THE EFFECT OF INTERVAL LENGTH AND MODEL BASIS ON FUZZY TIME SERIES ELECTRIC LOAD FORECASTING.

BY

ADEOLA KOLADE OMONIYI

(M.SC/ENG/47815/05-06)

A THESIS SUBMITTED TO THE POSTGRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE (M.Sc.) IN ELECTRICAL ENGINEERING

DEPARTMENT ELECTRICAL ENGINEERING
FACULTY OF ENGINEERING
AHMADU BELLO UNIVERSITY
ZARIA

2006

DECLARATION

I Adeola Kolade Omoniyi hereby declare that this thesis titled "The Effect of Interval Length and Model Basis on Fuzzy Time Series Electric Load Consumption Forecasting", presented to the Department of Electrical Engineering, Ahmadu Bello University, Zaria is the result of my own research and it has never been presented in any form, anywhere for the award of a higher degree.

All quotations and sources of information are acknowledged by means of references.

----------------------------------- ----------------------

Adeola Kolade Omoniyi Date

CERTIFICATION

This thesis titled "The Effect of Interval Length and Model Basis on Fuzzy Time Series Electric Load Consumption Forecasting" satisfies one of the requirements for the award of a Master of Science (M.Sc.) degree in Electrical Engineering and has been approved by the Department of Electrical Engineering Ahmadu Bello University, Zaria, for its contribution to knowledge and literary presentation.

------------------------ ----------------------

Dr M. B. Mu'azu Date
(Major Supervisor)

------------------------ ----------------------

Dr K. S. Farayola Date
(Supervisor)

-------------------------- ----------------------

Dr M. B. Mu'azu Date
(Head of Department)

----------------------------- -----------------------

Dean Postgraduate School. Date

DEDICATION

This project work is dedicated to God, my wife and my two wonderful children.

ACKNOWLEDGEMENT

All thanks to almighty God, the giver of life, wisdom and knowledge for making this dream a reality.

I am immensely grateful to my ever dependable parents Lt Col. and Mrs Adeola for their moral and financial support during the course of my educational pursuit.

My greatest profound gratitude to my wife, Mrs Omoboye Adeola and my two lovely children, (Erioluwa and Inioluwa Adeola). Your priceless love, support, care and prayers has brought me this far. My prayer is that you will not sow for others to reap. (Amen).

My deepest gratitude goes to my supervisors Dr. M.B. Mu'azu and Dr. K. S. Farayola who took the uphill task of supervising this work upon themselves out of their limited time. Your encouragement, advice and support have made this seemly impossible work possible. May the good Lord continue to shower his blessings on you and your families.

I also appreciate and would record my deep thanks to my friends Engr. Olanrewaju O. Matthew and Miss Nwankwo Charity for their unquantifiable support during the course of this research work. I wish them success in all their various endeavours.

Lastly, to all those who positively impacted on this research work but not individually mentioned I say thanks and our creator will stupendously reward them.

Adeola Kolade. O

JUNE, 2008.

TABLE OF CONTENTS

Chapter Two: Literature Review and Theoretical Background

Chapter Three

Chapter Four: Result and Analysis

Chapter Five: Conclusion and Recommendation

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

FTS:	Fuzzy Time Series
MBF:	Membership Function
PID:	Proportion-Integral-Derivative
MAX/MIN:	Maximum/Minimum Operator
MSE:	Mean Square Error
RMSE:	Root Mean Square Error
ARIMA:	Auto Regression Integration Moving Average
MA:	Moving Average
PHCN:	Power Holding Company of Nigeria
MW/MVA:	Megawatts/Mega Volt Ampere

ABSTRACT

Fuzzy Time Series (FTS) forecasting technique is the amalgamation of fuzzy logic and time series technique. The critical issue in FTS forecasting is the determination of the interval length. This paper therefore, is a research on the effect of varying interval length and model basis on electric load forecasting using Fuzzy Time Series Model. The methodology adopted is presented and the data used is the load (in MW/MVA) obtained from PHCN over a 24-week period. The data for 18 weeks is used as the test data while the remaining 6 weeks is the validation data. It is shown that varying interval length and model basis give different forecasting results and that interval length five gives a significantly better result than others based on the quantitative and qualitative performance test. Furthermore, the results obtained show that model basis of four gives better forecasting result when compared to model basis of five and six. The results obtained are presented and discussed from the standpoint of their degree of consistency exhibited by the two elements.

CHAPTER ONE

1.1 INTRODUCTION

The prediction of time series is an important problem in monitoring, diagnosis, control and decision support for technical and non-technical systems. While there have been many conventional time series models, one particular important group of model has been the family of Fuzzy Time Series. The advent of Fuzzy logic made it possible to tackle a many problems with Fuzzy input. One of them is the forecasting problem.

Studies have shown that this model seem to be more appropriate in forecasting, since it could address the problem of time-dependent actuating variables (e.g. temperature, global solar radiation, etc) which other traditional models such as statistical, adaptive, smooth dynamic series, auto regression integration moving average (ARIMA) models could not resolve without great inaccuracy. Furthermore, the necessary conditions for applying the conventional time series could be removed in Fuzzy Time Series model.

The Fuzzy Time Series model can be roughly categorized according to how they formulate the relationships among observations, that is, Fuzzy rules, Fuzzy function, Fuzzy relationship and others.

Prediction of electric load power consumption can also be resolved using Fuzzy model. The electricity consumption profile of a power generation and distribution company such as Power Holding Company of Nigeria (PHCN) can after elimination of possible long term trends be regarded as a stationary time series with seasonal characteristics. It is extremely important for an optimal management of generation and distribution of electric energy to have as precise as possible the prediction of the load to be expected.

However, in applying Fuzzy model to forecasting, the determination of the length of interval is critical. In many previous models, interval length and models basis were set arbitrarily. No explanations were provided to determine the length of the interval. However, one recent study [1] demonstrated that varying interval length and model basis could have great impact on the forecasting result. This study, therefore, is aimed at investigating the effect of varying interval length and model basis on electric load forecasting.

A time series which contains load values for a period of 6 months (24 weeks) is employed to do this.

1.2 PROJECT MOTIVATION

Determination of interval length and model basis in Fuzzy Time Series has been a critical issue. There has not been any empirical method of determining an appropriate interval length and as such it becomes very important that an investigation be carried out to determine the effect of varying the interval length on the forecasting result. Because of the critical nature of electricity in the economic growth and well being of any economy, any method that can enhance forecasting load demand or load consumption pattern becomes very important.

1.3 STATEMENT OF PROBLEM

The electricity consumption profile of a power distribution company can, after elimination of possible long-term trends, be regarded as a stationary time series. It is extremely important that for an optimal management of electric energy, the load prediction should be as precise as possible.

This thesis is aimed at investigating the effect of varying the interval length and model basis in a bid to improving the forecasting result of electric load power consumption.

1.4 METHODOLOGY

There are many methods in forecasting time series. They include:

i. Moving average

ii. Auto correlation

iii. Auto Regression Integration Moving Average (ARIMA)

iv. Fuzzy Time Series etc.

Moving average (MA): This is one of the statistical methods of forecasting. It provides a set of very powerful indicator for tracking trend and trend reversals. Moving average is a lagging indicator, or trend following formula, that smoothens the volatile swings in a market. It attempts to tone down the fluctuations of market prices to a smoothed trend, so that distortions are reduced to a minimum. Essentially, Moving averages is a method for estimating incidence density when the time period spans several years. The main drawback to using Moving averages is that broadly speaking markets spend more time locked in ranges than actually trending.

Autocorrelation: This is the correlation (relationship) between members of a time series of observations, such as weekly share prices or interest rates, and the same value at a fixed time interval inter. It is also a statistical method of forecasting. More technically speaking, autocorrelation occurs when residual error terms from observation of the same variable at different times are correlated (related).

Autocorrelation measures the association or mutual dependence between values of the same time series and different time lags.

Auto Regression Integration Moving Average (ARIMA): This is the integration of autoregressive process as well as moving average parameters and it explicitly includes differencing in the formulation of the model. Specifically, the types of parameters in the model are the autoregressive parameters (P) the number of differencing passes (d) and the moving average parameter (q).

However, ARIMA is a complex technique, it is not easy to use, as it requires a great deal of experience, and the method is appropriate only for a true series that is stationary. That is, its mean, variance, and autocorrelation should be approximately constant through time [24].

The method applied in this study is Fuzzy Time Series. Different research work on Fuzzy Time Series proposed different method but the method adopted by ***Abbasov et al*** is used in this study since it allows the prediction beyond the time available in the test data.

