DOKUZ EYLUL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

FUZZY TIME SERIES AND RELATED APPLICATIONS

by
Deniz GÜLER

October, 2011
İZMİR

FUZZY TIME SERIES AND RELATED APPLICATIONS

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Master of Science in Department of Statistics, Statistics Program

by
Deniz GÜLER

October, 2011
İZMİR

M. Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"FUZZY TIME SERIES AND RELATED APPLICATIONS"** completed by **DENİZ GÜLER** under supervision of **PROF. DR. EFENDİ NASİBOĞLU** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Efendi NASİBOĞLU

Supervisor

Doç.Dr. Esin Firuzan

(Jury Member)

Yrd.Doç.Dr. Burak ORDİN

(Jury Member)

Prof. Dr. Mustafa SABUNCU
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

It has been a harsh period, especially during documenting, putting all the facts and research outputs together in this thesis. I appreciate for giving the strength and motivation when I'm low to my supervisor Prof. Dr. Efendi NASİBOĞLU. Through his lessons the enlightenment helped me to realize and interpret the roots, also the aims of the theme.

At the pre- preparations of the thesis, Doç. Dr. Esin FİRUZAN also helped me a lot with her lectures to gain understanding the backup information, which lies behind the thesis.

Patiently waiting the results, without any judgments and ambition, my family needed the most intense thanks.

Deniz GÜLER

FUZZY TIME SERIES AND RELATED APPLICATIONS

ABSTRACT

Currently, inventing new approaches for modeling the classical time series analysis with last decade's favorites theme Fuzzy Logic and Sets Theory is going to be popular. In many different scientific models and research areas the Fuzzy Logic Systems are easy to integrate with. Forecasting the short/long distance of future is the main objective of Time Series Analysis and lately it evolves Fuzzy Logic Systems. The main aim in this thesis is evaluating the forecasting or estimation error rate on invented and also improved new models, if they have stronger or weaker affiliations.

At the introduction section, effects of Time Series Analysis and Fuzzy Logic Systems in human daily life are separately discussed. The second and third sections include the axioms, definitions of Time Series Analysis and Fuzzy Logic and Sets Theory. The following section after them defines and compares how the newly invented methodology of Fuzzy Time Series gathered. Also the pros and cons of the new system is discussed, so if the forecasting or estimating abilities are superior or not.

Keywords: Box-jenkins, fuzzy logic, fuzzy numbers, time series analysis, high-order fuzzy time series.

BULANIK ZAMAN SERİLERİ VE UYGULAMALARI

ÖZ

Bu çalışmada, uzun yıllardır süregelen klasik zaman serileri araştırmalarına yeni bilimsel yaklaşımların incelenmesi hedeflenmiştir. Son yılların gözde bilim alanı olan, Bulanık Mantık ve Kümeler Teorisi ile Zaman Serisi Analizi iç içe geçirilmiştir. Bu tezde birçok farklı bilim alanına veya araştırma konusuna kolaylıkla bütünleşmiş bulanık mantık sistemleri kullanılmıştır. Geleceği tahminlemede çok önemli rol oynayan Zaman Serileri'ne çeşitli yöntemler dahil edilmektedir. Tezin ana amacı; tezde önerilen yöntemlerin tahminlemede, modellemede hata payını azaltma ve/veya tahmincinin yeteneklerini geliştirme gücüne sahip olup olmadığını araştırmak. Düşük miktarda veri ile çalışma imkanı sağlayabildiğini sınamaktır.

Tezin ilk bölümü güncel hayatta bilimin etkileri, Zaman Serileri Analizi ve Bulanık Mantık Sistemlerinin yaşamımıza etkisini anlatmaktadır. İkinci ve üçüncü bölümler ise sırası ile Zaman Serileri ve Bulanık Kümeler Teorisi konularının prensip ve bilimsel temellerini açıklamaktadır. Tezin dördüncü bölümü, ikinci ve üçüncü bölümlerin nasıl birleşerek yepyeni bir bilimsel açılım olan, Bulanık Zaman Serilerini oluşturduğunu ve bu yeni sistemin, geçmiş yöntemlere olan olumlu ve olumsuz kabiliyetlerini incelemekte ve sorgulamaktadır.

Anahtar sözcükler : Box-jenkins, bulanık mantık, bulanık sayılar, zaman serisi analizi, yüksek mertebeli bulanık zaman serileri.

CONTENTS

CHAPTER ONE
INTRODUCTION

Time series analysis is a problem which has always attracted the attention of soft computing (SC) researchers. Forecasting future values of a series is usually a very complex task, and many SC methods and models have been faced with it, including fuzzy rule-based models (FRBM) in their various formulations. Notwithstanding, a common characteristic of those approaches is that they usually consider time series as just another data set which requires some small adaptations to be cast into the regression or classification format for which most SC models were created. However, time series analysis is a prominent field in Econometrics, which has been widely studied under a statistical perspective during the last centuries. In 1807, Fourier proved that a deterministic time series can be approximated by a sum of sine and cosine terms. But it was not until the beginnings of the 20th century when a stochastic approach for time series was first introduced, while the foundations for a general stochastic process theory were fixed in the 1930s by Khinchin (1934) & Kolmogorov (1931). Independently, in 1927 Yule (1927) stated that Fourier analysis is not suited for stochastic time series analysis and introduced second order autoregressive processes as theoretical schemes able to generate series with stochastic cyclic oscillations.

In 1970, the idea of forecasting future values of a time series as a combination of its past values received a strong impulse after Box & Jenkins (1970). In that work, Box & Jenkins proposed a modeling cycle for the autoregressive (AR) model, which assumes that future values of a time series can be expressed as a linear combination of its past values.

Of course this linearity assumption implies certain limitations, and in the last years much research has been devoted to nonlinear models. Nonlinear and non-stationary models are more flexible in capturing the characteristics of data and, in some cases, are better in terms of estimation and forecasting. These advances do not rule out linear models at all, because these models are a first approach which can be of great help to further estimate some of the parameters. Furthermore, modeling of any real-world problem by using nonlinear models must start by evaluating if the behavior of the series follows a linear or nonlinear pattern.

For some reason, SC researchers do not usually go deep into classical time series analysis, disregarding all the knowledge gathered through the years in the statistical field. In this thesis, we take a step forward in the quest for an SC-based time series research which integrates methods and models introducing a dynamical forecasting accuracy coming from fuzzy rule-based models.

By applying this test, practitioners will be able to determine if a series data generating process is linear, in which case it can be modeled by using a linear model or a single-rule fuzzy rule-based model. The experiments show that the test is robust against Type I errors (rejecting the null hypothesis when it is actually true) and very powerful against Type II errors (not rejecting the null hypothesis when it is false).

The structure of the thesis is as follows: in Chapter 2, a brief review of some statistical models of Time Series Analysis with Box & Jenkins (1970) methodology is offered, while in Chapter 3 their links with fuzzy rule-based models are recalled. In Chapter 4 the fuzzy rule-based methods are presented, both intuitively and in its mathematical formulation.

CHAPTER TWO
TIME SERIES ANALYSIS

2.1 The Concept of a Time Series

A time series is defined as a sequence of observations (measurements) ordered by time $\{x_t\}$, $t \in T$. We restrict ourselves to equidistant time series, i.e. the parameter set is a finite set of equidistant points of time: $T = \{1, 2, 3, \ldots, N\}$.

We distinguish two classes of time series analysis approaches:

- One class which represents a time series with a kinetic model (component analysis, classical analysis):

$$x_t = f(t) \qquad (1)$$

the measurements or observations are seen as a function of time;

- One class which represents a time series with a dynamical model ("ARIMA model", "Box & Jenkins procedure"):

$$x_t = f(x_{t-1}, x_{t-2}, x_{t-3}, \ldots) \qquad (2)$$

the measurements or observations are not seen as a function of time, but as a function of their own past (and, perhaps of the past of other measured or observed variables).

The classical procedure decomposes the time series function $x_t=f(t)$ into up to four components:

- The trend: a long-term monotonic change of the average level of the time series,
- The trade cycle: a long wave in the time series,
- The seasonal component: a yearly variation in the time series,

• The residual component which represents all the influences on the time series which are not explained by the other three components.

2.2 The Box & Jenkins Method

2.2.1 The Concepts of Box & Jenkins Method

The Box & Jenkins model is based on a combination of two different approaches which are used for modeling a univariate time series. Particularly Auto-regressive (AR) and Moving Averages (MA) models are used to decompose the time series into a trend, seasonal, cycle or residual components.

The Box & Jenkins model assumes that the time series is stationary, but models can be extended to include seasonal AR and seasonal MA terms. Although this complicates the notation and mathematics of the model, the underlying concepts for seasonal AR and MA terms are similar to the non-seasonal AR and MA terms.

The most general Box & Jenkins model includes difference operators; such as AR and MA terms, seasonal difference operators, seasonal AR and MA terms. As with modeling in general, however, only necessary terms should be included in the model.

As typically in classical time series, an effective fitting of Box & Jenkins models requires at least a moderately long series, which consists at least of 50 observations (Chatfield, 1996). Many other would recommend at least 100 observations.

There are three primary stages in building a Box & Jenkins time series model

1. Model Identification
2. Model Estimation
3. Model Validation (Diagnostics)

2.2.2 Box & Jenkins Model Identification

The first step in developing a Box-Jenkins model is to determine if the series is stationary and if there is any significant seasonality that needs to be modeled. Stationarity can be assessed from a run sequence plot. The run sequence plot should show constant location and scale. It can also be detected from an autocorrelation plot. Specifically, non-stationarity is often indicated by an autocorrelation plot with very slow decay.

In an additive time series model (3) the first two components are often aggregated into the *smooth* components. Component two and three are often aggregated into the *cyclic* component. The simplest case assumes that the four components add up to the time series:

$x_t = m(t)+k(t)+s(t)+u(t)$..(3)

- *m* is a monotonic function,
- *k* is a periodic function with period>*1* year,
- *s* is a periodic function with period=*1* year,
- *u* is a random function (stochastic process).

In many cases we can observe that the amplitude of *s(t)* and/or the variance of *u(t)* increase with *t* (or with *m(t)*). Hence it is a good idea to model the time series as follows:

$x_t = m(t)*k(t)*s(t)*u(t)$ (multiplicative model)..(4)

so it leads to,

$\log x_t = m(t)+k(t)+s(t)+u(t)$ (multiplicative model)..(5)

which is the same as,

$x_t = exp[m(t)]\ exp[k(t)]\ exp[s(t)]\ exp[u(t)]$ (multiplicative model)..................(6)

In both cases one will estimate the parameters of the functions *m, k,* and *s* with regression methods (making some assumptions about the period of the trade cycle component). The residual component *u(t)* is the regression residual (so-called global component model).

Seasonality (or periodicity) can usually be assessed from an autocorrelation plot, a seasonal sub series plot, or a spectral plot. Instead one could try to eliminate the residual component by some smoothing procedure such as moving averages (so-called local component model). Box & Jenkins recommend the differencing approach to achieve stationarity. However, fitting a curve and subtracting the fitted values from the original data can also be used in the context of Box & Jenkins models.

At the model identification stage, main goal is to detect seasonality, if it exists, and to identify the order for the seasonal autoregressive and seasonal moving average terms. For many series, the period is known and a single seasonality term is sufficient. For example, for monthly data we would typically include either a seasonal AR 12 term or a seasonal MA 12 term. For Box & Jenkins models, we do not explicitly remove seasonality before fitting the model. Instead, we include the order of the seasonal terms in the model specification to the ARIMA estimation software. However, it may be helpful to apply a seasonal difference to the data and regenerate the autocorrelation and partial autocorrelation plots. This may help in the model identification of the non-seasonal component of the model. In some cases, the seasonal differencing may remove most or all of the seasonality effect.

Once stationarity and seasonality have been addressed, the next step is to identify the order (i.e., the *p* and *q*) of the autoregressive and moving average terms. The primary tools for doing this are the autocorrelation plot and the partial autocorrelation plot. The sample autocorrelation plot and the sample partial autocorrelation plot are compared to the theoretical behavior of these plots when the order is known.

Specifically, for an AR(*1*) process, the sample autocorrelation function should have an exponentially decreasing appearance. However, higher-order AR processes are often a mixture of exponentially decreasing and damped sinusoidal components.

Table 2.1 Sample autocorrelation function for model identification.

Shape	**Indicated Model**
Exponential, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to identify the order of the autoregressive model.
Alternating positive and negative, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to help identify the order.
One or more spikes, rest are essentially zero	Moving average model, order identified by where plot becomes zero.
Decay, starting after a few lags	Mixed autoregressive and moving average model.
All zero or close to zero	Data is essentially random.
High values at fixed intervals	Include seasonal autoregressive term.
No decay to zero	Series is not stationary.

For higher-order autoregressive processes, the sample autocorrelation needs to be supplemented with a partial autocorrelation plot. The partial autocorrelation of an AR(p) process becomes zero at lag $(p+1)$ and greater, so we examine the sample partial autocorrelation function to see if there is evidence of a departure from zero. This is usually determined by placing a 95% confidence interval on the sample partial autocorrelation plot (most software programs that generate sample autocorrelation plots will also plot this confidence interval). If the software program does not generate the confidence band, it is approximately, with N denoting the sample size.

The autocorrelation function of a MA(q) process becomes zero at lag *(q+1)* and greater, so we examine the sample autocorrelation function to see where it essentially becomes zero. We do this by placing the 95% confidence interval for the sample autocorrelation function on the sample autocorrelation plot. Most software that can generate the autocorrelation plot can also generate this confidence interval. The sample partial autocorrelation function is generally not helpful for identifying the order of the moving average process.

In practice, the sample autocorrelation and partial autocorrelation functions are random variables and will not give the same picture as the theoretical functions. This makes the model identification more difficult. In particular, mixed models can be particularly difficult to identify.

As an example, there is a time series graph in Figure 2.1 of electricity consumption in F.R.G. (Federal Republic of Germany). This time series includes some of the component defined at section *2.2.1*.

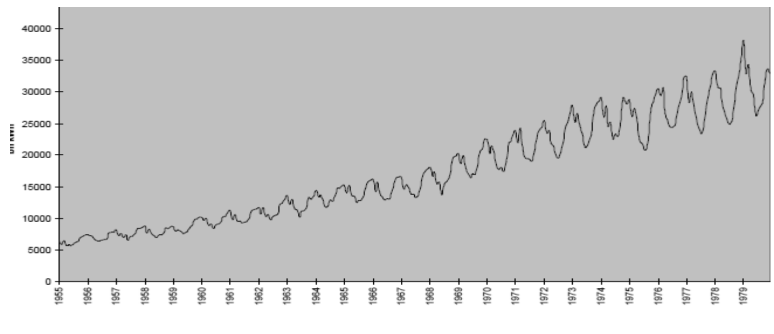


Figure 2.1 The yearly electricity consumptions between years 1955–1980 in F.R.G.

2.2.2.1 Stationary Processes

A stochastic process is a series $\{X_t\}$, $t \in T$ of random variables X_t. Here, t - the time parameter - is an element of the index set T which we will identify with the set of (positive) integers.

A random variable X is a mapping $X: \Omega \rightarrow R$ which attributes real numbers $X(\omega)$ to the outcomes ω of a random process. Thus, a result ω of a random process corresponds to the time series $\{X_t\}$, $t \in T$.

From one realization of a stochastic process mean function and variance function can only be estimated if we make certain assumptions about the process behind a time series. Note that we can never check whether these assumptions are met.

We will assume that empirical time series are realizations of stationary processes and test of the time series which we will analyze can be a realization of a stationary process. If this is not the case, then we will try to transform (filter) the time series in a manner that at least the filtered time series is stationary.

We call a stochastic process $\{X_t\}$, $t \in T$

- Stationary with respect to the mean if $\mu(t) = \mu$ for all $t \in T$,
- Stationary with respect to the variance if $\sigma^2(t) = \sigma^2$ for all $t \in T$,
- Stationary with respect to the covariance if $\gamma(s, t) = \gamma(s+r, t+r)$ for all $r,s,t \in T$,
- Weakly stationary if it is both stationary with respect to the mean and to the covariance.

In processes which are stationary with respect to their covariance we write the covariance and correlation functions

$$\gamma\ (s,\ t) = \gamma\ (s{+}r,\ t{+}r) = \gamma\ (s{-}t) = \gamma\ (\tau) = \gamma\ ({-}\tau) \qquad (7)$$

and

$$\rho\ (s,\ t) = \rho\ (s{+}r,\ t{+}r) = \rho\ (s{-}t) = \rho\ (\tau) = \rho\ ({-}\tau) \qquad (8)$$

2.2.2.2 Autocorrelation & Partial Autocorrelation

An important guide to the properties of a time series is provided by a series of quantities called sample autocorrelation coefficients, which measure the correlation between observations at different distances apart. Autocorrelation seems like the ordinary correlation coefficient, but the main difference is that autocorrelation uses x_t and x_{t+1}, instead of x and y. And it's given by

$$r_1 = \frac{\sum_{t=1}^{N-1}(x_t - \bar{x}_1)(x_{t+1} - \bar{x}_2)}{\sqrt{\sum_{t=1}^{N-1}(x_t - \bar{x}_1)^2 \sum_{t=1}^{N-1}(x_{t+1} - \bar{x}_2)^2}} \qquad (9)$$

x_1 is the mean of first *N-1* observations and x_2 is the mean of the last *N-1* observations.

