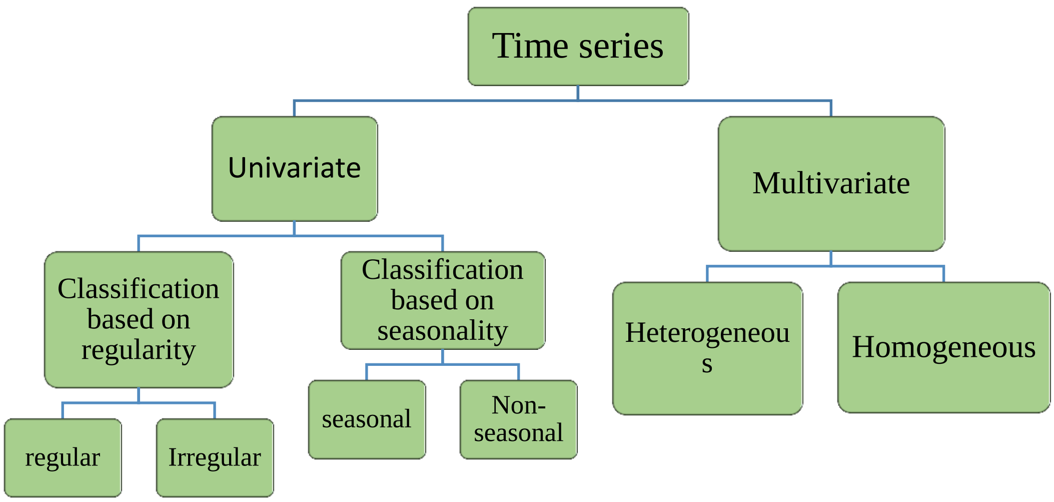

Figure 3.1: classification of time series

The following is a brief definition of the commonly used terminologies in describing time series.

Stationary Data: This describes a time series variable which exhibits no significant upward or downward trend over time.

Non-stationary Data: A non-stationary time series data is a data with variable exhibiting a significant upward or downward trend over time.

Seasonal Data: This describes a time series variable exhibiting repeating patterns at regular intervals over time.

Time series analysis: Time series analysis is the process of using statistical techniques to model and explain a time-dependent series of data points. This involves methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series analysis refers to problems in which observations which are collected at regular time intervals are correlated.

Time series forecasting: Time series forecasting is the process of using a model to generate predictions (forecasts) for future events based on known past events. The goal of forecasting is to project the underlying trend or pattern of the time series into the future as the most likely values for the data. Forecasting is the combining of knowledge from the past and future expectations with an estimated model to produce likely outcomes for the future. It enables more accurate predictions of the future to be made, reducing the uncertainty inherent in the decision-making process.

Regression analysis: Regression analysis is used essentially in such a way to test theories that the current values of one or more independent time series have some influences on the current value of another time series.

3.4 Time Series Properties

Time series have interesting features and properties that differentiate them from other data types. When compared to other data types, they behave in a different way. The following is a description of the most relevant properties of time series.

1. Time series data has a natural temporal ordering. They are generally written in a predefined order or some aggregated result based on the need of the user. This means access to data is usually alongside the time dimension, as, for instance, in the retrieval of all observations over a range of consecutive dates. Consequently, storage, retrieval, and update of time series data are not independent of each other. This differs from typical data mining/machine learning applications where each data point is an independent example of the concept to be learned.

2. The value of a time series in a time period is often affected by the values of variables in preceding periods, thus making the order in which the data occurs in the spreadsheet very important. In essence, time series data are not commonly altered. A typical example of time series data is the daily reading of average wind speed. The data are recorded in intervals which are predefined and are normally not modified once they have been recorded. The data may be aggregated for further analysis but the aggregation is linked to a predefined granularity and sequential order.

3. Time series data are usually manipulated as one single object i.e. As a collection of data. This is because the order in which the data occurs in time series is very important because, unlike other data

types, the ordering often represents the dependencies between the collected data. Thus, changing the order could change the meaning of the data. Consequently, manipulation of time series data puts more emphasis on aggregation operations on collections of data rather than on an individual data item .

4. Time series have a header i.e. A general description. The header contains all the metadata about the time series. Metadata can be information which allows the time series to be self-describing. The metadata can be information such as name, title, source, type of value, and type of time series. But more importantly is the date-time field which defines the dataset as a time series.

5. Data in time series are not necessarily identically distributed but they are dependent on preceding values. A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart.

6. In time series analysis, the past behavior of a variable is analyzed in order to predict its future behavior. By observing the different past states of a variable, the future states can be predicted through forecasting.

3.5 Time Series Analysis

Time series data occurrences are becoming extremely valuable to the operations and development of modern organizations. Financial institutions, for example, rely on analysis of time series for forecasting economic conditions, developing and using complicated financial decision support models, and conducting international financial transactions. Likewise, public and private institutions are using time series data to manage and project the loads on their networks. More and more time series are used in this type of investigation and hundreds of thousands of time series that contain valuable economic and financial information are nowadays available both on and off-line . Thus, an understanding of the standard practices for time series analysis is appropriate for the reader. This section expatiates on the processes, practices and objectives of time series analysis. It also describes the uses and relevance of time series models as well as categorizes them. Time series analysis accounts for the fact that data points taken over time may have an internal structure (such as autocorrelation, trend or seasonal variation) that should be accounted for. As defined earlier, time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. It involves the use of techniques for drawing inferences from time series data. Note however that one other main purpose for analyzing time series is forecasting.

Forecasting is the application of a model to predict future values based on previously observed time series values. In order to perform time series analysis, it is essential to set up hypothetical probability representation of the data. Such representation is called the Model. After the model has been appropriately determined, it is then possible to estimate parameters, check for goodness of fit to the data, and possibly to use the fitted model to enhance the understanding of the mechanism generating the series .

Once a satisfactory model has been developed, it may be used in a variety of ways depending on the particular field of application. The model may be used simply to provide a compact description of the data.

3.5.1 Basic Objectives of Time series Analysis

The major approach to time series analysis is usually to determine a model that describes the pattern of the time series. Uses for such a model are:

- To describe the important features of the time series pattern.

To explain how the past affects the future or how two time series can "interact".

- To forecast future values of the series.
- To possibly serve as a control standard for a variable that measures the quality of product in some manufacturing situations.

The goal of building a time series model is the same as the goal for other types of predictive models which is to create a model such that the error between the predicted value of the target variable and the actual value is as small as possible. The primary difference between time series models and other types of models is that lag values of the target variable are used as predictor variables, whereas traditional models use other variables as predictors, and the concept of a lag value doesn't apply because the observations don't represent a chronological sequence.

Thus, the aim of time series analysis is to describe and summarize time series data, determine most suitable models, and make forecasts.

3.5.2 Time series Models

Time series models are used to describe the underlying data-generating process of a time series. The usual process to time series analysis and forecasting is:

1. Preprocess data for analysis. The preprocessing stage may involve some initial analysis steps e.g. plotting the data, determining time series characteristics such as trends, seasonality etc.

2. Determine suitable model for the preprocessed time series

3. Apply/fit model to time series data. Additionally, after fitting the time series model to the time series data, the fitted model can be used to determine departures (outliers) from the (assumed) data-generating process or forecast function components (future trend, seasonal or cycle estimates).