The following is the suggested methodology in carrying out the Fuzzy Time Series forecasting:

i. Obtaining electric load power consumption data from a PHCN over a 24 – week period;

ii. Partitioning the data into training data (18-weeks) and validation data (6 weeks);

iii. Defining the universal set U containing the interval between least and greatest variation of load;

iv. Dividing the Universal set U into varying interval lengths (5, 6, and 7) containing variation values corresponding to different load consumed;

v. Determining the respective value of linguistic variable or the Fuzzy set (t) i.e. the qualitative description of variation values of total load as a linguistic variable;

vi. Fuzzifying the input data or the conversion of numerical value into Fuzzy value;

vii. Selecting the parameter $w>1$(model basis) corresponding to the time period prior to the concerned week;

viii. Calculating the Fuzzy matrix $p^w(T)$ and forecasting of the expected load for the proceeding week;

ix. DeFuzzying the obtained result or conversion of Fuzzy into quantitative (crisp) value; and

ix. Tabulating the result of different interval and model basis and comparing the results with the validation data.

1.5. PROJECT OUTLINE

The thesis is divided into five chapters: Chapter one introduces the research work where the objectives of the research are defined and the methodology applied is explained. The review of literature of similar research work with the theoretical background that forms the bases for the evaluation of the different available options are contained in chapter two. Chapter three discusses the methodology in achieving the thesis aims. While Chapter four deals with the analysis of the results obtained. Chapter five contains limitations, conclusion and recommendation. References are provided at the end of this thesis.

CHAPTER TWO

LITERATURE REVIEW AND THEORETICAL BACKGROUND

2.1 LITERATURE REVIEW

Since the introduction of Fuzzy set theory [2] by ***Zadeh*** in 1965, the concepts of Fuzzy logic has been widely applied. This can be attributed to the nature of Fuzzy systems having tools, which can cope with the vagueness of human languages. Based upon the works by ***Zadeh***, Fuzzy Time Series models were defined and studied. In the work of ***Song*** **et al**, Fuzzy relational equations were employed, definition of various Fuzzy time series was given, some important properties of Fuzzy time series was explored and a set by set procedure for implementation of Fuzzy Time Series was suggested to analyze the special dynamic process with linguistic values [3] [4].

Basically, Fuzzy set theory is non-parametric. The main advantage of Fuzzy approach is that human experience and knowledge can be applied from the start till the end of the forecasting procedure. Furthermore, Fuzzy Time Series model has been proved as a robust and effective tool for variety of forecasting problems. When the historical data on hand are linguistic values, the traditional time series methodologies fail to work. Under such conditions, the Fuzzy Time Series can be developed and applied. Since it has been successfully applied in different areas, such as decision making, control theory, business analysis and forecasting, more and more computer scientists and statisticians have developed interest in the computational potentials of Fuzzy theory [5] [6]. ***Hirota*** [7] used Fuzzy concepts of membership and vagueness to build up a law and other knowledge system. ***Konolige et al*** [8] applied Fuzzy controller in a hierarchical fashion in the field of multiple agent process control.

Song et al [9] – [11] proposed a time-invariant and time variant Fuzzy Time Series models to solve the university enrollment time series forecasting problem. ***Abbasov et al*** explored the possibilities of using Fuzzy set theory or the apparatus of sets, going by the name Fuzzy logic to model a demographic process of population forecasting [13]. ***Otto et al*** demonstrated electric load forecast based on Fuzzy algorithm.

Several other researchers have proposed different methods based either on improving the average forecasting error or increasing the speed of computation. ***Sullivan et al*** [14] have proposed a new algorithm based on Markov model using linguistic labels with probability distributions. ***Kunhuamg et al****,* in their work on the application of neural network to forecast Fuzzy Time Series suggested, using two models (a basic model and hybrid model) using neural network approach to forecast all of the observations [15].

The emphases on these models have been on the establishment of Fuzzy relationships among observations. It is common for these models to include the following steps:

i. Define the universe of discourse and interval for the observations;
ii. Partition the universe based on the interval;
iii. Define the Fuzzy set for the observation;
iv. Fuzzify the observations;
v. Establish the Fuzzy relationships;
vi. Perform the forecast; and
vii. Defuzzify the forecasting result.

However, one critical factor is the length of intervals for the observations to be fuzzified as in Fuzzy Time Series models; the very first step is to determine length of intervals. In many previous models, for example, enrollment forecasting, the length of the intervals were set the same and all were arbitrary (possibly depending on the sample population) [16] [17]. No explanations were provided as to how the length of these intervals was determined. However, the issues of length of interval had never been discussed until the effective lengths of interval were proposed. That study

[18] showed that different length of intervals may result in different Fuzzy relationships and in turn different forecasting results. And the effective length of interval did improve forecasting result in various empirical analyses. Thus, determining useful length of interval is certainly important for Fuzzy Time Series forecasting.

Furthermore, how the determined length of interval can be adjusted during the formulation of Fuzzy relationship so that the Fuzzy relationship can be captured more appropriately is also critical for improving Fuzzy Time Series forecasting.

2.2 FUZZY SETS THEORY AND FORECASTING

The Fuzzy sets theory can be defined as a mathematical formulation that enables the elimination of indefiniteness and deal with incomplete, inaccurate information of both qualitative and quantitative nature. The Fuzzy sets theory, advanced by L. **Zadeh** [8], one of the well known representative of modern applied mathematics, excluded any definite description of the task and doing this offers such a solution scheme to the problem such that subjective reasoning and evaluation plays a principal role. Thus, anyone, encountering indefinite, incomplete information or data, can form some conclusion, by passing through reasoning all these realities. The use of

Fuzzy verbal notation in every day speech such as, much, more, little, small, many, a number of, etc, enables one to give a qualitative description of the problems which must be tackled taking account of its indefinite nature as well as obtain explanation of the factors that can not be described qualitatively.

The advent of Fuzzy logic made it possible to tackle a great many problems with Fuzzy input data. One of them is the forecasting problem. Many of the structural elements of forecasting (input data and interdependence between its components, interval evaluation of indicators and their interdependence, expert evaluation and judgment etc) are either of a Fuzzy nature or by being in Fuzzy relationship; condition the Fuzzy description of the problem. The application of Fuzzy logic to the handling of forecasting problem was undertaken by researchers in which the mathematical model of time series was describe in a Fuzzy form for handling the problems with Fuzzy input data. This approach was developed later by other scientists to resolve analogous problem. The consumer energy consumption feature, functioning under indefinite, uncertain circumstances, conditions the fuzziness of input data or loads the task unto Fuzzy environment. Therefore, from both theoretical and practical stand point, handling the concerned problem based on Fuzzy time series will be more expedient.

2.2.1 INTRODUCTION TO FUZZY LOGIC.

Fuzzy Logic was designed to mathematically represent uncertainty and vagueness and to provide formalized tools for dealing with the imprecision intrinsic to many problems. [22]. The Fuzzy Logic System is a computing framework based on the concepts of Fuzzy Set theory, Fuzzy IF-THEN, and Fuzzy reasoning. It has applications in areas such as automatic control, data analysis, time series prediction, and robotics and pattern recognition. Its basic structure consists of three components: a Rule-Base, which contains a selection of Fuzzy rules; a data-base, which defines the membership function (MBF) used in the Fuzzy rules; and a Reasoning Mechanism, which performs the inference procedure [28].

Fuzzy logic systems are plain and simple to understand, as the 'intelligence' of a system is not hidden in differential equations or source code. The use of linguistic modeling greatly enhances system characteristic and the potential leading to quick development cycles. The primary building block of any Fuzzy Logic system is linguistic variable and the possible values of a linguistic variable are presented by linguistic terms. [30].The knowledge base of Fuzzy Logic System is composed of a database and a rule base. The database defines the Membership Functions for the linguistic terms and the rule base represents Fuzzy control rule. Therefore the Membership Function and the Fuzzy rules greatly affect control performance.[24]. Rules are influential in selecting the number of variables and Membership Functions to be modeled with Fuzzy Logic, because the model becomes exponentially more complex as the number of variables or Membership Function increases. This is because a rule must be available for each possible combination of the input variable Membership Function (MBFs) and potential output Membership Functions (MBFs).

Fuzzy Logic (controller) System is an appealing alternative to such conventional methods as PID (Proportional-Integral-Derivation) control and model-based control. This is especially true where a system follows some general operating characteristics and a detailed process understanding is unknown or traditional models become overly complex [24]. Conventional PID fails if the task is non-linear and model-based methods also suffer considering the kinds of assumption that may be made in deriving the model and the kinds of differential equations that may be derived [25].

2.2.2 FUZZY LOGIC OPERATORS

The basic operator: intersection (AND), Union (OR) and Complement (NOT) also exit for Fuzzy Set but defined differently as follows:

AND: The intersection of A and B;

$\boldsymbol{\mu(A \cap B)} = \mathrm{Min}\{\mu(A), \mu(B)\}$;

OR: The union of A and B;

$\boldsymbol{\mu(AUB)} = \mathrm{Max}\{\mu(A), \mu(B)\}$; and

NOT: The complement of a Fuzzy Set is defined as:

$\boldsymbol{\mu(A')} = 1 - \mu(A)$.

where μ = Membership Function. This is defined over the range of input and output variable values and linguistically describes the Universal of Discourse.