Sample partial ACF of Series (spacf) is a vector of length *nLags + 1* corresponding to lags *0, 1, 2, ... n* Lags. The first element of spacf is unity, that

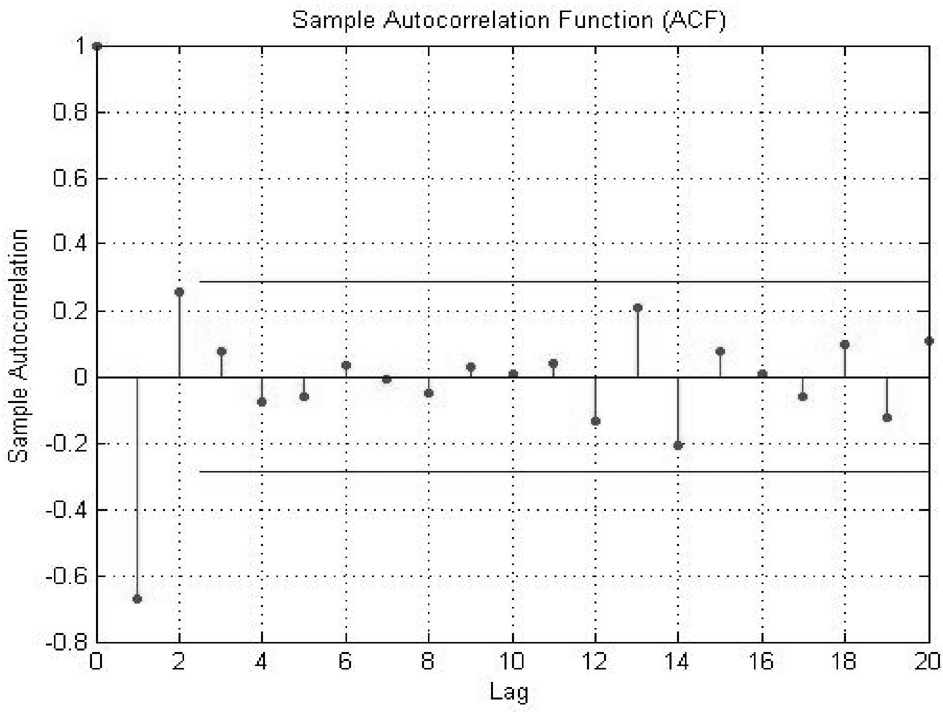
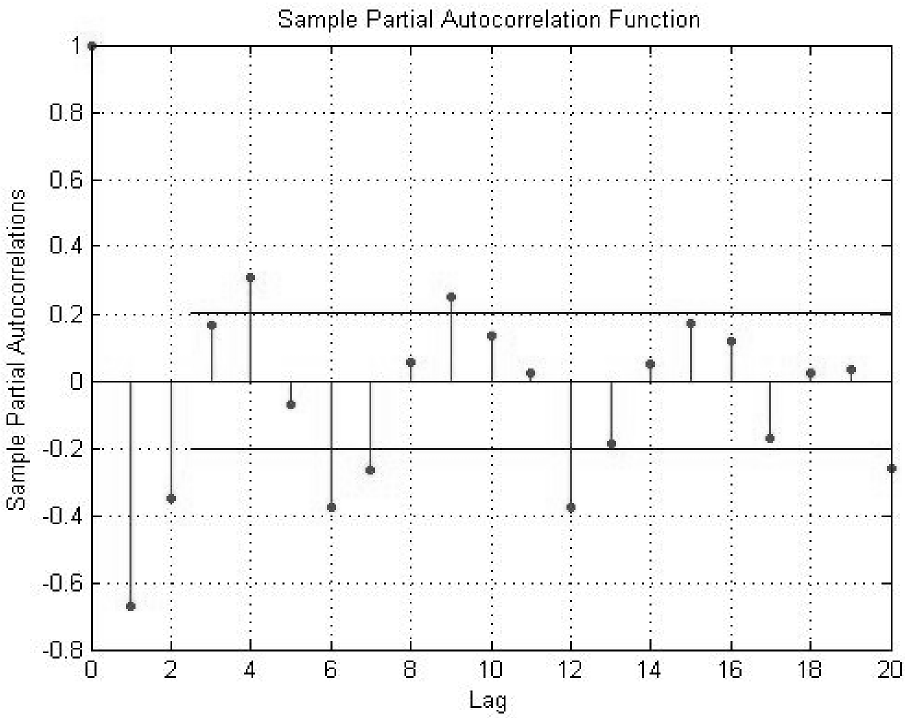


Figure 2.2 The sample autocorrelation function yearly electricity consumptions between years 1955–1980 in F.R.G.

Figure 2.3 The sample partial autocorrelation function yearly electricity consumptions



y the
ed to
d for

ndom
ndom
when
yclic
components.

There are two distinct groups of smoothing methods:

- Averaging Methods
- Exponential Smoothing Methods

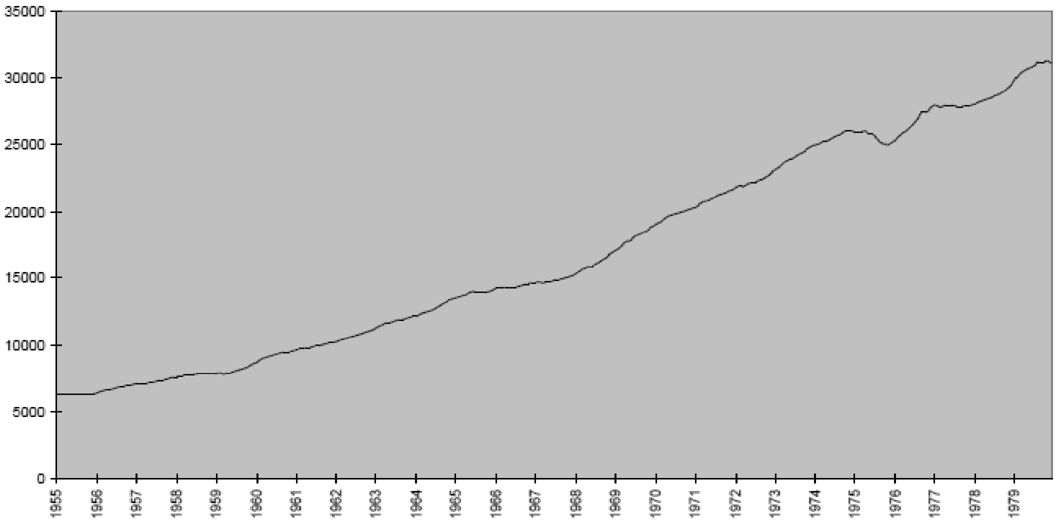


Figure 2.4 The smoothed yearly electricity consumptions between years 1955–1980 in Federal Republic of Germany.

The *n-th* order moving Average process

$$X_t = \mu + Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2} + \ldots + \beta_n Z_{t-n} \quad (11)$$

where μ, β_i are constants and Z_t denotes a purely random process.

2.2.3 Calculating the Trend Components

The trend component (or the "smooth" component as a whole) is mostly estimated by polynomial regression ($\sum u_t^2 = \min!$)

$$x_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t_3 + \ldots + u_t \quad (12)$$

2.2.4 Estimating the Trend Component: Prediction

If we use only the first 15 (instead of 25) years for parameter estimation, i.e. if we use only the knowledge available at the end of 1974, the time series model would be straighter, compare to parameters from full knowledge used model.

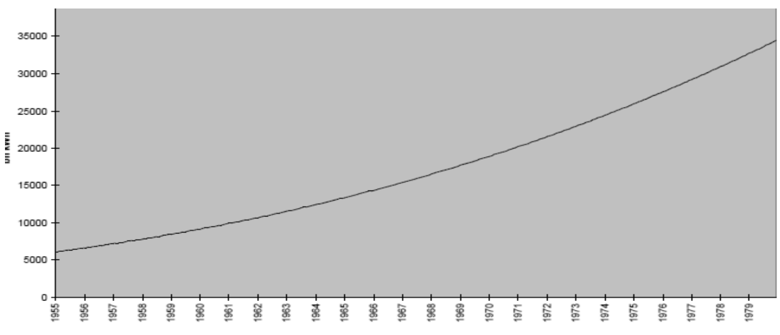


Figure 2.5 The smoothed yearly electricity consumptions between years 1955–1970 in Federal Republic of Germany.

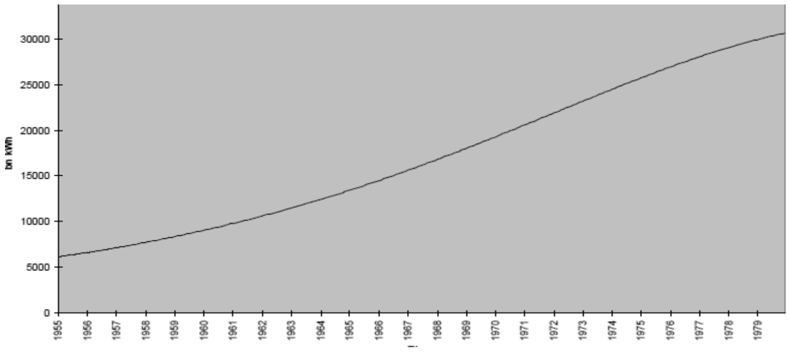


Figure 2.6 The smoothed yearly electricity consumptions between years 1955–1980 in Federal Republic of Germany.

2.3 Estimation

There are several different methods for estimating the parameters. All of them should produce very similar estimates, but may be more or less efficient for any given model. In general, during the parameter estimation phase a function minimization algorithm is used (the so-called *quasi-Newton* method) to maximize the likelihood of the observed series, given the parameter values. In practice, this requires the calculation of the sums of squares (SS) of the residuals, given the respective parameters. So the chosen model could be fitted best by using these methods.

2.3.1 Estimating the Autocovariance & Autocorrelation Functions

The autocorrelation coefficients describing very useful statistical information as it are noted in Section 2.4. Autocorrelation function (acf) of a stationary time series shows the main properties and characteristics of the set.

If $X(t)$ is a stationary time series and has the mean of μ and variance of σ^2 then,

$$c_k = \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) / N \text{ ....(13)}$$

A progressive method of estimating the acf is the jackknife estimation. In this procedure the time series is divided into two halves, and the sample acv.f. is estimated from each half of the series. The method is denoted as

$$\tilde{c}_k = 2c_k - \frac{1}{2}(c_{k1} + c_{k2}) \text{ ....(14)}$$

To advance the theoretical acf the jackknife method should be adapted to estimate a lesser biased equation (14). The jackknife estimator is given in an obvious notation by

$$\tilde{r}_k = 2r_k - \frac{1}{2}(r_{k1} + r_{k2}) \text{ ....(15)}$$

2.3.2 Fitting a Moving Average Process

First it needed to estimate the parameters of the process, and then the order of the process should be found. The theoretical first order autocorrelation coefficients equate by

$$r_1 = \hat{\beta}_1 / \left(1 + \hat{\beta}_1^2\right) \text{ ....(16)}$$

and choose the solution $\hat{\beta}_1$ such that $|\hat{\beta}_1| < 1$, because it can be shown that this gives rise to an inefficient estimator. The approach suggested by Box & Jenkins (1970). If $z=0$, we have

$$z_1 = x_1 - \mu,$$
$$z_2 = x_2 - \mu - \beta_1 Z_1, \ldots,$$
$$z_N = x_N - \mu - \beta_1 Z_{t-1}$$

The order of the moving average process is usually evident from the sample acf for a given set of data. The theoretical acf MA(q) process has a very simple form in that it 'cuts off' at lag q, and so the analyst should look for the lag beyond which the values of r_k are close to zero.

2.3.3 Fitting an Autoregressive Process

In autoregressive model is an observation of a time period depends on a number of past observations and that period random error. The order of the model p is the number of past observations included in the model. Suppose we have an Autoregressive (AR) process of order *p,* with mean μ, given by

$$X_t - \mu = \alpha_1 (X_{t-1} - \mu) + \cdots + \alpha_p (X_{t-p} - \mu) + Z_t \quad \text{...(17)}$$

In first order case, with $p=1$, we find

$$\hat{\mu} = \frac{\bar{x}_{(2)} - \hat{\alpha}_1 \bar{x}_{(1)}}{1 - \hat{\alpha}_1} \quad \text{...(18)}$$

and

$$\hat{\alpha}_1 = \frac{\sum_{t=1}^{N-1} (x_t - \hat{\mu})(x_{t+1} - \hat{\mu})}{\sum_{t=1}^{N-1} (x_t - \hat{\mu})^2} \quad \text{..(19)}$$

where $\bar{x}_{(1)}, \bar{x}_{(2)}$ are the means of the first and last *(N-1)* observations.

Determining only by looking at acf won't be enough for an AR process. Partial autocorrelation function (spacf) comes to an aid to determine the order of the AR process. To find out the data set, if it's a MA or AR process, simply using the table, which is suggested by Box & Jenkins (1970), might be appropriate.

Table 2.2 Box. & Jenkins MA or AR decision table.

Process	MA	AR
Autocorrelation function	Cuts off	Infinite. Tails off. Damped Exponentials and/or Cosine waves
Partial Autocorrelation function	Infinite. Tails off. Damped Exponentials and/or Cosine waves	Cuts off

2.4 Forecasting

One of the strongest powers of time series analysis is to forecast the future values of an observed time series. Forecasting is a very important procedure in many areas such as economics, stock control or in production to determine the planning of coming seasons.

Our data from the time series aren't always simple to be foretold how it's going to end. Mostly time series are sophisticated and are dependent more then one other series, which has to be defined by multivariate, rather than univariate. Forecasting of a time series might be analyzed under these two topics.

2.4.1 Univariate procedures

There are lots of common and uncommon methods used to forecast a univariate time series. Some of them are efficient at long-term forecasting, instead of short-term. The well known basic models for both terms are extrapolation of trend curves, exponential smoothing, the Box-Jenkins procedure, stepwise autoregression, the Holt-Winters forecasting procedure.

2.4.2 Multivariate procedures

Forecasting a multivariate time series is a more complicated and harder process. As in section 2.7.1 methods vary either for long-, short-term forecasting. There are two known basic models like multiple regression and economic models.

CHAPTER THREE
FUZZY SETS & FUZZY THEORY

After describing the fundamentals of Time Series Analysis, in this chapter the basics about Fuzzy Logic will be discussed. Under following topics the use and integration of fuzzy systems through other disciplines defined more clearly.

3.1 The Concept of a Fuzzy Time Series

Fuzzy logic (FL) was introduced to the scientific arena in 1965 by Prof. Lotfi A. Zadeh, who is a professor of computer science at the University of California, Berkeley, and the first industrial applications appeared in 1970s. The historical progress of the traditional fuzzy logic is given below following the documentation of fuzzy TECH 5.3 User's Manual. One of the early applications of Fuzzy Logic Controller (FLC) was developed by Ebrahim Mamdani in England for controlling a steam engine. In Germany, Hans Jurgen Zimmermann applied FL to decision support systems. Another important milestone is the use of FL for cement kiln control in 1975 in Denmark.

These successful applications in Europe drew the interest of Japanese scientists in the beginning of 1980s. One of the early applications in Japan was on a water treatment plant, realized by Michio Sugeno in 1983. In 1987, fuzzy logic control was also applied to Sendai railways. After these applications FL became prevalent in Japan, and used in many industrial and consumer products, such as washing machines, cameras, etc. Because of the technological advantages and the establishment of many companies, quite a number of fuzzy societies have been founded in Japan. These include:

- International Fuzzy Systems Association (IFSA)
- Japan Society for Fuzzy Theory and Systems (SOFT)
- Biomedical Fuzzy Systems Association (BMFSA)
- Laboratory for International Fuzzy Engineering Research (LIFE)

- Fuzzy Logic Systems Institute Iizuka (FLSI)
- Center for Promotion of Fuzzy Logic at TITech.

The rapid rise of FL in Japan also influenced Europe and a great number of industrial applications of FL started to appear. About the same time, US also responded to the competition between Japan and Europe, and FL was used in new areas, such as decision support systems, hard disk controllers, memory cache, echo cancellation, network routing, and speech recognition.

Traditional FLCs have widely been used in many control applications with great success for more than three decades. In real life applications, systems are confronted with many uncertainties and imprecise information due to the inner and outer dynamics of the systems, such as highly nonlinear systems, incomplete sensory information and noise from external environment. To overcome these uncertainties, Fuzzy Logic Systems (FLSs) work collectively with some optimization techniques that enable the tuning of the system to achieve the desired performance.

Several approaches are proposed in the literature to this end Jang & Sun & Mizutani, (1997), Mendel (2001a). However, when a system is affected by both inner and outer uncertainties, the traditional type-1 fuzzy logic systems may become inadequate, and the type of optimization that is done becomes irrelevant. To obtain the desired performance and come up with a minimum error response, some other approaches should be sought. This thesis has the goal of comparing the performance of various different approaches to fuzzy modeling on historical time series data, namely the traditional FLS with parameterized conjunctions. The historical backgrounds of these methods are briefly summarized.