4. Perform forecasting and prediction of future values of time series.

Models for time series data can have many forms and represent different stochastic processes. There are three broad classes of the models: the autoregressive (AR) models, the integrated (I) models, and the moving average (MA) models. These three classes depend linearly on previous data points. Combinations of these ideas produce autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models. The autoregressive fractionally integrated moving average (ARFIMA) model generalizes the former three. Extensions of these classes to deal with vector-valued data are available under the heading of multivariate time series models and sometimes the preceding acronyms are extended by including an initial "V" for "vector", as in VAR for vector auto-regression. An additional set of extensions of these models is available for use where the observed time series is driven by some "forcing" time series (which may not have a causal effect on the observed series): the distinction from the multivariate case is that the forcing series may be deterministic or under the experimenter's control. For these models, the acronyms are extended with a final "X" for "exogenous".

3.5.3 Classification of Time series Models

Two different types of models are generally used for a time series :

Multiplicative and Additive models.

Multiplicative Model: $$Y(t) = T(t) \times S(t) \times C(t) \times I(t) \quad (3.1)$$

Additive Model: $$Y(t) = T(t) + S(t) + C(t) + I(t) \quad (3.2)$$

Here $Y(t)$ is the observation and $T(t)$, $S(t)$, $C(t)$ and $I(t)$ are respectively the trend, seasonal, cyclical and irregular variation at time t. Multiplicative model is based on the assumption that the four components of a time series are not necessarily independent and they can affect one another; whereas in the additive model it is assumed that the four components are independent of each other .

Time series models are also classified based on their suitability for analyzing different types of time series data. Thus, models which are considered suitable for stationary time series are referred to as stationary time series models. This thesis classifies time series under the following heading (based on the prevalent classification in many literatures on time series):

- Stationary univariate time series models
- Non-stationary univariate time series models
- Stationary multivariate time series models
- Non-stationary multivariate time series models
- Structural Change and Nonlinear Models

3.5.4 A General Approach to Time Series Modeling

In this section, a review of the general approaches to time series modeling is made. This also includes some important underlying characteristics to consider for time series modeling based on. Below is a step-wise overview of the way in which time series modeling can be carried out.

• Plot the time series and examine the main features of the graph, checking in particular whether there is:

a. A trend, i.e. on average, do the measurements tend to increase (or decrease) over time?

b. A seasonal component, i.e. is there is a regularly repeating pattern of highs and lows related to calendar time such as seasons, quarters, months, days of the week, and so on?

c. Any apparent sharp changes in behavior,

d. Any outlying observations. In regression, outliers are far away from the fitted line. With time series data, outliers are far away from the other data

e. A long-run cycle or period unrelated to seasonality factors

f. A constant variance overtime, or whether the variance is non-constant.

• Remove the trend and seasonal components to get stationary residuals. To achieve this goal it may sometimes be necessary to apply a preliminary transformation to the data. For example, if the magnitude of the fluctuations appears to grow roughly linearly with the level of the series, and then the series will have fluctuations of more constant magnitude.

• Choose a model to fit the residuals, making use of various sample statistics including the sample autocorrelation function briefly mentioned earlier in this section.

• Forecasting will be achieved by forecasting the residuals and then inverting the transformations described above to arrive at forecasts of the original series $\{X_t\}$.

It is often not a direct process to choose the model to fit a time series and analyze it, since there are usually many time series analysis and forecasting techniques, each with different characteristics. It is thus usually difficult to know which technique will be best for a particular data set. It is customary to try out several different techniques and select the one that seems to work best.

3.5.5 Time series Tools

The increase in the amount of time series data generated daily in both public and private sectors has necessitated the need for more efficient processes, methods and tools for analyzing the huge datasets. While much progress has been made in the development of time series tools, analysis and forecasting of time series data are still among the most important problems that analysts face across many fields today. These problems cut across many fields ranging from the natural sciences to finance, production operations and economics.

As a result, there is a widespread need for efficient time series software to manage and analyze these data in efficient ways. Many software applications have thus evolved in the recent years to meet up the challenge of time series analysis and forecasting. There is some of time series software:

1. OpenEpi
2. MATLAB
3. SAS / Econometrics and Time series Software (ETS)
4. Microsoft Excel
5. GMDH Shell
6. R Language

The GMDH Shell was used in this thesis.

3.6 About GMDH

GMDH Shell Group Method of Data Handling (GMDH) Shell is a forecasting-software that enables users of all types to easily and accurately forecast their data. It's developed by GMDH LLC - a privately held company founded in 2009 with an idea to build the best forecasting software. GMDH is a state of the art predictive modeling technology with accuracy and reliability proven by over 40 years of scientific research. GMDH Shell is powerful for forecasting time series for small businesses, traders and scientists. The tool provides an easy-to-use way to accurately forecast time series, create classifiers and regression models. Based on artificial neural networks, it allows users to easily create predictive models, as well as preprocess data with simple point-and-click interface. (GMDH Shell LLC, 2013).

3.7 Supports for Time Series Analysis

GMDH Shell provides good analysis and forecasting, it also features comprehensive data-manager tool for quick entering of input data, rich visualization capabilities and templates ready for instant analysis of time series data. GMDH shell also has a user-friendly interface and easy to use wizards for performing data imports, analysis and visualization of time series. The interface of the program isn't overloaded with excessive details, so even a low experienced user can quickly begin using it. However, users may get confused while performing more complicated analysis (GMDH Shell LLC., 2013).

One the other hand, the tool is slightly expensive for students and small scale businesses. It is also slow for huge data sets. The tool responds slowly for datasets more than 5,000 records. For example, a dataset of about 10,000 rows is analyzed in 2 hours.

Overall, GMDH Shell provides about the most user-friendly interface and one of the most powerful end-user oriented time series analysis software on the market. With it, time series analysis and forecasting can be easily done by end users with little data analysis skills.

CHAPTER FOUR

SIMULATION RESULTS AND DISCUSSIONS

4.1 Background

GMDH is a learning machine based on the principle of heuristic self-organizing proposed by Ivakhnenko in the 1960s. It is an evolutionary computation technique, in which a series of operations of seeding, rearing, crossbreeding, selection and rejection of seeds correspond to the determination of the input variables, structure and parameters of model, and selection of model by principle of termination.

The classical GMDH algorithm can be represented as a set of neurons in which different pairs of them in each layer are connected through a quadratic polynomial and thus produce new neurons in the next layer.

Such representation can be used in modeling to map inputs to outputs. The formal definition of the identification problem is to find a function, $\hat{F}$, that can be approximately used instead of the actual one f . In order to predict output $\hat{Y}$ for a as close given input Vector $X = (X_1, X_2, X_3, \ldots. X_n)$ as possible to its actual output y. Therefore, given number of observations (M) of multi-input, single output data pairs so that:

$$Y_1 = f\,(X_{i1}, X_{i2}, X_{i3}, \ldots. X_{in}) \;,(i=1,2,3,\ldots ,M) \qquad (4.1)$$

It is now possible to train a GMDH to predict the output values $\hat{Y}_i$ for any given input vector

$= (X_{i1}, X_{i2}, X_{i3}, \ldots . X_{in})$ that is,

$$\widehat{Y}_1 = \widehat{F}(X_{i1}, X_{i2}, X_{i3}, \ldots . X_{in}) ,(i=1,2,3,\ldots ,M) \quad (4.2)$$

In order to determine a GMDH , the square of the differences between the actual output and the predicted one is minimized, that is

$$\sum_{i=1}^{M} [\widehat{Y}_1 - Y_1]^2 \longrightarrow \text{Min} \quad (4.3)$$

The general connection between the inputs and the output variables can be expressed by a complicated discrete form of the Volterra functional series in the form of:

$$Y=a_o+\sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\, x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ij}\, x_i x_j x_k + .. \quad (4.4)$$

4.2 General Description of Sudanese national grid and Calculation Methodology

The Sudanese national grid consists of four grids Khartoum grid, Aljazeera grid, northern state grid and the eastern state grid.