2.2.3 FUZZY MEMBERSHIP FUNCTION

The use of Fuzzy Set by Membership Function (MBF) in logical expression is called Fuzzy Logic. The Membership Function (MBFs) define the degree to which the value of a variable belongs to the universal set and is usually a linguistic term, for example Low, High etc. the linguistic terms normally describe a concept related to the value of the linguistic variable. The degree of membership in a set becomes the degree of truth of a statement [31]. This is the reason why in Fuzzy Logic, everything is a matter of ‘degree’ and such exact reasoning is viewed as a limiting case of approximate reasoning.

The main disadvantage of traditional (crisp) set theory is that it implies an aura of precision and definiteness for a decision that may not be warranted.

In real-world systems, crisp is an artificial result of digitizing analogue data. Data in a "boundary" condition, that is, samples that could be said to be a member of more that one class may be categorized differently if a different technique of digitizing the data is used. A boundary condition sample X may be assigned to class A using crisp technique T, but to class B using crisp technique T_1, and there is no way to gauge from the results which assignment is more appropriate. Also it is impossible to tell whether sample X is more or less useful to a calculation than a second sample Y, which is classified as A by both techniques, and therefore more strongly representative of that class.

Fuzzy set theory replaces the crisp "is a member or is not a member" classification by assigning each sample in value of "how closely it represents each given class". Fuzzifying the techniques means that X would be seen to have almost equal membership in A and B, no matter which Fuzzy technique is used, in turn clearly distinguishing it from Y.

Let us assume that $U = \{u_1, u_2 \ldots u_n\}$ is a universal time set. The Fuzzy set A universal set U is defined as follows:

$$A = \{(\mu_A (u_1)/u_1, \mu_A(u_2)/u_2 \ldots\ldots \mu_A(u_n)/u_n)\} \quad \ldots\ldots(2.1)$$

$$A = \{(\mu_A (u_i)/u_i), u_i [0, 1]\} \quad \ldots\ldots(2.2)$$

Where μ_A (u_i) – membership function, μ_A (u_i):u = [0, 1]: μ_A (u_i) is a degree of belonging of U_i to the set A μ_A $(u_i) \leq [0, 1]$, "/" is a division sign.

Let us assume that Y(t) (t = 0, 1, 2, …), which is a subset of set R of real numbers, is simultaneously a universal set on which is defined a Fuzzy set $u_i(t)$, (t = 1,2, ..), that is, the Membership Function is time dependent. Let us define a set F(t) arranged out of [u_i (t), t = 1, 2, …..] more precisely, F(t) is a set of Fuzzy sets F(t) = {u_i (t) t = 1, 2, …..}. Then F(t) is a Fuzzy Time Series defined as a linguistic variable of the Fuzzy Sets. It is evident, if F(t) is accepted as a linguistic variable, the Fuzzy Sets {u_i (t), t = 1, 2…} contained in F(t), will assume the possible corresponding values of F(t). Besides, as is evident, F(t) is time-dependent, which means the function F(t) will assume different value at different time moments.

2.3 TIME SERIES

A time series is sequence of observations which are ordered in time (or space). If observations are made on some phenomenon throughout time, it is most sensible to display the data in the order in which they arose, particularly, since successive observations will probably be dependent. Time series are best displayed in a scattered plot. Time is called the independent variable (in this case, however, something over which there is little control).

There are two kinds of time series data:

1. Continuous: where there is an observation at every instance of time, e.g. lie detectors, electrocardiograms. The observation x at time T is denoted at X (t); and
2. Discrete: Where there is an observation at (usually regular) spaced intervals. Denoted at X (t).

To increase the understanding of time series, its main futures are as follows:

a. Trend components – Trend is a long – term movement in a time series. It is the underlying direction (an upward or downward tendency) and rate of change in a time series, when allowance has been made for other components. A simple way of detecting trend in seasonal data is to take averages over a certain period. If these averages change with time, then there is evidence of the trend in the series;
b. Cyclical Components – In weekly or monthly data, the cyclical component describe any regular fluctuations. It is a non-seasonal component which varies in a recognizable cycle;
c. Seasonal components – The seasonal component of a weekly or monthly data which is often referred to as seasonality is the

component of variable in a time series which is dependent on time of the year. It describes any regular fluctuation with a period of less than one year; and

d. Irregular components – The irregular component is the left over when other components of the series (trend, seasonal and cyclical) have been accounted for [17].

Fuzzy time series represent a consecutive series of observations that is conducted, by equal time interval and lies at the root of exploring real processes in economics, metrology and natural sciences etc. The analyses of time series of observation consist of the following:

i. Constructing the mathematical models of time series of observation of real processes;

ii. Model identification or selection of quantitative evaluation or estimation method for accessing model parameters in order to test the extent to which the models are adequate to reflect the real process; and

iii. The conversion of identification model into series through the statistical evaluation parameters.

It is assumed that the time interval $0 \leq t \leq T$ of process X (t) is observed, that is, the parameters t varies along the term interval, [0, T] or assumes any integer belonging to the interval. For every fixed time moment, t = s, the value of function, beginning from this moment is generally determined by the function argument at all the time moment ranging from $t = 0$ to $t = s – 1$, and value of function at the time moment ranging from $t = 0$ to $t = s – 2$.

CHAPTER THREE

3.1 FORECASTING METHODOLOGY

The Fuzzy time series technique proposed by Abbasov et al: [28] will be applied to the energy consumption data obtained from PHCN on a weekly basis. The methodology described earlier on will be used and defined explicitly for this data. The complete data set used in developing the Fuzzy time series forecast is 24 Weeks data. The first 18 weeks is used as a training set, for the forecast while the test data (6 week) is used as the validation data as shown in Table 3. 1

Table 3.1: The variation of load energy consumption for the period of 18 weeks.

Weeks	Energy consumed in MW	Variation in MW
1	2357.1	
2	2137.7	-2194
3	2114.4	-23.3
4	2221.7	107.3
5	2323.6	101.9
6	2284.1	-39.5
7	2301	16.9
8	2210	-91
9	2220.6	10.6
10	2247.3	26.7
11	2155.1	-92.2
12	2249.6	94.5
13	2191.4	-58.2
14	2273.7	82.3
15	2127.3	-146.4
16	2131	3.7
17	2080	-51
18	2146	66

The methodology for the forecast is described in the following steps:

i. Determining the universal set. Table 3.1 gives the dynamics of the total consumed energy over a period of 18 weeks (input data for retrospective forecast) and variation in the total consumed energy between every next and previous week. Variation for the current week is understood to be the difference between the energy consumed in the current week and that consumed in the previous week. For example, variation for week 10 is equal to 2247.3 – 2220.6 = 26.7. To define a universal set U, first of all, the smallest and greatest variation values must be formed over the period (week 1 to week 18). To ensure the smoothness of boundaries of the interval, adequate values D_1 and D_2 (positive figures) are selected [28]. After that the universal set U can be defined as:

$$U:U = [V_{min} - D_1, V_{max} + D_2] \qquad (3.1)$$

where V_{min} =-219.4 is the smallest variation (week 2), V_{max}=107.4 is the greatest variation (week 4) and D_1 and D_2 are taken to be 0.6 and 2.6 respectively, thus, the universal set U will be as follows:

U = [-219.4 – 0.6, 107.4+2.6]

U = [-220, 110]

ii. **Dividing the universal set U into several equal length intervals**. Three cases would be considered to test and to draw up conclusions on the effect of interval length on the overall forecasting result.

Case A: 5 Equal length intervals

Case B: 6 Equal length intervals

Case C: 7 Equal length intervals

CASE A: FIVE EQUAL LENGTH INTERVAL

The universal set is partitioned into five equal length intervals as follows:

U_1 = [-220, -154]

U_2 = [-15, -88]

U_3 = [-88, -22]

U_4 = [-22, 44]

U_5 = [44, 110]

In order to account for the fact that forecasting with Fuzzy time series exhibits the least average error [28], it is necessary to find the middle points of the intervals e.g. the midpoint of U_1 is

$$U_m^1 = \frac{(-220)+(-154)}{2} = -187 \quad (3.2)$$

And

U^2_m = -121

U^3_m = -55

U^4_m = 11

U^5_m = 77

iii. Determining the Fuzzy sets. Fuzzy sets are defined on the universal set U. In this case “the variation in total load energy consumption” is a linguistic variable that assumes the following linguistic values:

A_1 = Significant Decrease in load consumption;

A_2 = Moderate Decision in load consumption;

A_3 = Minimal change in load consumption;

A_4 = Moderate increase in local consumption; and

A_5 = Significant increase in load consumption.

For example, the linguistics value “significant decrease in load consumption” is given by the variable < [-220, -154], A_1 >, where A_1 is the Fuzzy set defined on the domain [-220, 110] of the universal set U.

The Fuzzy sets, A_1 ……., A_5 are defined on U by the following relationship, according to **Abbasov et al** [28].

$$\mu_{Ai}(U_i) = \frac{1}{1 + \left[C\left(U - U^i{}_m\right)\right]^2} \qquad (3.3)$$

Where U = variation, U^1_m is the middle point of the corresponding interval; C is a constant.