A FS (Fuzzy Sets) has IF-THEN type of rules. During the optimization process, both the antecedent and the consequent part of the rules can be tuned. If the linguistic terms play a major role in the design of fuzzy controller, the tuning of the membership functions may not be desirable as the linguistic interpretation can be lost due to the membership functions moving out of the domain or having large intersections with each other. In applications where the interpretation of the linguistic variable, the expert knowledge, and the rule base are important, the membership functions should therefore not be modified, at least not drastically. In this thesis, Fuzzy Time Series Analysis is proposed as other approaches alternative to traditional Time Series Analysis.

3.2 Fuzzy Modeling

The most important feature of fuzzy logic is the ability to define human thinking and interpretation about the system by using various kinds of (e.g., Gaussian, Gbell, Triangular, Trapezoidal) membership functions and IF-THEN type of rules. In fuzzy models, in which the human expert knowledge is the key element of the design of the fuzzy model, tuning the membership functions can result in the loss or distortion of the expert knowledge. In such applications, another type of adaptation can be more appropriate than the adaptation of the membership functions Batyrshin & Kaynak & Rudas (2002).

First of all, when we consider the traditional fuzzy logic systems, there are four main components, which can be described as in the list below. The main structure of type-1 fuzzy logic systems is shown in Figure 2.1.

- Fuzzification
- Fuzzy Rule-Base
- Fuzzy Inference Engine
- Defuzzification

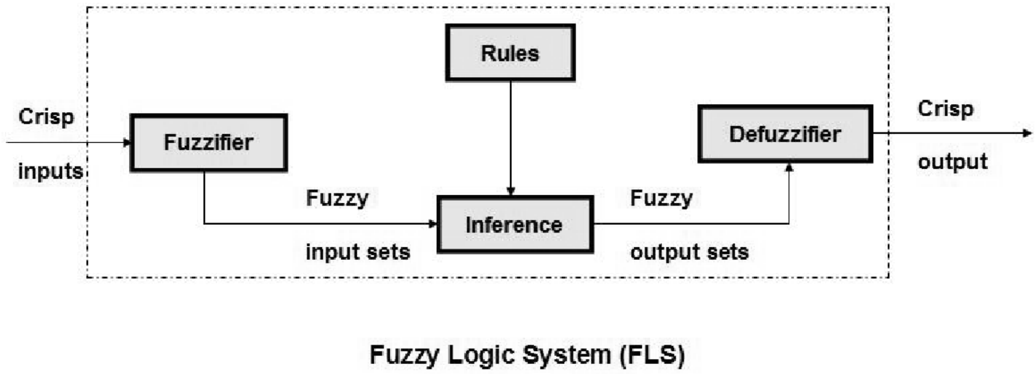


Figure 3.1 Type-1 Fuzzy Logic System

Another approach alternative to traditional fuzzy logic is type-2 fuzzy logic. In literature, type-2 fuzzy logic was first proposed by Prof. L. A. Zadeh in 1975 as an extension of type-1 fuzzy sets, and the basic mathematical and theoretical foundations were established by him John & Coupland (2007). One of the most important features of type-2 fuzzy sets is the ability to incorporate uncertainties into the membership functions, and this feature makes type-2 fuzzy sets preferable when there exist significant uncertainties.

The progress of type-2 fuzzy logic since 1975 is briefly summarized below and prepared by the help of the report "Type-2 Fuzzy Logic: A Historical View" published in 2007 John & Coupland (2007).

The emergence of fuzzy set theory goes back to the years 1975-1981. Some notable works are those carried out by Mizumoto & Tanaka (1981) and Dubois & Prade (1982) such as on logical connectives (AND and OR).

By the mid-1980s, type-2 interval fuzzy sets started to be developed by scientists, Gorzalczany, Turksen, Schwartz and Klir & Folger.

In the study of Prof. L. A. Zadeh (1996), fuzzy logic is defined as computing with words (CWW). In addition, Mendel (2001b, 2003) use type-2 fuzzy logic for CWW.

The number of publications from 1988 to today reported at http://www.type2fuzzylogic.org/publications/statistics.php can be seen in Figure 3.2. The numbers include all types of publications.

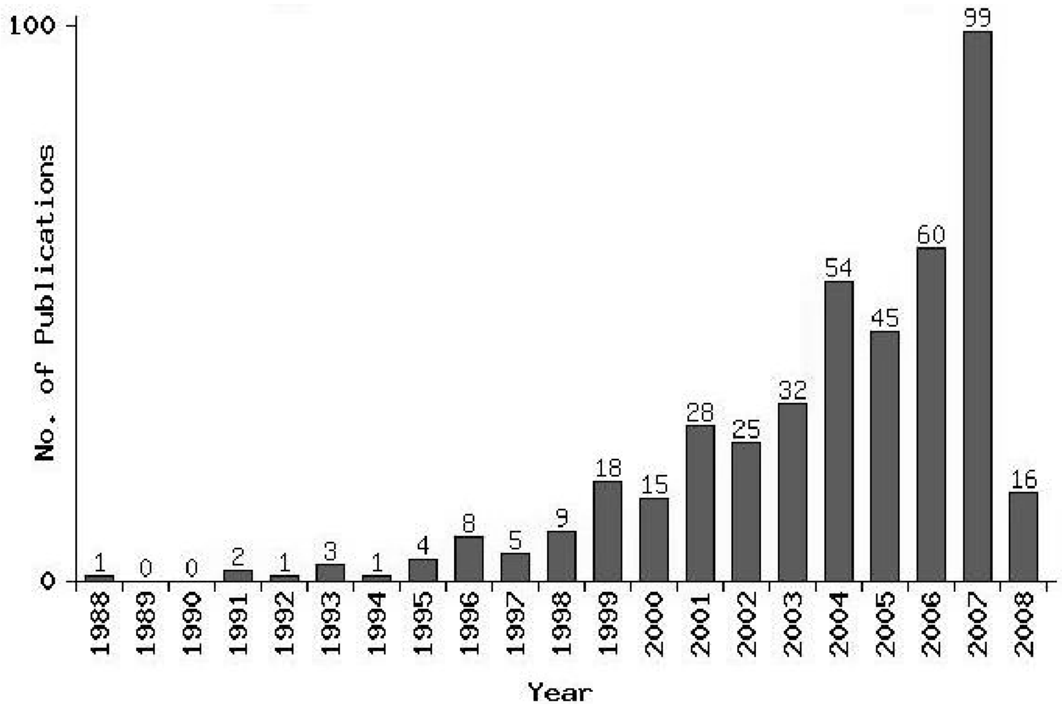


Figure 3.2 Number of publications in each year

A search in Web of Science done by entering “type-2 fuzzy” under the general search tab results in Figures 3.3 and 3.4. The number of publications those in journals cited by SCIE (Science Citation Index Expanded).

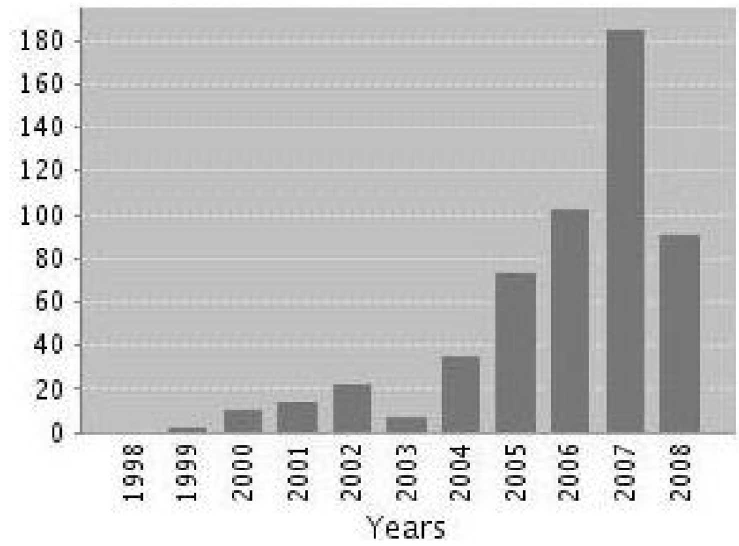


Figure 3.3 Citations in each year

Most of the applications in this topic are about in the area of control engineering and medical science. The milestones of the control applications are: Plant Control with type-2 interval fuzzy sets, type-2 interval fuzzy logic controller gives better results than type-1 under high uncertainties, control of complex multi-variable liquid level process with type-2 interval fuzzy controller, the control of non-autonomous robots in a football game with type-2 interval fuzzy logic controller.

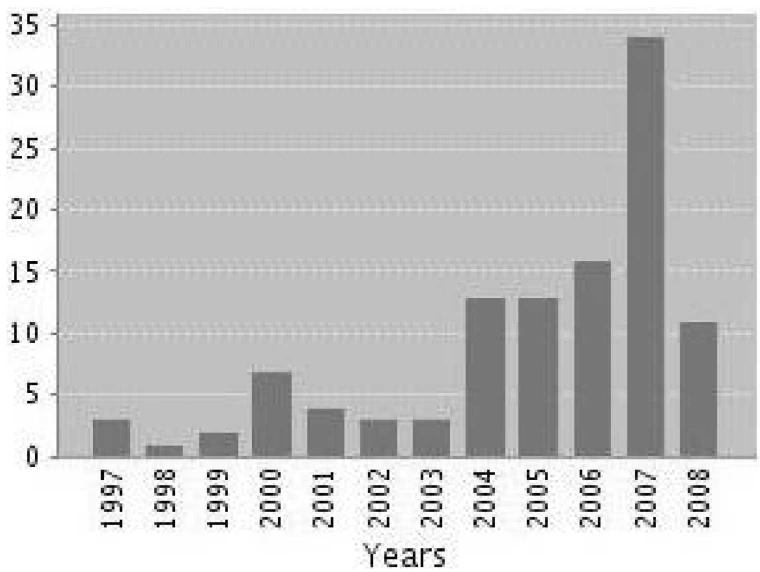


Figure 3.4 Published items in each year

As it is mentioned earlier, traditional Time Series Analysis is not efficient in many applications to problems containing great amount of uncertainty.

The aim of this thesis is a comparative study of fuzzy modeling methods which are used to forecast time series data more accurately. Based on such study, it is proposed to develop and improve alternative methods to traditional fuzzy logic and make these methods preferable in applications where the systems have great amount of uncertainty.

3.2.1 Fuzzification

Initially, the crisp inputs are fuzzified by using membership functions. A fuzzy set A is defined in universe of discourse *X* and is indicated by a membership grade, which takes values in the closed interval *0* and *1 ([0, 1])* Jang & Sun & Mizutani, (1997).

$$A = \{(x, \mu_A(x)) | x \in X\} \quad (3.1)$$

where *x* are the elements of *X*, and $\mu_A(x)$ is called the membership function, and indicates the degree of belonging . Every element of *X* maps to a membership grade taking the values between *0* and *1*. The fuzzy sets can be defined by using linguistic labels such as; SMALL, LARGE, MODERATE, YOUNG, SLOW, FAST, etc. These fuzzy sets are specified by membership functions, so that mathematical computations can be performed. There are several types of membership functions. For instance, gaussian, gbell, triangular, trapezoidal, etc. In the following several types of membership functions are shown Jang & Sun & Mizutani, (1997).

3.2.1.1 Gaussian Membership Function

A Gaussian membership function (mf) is defined as follows:

$$\text{Gaussian mf}\,(x, [sigma, center]) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \quad (3.2)$$

where c is the center and σ is the width of the membership function. x is the input of the system. The example of Gaussian mf is shown in Figure 3.5.

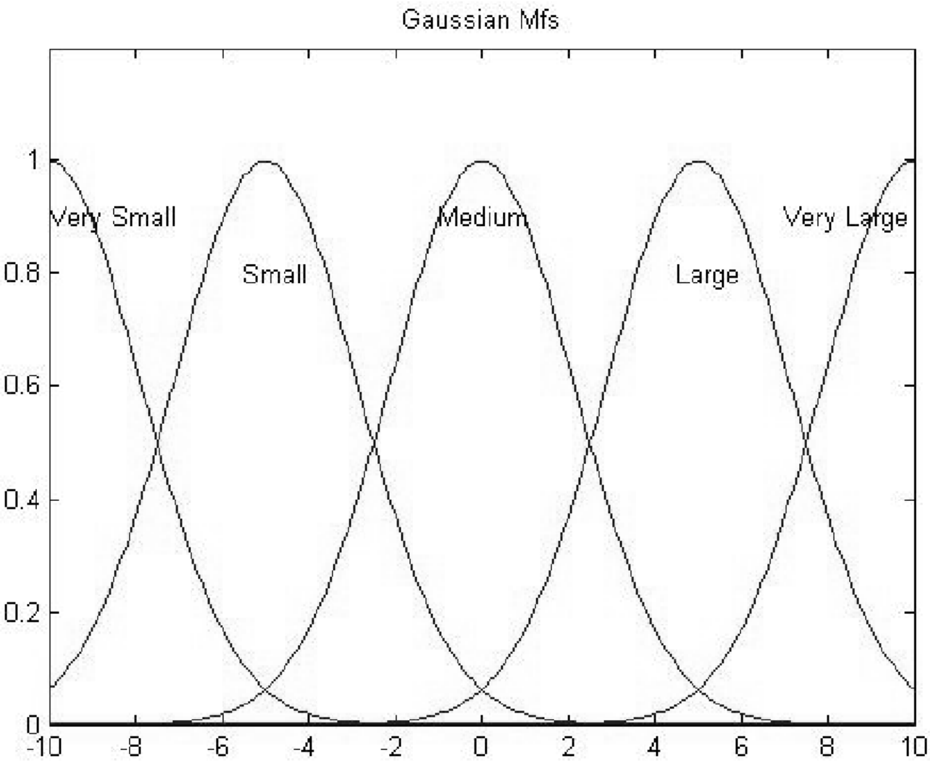


Figure 3.5 Gaussian membership functions with linguistic values “Very Small”, “Small”, “Medium”, “Large”, “Very Large”

3.2.1.2 Triangular Membership Function

$$\text{Trianglemf}(x,[a,b,c]) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ \dfrac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases} \quad \text{.......................................}(3.3)$$

where a, b, and c define the corners of the membership function and $a \le b \le c$. The example of triangle mf is shown in Figure 3.6.

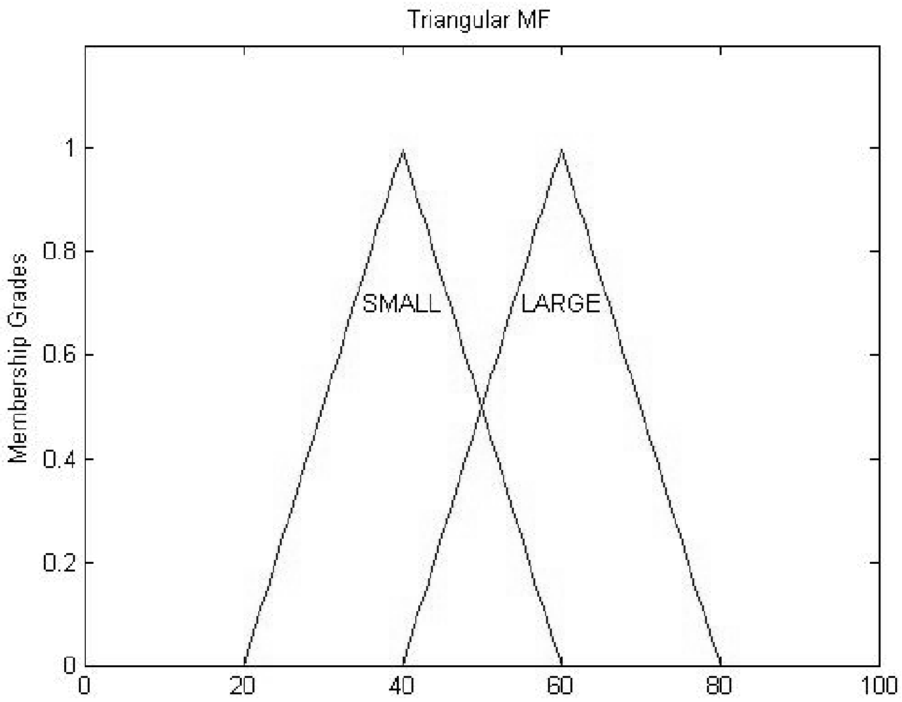


Figure 3.6 Triangular membership functions with linguistic values “Small”, “Large”

3.2.1.3 Gbell Shaped Membership Function

$$\text{Gbell mf }(x,[a,b,c]) = \frac{1}{1+\left|\frac{x-c}{s}\right|^{2b}} \quad \text{...(3.4)}$$

where *a* determines the width, *b* determines the slope and *c* determines the center of the membership function. The example of Gbell mf is shown in Figure 3.7.