In this thesis we will forecast:

- The peak load for the all Sudan grid next 24 hours (day).
- The peak load for every grid mentioned above for the next week.

The hourly and the daily peak load data was collected from the daily report of operation and control of network. The hourly data covers a period from 1^{st} January to 31^{th} January 2014 and it covers all Sudanese grid while the daily data covers a period from 1^{st} January to 31^{th} January 2014 and it covers Khartoum grid, Aljazeera grid, northern grid and eastern grid.

The hourly and daily data was compiled using MS Excel spread sheet. The excel file was imported to the modeling environment in

CSV/XLS/XLSX format which was then imported into the time series tool to forecast the peak load.

The results can be served as the criteria to evaluate the prediction performance relative to power load value including mean absolute percentage error (MAPE), the prediction performance is better when the loss function value is smaller. The loss function is expressed as follows:

$$\mathrm{MAPE} = \frac{1}{\mathrm{T}} \sum_{\mathrm{t}=1}^{\mathrm{T}} \left| \frac{\hat{\mathrm{y}}_{\mathrm{t}} - \mathrm{y}_{\mathrm{t}}}{\mathrm{y}_{\mathrm{t}}} \right| \times 100\% \tag{4.5}$$

Where T demonstrates the total time, y_{t} and $\hat{\mathrm{y}}_{1}$represents the original value and prediction value when time is t, respectively.

The analysis of results which carried out by GMDH software is discussed in next section.

4.3 Simulation result of next 24 hours for all Sudan grid

A figure (4.1) shows the actual data, the model fit and the Predicted peak load r respectively. The actual data is the historical data that was used for training and testing and it is grey in colour. The model fit is a model value fitted to the data and is blue in colour. The predicted value is the value forecasted by the GMDH-time series model and is red in colour.

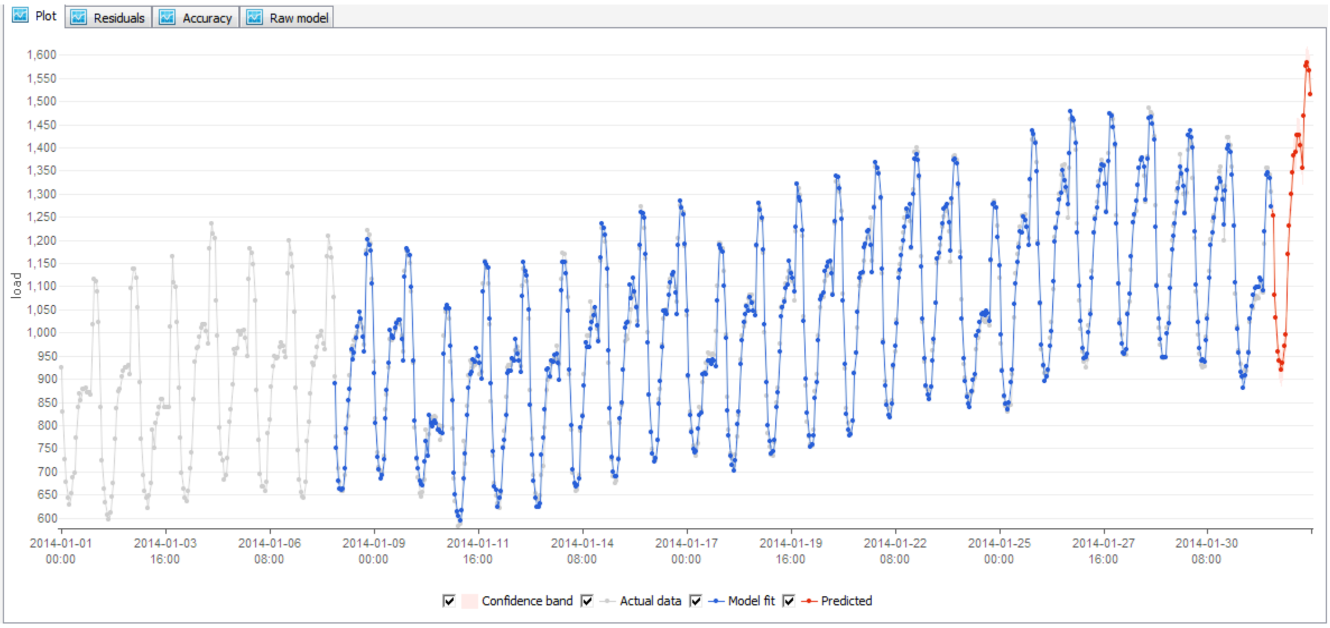

Figure 4.1: The actual peak load, the model fit and the predicted data for 1st February 2014

Table 4.1 Hourly Actual values of peak load and the predicted using the GMDH for 1st February 2014

Hour	Forecast	Actual
1	1162	1166
2	1058	1050
3	968.1	964
4	915.5	909
5	901.1	892
6	873.6	871
7	877.3	873
8	910	909
9	949	951
10	1053	1051
11	1144	1145
12	1180	1181
13	1251	1244
14	1278	1276
15	1290	1288
16	1301	1299
17	1302	1294

18	1271	1273
19	1246	1248
20	1345	1339
21	1443	1448
22	1453	1453
23	1450	1447
24	1416	1415

From table 4.1, the forecasted values are close to the actual values .The small variation between actual value and the predicted value due to the fact that in some hours of that day actual values of peak load are affected by presence of maintenance operation and shedding. The observation of line outage in that day is collected in table below:

Table 4.2: lines outage for 1st February

Area	Voltage level (KV)	NO. of outage lines
Khartoum	33	2
	11	5
Omdurman	33	1
	11	4
Khartoum north	33	0
	11	7
Aljazeera	33	4
	11	7