There is no known method of determining thus constant but according to **Abbasov et al** [27], it must be chosen in such a way that it ensures the conversion of definite quantitative value into Fuzzy values or belonging to the interval $A_i = (\mu_{Ai}(u_i/u_i), u_i \in u\ \mu_{Ai}(u_i) \in [0,1]$. A value of

C = 0.006 is chosen for this work.

Consequently,

$A_1 = 1/U_1 + 0.86/U_2 + 0.61/U_3 + 0.41/U_4 + 0.28/U_5$ …………….(3.4)

$A_2 = 0.86/U_1 + 1/U_2 + 0.86/U_3 + 0.61/U_3 + 0.41/U_5$ ....(3.5)

$A_3 = 0.61/U_1 + 0.86/U_2 + 1/U_3 + 0.86/U_4 + 0.61/U_5$ ....(3.6)

$A_4 = 0.41/U_1 + 0.61/U_2 + 0.86/U_2 + 1/U_4 + 0.86/U_5$ ....(3.7)

$A_5 = 0.28/U_1 + 0.41/U_2 + 0.61/U_3 + 0.86/U_4 + 1/U_5$ ....(3.8)

The continuous membership function of Fuzzy sets A_1 depicting the values of the linguistic values of the variation of load consumption are shown in Figure 3.1.

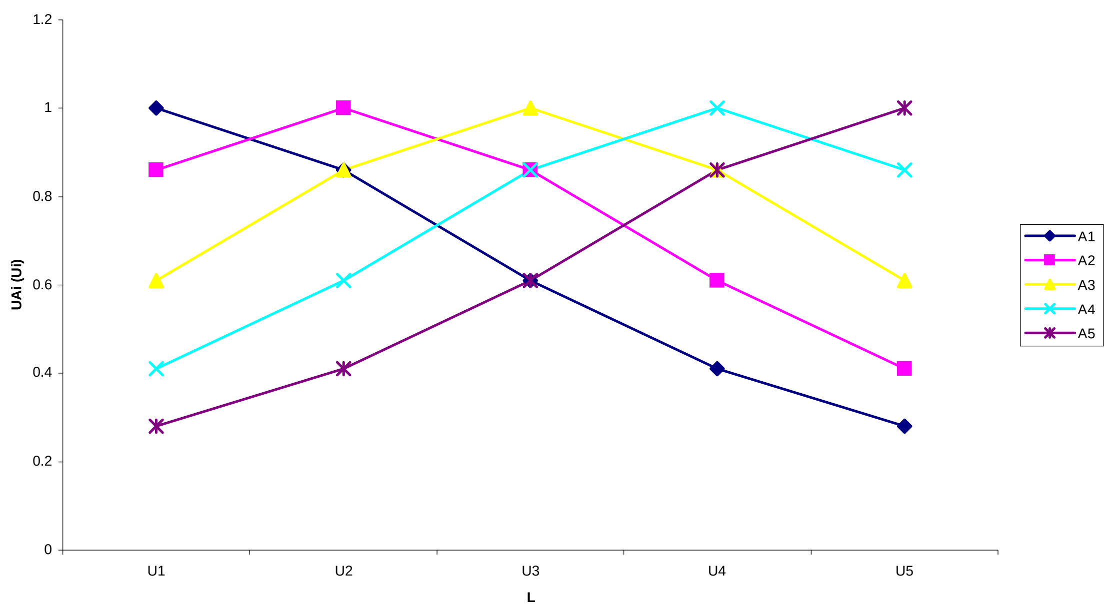

Figure 3.1 Member function (MBF) of set of Fuzzy interval

iv. **Fuzzifying the input data** that is, converting numerical values into Fuzzy values. This is also achieved using equation (3.3). This time, if B_n, $B_n \in y_i$ is a variation for the I-th week, then membership function for $\mu(\mu_i)$ is calculated by means of equation (3.3) by holding valid the equality $y = B_n$, that is, to say by separating the interval, to which belongs the considered variation from the universal set U. The result is shown in Table 3.2

Table3.2 Fuzzification table with interval length of 5

Week	Energy consumption in (MW)	Variation	Constant	U1	U2	U3	U4	U5
1	2357.1		0.006					
2	2137.7	-219.40	0.006	0.964	0.742	0.507	0.344	0.240
3	2114.4	-23.30	0.006	0.509	0.744	0.965	0.959	0.734
4	2221.7	107.30	0.006	0.243	0.348	0.513	0.750	0.968
5	2323.6	101.90	0.006	0.250	0.359	0.530	0.771	0.978
6	2284.1	-39.50	0.006	0.561	0.807	0.991	0.916	0.672
7	2301.0	16.90	0.006	0.401	0.594	0.843	0.999	0.885
8	2210.0	-91.00	0.006	0.751	0.969	0.955	0.728	0.496
9	2220.6	10.60	0.006	0.416	0.616	0.866	1.000	0.863
10	2247.3	26.70	0.006	0.378	0.560	0.806	0.991	0.917
11	2155.1	-92.20	0.006	0.756	0.971	0.953	0.723	0.492
12	2249.6	94.50	0.006	0.260	0.374	0.554	0.799	0.989
13	2191.4	-58.20	0.006	0.626	0.876	1.000	0.853	0.603
14	2273.7	82.30	0.006	0.277	0.402	0.596	0.845	0.999
15	2127.3	-146.40	0.006	0.944	0.977	0.769	0.529	0.358
16	2131.0	3.70	0.006	0.433	0.641	0.890	0.998	0.838
17	2080.0	-51.00	0.006	0.600	0.850	0.999	0.878	0.629
18	2146.0	66.00	0.006	0.303	0.443	0.655	0.902	0.996

v. Selecting the Parameter (w > 1) corresponding to the time period prior to the concerned week is used in the calculation of the relationship matrix.

Based on the choice of W (the model basis) or the past weeks, a Fuzzy relationship matrix $R^w(t)$ is calculated by means of which is given a forecast. The relationship matrix R^w (t) is then determined from the expression.

$$R^w(t) = 0^w(t) \ ^o \ C^w(t) \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (3.9)$$

Where 0^w (t) is the operation matrix (i x j), $C^w(t)$ is the criterion matrix (i x j) and '^{o}' is the operation Min (n). (Where i is the number of rows, which conforms to the sequence of week t – 2, t – 3 . . . , t – w and j is the number of columns conforming to the number of variation intervals (a row matrix corresponding to Fuzzy variation in load consumption for the week t – 1).

The choice of w is very critical to Fuzzy series analysis and as such a sensitivity analysis is carried out for the Fuzzy time series model with varying values of w in order to determine which one gives the least forecasting error. This is carried out using values of w = 4, 5, and 6 and the outcomes are as shown in Table 3.3

Table 3.3 forecasting Error with values of W

		w=4		w=5		w=6	
Interval length	Actual Load	Forecasted Load	Square Error	Forecasted Load	Square Error	Forecasted Load	Square Error
5	2119.90	2123.10	10.24	2127.90	64.00	2127.90	64.00
6	2119.90	2130.00	102.01	2136.50	275.56	2123.00	9.61
7	2119.90	2125.60	32.49	2130.70	116.64	2130.90	121.00
		RMSE	6.95	RMSE	12.33	RMSE	8.05

The Root Mean Square Error for w=5 and w=6 are similar considering the procedure as subsequently explained. On the basis of Table 3.3, the model basis, w=6 was selected.

As an example, for week 19, a 5 x 5 operation matrix will be formed as follows:

$$0^6 \text{ (week 19)} = \begin{vmatrix} \text{Fuzzy variation for week 13} \\ \text{Fuzzy variation for week 14} \\ \text{Fuzzy variation for week 15} \\ \text{Fuzzy variation for week 16} \\ \text{Fuzzy variation for week 17} \end{vmatrix}$$

$$0^6 \text{ (week 19)} = \begin{vmatrix} 0.663 & 0.874 & 1.000 & 0.859 & 0.642 \\ 0.327 & 0.401 & 0.592 & 0.839 & 0.998 \\ 0.945 & 0.978 & 0.773 & 0.534 & 0.362 \\ 0.434 & 0.640 & 0.886 & 0.999 & 0.847 \\ 0.601 & 0.849 & 0.999 & 0.884 & 0.638 \end{vmatrix}$$

A 1 x 5 criterion matrix will be formed as follows

$$C^6 \text{ (week 19)} = |\text{Fuzzy variation for week 18}|$$

$$C^6 \text{ (week 19)} = |0.303\ 0.442\ 0.651\ 0.896\ 0.997|$$

The relationship matrix, R^w (t) is then determined from equation (3.9). resulting in:

$$R^w \text{ (week 19)} = \begin{vmatrix} R_{11} & R_{12} & R_{13} & R_{14} & R_{15} \\ R_{21} & R_{22} & R_{23} & R_{24} & R_{25} \\ R_{31} & R_{32} & R_{33} & R_{34} & R_{35} \\ R_{41} & R_{42} & R_{43} & R_{44} & R_{45} \\ R_{51} & R_{52} & R_{53} & R_{54} & R_{55} \end{vmatrix}$$

$$\begin{vmatrix} 0.303 & 0.442 & 0.651 & 0.859 & 0.612 \\ 0.303 & 0.401 & 0.592 & 0.839 & 0.997 \\ 0.303 & 0.442 & 0.651 & 0.534 & 0.362 \\ 0.303 & 0.442 & 0.651 & 0.896 & 0.847 \\ 0.303 & 0.442 & 0.651 & 0.884 & 0.638 \end{vmatrix}$$