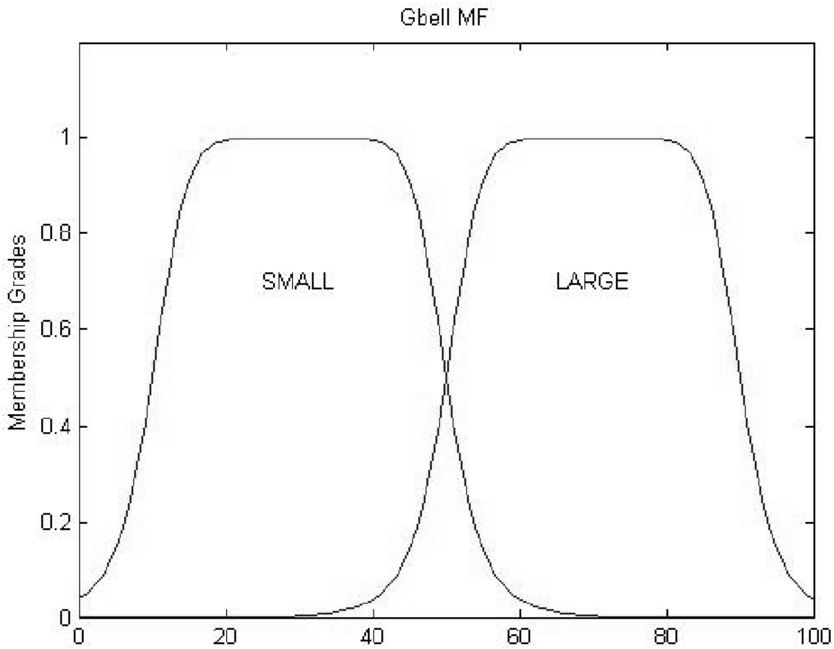


Figure 3.7 Gbell shaped membership functions with linguistic values “Small”, “Large”

3.2.2 Fuzzy Rule-Base and Fuzzy Inference System (FIS)

Fuzzy Inference Systems are prevalently applied in control engineering and in multidisciplinary areas. FIS involves nonlinear mapping from input data to output data and this nonlinear mapping is performed by using fuzzy if-then rules. Fuzzy Logic Systems are universal approximators and this property enables us to build optimal fuzzy models Batyrshin & Kaynak & Rudas (2002). Traditionally, to obtain an optimal fuzzy model, the membership function parameters are tuned.

The IF part of the rule is called antecedent or premise, and the THEN part of the rule is called consequent or conclusion part of the rule. The examples of fuzzy if-then rules that are used in daily life are as follows;

- IF temperature is HIGH and humidity is HIGH, THEN fan works fast.
- IF the soil is DRY and the temperature is HIGH, THEN open the valve ROUNDLY.
- IF X is POSITIVE LARGE and Y is POSITIVE LARGE, THEN Z is POSITIVE LARGE.

The fuzzy models differ by using different consequent membership functions, aggregation and defuzzification methods Batyrshin & Kaynak & Rudas (2002). There are various types of fuzzy models; but the most commonly used ones are:

- MAMDANI MODEL

R^i = IF X_1 is A_{i1} and ... and X_n is A_{in},

THEN $Z^i = C_i$

- SUGENO MODEL (a.k.a. TSK)

R^i = IF X_1 is A_{i1} and ... and X_n is A_{in},

THEN $z^i = a^i_n x_n + a^i_{n-1} x_{n-1} + \ldots + a^i_0$

where i $(i = 1,2,\ldots,M)$ indicates the number of rule. In these rule structures, A_{in} and C_i are the antecedent and consequent fuzzy sets, respectively. Z^i is the output of the Mamdani model and is a fuzzy set. z^i is the output of the Sugeno model, which is a first order polynomial at the consequent part of the rule structure. X_n is the input variable and n is the number of input variable.

Mamdani and Sugeno model are the same in the fuzzification block and in the antecedent part of the rules; they only differ in the consequent part of the if-then rules.

As it is seen, both in Mamdani and Sugeno model the antecedent parts of the rules are the same, which contains antecedent fuzzy sets A_{in}'s, and inputs X_n's. They differ in the consequent part of the rules. In Mamdani Model, the consequent part is a fuzzy set, C_i. On the other hand, in Sugeno Model, the consequent is a real valued function $z^i = a^i_n x_n + a^i_{n-1} x_{n-1} + \ldots + a^i_0$. Depending on the degree of the polynomial, the Sugeno model is called as zero order Sugeno model, first order Sugeno model, and so on Batyrshin & Kaynak & Rudas (2002), Jang & Sun & Mizutani, (1997). The antecedent part of the rules are combined with the fuzzy operators such as AND, OR, NOT. These operators determine the firing strength (ω^i) of the rules.

Now let's consider the traditional type-1 fuzzy logic operators and assume A_{in} are fuzzy sets where i indicate the number of rules and n indicates the number of antecedent fuzzy sets.

3.2.2.1 Intersection of Fuzzy Sets

The intersection is called as AND operator and is basically used for finding the minimum of the antecedent membership functions Jang & Sun & Mizutani, (1997)

$$\omega^i = \min\left(\mu_{A_{i1}}(x_1), \mu_{A_{i2}}(x_2)\right) \quad (3.5)$$

Generally, instead of minimum, one can use any t-norm. T-norm is defined as a function *T: [0,1] x [0,1]→ [0,1]* satisfying the four conditions monotonicity, commutativity, associativity, and boundary Jang & Sun & Mizutani, (1997)

Monotonicity:

$$T(x,y) \leq T(u,v) \text{ if } x \leq u \text{ and } y \leq v \quad (3.6)$$

Commutativity:

$$T(x,y) = T(y,x) \quad (3.7)$$

Associativity:

$$T(x,T(y,z)) = T(T(x,y),z) \quad (3.8)$$

Boundary:

$$T(0,0) = 0,$$
$$T(1,x) = T(x,1) = x \quad (3.9)$$

In literature, the most commonly used t-norm operations are minimum, algebraic product, bounded product, and drastic product that are calculated as follows, respectively Jang & Sun & Mizutani, (1997)

$T_c(x,y) = min(x,y)$ ....(3.10)

$T_p(x,y) = xy$...(3.11)

$T_b(x,y) = max\{0,(x+y-1)\}$ ....(3.12)

$$T_d(x,y) = \begin{cases} x & \text{if } y = 1 \\ y & \text{if } x = 1 \\ 0 & \text{if x, } y < 1 \end{cases} \qquad (3.13)$$

The corresponding surfaces of t-norms are given in Figure 3.8 where $0 \leq x,y \leq 1$

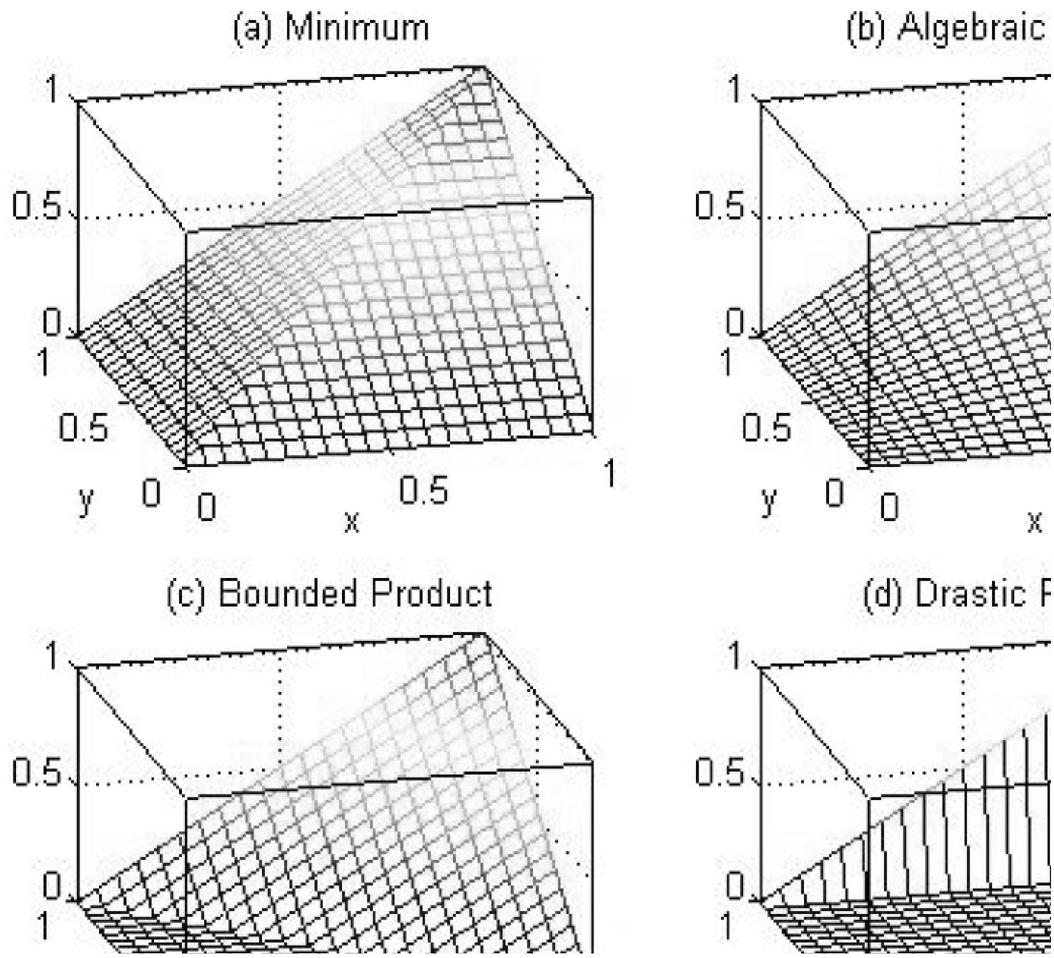


Figure 3.8 The corresponding surface of t-norms a. Minimum, b. Algebraic Product, c. Bounded Product, d. Drastic Product

3.2.2.2 Union of Fuzzy Sets

Union (disjunction) of the fuzzy sets is defined by OR operator and is calculated usually by finding the maximum of the antecedent membership functions:

$$\omega^i = \max\left(\mu_{A_{i1}}(x_1), \mu_{A_{i2}}(x_2)\right) \qquad (3.14)$$

Generally, instead of maximum, one can use any s-norm. S-norm is defined as a function $S{:}[0,1]x[0,1]\rightarrow[0,1]$ satisfying the four conditions monotonicity, commutativity, associativity, and boundary Jang & Sun & Mizutani, (1997)

<u>Monotonicity:</u>

$$S(x,y) \leq S(u,v) \text{ if } x \leq u \text{ and } y \leq u \qquad (3.15)$$

<u>Commutativity:</u>

$$S(x,y) = S(y,x) \qquad (3.16)$$

<u>Associativity:</u>

$$S(x,S(y,z)) = S(S(x,y),z) \qquad (3.17)$$

<u>Boundary:</u>

$$S(1,1) = 1,$$

$$S(x,0) = S(0,x) = x \qquad (3.18)$$

In literature, the most commonly used S-norms are maximum, algebraic sum, bounded sum, and drastic sum that are respectively calculated as follows Jang & Sun & Mizutani, (1997)

$$S_c(x,y) = max(x,y) \qquad (3.19)$$

$$S_p(x,y) = x+y-xy \qquad (3.20)$$

$S_b(x,y) = min\{1,(x+y)\}$ ....(3.21)

$$S_d(x,y) = \begin{cases} x & \text{if } y = 0 \\ y & \text{if } x = 0 \\ 1 & \text{if x, } y > 0 \end{cases} \quad \text{...(3.22)}$$

The corresponding surfaces of t-norms are given in Figure 3.9

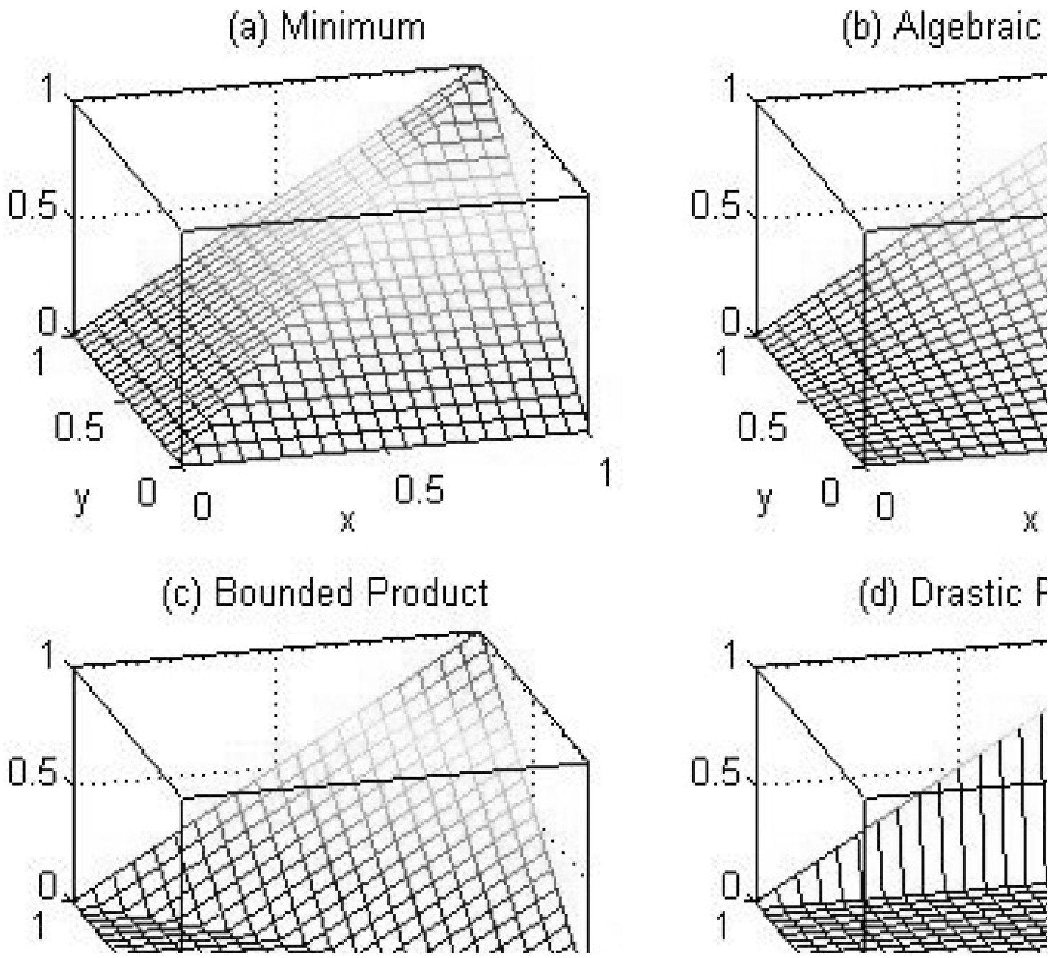


Figure 3.9 The corresponding surface of s-norms a. Minimum, b. Algebraic Sum, c. Bounded Sum, d. Drastic Sum

3.2.3 Weighted Average Calculation in TSK Model

In TSK FLS, there is no need to defuzzify the results of the rules; since they are already a crisp output Jang & Sun & Mizutani, (1997). Their weighted average is calculated as:

$$z = \frac{\sum_{i=1}^{M} \omega^i z^i}{\sum_{i=1}^{M} \omega^i} \qquad (3.23)$$

M is the number of rules ($i = 1, 2, ..., M$) and z is the actual output of the system.

3.2.4 Mamdani Fuzzy Inference and Defuzzification Methods

As it was stated earlier, the antecedent parts of the rules are same for both Mamdani and Sugeno model. However, in Mamdani model "compositional rule of inference" is carried out, and can be defined as max-min composition of fuzzy sets. If A_{in} are the antecedent membership functions and C_i is the consequent membership function, the max-min composition is calculated as:

$$\text{max-min composition} = max(min(A_{i1},...,A_{in},C_i)) \qquad (3.24)$$

In addition, compositional rule of inference can be used as the combination of max and product, for example, t-norm and t-conorm operators.After finding each result of the rule, these results are aggregated by using one of the aggregation methods; such as maximum, sum, probabilistic or MATLAB Fuzzy Logic Toolbox Tutorial.

Each result of the rule that is calculated by implication method is a fuzzy set. Defuzzification method converts the fuzzy sets into a crisp value. First of all, the qualified fuzzy sets are aggregated, and then by using appropriate defuzzification method the crisp output is derived Jang & Sun & Mizutani, (1997).

In Mamdani model, there are five types of defuzzification methods;

1. Center of Area
2. Bisector of Area
3. Small of Maximum
4. Middle of Maximum
5. Large of Maximum

3.2.4.1 Center of Area (Centroid) Defuzzification Method

Center of area method is the most commonly used defuzzification method in Mamdani models. In this method, the center of gravitiy of the aggregated output membership function is found and is calculated as follows:

$$z_0 = \frac{\int_z \mu(z) z dz}{\int_z \mu(z) dz} \quad \text{(3.25)}$$

where z_0 is the centroid of the area, a crisp value, z is the output variable, and $\mu(z)$ indicates the aggregated output of the membership functions. An example of centroid method is shown in Figure 3.10.