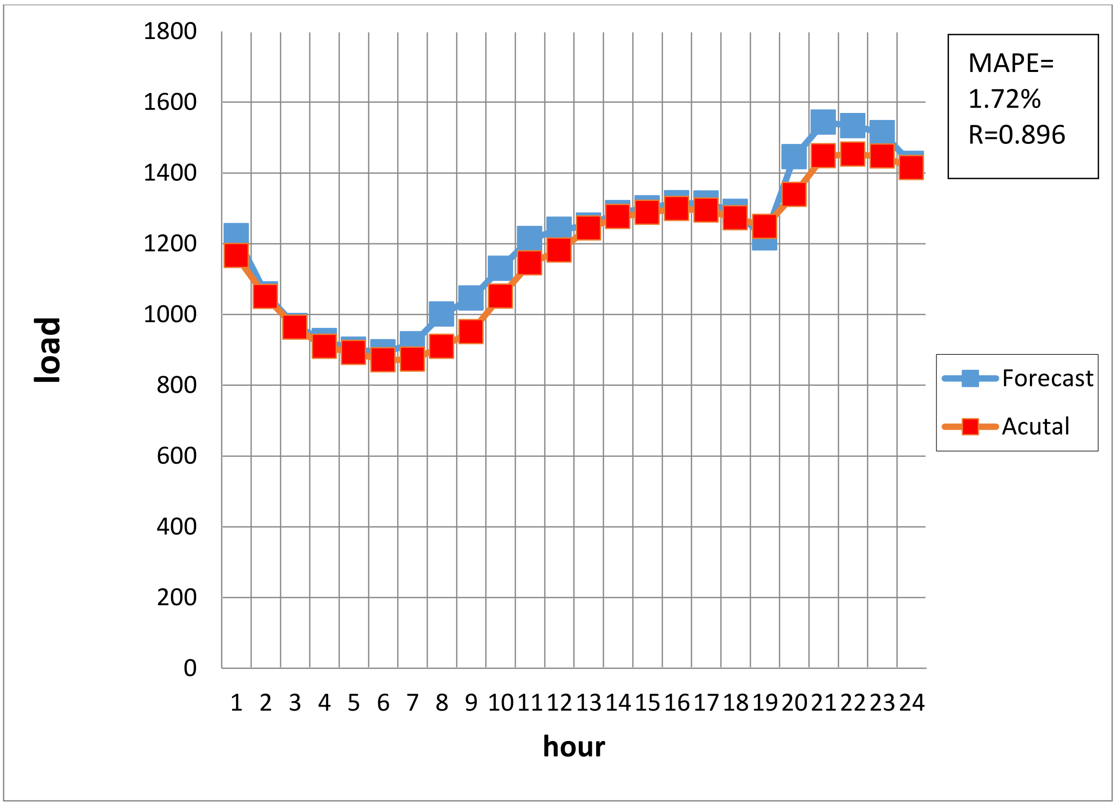

Figure 4.2: Plot of Hourly actual and forecasted peak load for 1st February 2014

Figure 4.2, represent the hourly actual peak load and forecasted load for 1st February 2014. Result showed a correlation (R) 0.896 and a mean absolute percentage error (MAPE) 1.72%. Since the R value is close to 1, it shows a close relationship between the actual and the predicted value and also the low error values indicate a high degree of forecasting accuracy.

4.4 Simulation result of next 7 days for Khartoum network

A figure (4.3) shows the actual data, the model fit and the Predicted peak load respectively. The actual data is the historical data that was used for training and testing and it is grey in colour. The model fit is a model value fitted to the data and is blue in colour. The predicted value is the value forecasted by the GMDH-time series model and is red in colour.

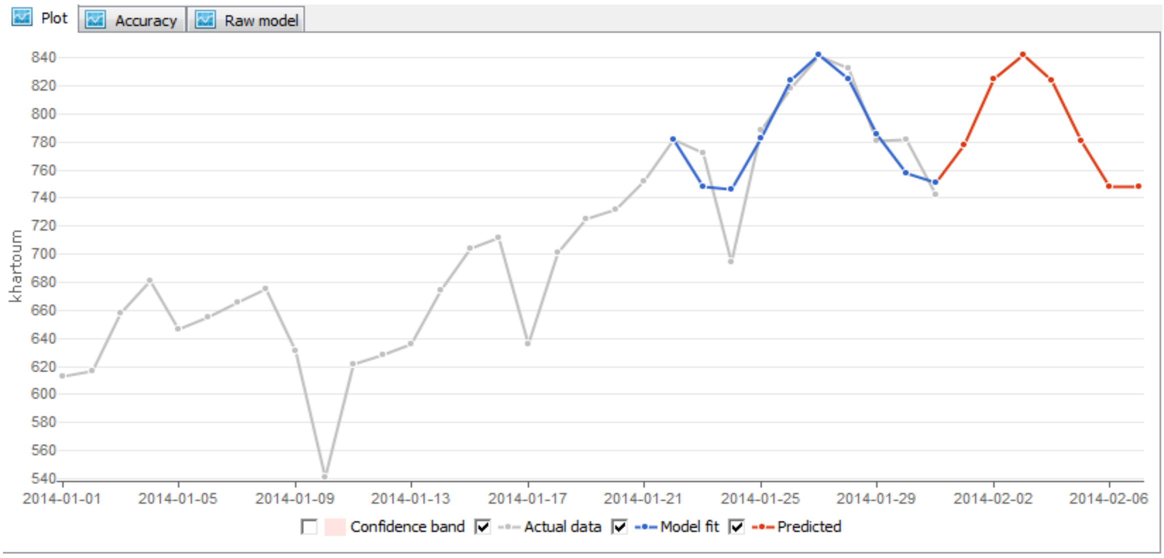

Figure 4.3: The actual peak load, the model fit and the predicted data for Khartoum network

Table 4.3 Actual values of peak load and the predicted using the GMDH for Khartoum network.

Date	Actual	Forecast
1-2-2014	810	815
2-2-2014	846	847
3-2-2014	809	810
4-2-2014	777	782
5-2-2014	752	753
6-2-2014	716	715
7-2-2014	617	621

From table 4.3, it obvious that there is small variation between actual value and the predicted value , it referred to the fact that in days of that week actual values of peak load are affected by presence of maintenance operation and shedding. The observation of outage lines in that week is collected in table below:

Table (4.4): The outage lines for Khartoum grid

Day of week	Voltage level	No. of outage lines
1	33	3
	11	16
2	33	3
	11	22
3	33	5
	11	9
4	33	7
	11	11
5	33	3
	11	10
6	33	3
	11	14
7	33	5
	11	13

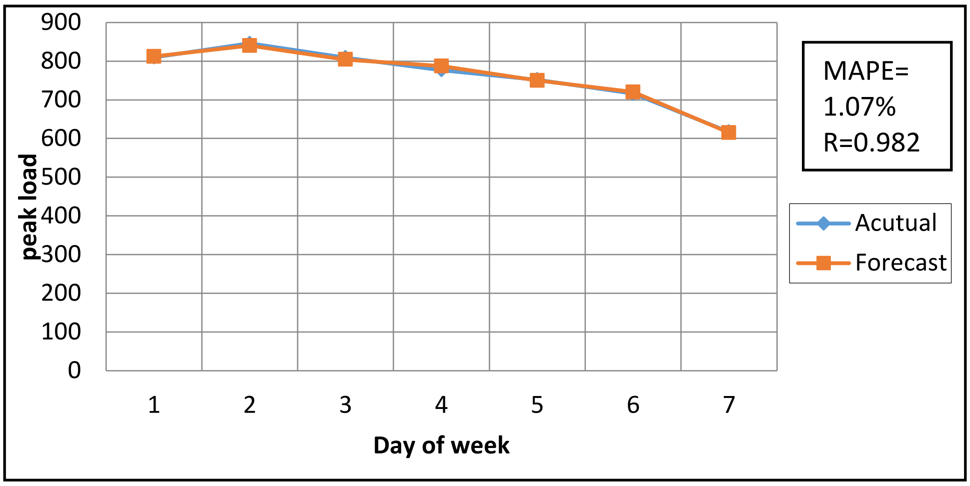

Figure 4.4: Plot of actual and forecasted peak load for Khartoum network

Figure 4.4, represent the daily actual peak load and forecasted load for 1^{st} week at February 2014. Result showed a correlation (R) 0.982 and a mean absolute percentage error (MAPE) 1.07%. Since the R value is close to 1, it shows a close relationship between the actual and the predicted value and also the low error values indicate a high degree of forecasting accuracy.

4.5 Simulation result of next 7 days for Aljazeera network

A figure (4.5) shows the actual data, the model fit and the Predicted peak load respectively. The actual data is the historical data that was used for training and testing and it is grey in colour. The model fit is a model value fitted to the data and is blue in colour. The predicted value is the value forecasted by the GMDH-time series model and is red in colour.