The final fuzzified predicted variation F^w (t) is given as:

$$F^6 \text{ (week 19)} = \begin{vmatrix} \text{Max } (R_{11} & R_{12} & R_{13} & R_{14} & R_{15}) \\ \text{Max } (R_{21} & R_{22} & R_{23} & R_{24} & R_{25}) \\ \text{Max } (R_{31} & R_{32} & R_{33} & R_{34} & R_{35}) \\ \text{Max } (R_{41} & R_{42} & R_{43} & R_{44} & R_{45}) \\ \text{Max } (R_{51} & R_{52} & R_{53} & R_{54} & R_{55}) \end{vmatrix}$$

F^6 (week 19) = |0.303 0.442 0.651 0.902 0.996|

vi. **Defuzzifying obtained result**, that is, conversion into quantitative values. In order to achieve this, **Abbasov et al** [28] proposed the following formula:

$$V(t)=\frac{\sum_{i=1}^{5}\mu_t(U_i).U_m^i}{\sum_{i=1}^{5}U_t(U_i)} \quad \ldots\ldots (3.10)$$

Where V (t) is the expected variation for the forecast week t, $\mu_t(U_i)$ is the calculated membership function (MBF) for the forecast week t and U^i_m the middle points of the interval (the size of the interval is 5).

Therefore, for week 19,

$$V(\text{week 19}) = \frac{0.303x-187+0.442x-121+0.65x-55+0.902x11+0.996x77}{0.33+0.442+0.65+0.902+0.996}$$

$$V(\text{week 19}) = \frac{-59.32}{3.294} = -18.04$$

This means that there will be a moderate decrease in load consumption from the value (V (week 19)) for the load consumption in week 18. Therefore,

R (week 19) = R (week 18) + V (week 19) = 2146 – 18.04 = 2127.96

Similarly, the load consumption value for week 20 is determined. The same approach is used to determine the forecasted load consumption for other weeks as shown in Table 3.4

Table 3.4 Forecasted Load consumption for week 19 to week 24 using interval length of 5

Week	Actual Load consumption in MW	Forecasted Load consumption in MW
19	2119.9	2127.5
20	2061.9	2079.9
21	1997.6	2030.9
22	1989	1976.1
23	1968.7	1960.6
24	1989.4	1994

CASE B: SIX EQUAL LENGTH INTERVAL

Similarly, for case B that the universal set is partitioned into six equal lengths, the same approach from step 1 to 6 is used.

Step (1)

U_1 = [-230, -170]

U2 = [-170, -110]

U_3 = [-110, -50],

U_4 = [-50, 10]

U_5 = [10, 70],

U_6 = [70, 130].

Also the midpoints, as defined in equation (3.2), are:

$U^1_m = -200$

$U^2_m = -140$

$U^3_m = -80$

$U^4_m = -20$

$U^5_m = 40$

$U^6_m = 100$

Table 3.5 shows the fuzzification of Table 3.1 with interval length of six

Table 3.5: Fuzzification Table with Interval Length of Six

Week	Energy consumption in MW	Variation	Constant	U1	U2	U3	U4	U5	U6
1	2357.1		0.006						
2	2137.7	-219.40	0.006	0.987	0.815	0.588	0.411	0.292	0.214
3	2114.4	-23.30	0.006	0.471	0.671	0.896	1.000	0.874	0.646
4	2221.7	107.30	0.006	0.227	0.312	0.442	0.632	0.860	0.998
5	2323.6	101.90	0.006	0.234	0.322	0.456	0.651	0.879	1.000
6	2284.1	-39.50	0.006	0.519	0.733	0.944	0.986	0.815	0.588
7	2301.0	16.90	0.006	0.371	0.530	0.747	0.953	0.981	0.801
8	2210.0	-91.00	0.006	0.700	0.920	0.996	0.846	0.618	0.432
9	2220.6	10.60	0.006	0.385	0.551	0.772	0.967	0.970	0.777
10	2247.3	26.70	0.006	0.351	0.500	0.709	0.927	0.994	0.838
11	2155.1	-92.20	0.006	0.705	0.924	0.995	0.842	0.614	0.429
12	2249.6	94.50	0.006	0.243	0.336	0.477	0.679	0.903	0.999
13	2191.4	-58.20	0.006	0.580	0.806	0.983	0.950	0.742	0.526
14	2273.7	82.30	0.006	0.258	0.360	0.513	0.726	0.939	0.989
15	2127.3	-146.40	0.006	0.906	0.999	0.863	0.635	0.444	0.314
16	2131.0	3.70	0.006	0.401	0.574	0.799	0.980	0.955	0.750
17	2080.0	-51.00	0.006	0.556	0.778	0.971	0.967	0.770	0.549
18	2146.0	66.00	0.006	0.282	0.396	0.566	0.790	0.976	0.960

Step 2: Selection of Parameter

$$0^6(\text{week } 19) = \begin{vmatrix} 0.580 & 0.806 & 0.983 & 0.950 & 0.742 & 0.526 \\ 0.258 & 0.360 & 0.513 & 0.726 & 0.939 & 0.989 \\ 0.906 & 0.999 & 0.863 & 0.635 & 0.444 & 0.314 \\ 0.401 & 0.574 & 0.799 & 0.080 & 0.955 & 0.750 \\ 0.556 & 0.778 & 0.971 & 0.967 & 0.770 & 0.549 \end{vmatrix}$$

$$C^6 (\text{week } 19) = |0.282 \; 0.396 \; 0.566 \; 0.790 \; 0.976 \; 0.960|$$

$$\begin{vmatrix} 0.282 & 0.396 & 0.566 & 0.790 & 0.742 & 0.526 \\ 0.258 & 0.360 & 0.513 & 0.726 & 0.939 & 0.960 \\ 0.282 & 0.390 & 0.566 & 0.635 & 0.444 & 0.314 \\ 0.282 & 0.396 & 0.566 & 0.790 & 0.955 & 0.750 \\ 0.282 & 0.396 & 0.566 & 0.790 & 0.770 & 0.549 \end{vmatrix}$$

$$C^6 (\text{week}) \quad |0.282 \; 0.396 \; 0.566 \; 0.790 \; 0.939 \; 0.960|$$

$$V(\text{week } 19) = \frac{0.282\text{x}-200+0.396\text{x}-140+0.566\text{x}-80+0.790\text{x}-20+0.939\text{x}40+0.960\text{x}100}{0.282+0.390+0.566+0.790+0.939+0.960}$$

V(week 19) = -23

R(week 19) = R(week 18) + V(week 19) = 2146 – 23 = 2123.

The same approach is used to determine the forecasted load consumption for other weeks as shown in Table 3.6

Table 3.6 Forecasted Load consumption for week 19 to week 24 using interval length of 6

Week	Actual load energy consumption	Forecasted Load energy consumption
19	2119.9	2136
20	2061.9	2093
21	1997.6	2027
22	1989	1988.4
23	1968.7	1960.8
24	1989.4	1991.2

CASE C: SEVEN EQUAL LENGTH INTERVAL

The universal set is partitioned into seven equal –length

U_1 = [-230, -180]

U_2 = [-180, -130]

U_3 = [-130, -80]

U_4 = [-80, -30]

U_5 = [-30, 20]

U_6 [20, 70]

U_7 = [70 120]