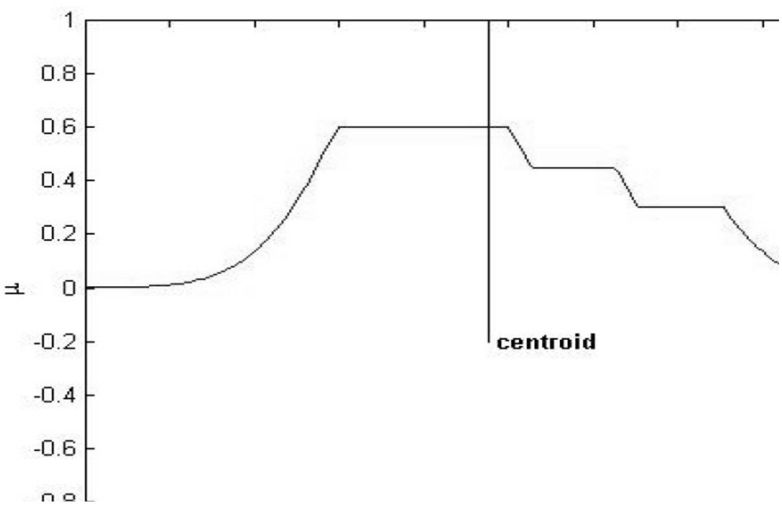


Figure 3.10 Center of area (Centroid) defuzzification method

3.2.4.2 Bisector Defuzzification Method

In bisector of area method the vertical line divides the aggregated region in two equal areas, and z_0 satisfies the following equation:

$$\int_{\alpha}^{z_0} \mu(z)dz = \int_{z_0}^{\beta} \mu(z)dz \quad \text{..........(3.26)}$$

where $\alpha = \min\{z | z \in Z\}$ and $\beta = \max\{z | z \in Z\}$. An example of bisector of area method is shown in Figure 3.11.

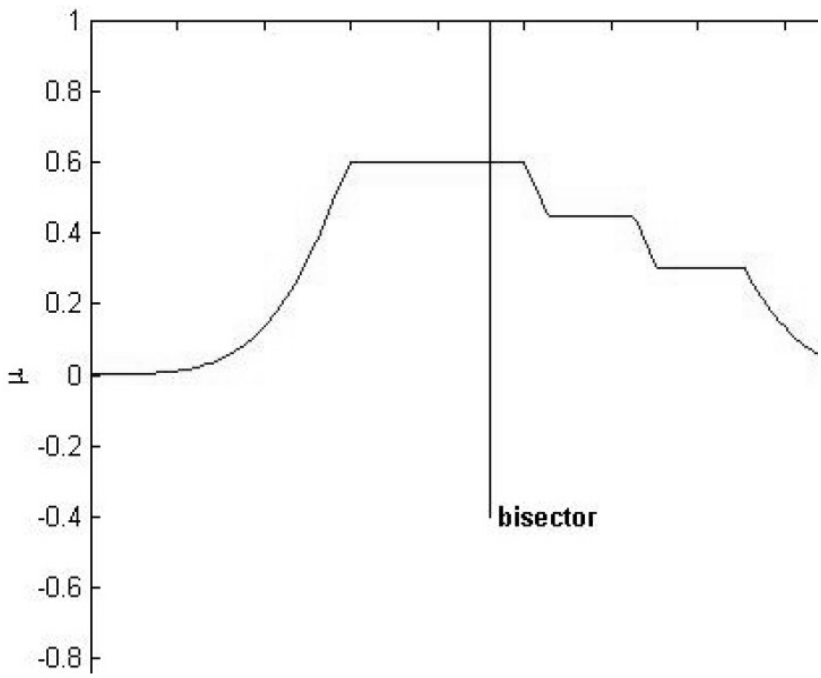


Figure 3.11 Bisector of area defuzzification method

3.2.4.3 Smallest of Maximum (SOM) Defuzzification Method

SOM, z_0, isthesmallest value where value z takes on maximum. An example of SOM defuzzification method is shown in Figure 3.12

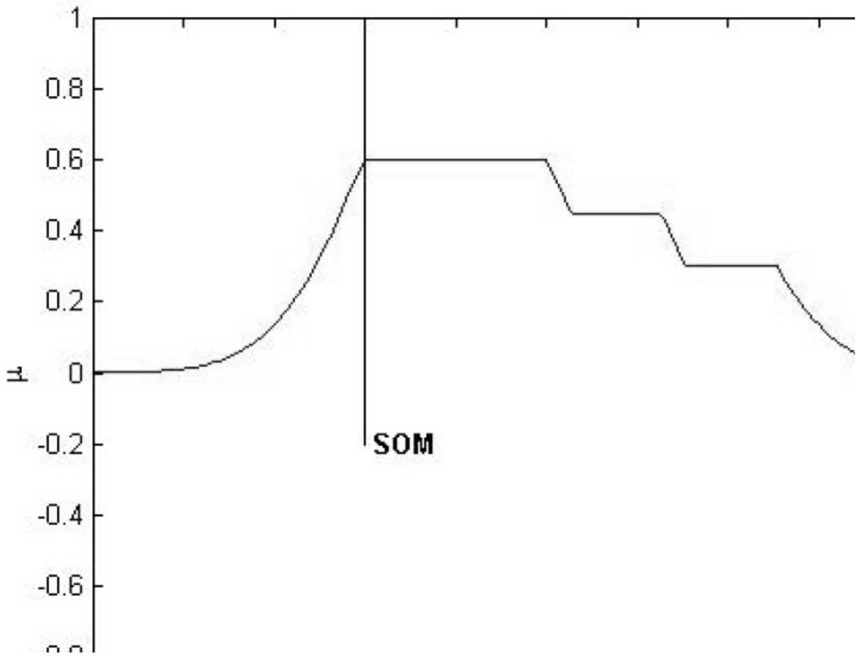


Figure 3.12 Smallest of Maximum (SOM) defuzzification method

3.2.4.4 Largest of Maximum (LOM) Defuzzification Method

The largest of the maximum, z_0, is the largest corresponding value to the largest z value. An example of LOM defuzzification method is given in Figure 3.13.

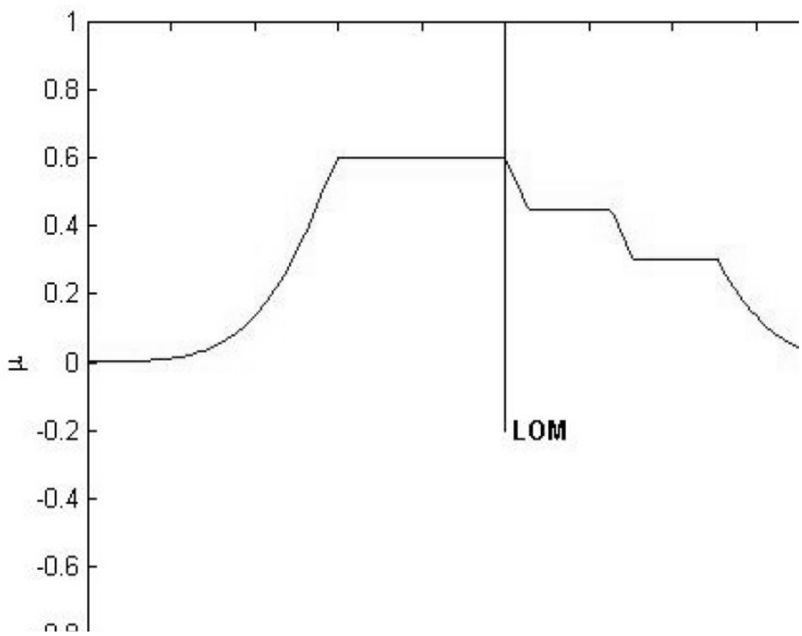


Figure 3.13 Largest of Maximum (LOM) defuzzification method

3.2.4.5 Mean of Maximum (MOM) Defuzzification Method

Mean of the maximum, is the mean value of the SOM and LOM. An example of mean of maximum method is shown in Figure 3.14.

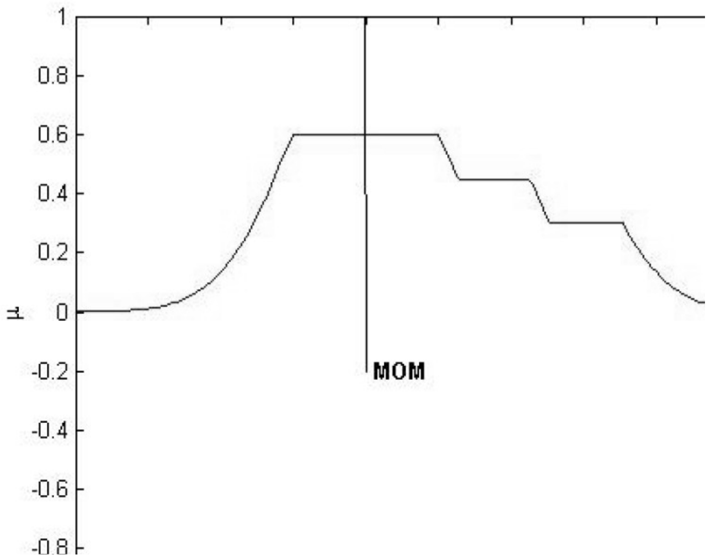


Figure 3.14 Mean of Maximum (MOM) defuzzification method

For better understanding, the defuzzification methods described above are shown in Figure 3.15.

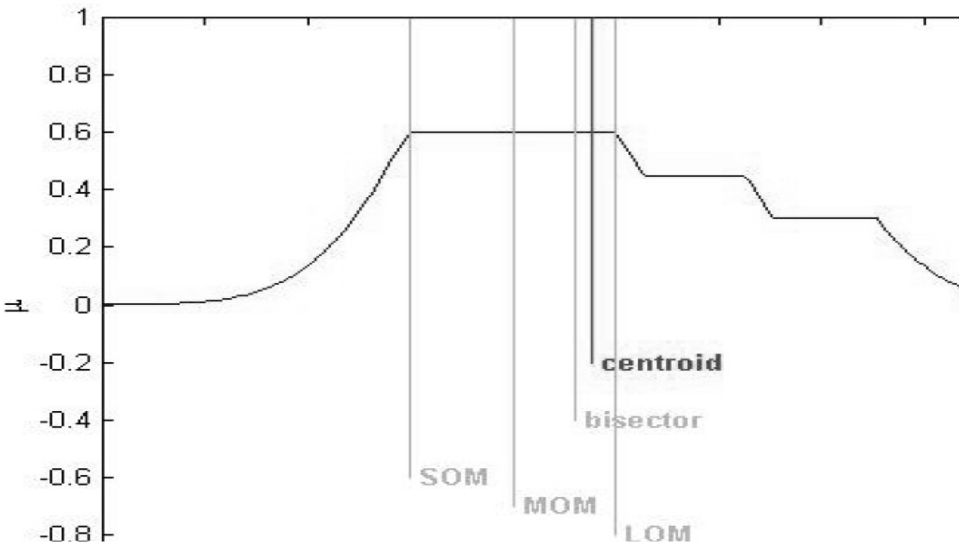


Figure 3.15 All five defuzzification methods

CHAPTER FOUR
FUZZY TIME SERIES ANALYSIS

4.1 The Concept of a Fuzzy Time Series

In recent years, many researchers have presented different forecasting methods to deal with forecasting problems based on classical time series analysis. While dealing with forecasting problems using classical time series analysis methods, it is important to decide the sufficient universe of discourse due to the fact that it will affect the forecasting accuracy. In fuzzy time series, it's been presented a new method to deal with the forecasting problems based on different orders of fuzzy time series, where the universe of discourse is tuned by using some algorithms, where Fuzzy Sets Theory and Fuzzy Reasoning is integrated to the historical observations of classical time series. The proposed methods can achieve a higher forecasting accuracy rate than some of the existing time series analysis methods.

The time series forecast has been a widely used forecasting method. Although time series forecast can deal with many forecasting problems, it cannot solve forecasting problems in which the historical data are vague, imprecise, or are in linguistic terms. To address this problem, Song and Chissom (1993a, b, 1994) presented the definitions of fuzzy time series by using fuzzy relational equations and approximate reasoning. Since then, a number of researchers have built on their research and developed different fuzzy forecasting methods (Chen (1996, 2002); Hwang, Chen & Lee (1998); Chen & Hwang (2000); Huarng (2001a,b); Lee & Chou (2004).

Generally, the existing fuzzy forecasting methods can be classified into two types: time-variant and time-invariant. In time-variant models Song & Chissom (1994), Hwang (1998), Chen & Hwang (2000) used fuzzy composition operations, such as $F(t)=F(t-1) \circ R_w(t, t-1)$ *or* $F(t)=F(t-1) \circ O_w(t)$, to calculate the forecasted values. On the other hand, time-invariant forecasting models by Song & Chissom (1993a), Chen (1996 & 2002), Huarng (2001a,b), Lee & Chou (2004) often form fuzzy logical relationships, such as $A_i \rightarrow A_m$ *or* $A_i, \ldots, A_k \rightarrow A_m$, based on historical data, and group them as heuristic rules to derive the forecasted values.

This chapter would be divided in three parts, to gain an understanding how Fuzzy Time Series methodology works. In first part the basic understanding and early times invention of Fuzzy Time Series will be discussed. The main idea and how the structure was build would be defined; hence the following two parts are going to be as classified above, time-invariant and time-variant models.

4.2 The Invention of Fuzzy Time Series

No one could deny the laudable accomplishment that time series techniques have achieved in the past decades in a wide range of areas Box & Jenkins (1970). Time series, defined as a collection of random variables indexed on time, can be employed to model many a phenomenon. As fuzzy set theory is enjoying wider and wider recognitions and acceptance, one has found it possible to consider the extension of the conventional concept of time series. One possibility is to assume that the values a time series takes are fuzzy sets while they are taken in a deterministic fashion. This has led to the concept of fuzzy time series Song & Chissom (1993). The possibility is to assume that both the values and the probability in which a time series takes its fuzzy values are fuzzy sets, and this is the motif of this chapter.

Fuzzy time series is quite common in our daily lives. For example, one usually uses linguistic terms such as "good", "bad", "not very good" and so on to express one's mood or feeling. If recording such observations, one will have a dynamic process whose observations are linguistic or fuzzy sets. This is a fuzzy time series. Through application Song & Chissom (1993), it has been found that FTS can be a good means to predict a variety of dynamic processes.

In our daily lives, it can be observed that one sometimes associates fuzzy events with a linguistic value as the probability with which the event takes place. These linguistic values are called linguistic probability in Zadeh (1975), or fuzzy probability. For example, in weather forecasting, the weatherman would associate a fuzzy probability with a certain weather condition, e.g., he may associate a high chance with a good weather, or a nearly thirty percent chance with heavy rains for the next day, and so forth. Here, the terms "a high chance", "'a good weather". "a nearly 30% chance" and "heavy rains" are fuzzy. If one recorded such weather forecasting for a period of time, one would have a dynamic process whose values are fuzzy sets and the probability with which this process assumes a given value is also a fuzzy set. Obviously, this phenomenon is not hard to encounter but how to model it mathematically needs special attention.

A natural question will be how to model or describe this process mathematically with a proper approach. Since what is involved here are fuzzy sets, fuzzy logic is of course the first candidate to be considered. As is the case of fuzzy time series, if we separated the fuzzy observations and the fuzzy probabilities, we would have two different fuzzy time series, and the methods employed in Song & Chissom (1993) can be borrowed here. But what we are more interested in is to model the process as a whole. Moreover, you would be curious to know if there is any relationship between the fuzzy observations and the fuzzy probabilities. To clear this curiosity the invention of Fuzzy Time Series should be understood.

4.2.1 Fuzzy Time Series and its Models by Q. Song & B.S. Chissom

Time series, defined as a collection of random variables indexed on time, can be employed to model many a phenomenon. As fuzzy set theory is enjoying wider and wider recognitions and acceptance, one has found it possible to consider the extension of the conventional concept of time series. One possibility is to assume that the values a time series takes are fuzzy sets while they are taken in a deterministic fashion. This has led to the concept of fuzzy time series Song & Chissom (1993, 1994). Another possibility is to consider the values a time series takes are fuzzy sets while the probability in which those values are taken is real. This is the concept of fuzzy stochastic processes Wang & Zhang (1992). The other possibility is to assume that both the values and the probability in which a time series takes its fuzzy values are fuzzy sets.

Fuzzy time series is quite common in our daily lives. For example, one usually uses linguistic terms such as "good", "bad", "not very good" and so on to express one's mood or feeling. If recording such observations, one will have a dynamic process whose observations are linguistic or fuzzy sets. This is a fuzzy time series. Through applications Song & Chissom (1993), it has been found that FTS can be a good means to predict a variety of dynamic processes.

In our daily lives, it can be observed that one sometimes associates fuzzy events with a linguistic value as the probability with which the event takes place. These linguistic values are called linguistic probability in Zadeh (1975). For example, in weather forecasting, the weatherman would associate a fuzzy probability with a certain weather condition, e.g., he may associate a high chance with a good weather or a nearly thirty percent chance with heavy rains for the next day and so forth. Here, the terms "a high chance", "'a good weather", "a nearly 30% chance" and "heavy rains" are fuzzy. If one recorded such weather forecasting for a period of time, one would have a dynamic process whose values are fuzzy sets and the probability with which this process assumes a given value is also a fuzzy set. Obviously, this phenomenon is not hard to encounter but how to model it mathematically needs special attention. It'll be define as dynamic process as Fuzzy Time Series (FTS). It is so named because of its two distinguishing characteristics: Its observations are fuzzy and the probabilities with which it assumes an observed value are fuzzy as well.

4.2.1.1 Definitions of the Fuzzy Time Series

A natural question will be how to model or describe this process mathematically with a proper approach. Since what is involved here are fuzzy sets, fuzzy logic is of course the first candidate to be considered. As is the case of fuzzy time series, if we separated the fuzzy observations and the fuzzy probabilities, we would have two different fuzzy time series, and the methods employed in Song & Chissom (1993) can be borrowed here. But what we are more interested in is to model the process as a whole. Moreover, we are curious to know if there is any relationship between the fuzzy observations and the fuzzy probabilities. The goal of this section is to give some preliminary results on FTS and its models.

