DYNAMIC LOAD FORECASTING FOR COMMERCIAL POWER NETWORK

ABDUSALAM RAJB ALZALET

A project report submitted in partial fulfillment of the requirement for the award of the Degree of Master of Electrical Engineering

Faculty of Electrical and Electronic Engineering
Universiti Tun Hussein Onn Malaysia

JUNE 2014
ABSTRACT

Load forecasting is an important component for power system energy management system. The electrical load is the power that an electric utility needs to supply in order to meet the demands of its customers. It is therefore very important to the utilities to have advance knowledge of their electrical load, so that they can ensure the load is met and thus minimising any interruptions to their service. It also plays a key role in reducing the generation cost, and also essential to the reliability of power systems. The electric power demand in Universiti Tun Hussein Onn Malaysia (UTHM) has increased as the power system network is getting larger with more consumption is to be expected. This loading trend is certain to continue in the near future. The aim of this project is to forecast the medium term loading of UTHM Linear regressions and polynomial based methods as well as artificial neural networks (ANN) approach have been adapted in the load forecasting from 2006 to 2012. The results attained are validated with the real data obtained from the Tenaga Nasional Berhad (TNB) which represents the monthly load electric consumption in UTHM. By comparing the forecasted results with the real data, the most suitable method has been proposed. When the approaches are compared according to their highest prediction error, the highest error for linear regression and Polynomial equation approaches are very high compared to the ANN approach. Generally the ANN approach has produced better results.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE</td>
<td>I</td>
</tr>
<tr>
<td>DECLARATION</td>
<td>II</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>III</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>IV</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>V</td>
</tr>
<tr>
<td>LIST OF CONTENTS</td>
<td>VI</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>VIII</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>IX</td>
</tr>
<tr>
<td>LIST OF SYMBOLS AND ABBREVIATIONS</td>
<td>XI</td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td>XIII</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Project Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Problem Statements</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Project Objectives</td>
<td>3</td>
</tr>
<tr>
<td>1.4 Project Scopes</td>
<td>4</td>
</tr>
<tr>
<td>CHAPTER 2 LITERATURE REVIEW</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Load models</td>
<td>6</td>
</tr>
<tr>
<td>2.3 Static load forecasting methods</td>
<td>9</td>
</tr>
<tr>
<td>2.3.1 Exponential Smoothing methods</td>
<td>9</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Year of each electric substation was built</td>
<td>24</td>
</tr>
<tr>
<td>4.1</td>
<td>Forecasted load demand by different approach for 2011-2012</td>
<td>32</td>
</tr>
<tr>
<td>4.2</td>
<td>Forecasting % errors by different approaches for 2011-2012</td>
<td>34</td>
</tr>
<tr>
<td>4.3</td>
<td>Forecasted load demand by different approach for 2012-2013</td>
<td>36</td>
</tr>
<tr>
<td>4.4</td>
<td>Forecasting % errors by different approaches for 2012</td>
<td>38</td>
</tr>
<tr>
<td>4.5</td>
<td>Forecasted load demand and its % errors by ANN for the year of 2007</td>
<td>39</td>
</tr>
<tr>
<td>4.6</td>
<td>Forecasted load demand and its % errors by ANN for the year of 2008-2009</td>
<td>42</td>
</tr>
<tr>
<td>4.7</td>
<td>Forecasted load demand and % errors by ANN for 2009-2010-2011</td>
<td>45</td>
</tr>
<tr>
<td>4.8</td>
<td>Forecasted load demand by ANN for 2013-2014</td>
<td>48</td>
</tr>
<tr>
<td>4.9</td>
<td>Forecasted load demand by different approaches for the year of 2011</td>
<td>50</td>
</tr>
<tr>
<td>4.10</td>
<td>Forecasting % errors by different approaches for the year of 2011</td>
<td>51</td>
</tr>
</tbody>
</table>
# List of Figures