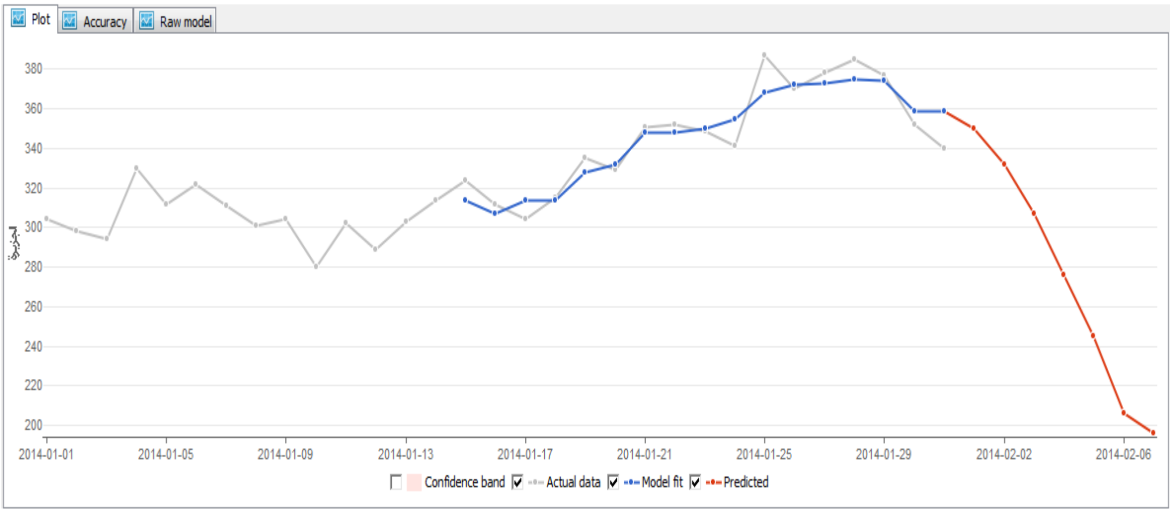

Figure 4.5: The actual peak load, the model fit and the predicted data for Aljazeera network

Table 4.5: Actual values of peak load and the predicted using the GMDH for Aljazeera network

Date	Actual	Forecast
1-2-2014	362	360
2-2-2014	330	332
3-2-2014	310	309
4-2-2014	277	276
5-2-2014	248	249
6-2-2014	207	206
7-2-2014	202	205

From table 4.5, the forecasted values are close to the actual values .The small variation between actual value and the predicted value due to the fact that in days of that week actual values of peak load are affected by presence of maintenance operation and shedding.

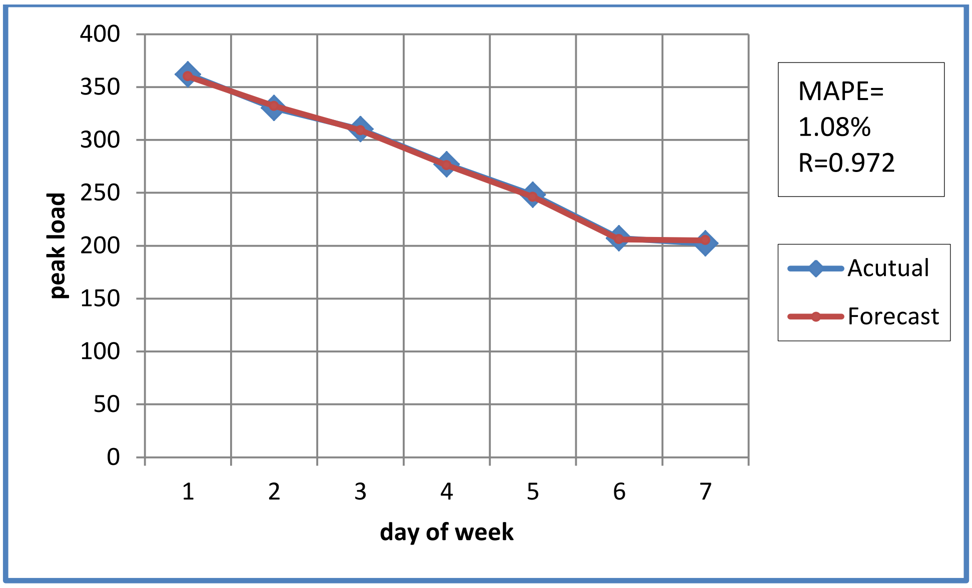

Figure 4.6: Plot of actual and forecasted peak load for Aljazeera network

Figure 4.6, represent the daily actual peak load and forecasted load for 1st week at February 2014. Result showed a correlation (R) 0.972 and a mean absolute percentage error (MAPE) 1.08%. Since the R value is close to 1, it shows a close relationship between the actual and the predicted value and also the low error values indicate a high degree of forecasting accuracy.

4.6 Simulation result of next 7 days for eastern network

A figure (4.7) shows the actual data, the model fit and the Predicted peak load respectively. The actual data is the historical data that was used for training and testing and it is grey in colour. The model fit is a model value fitted to the data and is blue in colour. The predicted value is the value forecasted by the GMDH-time series model and is red in colour.

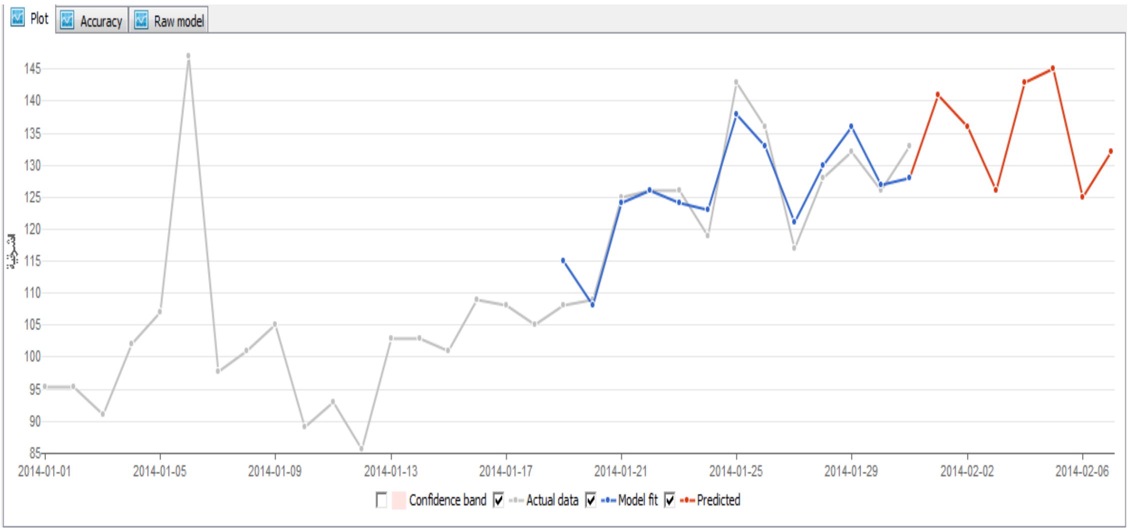

Figure 4.7: The actual peak load, the model fit and the predicted data for eastern network

Table 4.6: Actual values of peak load and the predicted using the GMDH for eastern network

Date	Actual	Forecast
1-2-2014	128	127
2-2-2014	191	196
3-2-2014	137	136
4-2-2014	121	123
5-2-2014	132	131
6-2-2014	121	122
7-2-2014	125	126

From table 4.6, the forecasted values are close to the actual values. The small variation between actual value and the predicted value due to the fact that in days of that week actual values of peak load are affected by presence of maintenance operation and shedding.