And the midpoints

U^1_m= -205,

U^2_m = -155,

$U^3{}_m = -105,$

$U^4{}_m = -55,$

$U^5{}_m = -5$

$U_6 = 45,$

$U_7 = 95$

Table 3.7 shows the fuzzification of Table 3.1 with interval length of seven

Table 3.7: Fuzzification table with interval length of seven

Week	Energy consumption in MW	Variation	Constant	U1	U2	U3	U4	U5	U6	U7
1	2357.1		0.006							
2	2137.7	-219.40	0.006	0.993	0.870	0.680	0.507	0.377	0.284	0.219
3	2114.4	-23.30	0.006	0.457	0.616	0.806	0.965	0.988	0.856	0.665
4	2221.7	107.30	0.006	0.222	0.288	0.381	0.513	0.688	0.877	0.995
5	2323.6	101.90	0.006	0.228	0.296	0.394	0.530	0.709	0.896	0.998
6	2284.1	-39.50	0.006	0.504	0.676	0.866	0.991	0.959	0.796	0.606
7	2301.0	16.90	0.006	0.361	0.485	0.651	0.843	0.983	0.972	0.820
8	2210.0	-91.00	0.006	0.681	0.871	0.993	0.955	0.790	0.600	0.445
9	2220.6	10.60	0.006	0.374	0.503	0.675	0.866	0.991	0.959	0.796
10	2247.3	26.70	0.006	0.341	0.457	0.616	0.806	0.965	0.988	0.856
11	2155.1	-92.20	0.006	0.686	0.876	0.994	0.953	0.785	0.596	0.442
12	2249.6	94.50	0.006	0.236	0.309	0.411	0.554	0.737	0.919	1.000
13	2191.4	-58.20	0.006	0.563	0.748	0.927	1.000	0.908	0.723	0.542
14	2273.7	82.30	0.006	0.252	0.330	0.442	0.596	0.785	0.952	0.994
15	2127.3	-146.40	0.006	0.890	0.997	0.942	0.769	0.581	0.431	0.323
16	2131.0	3.70	0.006	0.389	0.524	0.702	0.890	0.997	0.942	0.769
17	2080.0	-51.00	0.006	0.539	0.720	0.905	0.999	0.929	0.751	0.566
18	2146.0	66.00	0.006	0.274	0.363	0.487	0.655	0.846	0.984	0.971

$$0^6 \text{ (week 19)} = \begin{vmatrix} 0.563 & 0.748 & 0.927 & 1.000 & 0.908 & 0.723 & 0.542 \\ 0.252 & 0.330 & 0.442 & 0.596 & 0.785 & 0.952 & 0.994 \\ 0.890 & 0.997 & 0.942 & 0.769 & 0.551 & 0.431 & 0.323 \\ 0.389 & 0.524 & 0.702 & 0.890 & 0.997 & 0.942 & 0.769 \\ 0.539 & 0.720 & 0.905 & 0.999 & 0.929 & 0.751 & 0.566 \end{vmatrix}$$

$$C^6 \text{ (week 19)} = |0.274 \quad 0.363 \quad 0.487 \quad 0.655 \quad 0.846 \quad 0.984 \quad 0.971|$$

$$\begin{vmatrix} 0.274 & 0.363 & 0.487 & 0.655 & 0.841 & 0.723 & 0.542 \\ 0.252 & 0.330 & 0.442 & 0.596 & 0.785 & 0.952 & 0.971 \\ 0.274 & 0.363 & 0.487 & 0.655 & 0.846 & 0.431 & 0.323 \\ 0.274 & 0.363 & 0.487 & 0.655 & 0.846 & 0.942 & 0.769 \\ 0.274 & 0.363 & 0.487 & 0.655 & 0.846 & 0.751 & 0.566 \end{vmatrix}$$

$$F^6 \text{ (week 19)} = |0.274 \quad 0.363 \quad 0.487 \quad 0.655 \quad 0.846 \quad 0.952 \quad 0.971|$$

$$V(\text{week } 19) = \frac{0.274x\text{-}205+0.363x -155+0.487x -105+0.655x\text{-}55+0.846x\text{-}5+0.952x45+0.971x95}{0.274 + 0.363 + 0.487 + 0.655 + 0.846 + 0.952 + 0.971}$$

V (week 19) = - 15.14

R(week 19) = R(week 18) + V(week 19) = 2146 – 15.14 = 2130.9 .

The same approach is used to determine the forecasted load consumption for other weeks as shown in Table 3.8

Table 3.8 Forecasted Load consumption for week 19 to week 24 using interval length of 7

Week	Actual load energy consumption in MW	Forecasted Load energy consumption in MW
19	2119.9	2130.9
20	2061.9	2083
21	1997.6	2041
22	1989	1987.9
23	1968.7	1934.2
24	1989.4	1879.7

CHAPTER FOUR

RESULT AND ANALYSIS

4.1 INTRODUCTION

In evaluating the effectiveness or the effect of interval length on forecasting result, the load consumption over a certain period (six weeks) has been calculated. Tables 3.3, 3.5 and 3.7 show the results with different interval length of 5, 6 and 7 respectively.

The training data, for the time interval [week 1, week 18] was selected as an experimental base. It is clear that this period statistical data relating to the total load consumption is available.

The essence of this investigation consists of the following

i. The dynamics of the total load consumption for the examined period is considered to be unknown;

ii. With the aid of **Abbasov et al** [28] model, and using different interval length, the total load consumption was forecasted for other weeks from week 19 to week 24 based on the changes in the consumption for the previous weeks; and

iii. In order to show the effect of varying interval length on the overall result, a comparison was done for each interval length with actual load for the time period [week 19, week 24] as shown Tables 4.1 to 4.3.

Table 4.1 Comparison Table of Actual Load consumption and Forecasted Load consumption with interval length five

WEEK	Actual Load consumption in MW	Forecasted Load consumption in MW	Square Error
19	2119.9	2127.5	57.76
20	2061.9	2079.9	324
21	1997.9	2030.9	1108.9
22	19989	1976.1	166.41
23	1968.7	1960.6	65.61
24	1989.4	1994	21.16
		MSE	307.86
		RMSE	17.1
		R	0.98

Table 4.2 Comparison Table of Actual Load consumption and Forecasted Load consumption with interval length six

WEEK	Actual Load consumption in MW	Forecasted Load consumption in MW	Square Error
19	2119.9	2136	259.21
20	2061.9	2093	967.21
21	1997.9	2027	864.36
22	19989	1988.4	0.36
23	1968.7	1960.8	62.41
24	1989.4	1991.2	3.24
		MSE	273.87
		RMSE	19.0
		R	0.87

Table 4.3 Comparison Table of Actual Load consumption and Forecasted Load consumption with interval length seven

WEEK	Actual Load consumption in MW	Forecasted Load consumption in MW	Square Error
19	2119.9	2130.9	141.61
20	2061.9	2083	445.21
21	1997.9	2041	1883.6
22	1998.9	1987.9	121
23	1968.7	1934.2	1190.3
24	1989.4	1879.7	12034
		MSE	2635.9
		RMSE	51.3
		R	0.89

4.2 ANALYSIS

The analyses of the result obtained consist of forecast plots and statistical error measures in the estimation period like the Root Mean Square Error (**RMSE**). Forecast plots are qualitative indicators of the performance of a model as it allows for virtual comparison of the actual and predicted values. Plots of the actual load consumption and the forecasted are shown in Figures 4.1 to 4.3