2.1 Component-based approaches to load modelling. 8
2.2 Load curves for the day-ahead load forecasting using the Holt-Winters mechanism 14
2.3 Feed forward multi-layer ANN 17
2.4 Weekly BP/USD exchange rate series (1980-1993) 19
2.5 Monthly international airline passenger series (Jan. 1949-Dec. 1960) 19
3.1 ANN training procedure 25
3.2 Project flow chart 28
4.1 The load demand for UTHM power system from 2006-2012 30
4.2 Actual and forecasted load demand values using polynomial equation for 2011-2012 33
4.3 Actual and forecasted load demand values using linear regression method for 2011-2012 33
4.4 Actual and forecasted load demand values using linear regression method for 2012-2013 37
4.5 Actual and forecasted load demand values using polynomial equation for 2012-2013 37
4.6 Actual and forecasted load demand values using ANN for 2007 40
4.7 Errors by ANN for 2007 40
4.8 Forecasted and targeted values by ANN for 2007 41
4.9 Actual and forecasted load demand values using ANN for 2008-2009 43
4.10 Errors by ANN for 2008-2009
4.11 Forecasted and targeted by ANN for 2008-2009
4.12 Actual and forecasted load demand values using ANN for 2009-2010-2011
4.13 Errors by ANN for 2009-2010-2011
4.14 Forecasted and targeted values by ANN for 2009-2010-2011
4.15 Forecasted load demand values using ANN for 2013-2014
4.16 Actual and forecasted load demand values using different approaches for 2011
4.17 Forecasting % errors for 2011
LIST OF SYMBOLS AND ABBREVIATIONS

\( T \) - Estimated time
\( y_T \) - Real value for moment \( T \)
\( \lambda \) - Discount factor
\( \hat{y}_t \) - Forecasted value
\( a_t \) - Value of the intercept
\( b_t \) - Value of the slope
\( h \) - Time horizon
\( \alpha, \beta \) - Discount factors (constants)
\( e_t \) - Forecast error
\( SS_E \) - Sum of squared forecast errors
\( S_{t-p+h} \) - Seasonal adjustment
\( x(t) \) - Variable
\( t \) - Time
\( T(t) \) - Trend variation at time \( t \)
\( S(t) \) - Seasonal variation at time \( t \)
\( C(t) \) - Cylcical variation at time \( t \)
\( I(t) \) - Irregular variation at time \( t \)
\( Y \) - Dependent variable
\( X \) - Independent variable
\( N \) - Number of variable
\( Y_{tr} \) - Average value for the dependent variable
\( W \) - Weighting matrix
\( Kth \) - Output node
\( Jth \) - Hidden node
\( STLF \) - Short-term load forecasting
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTLF</td>
<td>Medium-term load forecasting</td>
</tr>
<tr>
<td>LTLF</td>
<td>Long-term load forecasting</td>
</tr>
<tr>
<td>UTHM</td>
<td>Universiti Tun Hussein Onn Malaysia</td>
</tr>
<tr>
<td>kWh</td>
<td>Kilowatt hour</td>
</tr>
<tr>
<td>MWh</td>
<td>Megawatt hour</td>
</tr>
<tr>
<td>TNB</td>
<td>Tenaga Nasional Berhad</td>
</tr>
<tr>
<td>MAPE</td>
<td>Mean Absolute Percentage Error</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural networks</td>
</tr>
<tr>
<td>PSS</td>
<td>Statistical package for the social sciences</td>
</tr>
<tr>
<td>QSAR</td>
<td>Quantitative Structure-Activity Relationship</td>
</tr>
</tbody>
</table>
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Project Planning For master project 1 and master project 2</td>
<td>60</td>
</tr>
<tr>
<td>B</td>
<td>Data of Load Consumption</td>
<td>62</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Project Background

Accurate electrical load forecasting is a crucial issue for resource planning and management of electrical power generation utilities. Load forecasting can be divided into three categories. Short-term load forecasting (STLF) that covers a period of one hour to one month. Medium-term load forecasting (MTLF) covers a period of one month up to one year. It is essential for scheduling fuel supplies and maintenance operation. Long–term load forecasting (LTLF) predicts the requirements of energy for more than one year. Capital assignment and infrastructure plans draw upon long-term forecasts.