In probability theory, if *Ω*, a non-empty set, is the sample space, and *A* is a *α*-algebra of subsets of *Ω,* then any element *A* in *A* is called an event. The probability of event *A*, *P(A)*, is a measure over a measurable space *(Ω, A)*, satisfying certain conditions. *(Ω, A, P)* is usually called a probability space. When a given event is not well-defined, we may encounter the so-called fuzzy event which is defined by Zadeh as follows Zadeh (1968).

Definition 1

Let *(Ω, A,P)* be a probability space in which *A* is the *α*-algebra of Borel sets in *Ω* and *P* is a probability measure over *Ω*. Then, a fuzzy event in *Ω* is a fuzzy set *A* in *Ω* whose membership function $\mu_A(\mu_A : \Omega \to [0,1])$ is Borel-measurable.

The probability of a fuzzy event *A* is defined by Zadeh with the Lebesgue-Stieltjes integral as follows Zadeh (1968):

$$P(A) = \int_{R^n} \mu_A(x)dP$$

which is the expectation of its membership function.

Klement generalized Zadeh's definition of fuzzy events by means of the fuzzy α-algebras which is stated as follows Klement (1980).

Definition 2

(Fuzzy σ-algebra). Let X be a non-empty set, *I* the unit interval [0,1] and *B* the *σ*-algebra of Borel subsets of *I*. The subset *α* of l^X is a fuzzy *σ*-algebra if

(1) $\forall x$ constant $\alpha \in \sigma$

(2) $\forall \mu \in \sigma \Rightarrow 1-\mu \in \sigma$

(3) $\forall (\mu_n)_{n \in N} \subset \sigma \Rightarrow \sup_{n \in N} \mu_n \in \sigma$

With such a definition, any element in a is also a fuzzy event. The advantage of this generalization is that fuzzy valued probability (or fuzzy probability for short) can be associated with a fuzzy event. It will adopt this generalized concept of fuzzy events.

Fuzzy probabilities are fuzzy sets defined on I=[0,1] whose membership functions are Borel-measurable. Just as probability is a measure, so is fuzzy probability. In this case, it is a fuzzy valued measure. Many authors have contributed to the development of fuzzy valued measures. Klement (1980) defined the fuzzy-valued measure in an axiomatic way where the fuzzy measure takes values on non-negative fuzzy numbers. Ralescu & Nikodym (1996) also proposed a definition of fuzzy valued measures. Other variants can be found in the literature Zhang & Li & Ma & Li (1990) and Stojakovic (1994). Here, in this section we will only consider the fuzzy probability which takes values on fuzzy sets with the understanding that fuzzy numbers may be regarded as fuzzy subsets. Similar to probability distribution, we can develop the concept of fuzzy probability distributions as a fuzzy mapping from a fuzzy *α*-algebra to a set of fuzzy probabilities, i.e., its domain is a fuzzy *α*-algebra and its range is a class fuzzy subsets defined on the interval I=[0,1]. Denote the fuzzy probability distribution as G.

Definition 3

(Fuzzy probability distribution). If a fuzzy mapping G satisfies the following conditions:

(1) $G(\Omega) = \Omega, \quad G(\emptyset) = \emptyset,$

(2) If $A \supset B$, then $G(A) \supset G(B)$;

(3) $G(A^c) = G^c(A),$

(4) $G\left(\bigcup A_i\right) = \bigcup G(A_i)$, where $\{A_i\} \in \sigma$;

then it is called a fuzzy probability distribution.

In the above, condition (1) is the boundary condition which is analogous to *P(Ω)*=1 and *P(Ø)*=0; condition (2) is simply the monotonicity of a measure; condition (3) is quite unique but necessary. For example, if we know that the probability of having a "Hot" day is "Likely", then the probability of having a day "Not Hot", according to (3), will be "Not Likely". Condition (4) says that *G* is closed under countable unions where A_i and A_j $(i \neq j)$ need not be disjoint. The necessity for (4) can be seen from an example. Suppose that the fuzzy probability of having "Hot Day" is "Likely", and that of "Very Hot Day" is "Very Likely". Then, the fuzzy probability of having "Either A Hot or A Very Hot Day" will be "Likely". This should be regarded as being consistent with what we can observe in daily lives. According to Definition 2, a fuzzy subset is characterized by its membership function. If $G(A_i)$ is a fuzzy probability, then its membership function is Borel-measurable, and therefore $\bigcup G(A_i)$ has a Borel-measurable membership function. In addition, it can be shown that the following properties can be derived from these four conditions:

(a) If $A \cap B = \emptyset$, then $G(A) \cap G(B) = \emptyset$;

(b) $G(\bigcap A_i) = \bigcap G(A_i)$, where $\{A_i\} \in \sigma$

(c) Let $\{A_i\} \in \sigma$, and $A_i \subseteq A_j$ if $i \leq j$. Then $G(\bigcup A_i) = \lim_{n \to \infty} G(A_n)$.

Several remarks are in order. Fuzzy Mapping *G* assigns a fuzzy probability to each fuzzy event in *σ*. It seems that conditions that *G* should satisfy can be proposed in an axiomatic fashion, and these conditions may not be unique, for basically *G* mimics the process how one assigns a fuzzy probability to a fuzzy event. The process of assigning a fuzzy probability to a fuzzy event is, unfortunately, influenced by one's preferences, experiences, emotion, and several other subjective and psychological factors. Thus, we would rather say that the conditions that *G* should satisfy are normative than descriptive. It is believed that when assigning a fuzzy probability, one should follow a certain set of rules although one can do it otherwise. Whether or not the conditions proposed above are meaningful can only be justified through observations. It can be seen that *G* defines a fuzzy valued measure on a. Its range, instead of in the interval [0,1], is in a class of fuzzy sets, i.e., its value can be a fuzzy subset defined on the interval [0,l]. Although there exist many open questions about this measure, we will proceed without touching upon these questions in the sequel. To define a fuzzy time series and the fuzzy stochastic fuzzy time series, we will employ the concept of fuzzy mappings proposed by Dubois & Prade (1982), although several other versions are also applicable:

Definition 4

(Fuzzy mappings). Dubois & Prade (1982) proposed a fuzzy mapping *f* from a set *U* to a set *V* is a mapping from *U* to the set of non-empty fuzzy sets of *V,* namely $P(V) \rightarrow \{\text{Ø}\}$.

With all the above definitions, we are ready to discuss fuzzy time series now. First, a new definition of fuzzy time series which is different from Song & Chissom (1993a) should be given to improve the process.

Definition 5

(Fuzzy time series). Let *M* be a fuzzy mapping from *T* to *F*:

$M : T \rightarrow F$

where $T = \{t | t = \ldots, 0, 1, 2, \ldots\}, F\{f_1, f_2, \ldots\}$, and f_i's are fuzzy sets. Then *M* is said to be a fuzzy time series, and is denoted as *F(t).* Since in Definition 5, each observation f_i is assumed implicitly to be deterministic, *F(t)* should be called a deterministic fuzzy time series.

Definition 6

(Fuzzy time series). If there exists a fuzzy relationship *R(t− 1,t)*, such that *F(t)=F(t−1)∘R(t−1,t)*, where "∘" is an arithmetic operator, then *F(t)* is said to be caused by *F(t−1)*. The relationship between *F(t)* and *F(t−1)* can be denoted by *F(t−1)→F(t).*

Definition 7

(Fuzzy time series). Suppose *F(t)* is calculated by *F(t−1)* only and *F(t)=F(t−1)∘R(t−1,t)*. For any *t*, if *R(t−1,t)* is independent of *t*, then *F(t)* is considered a time-invariant fuzzy time series. Otherwise, *F(t)* is time-variant.

Definition 8

(Fuzzy time series). Suppose $F(t-1)=\widetilde{A}_i$ and $F(t)=\widetilde{A}_j$, a fuzzy logical relationship can be defined as

$$\widetilde{A}_i \rightarrow \widetilde{A}_j$$

where are called the left-hand side and the right-hand side of the fuzzy logical relationship, respectively.

4.2.1.2 Major Steps of the Fuzzy Time Series

Chen (1996) revised the time-invariant models in Song & Chissom (1993 a, b) to simplify the calculations. In addition, Chen's method can generate more precise forecasting results than those of Song and Chissom (1993 a, b). Chen's method consists of the following major steps:

Step 1: Define the universe of discourse U.
Step 2: Divide U into several equal-length intervals.
Step 3: Define the fuzzy sets on U and fuzzify the historical data.
Step 4: Derive the fuzzy logical relationships based on the historical data.
Step 5: Classify the derived fuzzy logical relationships into groups.
Step 6: Utilize three defuzzification rules to calculate the forecasted values.

These major steps of solving a Time Series collection in a fuzzy way would be defined more detailed and used in different applications in data sets at following sections with various methodologies.

4.2.2 Handling Forecasting Problems Using Fuzzy Time Series

In this section, a method to deal with the forecasting problems is presented. Assume that the enrollment of year t is x and assume that the enrollment of year $t-1$ is y, then the variation of the enrollments between year t and year $t-1$ is equal to $x-y$. Firstly, we describe some heuristic rules which are similar to the human thought:

Rule 1: The variation of the enrollments between this year and last year is related to the variations of the enrollments between this year and the past years, and the relationship of the enrollments between this year and last year is closer than the one between this year and the other past years.

Rule 2: If the trend of the number of enrollments of the past years is increasing, then the number of enrollments of this year is increasing. If the trend of the number of enrollments of the past years is decreasing, then the number of enrollments of this year is decreasing.

From Rules 1 and 2, we might have two problems. Firstly, if the trend of the variations of the enrollments of the past years is not so obvious, how can we know the trend of the variation of the enrollment this year? Secondly, how to define the degree of variation of this year'? The solutions of these two problems are described by the following heuristic rule:

Rule 3: Let the variation of last year be a criterion. Compute the fuzzy relationships between last year and the other past years based on data variations. From the derived fuzzy relationships, we can know the degrees of relationships between the variation of last year and the variations of other past years. The variation of this year can be obtained from the derived fuzzy relationships.

Based on these heuristic rules, firstly we can fuzzify the historical enrollment data. In Song & Chen (1993a, 1993b) used the linguistic values (not many), (not too many), (many), (many many), (very many), (too many), (too many many) to describe the enrollments of the historical data. In this paper, we use the fuzzified variation of the historical enrollments and the linguistic values (big decrease), (decrease), (no change), (increase), (big increase), (too big increase) to forecast the university enrollments. The fuzzified variation of the historical enrollments between year *t* and year *t-1* can be described as follows:

$$F(t) = u_1 /(\text{big decrease}) + u_2 /(\text{decrease}) + \cdots + u_i /(\text{L}) + u_m /(\text{big increase})$$

where *F(t)* denotes the fuzzified variation of the enrollments between year *t* and year *t - 1*, u_i is the grade of membership to the linguistic value *L, m* is the number of the elements in the universe of discourse, and $1 \le i \le m$.

To forecast the enrollment of year *t*, we must decide how many years of the enrollments data will be used, where the number of years of the enrollments data we used is called the window basis. Suppose we set a window basis to w years, then the variation of last year is used to be a criterion and the other variations of w past years are used to form a matrix which is called the operation matrix. The criterion matrix *C(t)* and the operation matrix $O^W(t)$ at year *t* are expressed as follows:

$$C(t) = F(t-1) = \begin{bmatrix} (\text{big decrease}) & (\text{decrease}) & \cdots & (\text{too big increase}) \\ C_1 & C_1 & \cdots & C_1 \end{bmatrix}$$

$$O^W(t) = \begin{bmatrix} F(t-2) \\ F(t-3) \\ \vdots \\ F(t-w-1) \end{bmatrix} = \begin{bmatrix} (\text{big decrease}) & (\text{decrease}) & \cdots & (\text{too big increase}) \\ \mathrm{O}_{11} & \mathrm{O}_{12} & \cdots & \mathrm{O}_{1m} \\ \mathrm{O}_{21} & \mathrm{O}_{22} & \cdots & \mathrm{O}_{1m} \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{O}_{w1} & \mathrm{O}_{w2} & \cdots & \mathrm{O}_{wm} \end{bmatrix}$$

We can calculate the relation between the operation matrix $O^W(t)$ and the criterion matrix *C(t),* and we can get a relation matrix *R(t)[w,m]* by performing $R(t)=O^W(t)\circ C(t)$, where

$$R(t)=\begin{bmatrix} O_{11}\times C_1 & O_{12}\times C_2 & \cdots & O_{1m}\times C_m \\ O_{21}\times C_1 & O_{21}\times C_2 & \cdots & O_{2m}\times C_m \\ \vdots & \vdots & \vdots & \vdots \\ O_{w1}\times C_1 & O_{w2}\times C_2 & \cdots & O_{wm}\times C_m \end{bmatrix}=\begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ R_{21} & R_{22} & \cdots & R_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ R_{w1} & R_{w2} & \cdots & R_{wm} \end{bmatrix} \text{..........(1)}$$

where $R_{ij}=O_{ij}\times c_j, 1\le i\le w, 1\le j\le m$, and *"x"* is the multiplication operation. From the relation matrix *R(t),* we can know the degree of relationships between last year and the other past years in data variations. Then, we can get the forecasting variation of the enrollment of year *t*, where

$$F(t)=[Max(\mathrm{R}_{11},\mathrm{R}_{21},\ldots,\mathrm{R}_{\mathrm{w1}},)\cdots Max(\mathrm{R}_{12},\mathrm{R}_{22},\ldots,\mathrm{R}_{\mathrm{w1}},)\cdots Max(\mathrm{R}_{\mathrm{1m}},\mathrm{R}_{\mathrm{2m}},\ldots,\mathrm{R}_{\mathrm{wm}},)]$$

The methods algorithm steps are presented as follows:

Step 1: Calculating the variations of the historical data.

From the historical enrollment data shown in Song & Chissom (1993 a, b), compute the variations of the enrollments between any two continuous years. The variation of this year is the enrollment of this year minus the enrollment of last year. For example, if the enrollment of *1972* is *13,563* and the enrollment of *1971* is *13,055* then the variation of year *1972=13,563–13,055=508*. Based on the historical enrollment data shown in Song & Chissom (1993 a, b), we can obtain the variations of the enrollments between any two continuous years as shown in Table 4.1.

We can find the minimum increase D_{min} and maximum increase D_{max}. Then we define the universe of discourse *U*, $U=[D_{min} - D_1, D_{max} + D_2]$, where D_1 and D_2 are suitable positive numbers. In this section, we set D_{min}= *-955*, D_{max}=*-291*, D_1=*45*, D_2=*109,* so *U* can be represented as *U = [-1000, 1400].*

Table 4.1 Alabama enrollments differentiations

	Actual Enrollments	Differentiation
1971	*13,055*	
1972	*13,563*	*+508*
1973	*13,867*	*+304*
1974	*14,696*	*+829*
1975	*15,460*	*+764*
1976	*15,311*	*-149*
1977	*15,603*	*+292*
1978	*15,861*	*+258*
1979	*16,807*	*+946*
1980	*16,919*	*+112*
1981	*16,388*	*-531*
1982	*15,433*	*-955*
1983	*15,497*	*+64*
1984	*15,145*	*-352*
1985	*15,163*	*+18*
1986	*15,984*	*+821*
1987	*16,859*	*+875*
1988	*18,150*	*+1291*
1989	*18,970*	*+820*
1990	*19,328*	*+358*
1991	*19,337*	*+9*
1992	*18,876*	*-461*

Step 2: Partition the universe of discourse U into several even length intervals u_1, u_2, …, u_m.

In this section, the universe of discourse U partitioned into six intervals, where u_1=[-1000,-600], u_2=[-600,-200], u_3=[-200,200], u_4=[200,600], u_5=[600,1000], and u_6=[1000,1400].

Step 3: Define fuzzy sets on the universe of discourse U.

First, determine some linguistic values represented by fuzzy sets to describe the degree of variation between two continuous years. In this paper, we consider six fuzzy sets which are A_1=(big_decrease), A_2=(decrease), A_3=(no_change), A_4=(increase), A_5=(big_increase), A_6=(too_big_increase). Then, define fuzzy sets A_1, A_2... A_6 on the universe of discourse U as follows:

$$
\begin{aligned}
A_1 &= 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 \\
A_2 &= 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + 0/u_6 \\
A_3 &= 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + 0/u_5 + 0/u_6 \\
A_4 &= 0/u_1 + 0/u_2 + 0/u_3 + 0.5/u_4 + 1/u_5 + 0.5/u_6 \\
A_5 &= 0/u_1 + 0/u_2 + 0/u_3 + 0.5/u_4 + 1/u_5 + 0.5/u_6 \\
A_6 &= 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0.5/u_5 + 1/u_6
\end{aligned}
$$

Step 4: Fuzzify the values of historical data.

If the number of variation of the enrollment of year i is p, where $p \in u_i$, and if there is a value represented by fuzzy set A_k in which the maximum membership value occurs at u_j, then p is translated to A_k. The fuzzified variations of the enrollment data are shown in Table 4.2.