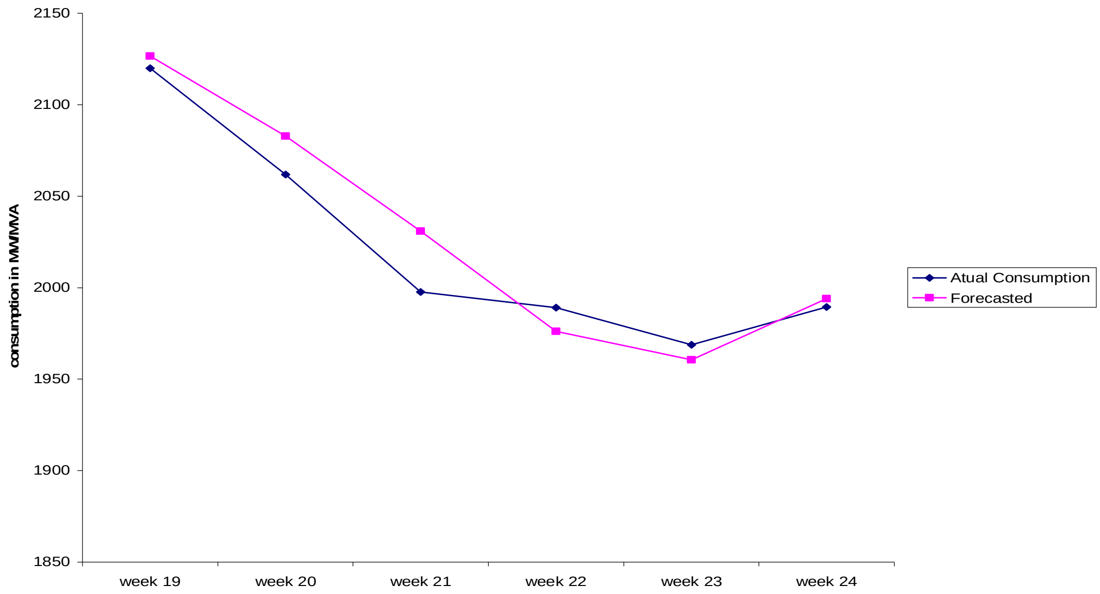

Figure 4.1 Plot of actual consumption and forecasted consumption with length 5

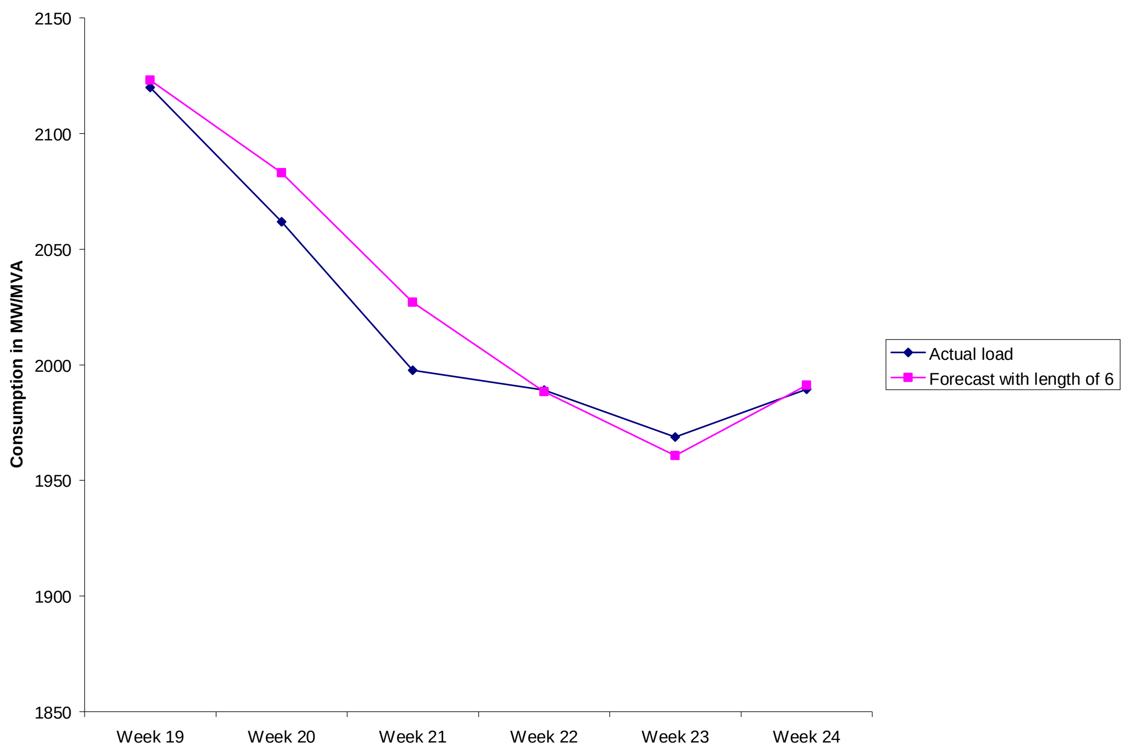

Figure 4.2 Plot of actual consumption and forecasted consumption with length 6

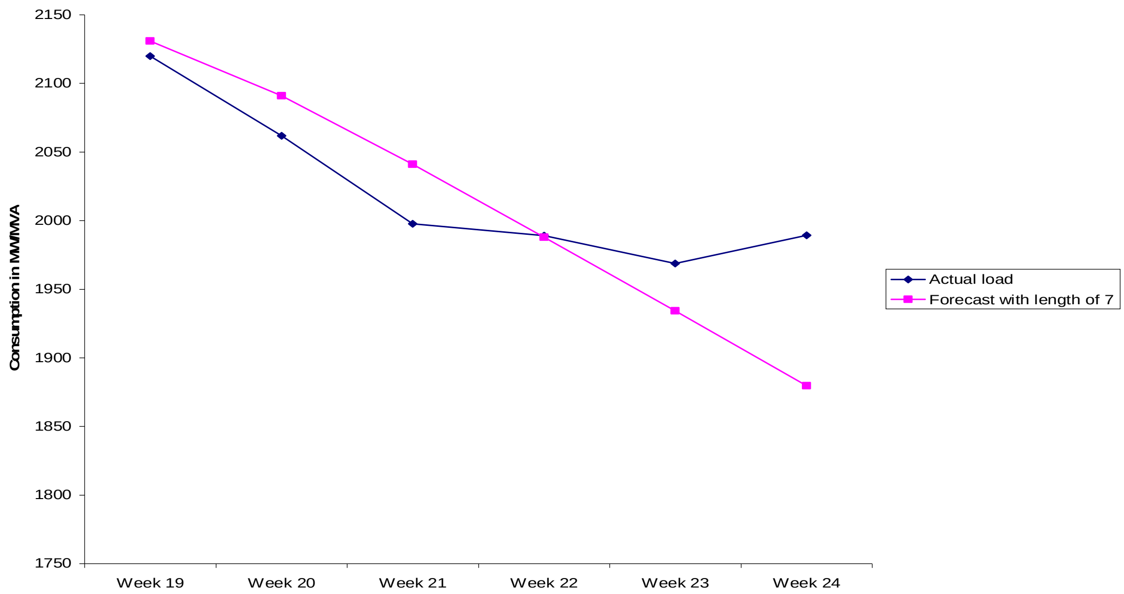

Figure 4.3 Plot of actual consumption and forecasted consumption with length 7

The Root Mean Square Error (RMSE) between the actual and forecasted load consumption is used as the performance function of the forecast model and the outcome is as tabulated in Table 4.1- 4.3.The root Mean Square Error (RMSE) is determined from:

RMSE= $\sqrt{MSE}$ ……………………………………………………….(4.1)

Where MSE = $\dfrac{\sum_{t=1}^{n}\left(Actual\_load - forecasted\_load\right)^2}{n}$ ………………………(4.2)

In addition to this, the Pearson's correlation factor R between the actual load and the forecasted load has been determined. The simple linear correlation coefficient R is a number between -1and +1 that tells how well a linear equation describe the relationship between two variables. (X and Y). R is designated as positive if Y increases as X increases and negative if Y decreases as X increases. An R or zero indicates an absence of relationship between the two variables. As R tends towards 1 the relationship gets better. In this work the correlation coefficient R tells how well trends in the validation value using Fuzzy Time Series follow trends in the actual data.

Table 4.4 shows the summary of the statistical performance function and the correlation function for the different interval length.

Table 4.4 Summary of results

Interval length 5		Interval length 6		Interval length 7	
RMSE	R	RMSE	R	RMSE	R
17.1	0.96	19.0	0.87	51.3	0.89

The interval length with the smallest RMSE is interpreted to be better than the ones with higher values. Also it is safe to assume the fact that R> 0.9 is an indication of a fairly good correlation between the actual load and the forecasted load.

In order to further show what effect interval length has on the overall forecast results, the length was increased further to equal length 8 and equal length 9 to forecast for week 19 to week 21. The results are summarized in table 4.5

Table 4.5 Comparison Table of Forecasted Load with different interval lengths.

Week	Actual Load	Forecasted Load with									
		Interval length 5		Interval length 6		Interval length 7		Interval length 8		Interval length 9	
		w=4	w=5	w=4	w=5	w=4	w=5	w=4	w=5	w=4	w=5
19	2119.9	2123.1	2127.9	2130.0	2136.2	2125.6	2130.9	2127.2	2135.1	2127.0	2136.1
20	2061.9	2079.5	2079.5	2093.0	2093.0	2084	2084.0	2085	2085.0	2082	2082.0
21	1997.6	2030.9	2030.9	2027	2027	2041	2041	2047	2047	2042.6	2042.6
	RMSE	21.82	22.23	30.82	31.44	28.31	28.82	31.76	32.68	28.74	29.95
	R	0.96		0.87		0.94		0.87		0.90	

From the analysis of the above table, it can be concluded that odd interval lengths give better forecasting results than even interval length which serves as a justification why in literature [10][28], its more common to see odd number interval length used.

4.3 SIGNIFICANCE OF RESULT

The comparative analyses of the actual load and the forecasted load and the consequent error of the different interval length have confirmed the high effect varying interval length has on the overall forecasting result.

From both qualitative performance indicator and the statistical error evaluation, it can be concluded that:

i. Different interval length has significant effect on the overall forecasting result [1].
ii. Interval length 5 gives a better forecasting result for this particular case study when compared interval length 6 and interval length of 7 gave a better result when compared interval length 8.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 INTRODUCTION

This work is aimed at investigating the effect of interval length and model basis on electric load consumption forecasting using Fuzzy time series method. Electricity is an important factor in the economic growth and well being of any economy; any method that can enhance forecasting the load demand or load consumption pattern becomes very important.

Fuzzy time series has enormous advantage over traditional model in that human knowledge can be applied from the start till the end of the forecasting procedure. It also provides a powerful and a qualitative framework to cope with the vagueness of human language. The use of Fuzzy verbal notation in every day speech such as, much, more, little, small, many, a number of, etc, enables one to give a qualitative description of the problems which must be tackled taking account of its indefinite nature as well as obtain explanation of the factors that can not be described qualitatively.

One critical issue in Fuzzy time series is the determination of interval length. There has not been any empirical method of determining an appropriate interval length and as such it becomes very important that an investigation be carried out to determine the effect of varying the interval length on the forecasting result. However the Fuzzy time series method adopted have a drawback, it requires a large amount of computation to derive the Fuzzy relationship matrix, $R^w(t)$, as a function w. **Abbasov et al** [28] model was adopted in this work because this model allows prediction into far future beyond the time available in the test data.