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 [1]. 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 [2].

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 a head [1].

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 [3]. 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 [2].
1.2 Problem Statements

Load forecasting problem is receiving great and growing attention as being an important and primary tool in power system planning and operation. Importance of load forecasting becomes more significant in developing countries with high growth rate. In recent years the electrical energy consumption is increased. A noticeable increase of electric energy consumption in UTHM University and especially after 2008 observed a significant increase in energy consumption due to developments that take place in all university facilities in terms of new section, library, and other new buildings. Since this development will be accompanied by increasing demand for energy, so it is necessary to perform the forecasting study to estimate the increase of energy demand that meet the needs of the future development plans, and helps the authority to take the right decisions regarding the investment and future plans.

1.3 Project Objectives

The objective of the project is to study the possible use of forecasting technique for UTHM power system loads, and estimating the annually load demand. It measurable objectives are as follows:

a) To analysis the historical data collected for UTHM in past years.

b) To propose dynamic load forecasting method for power consumption in UTHM.

c) To simulate the power system load forecast and determine the medium term power load demand for the next years.
1.4 Project Scopes

a) Collecting data for UTHM power loads since 2006 to 2012 and analysing it to determine the growth rate kWh.

b) Study the load forecasting techniques and choose the suitable techniques and choose the suitable techniques used.

c) Calculate the power loads using static methods to predict loads, to study the behaviour of future loads resulting using Excel.

d) Simulate the power loads using dynamic method to forecast the load demand of UTHM network by using ANN, MATLAB software.

e) Comparison between the electrical loads resulting by different approaches methods according to their highest predictions % error.
CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Load forecasting is of the most difficult problems in distribution system planning and analysis. However, not only historical load data of a distribution system play a very important role on peak load forecasting, but also the impacts of meteorological and demographic factors must be taken into consideration. Generally, load forecasting methods are mainly classified into two categories: classical approaches and ANN based techniques. Classical approaches are based on dynamic methods and Static methods forecast future value of a variable by using a mathematical combination of the historic information. Section 2.2 describes the study of the load models of electricity supply systems. Section 2.3 discusses Static load forecasting methods, section 2.4 describes the Dynamic load forecasting methods used and section 2.5 describes the Model validation methods.
2.2 Load models

For a long time, it has been recognized that the operation and performance of electricity supply systems are strongly influenced by the characteristics of the supplied load as given in Figure 2.1. Accordingly, the selection of the load model in a particular power system study will have a significant effect on the results of the study and, therefore, corresponding design decisions. Too optimistic load models can lead to inadequate system design or reinforcement, and may result in either costly upgrades or insecure systems, more vulnerable to various types of disturbances and collapse. Too pessimistic load models can, on the other hand, lead to unnecessary capital expenditure and uneconomical operation of the supply system.

In the past, more pessimistic load models have often been favored, in order to accommodate conservative safety and design margins. However, the most pessimistic load representation is sometimes hard to determine. Supply systems are increasingly being operated near to their operational margins, due to growing demand, economic and environmental pressures to run these systems close to their maximum capacity. Representative and accurate load models are thus becoming increasingly important for correct assessment of network performance. In addition, accurate load models can facilitate better decision making in relation to financial investment.