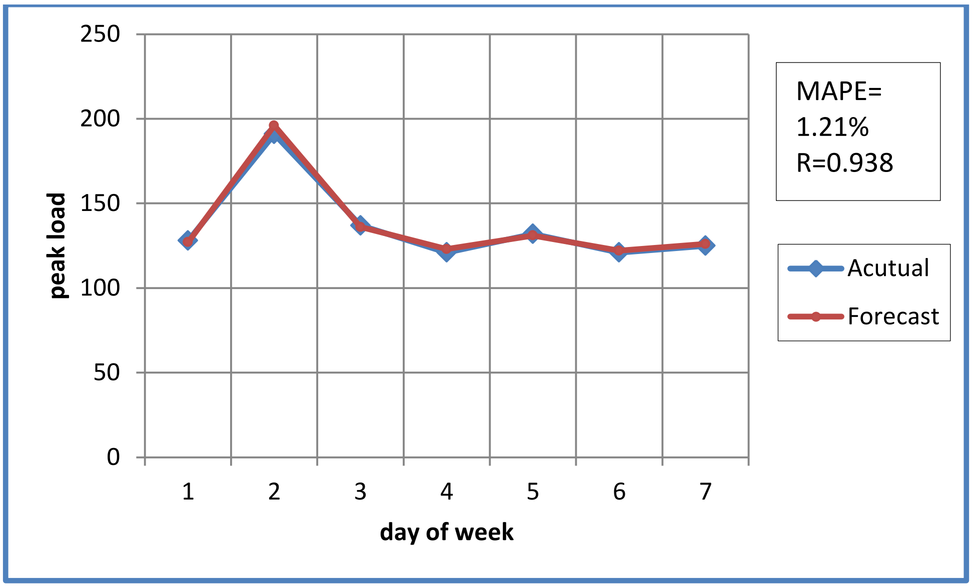

Figure 4.8: Plot of actual and forecasted peak load for eastern network

Figure 4.8, represent the daily actual peak load and forecasted load for 1st week at February 2014. Result showed a correlation (R) 0.938 and a mean absolute percentage error (MAPE) 1.21%. Since the R value is close to 1, it shows a close relationship between the actual and the predicted value and also the low error values indicate a high degree of forecasting accuracy.

4.7 Simulation result of next 7 days for northern network

A figure (4.9) shows the actual data, the model fit and the Predicted peak load respectively. The actual data is the historical data that was used for training and testing and it is grey in colour. The model fit is a model value fitted to the data and is blue in colour. The predicted value is the value forecasted by the GMDH-time series model and is red in colour.

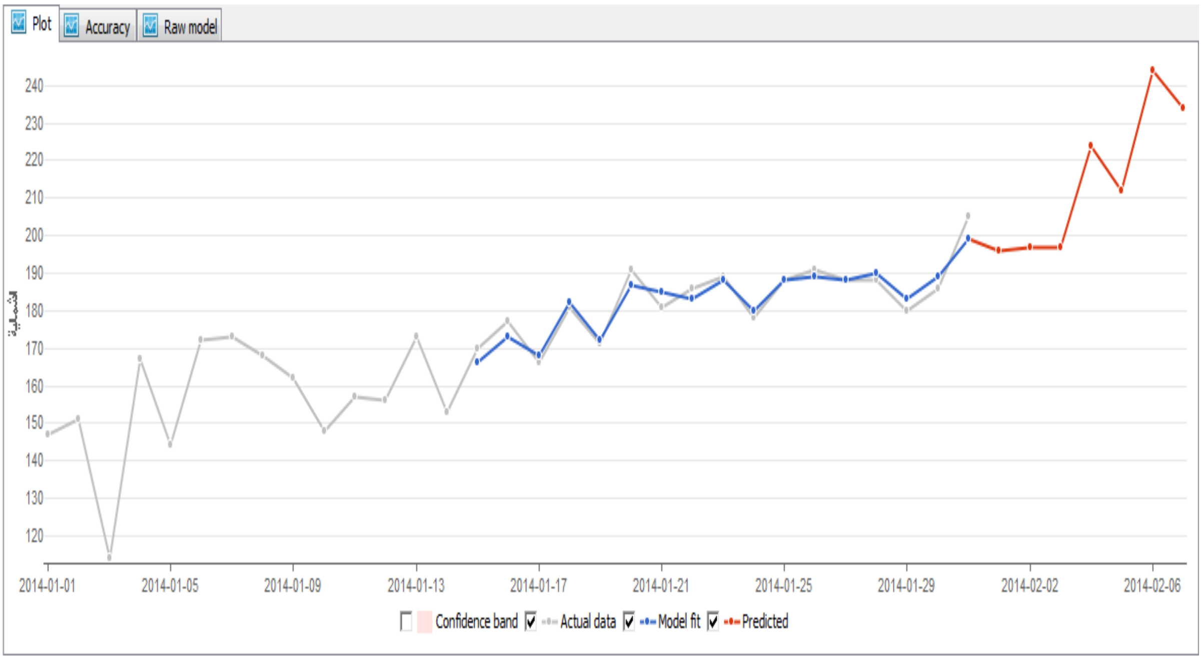

Figure 4.9: The actual peak load, the model fit and the predicted data for northern network

Table 4.7: Actual values of peak load and the predicted using the GMDH for northern network

Date	Actual	Forecast
1-2-2014	199	200
2-2-2014	218	219
3-2-2014	193	195
4-2-2014	204	204
5-2-2014	206	207
6-2-2014	204	204
7-2-2014	189	191

From table 4.7, the forecasted values are close to the actual values .The small variation between actual value and the predicted value due to the fact that in days of that week actual values of peak load are affected by presence of maintenance operation and shedding.

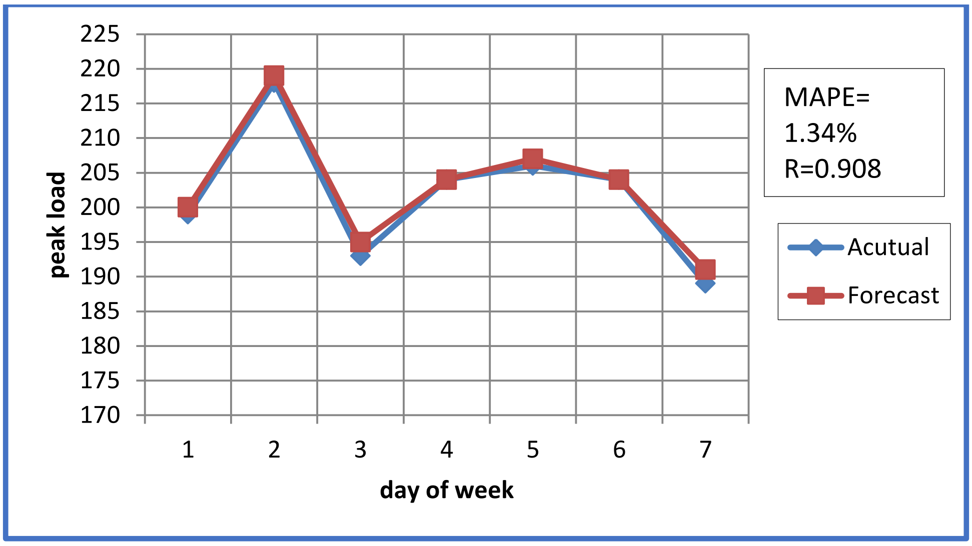

Figure 4.10: Plot of actual and forecasted peak load for northern network

Figure 4.10, represent the daily actual peak load and forecasted load for 1st week at February 2014. Result showed a correlation (R) 0.908 and a mean absolute percentage error (MAPE) 1.34%. Since the R value is close to 1, it shows a close relationship between the actual and the predicted value and also the low error values indicate a high degree of forecasting accuracy.