5.2 LIMITATIONS

Forecasting using Fuzzy time series method is a data driven technique. This means that the quality of data is very important to the working of this technique. This essentially means that the quality of the output is highly dependent on the quality of the input data. In carrying out this work, therefore one of the main constraints is ascertaining the quality of the data used.

Another limitation is the high computational requirement of the Fuzzy time series technique. The computational requirement increases sharply with increase in the model basis. It is worth noting that high computation implies high cost.

5.3 CONCLUSION

Computational intelligence techniques (Fuzzy logic, genetic algorithm, artificial neural networks) are useful for varied applications such as forecasting, data mining, image processing etc.The **Abbasov et al** [28] model in Fuzzy time series technique was applied in this work to test the effect of interval length on the forecast of electric load consumption. This model and technique can also be used to forecast various situations like temperature, rainfall, university enrollment etc once the input data are time series. That is, they are presented as they appear in time. The analysis of the result shows that varying interval length gives different forecasting results. Forecast verification is the process of determining the quality of forecasts. Various procedures exit but all involves measurement of the relationship between a forecast and corresponding observations. Statistical performance functions of Root Mean Square Error (**RMSE**) and the Correlation Factor (**R**) are used to measure the performance of different interval lengths and are summarized in Table 4.4. In addition to these, forecast plots were used as qualitative performance functions.

The values of Root Square Error (**RMSE**) obtained are **17.1, 19.0** and **51.3** corresponding to interval length 5, 6 and 7 respectively. The correlation factor **R** for interval length equal to 5, is shown to be **0.96**. While for length size 6 and 7 are **0.87** and **0.89** respectively. These are measurements of the standard error and how well the validation value of the model used trends in the actual data of the forecast. It can therefore be concluded that using interval length of 5 gives the best result in this case, judging from the qualitative and quantitative performance function. It can also be concluded that odd interval lengths give better forecasting results than even interval length which serves as a justification why in literature [10][28], its more common to see odd number interval length used.

5.4 RECOMMENDATION

The followings are suggested ways of improving this work.

i. Since the quality of data used plays a critical role to the overall forecasting result, the method used in obtaining the data used should be improved upon. Using on-line real time data acquisition method can enhance the accuracy and authenticity of the data used in developing the forecasting model.

ii. Further research work should exploit the possibility of using other developed model for Fuzzy time series order than the one used to test the effect of interval length on the overall forecasting result of load forecast.

iii. Further research should be carried out to code the technique in a programming language, especially C or C^{++}.

REFERENCE

1. Tian-shyug, L. (2001), "Prediction of Unemployment Rate Using Fuzzy Time Series with Box-Jenkro Methodology", International Journal of Fuzzy System Vol.3 No 4, December.

2. Dubois, D. and Prede, H. (1991), "Fuzzy Sets in Approximate Reasoning Part I: Inference with Possibility Distribution'', Fuzzy Sets and System Vol. 40, pp 143 – 202.

3. Song, Q. and Chisson, B. S. (1993), "Forecasting Enrollment with Fuzzy Time Series'', Part I Fuzzy Sets and System vol. 54, pp 1 – 9.

4. Song, Q. and Chisson, B. S. (1994), "Forecasting Enrollment with Fuzzy Time Series", Part II Fuzzy Sets and System, vol 62, pp 1 – 8.

5. Hirotu, K. (1979), "Extended Fuzzy Expression of Probabilistic Sets Advance in Fuzzy Set Theory and Applications", North-Holland, pp 201-214.

6 Konolige, K. and Meyers, (1992), "Flakey and Autonomous Mobile Robot",Technical Document Stanford Research Institute International.

7. Zadeh, L.A. (1965), "Fuzzy Sets" Information and Control, Vol.8, pp 338– 353.

8. Sullivan, S.J and Woddall, W. H. (1994), "A Comparison of Fuzzy Forecasting and Markov Modeling", Fuzzy Sets System vol. 64 pp 279 – 293.

9. Chem, S. M. (1996), "Forecasting Enrollment Base on Fuzzy Time Series", Fuzzy Set and System vol. 81, pp 311 – 319.

10. Mu'azu, M.B. "Fuzzy Time Series Forecasting: A Case Study of Temperature Forecasting for Zaria", Department of Electrical Engineering – A. B. U Zaria. Unpublished paper.

11. Su, S.F. and Li, S.H. (2003), "Neural Network Based Fusion of Global and Local Information in Predicting Time Series", Proceedings

of the 2003 IEEE International Joint Conference on System Man and Cybernetic.

12. Huarng, K. (2001), Effective Length of Interval to Improve Forecasting in Fuzzy Time Series.

13. Shy-ming, A., Chien, C. and Cling, H. (2004), “A New Method to Forecast Enrollment using Fuzzy Time Series” International Journal of Applied Science and Engineering, vol. 2 No. 3 pp234-244.

14. Chao-Chih,T. and Shun, J.W. (1999), “A Study for Second-order Modeling of Fuzzy Time Series”, IEEE International Fuzzy System Conference Proceeding.

15. Grigore, A. and Florin, P. (2005), “On Using the Fuzzy Nearest Neighbour Methods for Time Series Forecasting in Software Reliability”, Proceeding of SIG Conference, pp 28 – 30.

16. Dug, H. H. (2005), “A Note on Fuzzy Time Series Model”, Fuzzy Sets and System Vol 155 pp 309 – 315.

17. Hui-Kuagng, Y. (2005), “A Refined Fuzzy Time Series Model for Forecasting”, physica A Vol. 346, pp 657 – 681.

18. Kunhuang, H. and Yu, T.H. (2006), “The Application of Neural Networks to Forecast Fuzzy Time Series” physica A vol 363 pp 481 – 491.

19. Kunhuang, H. and Yu, T.H. (2005), “A Type 2 Fuzzy Time Series Model for Stock Index Forecasting”, Physica A Vol. 353 pp 445 – 462.

20. Kunhuang, H. and Yu, T.H. (2006), “Ratio Based Lengths of Intervals to Improve Fuzzy Time Series Forecasting”, IEEE Transactions on System, Man and Cybernetic – Part B Cybernetic Vol. 36 No. 2.

21. Hui-Kuang, Y. (2005), “Weighted Fuzzy Time Series model for Taiex Forecasting”, physica A vol. 349 pp 609 – 624.

22. Yu-Yun, H., Sze-man, T. and Berlin, W. (2003), "A New Approach of Bivariate Fuzzy Series Analysis to the Forecasting of Stock Index'', International Journal of Uncertainty, Fuzziness and Knowledge Based System. Vol.11 No.6, pp 671-690.

23. Song, Q. (2003), "A Note on Fuzzy Time Series Model Selection with Sample Autocorrelation Functions", International Journal on Systems and Cybernetics Vol.34, pp: 93-107.

24. Shyi-Ming, C. (2002), "Forecasting Enrollment Based on High-Order Fuzzy Time Series", International Journal on Systems and Cybernetics Vol. 33, pp: 1 – 16.

25. Kunhuang, H. and Yu, T.H., (2004), "A Dynamic Approach To Adjusting Lengths Of Intervals In Fuzzy Time Series Forecasting", Intelligent Data Analysis Vol. 8 pp: 3-27.

26. Shyi-Ming, C. and Chia-ching, H. (2004), "A New Methods to Forecast Enrollment Using Fuzzy Time Series", International Journal of Applied Science and Engineering Vol. 2 No3. pp: 233 – 244.

27. Dash, P. K., Ramakrishn, G., Liew. A. C. and Rahman, S. (1995), "Fuzzy Neural Networks for Time-Series Forecasting of Electric Load", IEE Proceeding on Generation, Transmission Distribution, Vol 142, No. 5.

28. Abbasov, A. M. and Mamedova, M.H., "Application of Fuzzy Time Series to Population Forecasting", Vienna University of Technology. www.corp.at/corp/archive/papers/2003/corp2003abbasov.pdf

29. Singh, S. (1993) "Forecasting Using a Fuzzy Nearest Neighbour Method", Proceedings 6th International Conference on Fuzzy Theory and Technology, Fourth Joint Conference on Information Science (JCIS '98), North Carotinu, Vol. 1, pp 80-83.

30. La Pensee, A. C. and Mort, N. (1999), "A Nenro Fuzzy Time Series Prediction in Telephone Banking", Third International Conference on Knowledge Based Intelligent Information Engineering Systems, Adelaide, Australia.

APPENDIX A
COMPLETE TRAINING DATA (24 WEEKS)

Weeks	Energy consumption in MW	Variation in MW
1	2357.1	
2	2137.7	-2194
3	2114.4	-23.3
4	2221.7	107.3
5	2323.6	101.9
6	2284.1	-39.5
7	2301	16.9
8	2210	-91
9	2220.6	10.6
10	2247.3	26.7
11	2155.1	-92.2
12	2249.6	94.5
13	2191.4	-58.2
14	2273.7	82.3
15	2127.3	-146.4
16	2131	3.7
17	2080	-51
18	2146	66
19	2119.9	26.1
20	2061.9	58
21	1997.6	64.3
22	1989	8.6
23	1968.7	20.3
24	1989.4	-20.7