Traditionally, load models have been developed for use in high and medium voltage studies, with low voltage distribution systems and associated loads represented as a combined single (i.e. aggregate) load model. Accordingly, these aggregate loads are relatively insensitive to variations in characteristics of individual loads. However, the move away from centralized to distributed generation of electricity requires distribution networks to be considered in more detail. This also means that distribution system load should be considered in greater detail. When representing these smaller aggregate distribution system loads, the characteristics of individual load components become more important, and should be given more attention than has been the case previously.
Load models will play an increasingly important role in system design and planning. A major reason for this is the anticipated higher penetration of distributed generation and subsequent need for more efficient management of distribution networks. That system load characteristics will strongly affect optimal location, size and financial feasibility of distributed generation. Furthermore, the types and numbers of loads commonly found in the distribution system have changed significantly in the last few decades, with most significant and most prominent increase of non-linear loads. This represents a major change in load inventory, which should be taken into account when considering load models for use in studies of modern power supply systems. Although the importance of accurate representation of system load is generally known, certain assumptions are often made about load models. This is due to the lack of or uncertainty in available load modeling data. Therefore, efforts in clarifying or presenting new load models are of general importance to all aspects of supply system design and operation [4].
In this report, the term “load” is defined as either a single device or load type connected to the supply system, or an aggregation of load types connected to the supply system. A load model is mathematical or analytical representation of the changes in active and/or reactive power demand of a load, usually as a function of the changes in applied voltage and, in some cases, frequency. Load models can be generally defined as either static or dynamic. A static model is used to represent the active and reactive power demand of the load as a function of voltage and frequency at a particular instant time. A dynamic load model represents active and reactive power demand as a function of voltage, frequency and time.

Dynamic models are typically represented in the form of differential equations and used in voltage stability studies, studies of inter-area oscillations and long term stability studies. A commonly used dynamic load model is based on the
equivalent circuit of an induction motor [5], this is because induction motor load represents a significant proportion of the total system load. According to [6], the lack of dynamic motor models in power system studies is often thought to be the main cause of differences in results between field-measurements and large-scale simulations. When static and dynamic models are used together, this is known as a composite dynamic load model [7].

2.3 Static load forecasting methods

These methods forecast future load based on its historical values. The goal is to infer the pattern in the historical data series and extrapolate that pattern into the future. The load is considered as a time series embedding hourly, daily and seasonal patterns. Several techniques have been used for the analysis of these methods such as multiple linear regressions, moving average process, general exponential smoothing. Problems encountered with this approach include the inaccuracy of prediction and numerical instability. In general, statistical methods work well unless there is a drastic change in the variables that have a potential effect on the load pattern.

2.3.1 Exponential Smoothing methods

The method used for the load forecast is based on time series and takes into consideration only the history of the consumption in order to establish a pattern in the past that might be useful and similar with the present load curves. This technique uses exponentially decreasing weights as the observation get older. Recent observations are given relatively more weight in forecasting than the older observations. Exponential Smoothing is used to generate the smoothed values in order to obtain estimates power load. Exponential smoothing types currently used are [8]:

The first order exponential smoothing;
The second order exponential smoothing;
The higher order exponential smoothing;
The Holt-Winters mechanism.

2.3.1.1 The first-order exponential smoothing

It uses a recursive equation that can also be seen as the linear combination of the current observation and smoothed observation of the previous time unit. As the latter contains the data from all previous observations, the smoothed observation at moment $T$ (estimated time) is in fact the linear combination of the current observation and the discounted sum of all previous observations [8]:

$$\tilde{Y}_{T+1} = \lambda \cdot y_T + (1 - \lambda) \cdot \tilde{Y}_T$$

(2.1)

Where:

- $\tilde{Y}_{T+1}$ is the estimated value for moment $T+1$;
- $y_T$ - the real value for moment $T$;
- $\lambda \in [0, 1]$ - discount factor
- $\tilde{Y}_t$ - forecasted value;

The discount factor represents the weight put on the previous observation while $(1 - \lambda)$ is the weight put on the smoothed value of the previous observations. The most important issue for the exponential smoothers is the choice of the discount factor, $\lambda$ [8, 9].