4.8 Simulation result of next 7 days for all Sudan network

A figure (4.11) shows the actual data, the model fit and the Predicted peak load respectively. The actual data is the historical data that was used for training and testing and it is grey in colour. The model fit is a model value fitted to the data and is blue in colour. The predicted value is the value forecasted by the GMDH-time series model and is red in colour.

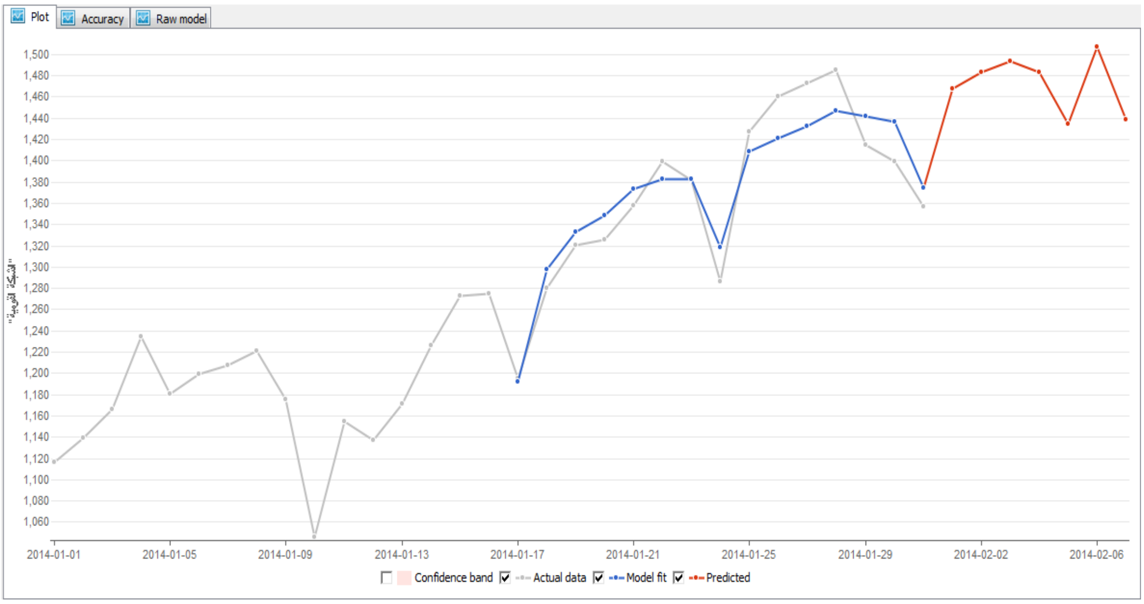

Figure 4.11: The actual peak load, the model fit and the predicted data for all Sudan network

Table 4.8: Actual values of peak load and the predicted using the GMDH for all Sudan network

Date	Actual	Forecast
1-2-2014	1504	1499
2-2-2014	1585	1587
3-2-2014	1449	1450
4-2-2014	1379	1377
5-2-2014	1338	1335
6-2-2014	1246	1248
7-2-2014	1133	1139

From table 4.8, the forecasted values are close to the actual values .The small variation between actual value and the predicted value due to the fact that in days of that week actual values of peak load are affected by presence of maintenance operation and shedding.

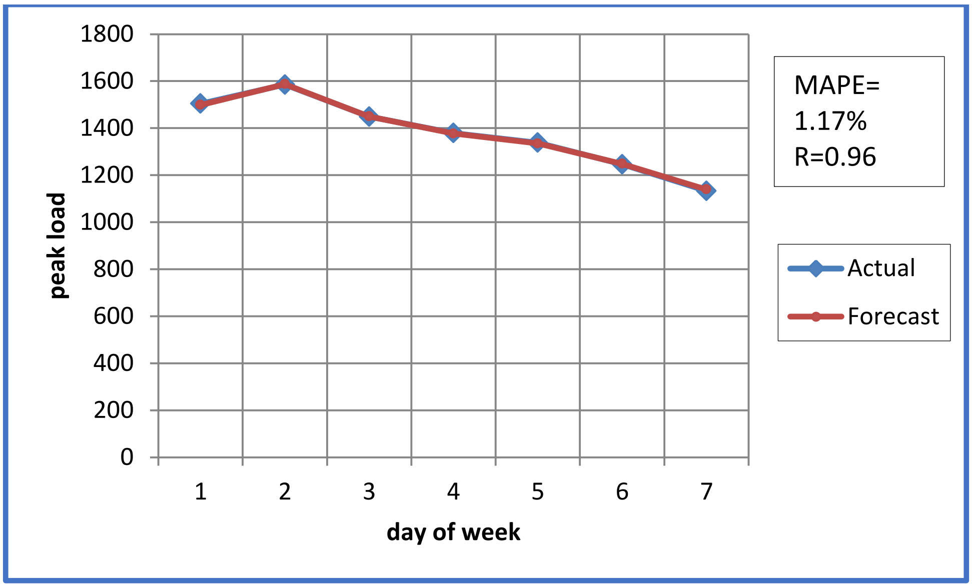

Figure 4.12: Plot of actual and forecasted peak load for all Sudan network

Figure 4.12, represent the daily actual peak load and forecasted load for 1st week at February 2014. Result showed a correlation (R) 0.96 and a mean absolute percentage error (MAPE) 1.17%. Since the R value is close to 1, it shows a close relationship between the actual and the predicted value and also the low error values indicate a high degree of forecasting accuracy.

Table 4.9: The predicted value of all Sudan grids

Forecasted values for Sudan grid By collecting the predicted peak load for each individual grids	Forecasted values for Sudan grid by program directly
1499	1499
1587	1587
1448	1450
1379	1377
1334	1335
1247	1248
1137	1139

Table 4.9 shows the predicted peak load for Sudan grid, the first column shows the predicted peak load for Sudan grid by collecting the predicted peak load for each individual grids comprising the whole grid of Sudan and the second column shows the predicted peak load for Sudan grid directly from the program. By comparing these two columns we find that the results are so closed to each other.

4.9 Analysis Results

Result obtained from the time series using GMDH showed a correlation (R) range between 0.896 and 0.982. Since the R value is close to 1, it shows a close relationship between the actual and the predicted value, hence a good forecast. A Mean absolute percentage error (MAPE) range between 1.07 % and 1.72% was obtained. The low error values indicate a high degree of forecasting accuracy.

The small variation between actual value and the predicted value due to the fact that in some hours and days actual value in peak load are affected by presence of maintenance operation and shedding.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

Electricity demand forecasting represents the main task in the Planning of electricity production because it determines the required resources to operate the electricity plants such as daily Consumption of fuels. Furthermore, it is the corner stone of Planning for electric plants and networks. The literature re-ports that the electric load pattern is very complex. Short-term load forecasts are required for the control and scheduling of power systems. Also required by transmission companies when a self-dispatching market is in operation.

Accurate daily load forecasts are significant for secure and profitable operation of modern power utilities. In this research GMDH is used to forecast the hourly and daily peak load. The results are computed and the performance is measured using correlation (R) and MAPE .The GMDH results show low MAPE values and gives a high correlation (R). These low values in MAPE and high correlation show a high degree of forecasting accuracy, it is obvious that GMDH is an efficient and straightforward approach.