Table 4.2 Fuzzified historical enrollments

Year	Differentiation	Fuzzified variations	Year	Differentiation	Fuzzified variations
1971			*1982*	*-955*	A_1
1972	*+508*	A_4	*1983*	*+64*	A_3
1973	*+304*	A_4	*1984*	*-352*	A_2
1974	*+829*	A_5	*1985*	*+18*	A_3
1975	*+764*	A_5	*1986*	*+821*	A_5
1976	*-149*	A_3	*1987*	*+875*	A_5
1977	*+292*	A_4	*1988*	*+1291*	A_6
1978	*+258*	A_4	*1989*	*+820*	A_5
1979	*+946*	A_5	*1990*	*+358*	A_4
1980	*+112*	A_3	*1991*	*+9*	A_3
1981	*-531*	A_2	*1992*	*-461*	A_2

Step 5: Choose a suitable window basis w, and calculate the output from the operation matrix $O^{W}S(t)$ and the criterion matrix $C(t)$, where t is the year for which we want to forecast the enrollment. For example, if we set w=5, then we can set a 4x6 operation matrix $0^{5}(t)$ and a 1x6 criterion matrix $C(t)$. Because w=5, we must use six past years enrollment data, so we begin to forecast in 1977. In this case, the operation matrix $0^{5}(t)$ and the criterion matrix $C(t)$ are as follows:

$$O^{5}(1977)=\begin{bmatrix}\text{fuzzy variation of the enrollment of 1975}\\ \text{fuzzy variation of the enrollment of 1974}\\ \text{fuzzy variation of the enrollment of 1973}\\ \text{fuzzy variation of the enrollment of 1972}\end{bmatrix}=\begin{bmatrix}A_5\\ A_5\\ A_4\\ A_4\end{bmatrix}$$

$$=\begin{bmatrix}(\text{big decrease}) & (\text{big decrease}) & (\text{no change}) & (\text{increase}) & (\text{big increase}) & (\text{too big increase})\\ 0 & 0 & 0 & 0.5 & 1 & 0.5\\ 0 & 0 & 0 & 0.5 & 1 & 0.5\\ 0 & 0 & 0.5 & 1 & 0.5 & 0\\ 0 & 0 & 0.5 & 1 & 0.5 & 0\end{bmatrix}$$

$$C(1977) = \text{fuzzy variation of the enrollment of 1976} = [A_3]$$

$$= \begin{array}{cccccc} (\text{big decrease}) & (\text{decrease}) & (\text{no change}) & (\text{increase}) & (\text{big increase}) & (\text{too big increase}) \\ \end{array}$$
$$\begin{bmatrix} 0 & 0.5 & 1 & 0.5 & 0 & 0 \end{bmatrix}$$

Calculate the relation matrix *R(t)* by $R(t)[i,j] = O^W(t)[i,j] \times C(t)[j]$, where $1 \le i \le 4$, and $1 \le j \le 6$. Then, based on formula (1), we can get

$$R(1977) = \begin{array}{cccccc} (\text{big decrease}) & (\text{big decrease}) & (\text{no change}) & (\text{increase}) & (\text{big increase}) & (\text{too big increa} \\ \end{array}$$
$$\begin{bmatrix} 0 & 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \end{bmatrix}$$

Based on formula (1), we can get the fuzzified forecasting variation *F(1977)* of year 1977 shown as follows:

$$F(1977) = \begin{array}{cccccc} (\text{big decrease}) & (\text{decrease}) & (\text{no change}) & (\text{increase}) & (\text{big increase}) & (\text{too big increase}) \\ \end{array}$$
$$\begin{bmatrix} 0 & 0 & 0.5 & 0.5 & 0 & 0 \end{bmatrix}$$

The fuzzified forecasted variations for the remaining years can be calculated by the same way and all the results are listed in Table 4.3.

Step 6: Defuzzify the fuzzy forecasted variations derived in Step 5. The following principles to defuzzify the fuzzified forecasted variations are:

(1) If the grades of membership of the fuzzified forecasted variation have only one maximum u_i, and the midpoint of u_i is m_i, then the forecasted variation is m_i. If the grades of membership of the fuzzified forecasted variation have more than one maximum $u_l, u_2, \ldots, u_k$, and their midpoints are $m_l, m_2, \ldots, m_k$ respectively, then the forecasted variation is $(m_1 + rn_2 + \ldots + m_k)/k$. For example, from Table 4.3, we can see that the maximum membership value of *F(1977)* is *0.5* which occurs at u_3 and u_4, where the midpoint of u_3 is *0* and the midpoint of u_4 is *400*. The forecasted variation of year 1977 *is (0 + 400)/2 = 200*.

Table 4.3 Membership functions of the forecasted variations

Year	U_1	U_2	U_3	U_4	U_5	U_6
1977	*0*	*0*	*0.5*	*0.5*	*0*	*0*
1978	*0*	*0*	*0.5*	*1*	*0.5*	*0*
1979	*0*	*0*	*0.5*	*1*	*0.5*	*0*
1980	*0*	*0*	*0*	*0.5*	*1*	*0.25*
1981	*0*	*0.25*	*1*	*0.5*	*0*	*0*
1982	*0*	*0.5*	*0.5*	*0*	*0*	*0*
1983	*0.5*	*0.5*	*0*	*0*	*0*	*0*
1984	*0*	*0.5*	*1*	*0.25*	*0*	*0*
1985	*0.5*	*1*	*0.5*	*0*	*0*	*0*
1986	*0*	*0.5*	*1*	*0.25*	*0*	*0*
1987	*0*	*0*	*0*	*0.25*	*0*	*0*
1988	*0*	*0*	*0*	*0.25*	*1*	*0.25*
1989	*0*	*0*	*0*	*0*	*0.5*	*0.5*
1990	*0*	*0*	*0*	*0.25*	*1*	*0.5*
1991	*0*	*0*	*0*	*0.5*	*0.5*	*0*
1992	*0*	*0*	*0.5*	*0.5*	*0*	*0*

(2) If the grades of membership of the fuzzified forecasted variation are all *0*, then we set the forecasted variation to *0*.

Table 4.4 Forecasting results of the fuzzy time series method

	Actual Enrollments	Forecasted Enrollments	Errors
1977	*15,603*	*15,511*	*0.59%*
1978	*15,861*	*16,003*	*0.90%*
1979	*16,807*	*16,261*	*3.25%*
1980	*16,919*	*17,607*	*4.04%*
1981	*16,388*	*16,919*	*3.24%*
1982	*15,433*	*16,188*	*4.89%*
1983	*15,497*	*14,833*	*4.28%*
1984	*15,145*	*15,497*	*2.32%*
1985	*15,163*	*14,745*	*2.76%*
1986	*15,984*	*15,163*	*5.14%*
1987	*16,859*	*16,384*	*2.82%*
1988	*18,150*	*17,659*	*2.71%*
1989	*18,970*	*19,150*	*0.95%*
1990	*19,328*	*19,770*	*2.29%*
1991	*19,337*	*19,928*	*3.06%*
1992	*18,876*	*19,537*	*3.50%*

Step 7: Calculate the forecasted enrollments. The forecasted enrollment is forecasted variation plus the number of actual enrollment of last year. For example, if the forecasted variation in 1977 is *200*, and the actual enrollment in 1976 is *15,311*, then the forecasted enrollment of 1977 is *15,311+200=15,511*. The results of the forecasted enrollment of the University of Alabama are shown in Table 4.4. The following error of each year by the proposed method under the window basis $w=5$ is also shown in Table 4.4. The curve of the actual enrollments and the forecasted enrollments are shown in Fig. 4.1 where the window basis is 5.

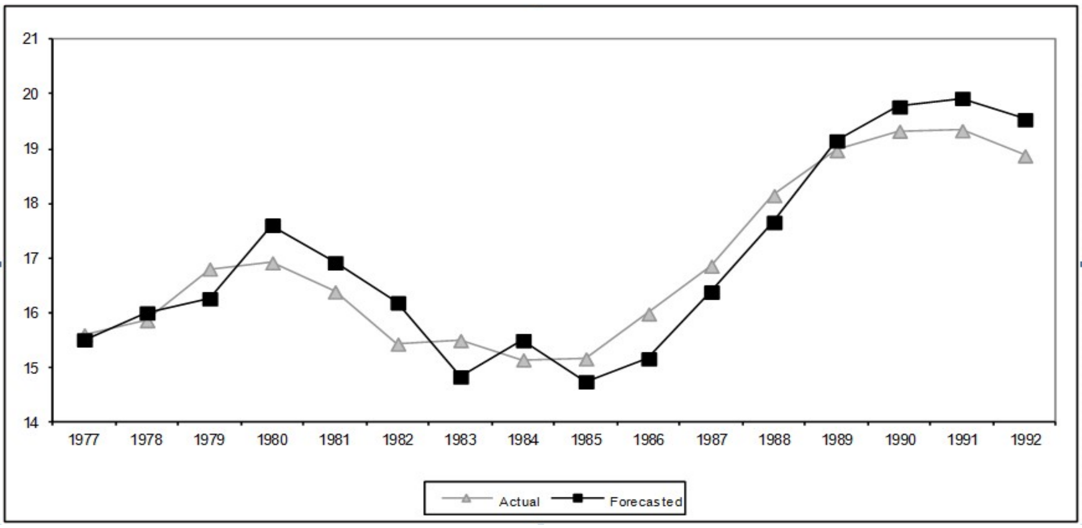


Figure 4.1 The Curve of the Actual Enrollments & Forecasted Enrollments

4.2.3 Fuzzy Time Series Modeling Using Trapezoidal Fuzzy Numbers

The fuzzy time series forecast Song & Chissom (1993a, b) was a widely used forecasting method, which can solve forecasting problems in which the historical data are vague, imprecise, or are in linguistic terms. Since then, a number of researchers have built on their research and developed different fuzzy forecasting methods Chen (1996 & 2002), Hwang, Chen & Lee (1998), Chen & Hwang (2000), Huarng (2001a, b), Lee & Chou (2004). As discussed previously, the existing fuzzy forecasting methods can be classified into two types: time-variant and time-invariant.

However, the drawback of both time-variant and time-invariant forecasting lies in the fact that their forecasting value is a single-point value. In some way, the forecasting results are just like the output of the traditional time series forecasting methods. Nevertheless, the single-point value cannot provide a decision analyst more useful information. To resolve this problem, the present study intends to develop an improved fuzzy time series method based on Chen's method (1996) because it provides an efficient forecasting algorithm and generates better forecasting results. The present research can achieve two major goals. The first goal is to provide the forecasting values with a trapezoidal fuzzy number instead of a single-point value. By doing so, the decision analyst can gather the information about the possible forecasted ranges under different degrees of confidence. The second goal is to revise Chen's algorithm to improve the accuracy in forecasting values. Two numerical examples were employed to effectively compare the proposed method with three fuzzy time series methods Chen (1996), Hwang (1998), Lee & Chou (2004) as well as to illustrate the proposed method and evaluate its forecasting performance.

Song & Chissom (1993a, b, 1994) defined their fuzzy time series by means of discrete fuzzy sets. The discrete fuzzy sets can be defined as follows: Let U be the universe of discourse, where $U = \{u_1, u_2, ..., u_n\}$. A fuzzy set A_i of U is defined by

$$\widetilde{A}_i = \mu_{\widetilde{A}_i}(\mu_1)/u_1 + \mu_{\widetilde{A}_i}(\mu_2)/u_2 + \cdots + \mu_{\widetilde{A}_i}(\mu_n)/u_n \text{ .... } (2)$$

where $\mu_{\widetilde{A}_i}$ is the membership function of $\widetilde{A}_i$, $\mu_{\widetilde{A}_i}: U \rightarrow [0,1]$. $\mu_{\widetilde{A}_i}(u_i)$ denotes the membership value of u_i in $\widetilde{A}_i$, $\mu_{\widetilde{A}_i}(u_i) \in [0,1]$ and $1 \leq i \leq n$.

Chen (1996) revised the time-invariant models in Song & Chissom (1993a, b) to simplify the calculations. In addition, Chen's method can generate more precise forecasting results than those of Song & Chissom (1993a, b). Chen's method consists of the following major steps:

Step 1: Define the universe of discourse U.
Step 2: Divide U into several equal-length intervals.
Step 3: Define the fuzzy sets on U and fuzzify the historical data.
Step 4: Derive the fuzzy logical relationships based on the historical data.
Step 5: Classify the derived fuzzy logical relationships into groups.
Step 6: Utilize three defuzzification rules to calculate the forecasted values.

The aim of the present research is to develop an improved fuzzy time series method that can both provide the forecasting values in terms of trapezoidal fuzzy numbers and generate more accurate forecasting results at the same time. As mentioned in Sect. 1, we chose Chen's method as a foundation to develop the proposed method. Several modifications between the proposed method and Chen's method are listed below:

1. Use a more advanced method to determine the number of equal-length intervals.
2. Use trapezoidal fuzzy numbers to define the fuzzy sets in fuzzy time series.
3. Apply the arithmetic operations of trapezoidal fuzzy numbers to compute the forecasted values.

First, the number and the length of intervals are assigned subjectively in Chen's method. However, Huarng (2001b) argued that the different number of intervals could affect the accuracy of the forecasting results. To resolve this problem, Huarng designed an average-based length method that can effectively determine the appropriate interval length in order to improve the forecasting results. Hence, the first modification is to employ the average-based length method to determine the appropriate length and number of intervals. Second, the current fuzzy time series models Song & Chissom (1993a, b, 1994), Chen (1996, 2002), Hwang (1998), Chen & Hwang (2000), Huarng (2001a, b), Lee & Chou (2004) utilize discrete fuzzy sets to define their fuzzy time series. Their discrete fuzzy sets are defined as follows:

Assume there are *m* intervals, which are $u_1=[d_1,d_2]$, $u_2=[d_2,d_3]$, $u_3=[d_3,d_4]$, $u_4[d_4,d_5]$,..., $u_{m-3}=[d_{m-3},d_{m-2}]$, $u_{m-2}=[d_{m-2},d_{m-1}]$, $u_{m-1}=[d_{m-1},d_m]$ *and* $u_m=[d_m,d_{m+1}]$. Thus, the fuzzy sets are defined such;

$$\begin{aligned}
\widetilde{A}_1 &= 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + \cdots + 0/u_m \\
\widetilde{A}_2 &= 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + \cdots + 0/u_m \\
\widetilde{A}_3 &= 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + 0/u_5 + \cdots + 0/u_m \\
&\vdots \\
\widetilde{A}_{m-1} &= 0/u_1 + 0/u_2 + \cdots + 0/u_{m-3} + 0.5/u_{m-2} + 1/u_{m-1} + 0.5/u_m \\
\widetilde{A}_m &= 0/u_1 + 0/u_2 + \cdots + 0/u_{m-3} + 0/u_{m-2} + 0.5/u_{m-1} + 1/u_m
\end{aligned}$$

The present study attempts to replace the above discrete fuzzy sets with trapezoidal fuzzy numbers, which could be defined as

$$\mu_{\widetilde{A}}(x) = \begin{cases} 0 & , x < a \\ \dfrac{x-a}{b-a} & , a \le x \le b \\ 1 & , b \le x \le c \\ \dfrac{d-x}{d-c} & , c \le x \le d \\ 0 & , x > d \end{cases}$$

According to the above definition, the discrete fuzzy sets can be replaced with the following trapezoidal fuzzy numbers:

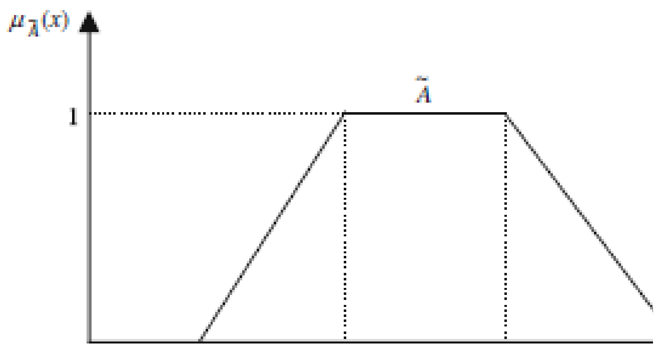


Figure 4.2 A Trapezoidal Fuzzy Number

$$\begin{aligned}
\widetilde{A}_1 &= (d_0, d_1, d_2, d_3) \\
\widetilde{A}_2 &= (d_1, d_2, d_3, d_4) \\
\widetilde{A}_3 &= (d_2, d_3, d_4, d_5) \\
&\vdots \\
\widetilde{A}_{m-1} &= (d_{m-2}, d_{m-1}, d_m, d_{m+1}) \\
\widetilde{A}_m &= (d_{m-1}, d_m, d_{m+1}, d_{m+2})
\end{aligned}$$

Chen (1996) developed three heuristic rules to calculate the forecasted values. These three heuristic rules use the midpoints of intervals to derive the forecasted values. To maintain the complete forecasting information, the present study intends to replace the midpoints of intervals with the trapezoidal fuzzy numbers. Specifically, we can apply the addition operation and the scalar multiplication operation of the trapezoidal fuzzy numbers to compute the forecasted values.

Step 1: Collect the historical data D_{vt}.