2.3.1.2 The second-order exponential smoothing

The first-order exponential smoothing method was extended by Holt for presenting time series with trend (and random component). This approach adjusts the time series considering the trend to be linear [8]:

$$\tilde{Y}_{T+h} = a_t + h \cdot b_t$$

(2.2)
Where

\( a_t \) - the value of the intercept;

\( b_t \) - the value of the slope;

\( h \) - time horizon.

For the one step ahead forecasting, the value of \( h \) is 1. Parameters \( a_t \) and \( b_t \) are calculated as follows:

\[
a_t = \alpha \cdot y_t + (1 - \alpha) \cdot (a_{t-1} + b_{t-1})
\]  
(2.3)

\[
b_t = \beta \cdot y_t + (1 - \beta) \cdot (a_t + a_{t-1}) + (1 - \beta) \cdot b_{t-1}
\]  
(2.4)

Where

\( \alpha, \beta \in [0, 1] \) are the discount factors (constants);

\( a_t, b_t \) - values for this parameters at time \( t \);

\( a_{t-1}, b_{t-1} \) - values for this parameters at time \( t - 1 \).

The constants \( \alpha, \beta \) will be chosen for the smallest sum of the squared forecast errors (the value of \( \lambda \) from the first-order exponential smoothing will be determined in same manner) [8, 9],

\[
e_t = \tilde{y}_t - y_t
\]  
(2.5)

\[
SS_E = \sum_{t=1}^{T} e_t^2
\]  
(2.6)

Where

\( e_t \) - the forecast error;

\( \tilde{y}_t \) - forecasted value;

\( y_t \) - real value;

\( SS_E \) - sum of squared forecast errors.
2.3.1.3 The higher-order exponential smoothing

The first and second order exponential smoothing can be extended to the general $n$-th degree polynomial model presented in the equation below:

$$y_t = \beta_0 + \beta_1 t + \frac{\beta_2}{2!} t^2 + ... + \frac{\beta_n}{n!} t^n + \epsilon_t$$

(2.7)

Where $\epsilon_t$ are assumed to be independent with mean 0 and constant variance $\sigma^2$. In order to estimate the parameters the next equations will be used:

$$\hat{y}_T^{(1)} = \lambda \cdot \hat{y}_T + (1 - \lambda) \cdot \hat{y}^{(1)}_{T-1}$$

$$\hat{y}_T^{(2)} = \lambda \cdot \hat{y}_T^{(1)} + (1 - \lambda) \cdot \hat{y}^{(2)}_{T-1}$$

$$\hat{y}_T^{(n)} = \lambda \cdot \hat{y}^{(n-1)}_T + (1 - \lambda) \cdot \hat{y}^{(n)}_{T-1}$$

(2.8)

Where $\hat{y}_T^{(n)}$ is the estimated value for the $n$-th order exponential smoothing [8].

2.3.1.4 The Holt-Winters mechanism for seasonal time series

Some time series data exhibit cyclical or seasonal patterns that cannot be effectively modeled using the polynomial model as given in equations (2.8). Several approaches are available for the analysis of this data as given in Figure 2.2. The methodology of the Day-Ahead forecast by Holt and winters and is generally known as winters’ method; in this case, a seasonal adjustment is made to the linear trend model. Two types of adjustments are used, namely the multiplicative and the additive model a multiplicative model and the estimated values are calculated as in the following equation [9]:

$$\hat{y}_{T+h} = (a_t + h \cdot b_t) \cdot S_{T-p+h}$$

(2.9)

Where
\( h \) - the time horizon;

\( \tilde{y}_{T+h} \) - The forecasted value;

\( a_t \) - The level of the time series;

\( b_t \) - The trend of the time series;

\( S_{t-p+h} \) - The seasonal adjustment. The parameters mentioned above are determined in the next recursive equations:

\[
a_t = \alpha \cdot \left( \frac{y_t}{S_{t-p}} \right) + (1-\alpha) \cdot (a_{t-1} + b_{t-1})
\]

(2.10)

\[
b_t = \beta \cdot \left( \frac{a_t}{a_{t-1}} \right) + (1-\beta) \cdot b_{t-1}
\]