5.2 Recommendations

1. For more accurate load forecasting models other weather information, such as wind speed, humidity, dew point, cloud cover, and rainfall, may possibly be included.
2. Other methods can be used like:
   - Neural Network for short term load forecasting
   - Neural-Fuzzy System for Short Term Load Forecast

REFERENCES

[1] D. W. Bunn, “Forecasting loads and prices in competitive

Power markets,” Proc. of the IEEE, vol. 88, pp. 163-169, Feb. 2000.

[2] M. P. Garcia and D. S. Kirschen, “Forecasting system Imbalance volumes in competitive electricity markets,” Power Systems, IEEE Transactions, vol. 21, pp. 240-248, Feb. 2006.

[3] H. L. Willis, “Power Distribution Planning Reference Book,” Second Edition, Revised and Expanded. New York: Marcel Dekker, 2002.

[4] J.Dudhia and J.F.Bresch. “A Global Version of the PSU-NCAR Mesoscale Model”. Monthly Weather Review, 2002.

[5] H.S. Hippert, C.E. Pedreira, and R.C. Souza. “Neural Networks for Short-Term Load Forecasting”: A Review and Evaluation. IEEE Transactions on Power Systems, 2001.

[6] H.L. Willis. Spatial “ Electric Load Forecasting”. Marcel Dekker, New York, 1996.

[7] R.F. Engle, C. Mustafa, and J. Rice. “Modeling Peak Electricity Demand. Journal of Forecasting, ” 1992.

[8] Robert Fich1, "Security Assessment and Enhancement, in Application Of Artificial Neural Network to Power System", M. A. El-Sharkawy And Degmar Neural (eds.), pp. 104-127, New Jersey: The Institute of Electrical and Electronic Engineers, Inc., 1996

[9] Andrew P. Douglas et al., "Risk Due Load forecast Uncertainty in Short Term Power System Planning", IEEE Transactions on Power System, Vol. 13, No. 4, PP. 1499, November 1998.

[10] D. W. Bunn and E. D. Farmer, "Economic and Operation Context of Electric load prediction", in D. W. Bunn and E. D. Farmer, (ed.), Comparative models for Electrical load forecasting, pp.3-11, New York: John Wiley and Sons Ltd., 1985.

[11] Andrew P. Douglas et al., "Risk Due Load forecast Uncertainty in Short Term Power System Planning", IEEE Transactions on Power System, Vol. 13, No. 4, PP. 1499, November 1998.

[12] Bunnoon, Pituk. "Mid-Term Load Forecasting Based on Neural Network Algorithm: a Comparison of Models."

[13] Zadeh, L. (1965). ‘‘Fuzzy sets’’, Inf. Control, vol. (1975). ‘‘The concept of a linguistic variable and its application to approximate reasoning.

[14] Rustum Mamlook, , Jordan, “A fuzzy inference model for short-term load forecasting, January 2009.

[15] Park, Dong C.,et al."Electric load forecasting using an artificial neural network." Power Systems, IEEE Transactions on 6.2 (1991): 442-449.

[16] Ghods, Ladan, and Mohsen Kalantar. "Different Methods of Long-Term Electric Load Demand Forecasting; A Comprehensive Review." Iranian Journal of Electrical & Electronic Engineering 7.4 (2011): 249.

[17] Adhikari, Ratnadip, and R. K. Agrawal. "An Introductory Study on TimeSeries Modeling and Forecasting." arXiv preprint arXiv:1302.6613 (2013).

[18] Power Generation, Operatoin, And Control Third Edition Allen J. Wood Bruce F. Wollenberg Gerald B. Sheblé

APPENDICES

A- All Sudan grid

DATE	Actual
2014-01-01	1116
2014-01-02	1139
2014-01-03	1166
2014-01-04	1235
2014-01-05	1181
2014-01-06	1199
2014-01-07	1208
2014-01-08	1221
2014-01-09	1176
2014-01-10	1046
2014-01-11	1155
2014-01-12	1137
2014-01-13	1171
2014-01-14	1226
2014-01-15	1273
2014-01-16	1275
2014-01-17	1195
2014-01-18	1280
2014-01-19	1321
2014-01-20	1326
2014-01-21	1358
2014-01-22	1399
2014-01-23	1382
2014-01-24	1286
2014-01-25	1427
2014-01-26	1460
2014-01-27	1473
2014-01-28	1485
2014-01-29	1415
2014-01-30	1399
2014-01-31	1357

B- Khartoum grid

date	Actual
2014-01-01	613
2014-01-02	617
2014-01-03	658
2014-01-04	681
2014-01-05	646
2014-01-06	655
2014-01-07	666
2014-01-08	675
2014-01-09	631
2014-01-10	541
2014-01-11	622
2014-01-12	628
2014-01-13	636
2014-01-14	674
2014-01-15	704
2014-01-16	712
2014-01-17	636
2014-01-18	701
2014-01-19	725
2014-01-20	732
2014-01-21	752
2014-01-22	782
2014-01-23	772
2014-01-24	694
2014-01-25	788
2014-01-26	818
2014-01-27	841
2014-01-28	832
2014-01-29	781
2014-01-30	782
2014-01-31	742

C- Aljazeera grid

ID	Actual
2014-01-01	304
2014-01-02	298
2014-01-03	294
2014-01-04	330
2014-01-05	312
2014-01-06	322
2014-01-07	311
2014-01-08	301
2014-01-09	304
2014-01-10	280
2014-01-11	302
2014-01-12	289
2014-01-13	303
2014-01-14	314
2014-01-15	324
2014-01-16	312
2014-01-17	304
2014-01-18	315
2014-01-19	335
2014-01-20	329
2014-01-21	351
2014-01-22	352
2014-01-23	349
2014-01-24	341
2014-01-25	387
2014-01-26	370
2014-01-27	378
2014-01-28	385
2014-01-29	377
2014-01-30	352
2014-01-31	340

D- Northern grid

date	Actual
2014-01-01	147
2014-01-02	151
2014-01-03	114
2014-01-04	167
2014-01-05	144
2014-01-06	172
2014-01-07	173
2014-01-08	168
2014-01-09	162
2014-01-10	148
2014-01-11	157
2014-01-12	156
2014-01-13	173
2014-01-14	153
2014-01-15	170
2014-01-16	177
2014-01-17	166
2014-01-18	181
2014-01-19	171
2014-01-20	191
2014-01-21	181
2014-01-22	186
2014-01-23	189
2014-01-24	178
2014-01-25	188
2014-01-26	191
2014-01-27	188
2014-01-28	188
2014-01-29	180
2014-01-30	186
2014-01-31	205

E- Eastern grid

date	Actual
2014-01-01	95.4
2014-01-02	95.4
2014-01-03	91
2014-01-04	102
2014-01-05	107
2014-01-06	147
2014-01-07	97.7
2014-01-08	101
2014-01-09	105
2014-01-10	89.1
2014-01-11	93
2014-01-12	85.6
2014-01-13	103
2014-01-14	103
2014-01-15	101
2014-01-16	109
2014-01-17	108
2014-01-18	105
2014-01-19	108
2014-01-20	109
2014-01-21	125
2014-01-22	126
2014-01-23	126
2014-01-24	119
2014-01-25	143
2014-01-26	136
2014-01-27	117
2014-01-28	128
2014-01-29	132
2014-01-30	126
2014-01-31	133