(2.11)

\[
S_t = \delta \cdot \left( \frac{y_t}{a_t} \right) + (1-\delta) \cdot S_{t-1}
\]

(2.12)

where \( \beta, \delta \) are the discount factors constants which must be chosen for the smallest sum of the squared forecast errors as given in equation (2.6), and \( p \) represents the period of the season. The discount can be determined also by the smallest MSE (mean square error – equation (2.13), or by any statistical indicator taken into account [8, 9],

\[
\text{MSE} = \frac{1}{n} \sum_{t=1}^{T} e_t^2
\]

(2.13)

Where \( n \) is the number of observations [8],
2.3.2 Multiple Linear Regressions

Multiple linear regression analysis presume that a variable, $Y$, is linearly related to multiple independent variables, $X_1, X_2, ..., X_k$. Thus, multiple linear regressions presumes that

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + ... + \hat{\beta}_k X_k + \varepsilon$$  \hspace{1cm} (2.14)

where $\beta_0$, $\beta_1$, $\beta_2$, $\beta_k$ are constants and $\varepsilon$ is an error term that is a normal random variable with mean 0 and unknown variance $\sigma^2$. The parameter $\beta_0$ is the intercept and the parameters $\beta_1$, $\beta_2$... $\beta_k$ are the slopes.

Regression analysis provides estimators $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$... $\hat{\beta}_k$ so that the estimated regression line is

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + ... + \hat{\beta}_k X_k.$$  \hspace{1cm} (2.15)

As in the case of simple linear regression, multiple linear regression analysis generates the estimators $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$... $\hat{\beta}_k$ by minimizing the sum of squared
errors,$\sum_{i=1}^{n} e_i^2$ We Will not present equations for the estimators because we will be using Excel to generate Estimates. The key point for multiple linear regressions is that the independent variables, i.e., X1, X2… Xk must be independent of each other in the sense of being uncorrelated with Each other. If two or more of the independent variables are correlated with each other, then at least one of the variables is redundant. More importantly, if two or more of the independent variables are correlated, and then we face the problem of multicollinearity.

The PHStat system includes a stepwise regression routine. This system runs through all possible regressions, given a data set, and selects the “best” regression. You will find the system under PHStat/Regression/Stepwise Regression. Stepwise regression is not a solution to the problem of multicollinearity, but it is any easy way to avoid the problem [10].

2.4 Dynamic load forecasting methods

When the traditional static load models are not sufficient to represent the behavior of the load, the alternative dynamic load models are necessary. The parameters of these load models can be determined either by using a measurement-based approach, by carrying field measurements and observing the load response as a result of alterations in the system, or by using a component-based approach; first by identifying individual and then by aggregating them in one single load. The literature for dynamic load models is quite large depending on the results from different field measurements and their purposes.

Many studies have shown the importance of the load representation in voltage stability analysis. Static load models are not accurate enough for capturing the dynamics of the network. Therefore dynamic load models are needed even if voltage collapse, in many cases, is a slow phenomenon [11].
2.4.1 Artificial Neural Networks

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 [13].

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 [14, 15].

![Feed forward multi-layer ANN diagram]

**Figure 2.3: Feed forward multi-layer ANN**

### 2.4.2 Time series

Time series modeling is a dynamic research area which has attracted attentions of researchers’ community over last few decades. The main aim of time series modeling is to carefully collect and rigorously study the past observations of a time series to develop an appropriate model which describes the inherent structure of the series. This model is then used to generate future values for the series, i.e. to make forecasts. Time series forecasting thus can be termed as the act of predicting the future by understanding the past. Due to the indispensable importance of time series forecasting in numerous practical fields such as business, economics, finance, science and engineering, etc. proper care should be taken to fit an adequate model to the underlying time series. It is obvious that a successful time series forecasting depends on an appropriate model fitting. A lot of efforts have been done by researchers over many years for the development of efficient models to improve the forecasting accuracy. As a result, various important time series forecasting models
have been evolved in literature [16, 17].

A time series is a sequential set of data points, measured typically over successive times. It is mathematically defined as a set of vectors $x(t), t = 0,1,2,...$ where $t$ represents the time elapsed. The variable $x(t)$ is treated as a random variable. The measurements taken during an event in a time series are arranged in a proper chronological order a time series containing records of a single variable is termed as univariate. But if records of more than one variable are considered, it is termed as multivariate. A time series can be continuous or discrete. In a continuous time series observations are measured at every instance of time, whereas a discrete time series contains observations measured at discrete points of time. For example temperature readings, flow of a river, concentration of a chemical process etc. can be recorded as a continuous time series. On the other hand population of a particular city, production of a company, exchange rates between two different currencies may represent discrete time series. Usually in a discrete time series the consecutive observations are recorded at equally spaced time intervals such as hourly, daily, weekly, monthly or yearly time separations in Figure 2.4, Figure 2.5. As mentioned in, the variable being observed in a discrete time series is assumed to be measured as a continuous variable using the real number scale. Furthermore a continuous time series can be easily transformed to a discrete one by merging data together over a specified time interval.

Two different types of models are generally used for a time series viz. Multiplicative and Additive models.

\[
Y(t) = T(t) \times S(t) \times C(t) \times I(t) \quad (2.16)
\]

\[
Y(t) = T(t) + S(t) + C(t) + I(t) \quad (2.17)
\]

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 [16].
2.5 Model validation methods

Validation is the task of demonstrating that the model is a reasonable representation of the actual system: that it reproduces system behavior with enough fidelity to
satisfy analysis objectives. Whereas model verification techniques are general the approach taken to model validation is likely to be much more specific to the model, and system, in question. Indeed, just as model development will be influenced by the objectives of the performance study, so will model validation be. A model is usually developed to analyze a particular problem and may therefore represent different parts of the system at different levels of abstraction.

As a result, the model may have different levels of validity for different parts of the system across the full spectrum of system behavior. For most models there are three separate aspects which should be considered during model validation:

(i) Assumptions
(ii) input parameter values and distributions
(iii) Output values and conclusions.

However, in practice it may be difficult to achieve such a full validation of the model, especially if the system being modelled does not yet exist. In general, initial validation attempts will concentrate on the output of the model, and only if that validation suggests a problem will more detailed validation be undertaken. Broadly speaking there are three approaches to model validation and any combination of them may be applied as appropriate to the different aspects of a particular model. These approaches are:

(i) Expert intuition
(ii) Real system measurements
(iii) Theoretical results/analysis.

In addition, as suggested above, ad hoc validation techniques may be established for a particular model and system [18].

2.5.1 Internal Validation

The most common internal method of validating the model is least squares fitting. This method of validation is similar to linear regression and is the $R^2$ (squared correlation coefficient) for the comparison between the predicted and experimental
activities. An improved method of determining $R^2$ is the robust straight line fit, where data points are away from the central data points (essentially data points a specified standard deviation away from the model) are given less weight when calculating the $R^2$. An alternative to this method is the removal of outliers (compounds from the training set) from the dataset in an attempt to optimize the QSAR model and is only valid if strict statistical rules are followed. The difference between the $R^2$ value is less than 0.3 indicates that the number of descriptors involved in the QSAR model is acceptable. The number of descriptors is not acceptable if the difference is more than 0.3 [19].

$$R^2 = \left( \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\left( N \sum X^2 - (\sum X)^2 \right) \left( N \sum Y^2 - (\sum Y)^2 \right)}} \right)$$

(2.18)

Where:

$Y$ Dependent variable;

$X$ Independent variable;

$N$ Number of variable;

2.5.2 External validation

Several authors have suggested that the only way to estimate the true predictive power of a QSAR model is to compare the predicted and observed activities of an sufficiently large external test set of compounds. The problem in external validation is how can we select the training and test set. clearly discussed that how we can solve this problem in one of their article22. To estimate the predictive power of a QSAR model, Golbraikh and Tropsha recommended use of the following statistical characteristics of the test set14: (i) correlation coefficient $R$ between the predicted and observed activities; (ii) coefficients of determination $R^2$ predicted vs. observed activities, and observed vs. predicted activities; (iii) slopes $k$ and $k'$ of the regression lines through the origin. They consider a QSAR. Model is predictive, if the following conditions are:
\( R^2 \text{ Pred} > 0.6, \)

\( r^2 - r_0^2 / r^2 < 0.1, r_2 - r_0^2 / r^2 < 0.1 \) and

\( 0.85 \leq k \leq 1.15 \) or \( 0.85 \leq k' \leq 1.15 \)

(2.19)

The predictive ability of the selected model was also confirmed by external \( R^2 \) pred. A value of \( R^2 \) pred is greater than 0.6 may be taken as an indicator of good external predictability.

\[
R_{\text{pred}}^2 = 1 - \frac{\sum_{i=1}^{\text{test}} (Y_{\text{exp}} - Y_{\text{pred}})^2}{\sum_{i=1}^{\text{test}} (Y_{\text{exp}} - Y_{\text{tr}})^2}
\]

(2.20)

Where

\( Y_{\text{tr}} \) is the average value for the dependent variable for the training set [19].
CHAPTER 3

METHODOLOGY

3.1 Power load data collection

The previous power loadings of UTHM are obtained from the Department of Property and Development. The power loads are measured in kWh for each available electric substation. The power load readings are based on the TNB tariff bill in which stated the amount of active power consumed per electric substation. The load data are collected for each month from the year of 2006 to 2012 and these data are taken as the reference data for the entire forecasting works. This project intents to forecast the UTHM electricity consumption for the following year of 2013 and 2014. Table 3.1 shows the date of each electric substation was built.
Table 3.1: Year of each electric substation was built

<table>
<thead>
<tr>
<th>Electric substation</th>
<th>Year built</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES01</td>
<td>Before 2001</td>
</tr>
<tr>
<td>ES02</td>
<td>Before 2001</td>
</tr>
<tr>
<td>ES03</td>
<td>Before 2001</td>
</tr>
<tr>
<td>ES04</td>
<td>2001</td>
</tr>
<tr>
<td>ES05</td>
<td>2001</td>
</tr>
<tr>
<td>ES06</td>
<td>2004</td>
</tr>
<tr>
<td>ES07</td>
<td>2004</td>
</tr>
<tr>
<td>ES08</td>
<td>2005</td>
</tr>
<tr>
<td>ES09</td>
<td>2005</td>
</tr>
<tr>
<td>ES10</td>
<td>2005</td>
</tr>
<tr>
<td>ES11</td>
<td>2005</td>
</tr>
<tr>
<td>ES12</td>
<td>2007</td>
</tr>
</tbody>
</table>

3.2 Analysis

The actual data as mentioned in section 3.1 is used to determine the future load demand of UTHM network. The work is performed using EXCEL software to derive an equation based on the existing data. EXCEL curve fitting tool known in the insert and layout has the ability to generate various types of equations such as Linear regressions and polynomial. The next step is to use MATLAB to derive an ANN approach by training the neurons in the specific layers. These layers consists of single input layer, one or more hidden layers and single output layer. The input layer contains of a number of neurons equals to the number of input variables in the training network by an iterative process. The weights are adjusted using some learning algorithms. For the purpose of forecasting in this project, two types of equations have been chosen linear regression and polynomial equations, ANN will be used in dynamic load forecasting after the static load forecasting is conducted. Ten hidden layers is to be introduced with Levenberg-Marquardt algorithm. The forecasting work is conducted by substituting the variable parameters into each equation. EXCEL will be utilized to plot column graphs and line graphs for all
REFERENCES


[9] Introduction to Time Series Analysis


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