Step 2: Define the universe of discourse U. Find the maximum D_{max} and the minimum D_{min} among all D_{vt}. For easy partitioning of U, two small numbers D_1 and D_2 are assigned. The universe of discourse U is then defined by:

$$U = [D_{\min} - D_1, D_{\max} + D_2]$$

Step 3: Determine the appropriate length of interval l. Here, the average-based length method by Huarng (2001b) can be applied to determine the appropriate l. The length of interval l is computed by the following steps:

Table 4.5 Base length coefficients of fuzzy numbers

Range	Base
0.1 – 1.0	0.1
1.1 – 10	1
11 - 100	10
101 – 1,000	100
1,001 – 10,000	1,000

1. Calculate all the absolute differences between the values D_{vt-1} and D_{vt} as the first differences, and then compute the average of the first differences.
2. Take one-half of the average as the length.
3. Find the located range of the length and determine the base Table 4.5.
4. According to the assigned base, round the length as the appropriate l.

Step 4: Define fuzzy numbers. The number of intervals (fuzzy numbers), *m*, is computed by

$$m = [D_{\max} + D_2 - D_{\min} + D_1]/l$$

Thus, there are m intervals and m fuzzy numbers.

Assume there are m intervals, and *m* fuzzy numbers, which are $u_1=[d_1,d_2]$, $u_2=[d_2,d_3]$, $u_3=[d_3,d_4]$, $u_4[d_4,d_5]$, ..., $u_{m-3}=[d_{m-3},d_{m-2}]$, $u_{m-2}=[d_{m-2},d_{m-1}]$, $u_{m-1}=[d_{m-1},d_m]$ *and* $u_m=[d_m,d_{m+1}]$. And the fuzzy number are defined like

$$\begin{aligned}
\widetilde{A}_1 &= (d_0, d_1, d_2, d_3) \\
\widetilde{A}_2 &= (d_1, d_2, d_3, d_4) \\
\widetilde{A}_3 &= (d_2, d_3, d_4, d_5) \\
&\vdots \\
\widetilde{A}_{m-1} &= (d_{m-2}, d_{m-1}, d_m, d_{m+1}) \\
\widetilde{A}_m &= (d_{m-1}, d_m, d_{m+1}, d_{m+2})
\end{aligned}$$

Step 5: Fuzzify the historical data. If the value of D_{vt} is located in the range of u_j, then it belongs to fuzzy number A_j. All D_{vt} must be classified into the corresponding fuzzy numbers.

Step 6: Generate the fuzzy logical relationships. For all fuzzified data, derive the fuzzy logical relationships based on Definition 8 in Section 4.2.1.1. The fuzzy logical relationship is like $\widetilde{A}_j \rightarrow \widetilde{A}_k$ which denotes that "if the D_{vt-1} value of time *t-1* is A_j, then that of time *t* is A_k.

Step 7: Establish the fuzzy logical relationship groups. The derived fuzzy logical relationships can be arranged into fuzzy logical relationship groups based on the same fuzzy numbers on the left-hand sides of the fuzzy logical relationships. The fuzzy logical relationship groups are like the following;

$$\mathrm{A_j} \longrightarrow A_{k1}$$
$$\mathrm{A_j} \longrightarrow A_{k2}$$
$$\vdots$$
$$\mathrm{A_j} \longrightarrow A_{kp}$$

Step 8: Calculate the forecasted outputs. The forecasted value at time t, F_{vt} is determined by the following three heuristic rules. Assume the fuzzy number of D_{vt-1} at time t-1 is A_j.

Rule 1: If the fuzzy logical relationship group of A_j is empty; then the value of F_{vt} is A_j, which is $(d_{j-1}, d_j, d_{j+1}, d_{j+2})$.

Rule 2: If the fuzzy logical relationship group of A_k is on-to-one; then the value of F_{vt} is A_j, which is $(d_{k-1}, d_k, d_{k+1}, d_{k+2})$.

Rule 3: If the fuzzy logical relationship group of A_j is one-to-many; then the value of F_{vt} is calculated as follows,

$$\begin{aligned} F_{vt} &= \frac{\widetilde{A}_{k1} + \widetilde{A}_{k2} + \cdots + \widetilde{A}_{kp}}{p} \\ &= \left(\frac{d_{k1-1} + \cdots + d_{kp-1}}{p}, \frac{d_{k1} + \cdots + d_{kp}}{p}, \frac{d_{k1+1} + \cdots + d_{kp+1}}{p}, \frac{d_{k1+2} + \cdots + d_{kp+2}}{p} \right) \end{aligned}$$

4.2.4 Chen's Enhanced Forecasting Enrollments Model

By this Method, a new method to forecast the enrollments of the University of Alabama based on fuzzy time series is presented. The historical enrollments of the University of Alabama are shown in Table 4.6.

Table 4.6 The historical enrollments of the university of alabama

Year	Actual Enrollments	Year	Actual Enrollments
1971	*13,055*	*1982*	*15,433*
1972	*13,563*	*1983*	*15,497*
1973	*13,867*	*1984*	*15,145*
1974	*14,696*	*1985*	*15,163*
1975	*15,460*	*1986*	*15,984*
1976	*15,311*	*1987*	*16,859*
1977	*15,603*	*1988*	*18,150*
1978	*15,861*	*1989*	*18,970*
1979	*16,807*	*1990*	*19,328*
1980	*16,919*	*1991*	*19,337*
1981	*16,388*	*1992*	*18,876*

First, this method defines the universe of discourse and partitions the universe of discourse into some even and equal length intervals. Then, it gets the statistical distributions of the historical enrollment data in each interval and re-divided each interval.

Then, it defines linguistic values represented by fuzzy sets based on the re-divided intervals and fuzzifies the historical enrollments to get fuzzified enrollments. Then, it establishes fuzzy logical relationships based on the fuzzified enrollments. Finally, it uses a set of rules to determine whether the trend of the forecasting goes up or down and to forecast the enrollments.

Assume that we want to forecast the enrollment of year *n*, then the "difference of differences" of the enrollments between years *n-1* and *n-2* and between years *n-2* and *n-3,* such as, *[(the enrollment of year (n-1)) - (the enrollment of year (n-2))] - [(the enrollment of year (n-2)) - (the enrollment of year (n-3))]*. This method is now presented as follows:

Step 1: Define the universe of discourse *U* and partition it into several even and equal length intervals $u_1, u_2... ,u_n$. For example, assume that the universe of discourse $U=[13,000, 20,000]$ is partitioned into seven even and equal length intervals $u_1, u_2, u_3, u_4, u_5, u_6, u_7$, where $u_1=[13,000, 14,000]$, $u_2=[14,000, 15,000]$, $u_3=[15,000,$

16,000], u_4*=[16,000, 17,000],* u_5*=[17,000, 18,000],* u_6*=[18,000, 19,000] and* u_7*=[19,000, 20,000].*

Step 2: Get a statistics of the distribution of the historical enrollments in each interval. Sort the intervals based on the number of historical enrollment data in each interval from the highest to the lowest. Find the interval having the largest number of historical enrollment data and divide it into four sub-intervals of equal length. Find the interval having the second largest number of historical enrollment data and divide it into three sub-intervals of equal length. Find the interval having the third largest number of historical enrollment data and divide it into two sub-intervals of equal length. Find the interval with the fourth largest number of historical enrollment data and let the length of this interval remain unchanged. If there are no data distributed in an interval, and then discard this interval.

For example, the distributions of the historical enrollment data in different intervals are summarized as shown in Table 4.7.

Table 4.7 The distribution of the historical enrollment data

Intervals	Number of historical enrollment
[13,000, 14,000]	*3*
[14,000, 15,000]	*1*
[15,000, 16,000]	*9*
[16,000, 17,000]	*4*
[17,000, 18,000]	*0*
[18,000, 19,000]	*3*
[19,000, 20,000]	*2*

After executing this step, the universe of discourse *[13,000, 20,000]* is re-divided into the following intervals in Table 4.8

Table 4.8 The intervals of the historical enrollment data

$U_{1,1}$=[13,000, 13,500]	$U_{1,2}$=[13,500, 14,000]
U_2=[14,000, 15,000]	$U_{3,1}$=[15,000, 15,250]
$U_{3,2}$=[15,250, 15,500]	$U_{3,3}$=[15,500, 15,750]
$U_{3,4}$=[15,750, 16,000]	$U_{4,1}$=[16,000, 16,333]
$U_{4,2}$=[16,333, 13,500]	$U_{4,3}$=[16,667, 17,000]
$U_{6,1}$=[18,000, 18,500]	$U_{6,2}$=[18,500, 19,000]
U_7=[19,000, 20,000]	

Step 3: Define each fuzzy set A_i based on the re-divided intervals and fuzzify the historical enrollments shown in Table 4.6, where fuzzy set A_i denotes a linguistic value of the enrollments represented by a fuzzy set, and $1 \le i \le 13$. For example, A_1=(very_very_very_very_few), A_2=(very_very_very_few), A_3=(very_very_few), A_4=(very_few), A_5=(few), A_6=(moderate), A_7=(many), A_8=(many_many), A_9=(very_many), A_{10}=(very_few), A_{11}=(too_many_many), A_{12}=(too_many_many_many), A_{13}=(too_many_many_many_many), defined as follows:

$$A_1 = 1/u_{1,1} + 0.5/u_{1,2} + 0/u_2 + 0/u_{3,1} + 0/u_{3,2} + 0/u_{3,3} + \\ 0/u_{3,4} + 0/u_{4,1} + 0/u_{4,2} + 0/u_{4,3} + 0/u_{6,1} + 0/u_{6,2} + 0/u_7$$

$$A_2 = 0.5/u_{1,1} + 1/u_{1,2} + 0.5/u_2 + 0/u_{3,1} + 0/u_{3,2} + 0/u_{3,3} + \\ 0/u_{3,4} + 0/u_{4,1} + 0/u_{4,2} + 0/u_{4,3} + 0/u_{6,1} + 0/u_{6,2} + 0/u_7$$

$$A_3 = 0/u_{1,1} + 0.5/u_{1,2} + 1/u_2 + 0.5/u_{3,1} + 0/u_{3,2} + 0/u_{3,3} + \\ 0/u_{3,4} + 0/u_{4,1} + 0/u_{4,2} + 0/u_{4,3} + 0/u_{6,1} + 0/u_{6,2} + 0/u_7$$

$$A_4 = 0/u_{1,1} + 0/u_{1,2} + 0.5/u_2 + 1/u_{3,1} + 0.5/u_{3,2} + 0/u_{3,3} + \\ 0/u_{3,4} + 0/u_{4,1} + 0/u_{4,2} + 0/u_{4,3} + 0/u_{6,1} + 0/u_{6,2} + 0/u_7$$

$$A_5 = 0/u_{1,1} + 0/u_{1,2} + 0/u_2 + 0.5/u_{3,1} + 1/u_{3,2} + 0.5/u_{3,3} + \\ 0/u_{3,4} + 0/u_{4,1} + 0/u_{4,2} + 0/u_{4,3} + 0/u_{6,1} + 0/u_{6,2} + 0/u_7$$

$$A_6 = 0/u_{1,1} + 0/u_{1,2} + 0/u_2 + 0/u_{3,1} + 0.5/u_{3,2} + 1/u_{3,3} + \\ 0.5/u_{3,4} + 0/u_{4,1} + 0/u_{4,2} + 0/u_{4,3} + 0/u_{6,1} + 0/u_{6,2} + 0/u_7$$

$$A_7 = 0/u_{1,1} + 0/u_{1,2} + 0/u_2 + 0/u_{3,1} + 0/u_{3,2} + 0.5/u_{3,3} + \\ 1/u_{3,4} + 0.5/u_{4,1} + 0/u_{4,2} + 0/u_{4,3} + 0/u_{6,1} + 0/u_{6,2} + 0/u_7$$

$$A_8 = 0/u_{1,1} + 0/u_{1,2} + 0/u_2 + 0/u_{3,1} + 0/u_{3,2} + 0/u_{3,3} + \\ 0.5/u_{3,4} + 1/u_{4,1} + 0.5/u_{4,2} + 0/u_{4,3} + 0/u_{6,1} + 0/u_{6,2} + 0/u_7$$

$$A_9 = 0/u_{1,1} + 0/u_{1,2} + 0/u_2 + 0/u_{3,1} + 0/u_{3,2} + 0/u_{3,3} + \\ 0/u_{3,4} + 0.5/u_{4,1} + 1/u_{4,2} + 0.5/u_{4,3} + 0/u_{6,1} + 0/u_{6,2} + 0/u_7$$

$$A_{10} = 0/u_{1,1} + 0/u_{1,2} + 0/u_2 + 0/u_{3,1} + 0/u_{3,2} + 0/u_{3,3} + \\ 0/u_{3,4} + 0/u_{4,1} + 0.5/u_{4,2} + 1/u_{4,3} + 0.5/u_{6,1} + 0/u_{6,2} + 0/u_7$$

$$A_{11} = 0/u_{1,1} + 0/u_{1,2} + 0/u_2 + 0/u_{3,1} + 0/u_{3,2} + 0/u_{3,3} + \\ 0/u_{3,4} + 0/u_{4,1} + 0/u_{4,2} + 0.5/u_{4,3} + 1/u_{6,1} + 0.5/u_{6,2} + 0/u_7$$

$$A_{12} = 0/u_{1,1} + 0/u_{1,2} + 0/u_2 + 0/u_{3,1} + 0/u_{3,2} + 0/u_{3,3} + \\ 0/u_{3,4} + 0/u_{4,1} + 0/u_{4,2} + 0/u_{4,3} + 0.5/u_{6,1} + 1/u_{6,2} + 0.5/u_7$$

$$A_{13} = 0/u_{1,1} + 0/u_{1,2} + 0/u_2 + 0/u_{3,1} + 0/u_{3,2} + 0/u_{3,3} + \\ 0/u_{3,4} + 0/u_{4,1} + 0/u_{4,2} + 0/u_{4,3} + 0/u_{6,1} + 0.5/u_{6,2} + 1/u_7$$

For simplicity, the membership values of fuzzy set A_i either are *0, 0.5* or *1*, where $1 \le i \le 13$. Then, fuzzify the historical enrollments shown in Table 4.6 and the linguistic values of the enrollments A_1, A_2... A_{13}. The reason for fuzzify the historical enrollments into fuzzified enrollments is to translate crisp values into fuzzy sets to get a fuzzy time series.

Step 4: Establish fuzzy logical relationships based on the fuzzified enrollments:

where the fuzzy logical relationship "$A_j \rightarrow A_q$ " denotes " if the fuzzified enrollments of year *n-1* is A_j , then the fuzzified enrollments of year n is A_q ". For example, based on the fuzzify historical enrollments obtained in Step 3, we can get the fuzzy logical relationships as shown in Table 4.9.

Table 4.9 Fuzzy logical relationships

$A_1 \rightarrow A_2$	$A_2 \rightarrow A_2$	$A_2 \rightarrow A_3$
$A_3 \rightarrow A_5$	$A_5 \rightarrow A_5$	$A_5 \rightarrow A_6$
$A_6 \rightarrow A_7$	$A_7 \rightarrow A_{10}$	$A_{10} \rightarrow A_{10}$
$A_1 \rightarrow A_2$	$A_9 \rightarrow A_5$	$A_5 \rightarrow A_5$
$A_{10} \rightarrow A_9$	$A_4 \rightarrow A_4$	$A_4 \rightarrow A_7$
$A_7 \rightarrow A_{10}$	$A_{10} \rightarrow A_{11}$	$A_{11} \rightarrow A_{12}$
$A_{12} \rightarrow A_{13}$	$A_{13} \rightarrow A_{12}$	$A_{13} \rightarrow A_{12}$

Step 5: Divide each interval derived in Step 2 into four subintervals of equal length, where the 0.25-point and 0.75-point of each interval are used as the upward and downward forecasting points of the forecasting. Use the following rules to determine whether the trend of the forecasting goes up or down and to forecast the enrollment. Assume that the fuzzy logical relationship is $A_i \rightarrow A_j$, where A_i denotes the fuzzified enrollment of year *n-1* and A_j denotes the fuzzified enrollment of year *n*, then (1) If $j>i$ and the difference of the differences of the enrollments between years *n-1* and *n-2* and between years *n-2* and *n-3* is positive, then the trend of the forecasting will go up, and we use the following Rule 2 to forecast the enrollments; (2) If $j>i$ and the difference of the differences of the enrollments between years *n-1* and *n-2* and between years *n-2* and *n-3* is negative, then the trend of the forecasting will go down, and we use the following Rule 3 to forecast the enrollments; (3) If $j < i$ and the difference of the differences of the enrollments between years *n-1* and *n-2* and between years *n-2* and *n-3* is positive, then the trend of the forecasting will go up, and we use the following Rule 2 to forecast the enrollments; (4) If $j<i$ and the difference of the differences of the enrollments between years *n-1* and *n-2* and between years *n-2* and *n-3* is negative, then the trend of the forecasting will go down, and we use the following Rule 3 to forecast the enrollments; (5) If $j=i$ and the difference of the differences of the enrollments between years *n-1* and *n-2* and

between years *n-2* and *n-3* is positive, then the trend of the forecasting will go up, and we use the following Rule 2 to forecast the enrollments; (4) If $j<i$ and the difference of the differences of the enrollments between years *n-1* and *n-2* and between years *n-2* and *n-3* is negative, then the trend of the forecasting will go down, and we use the following Rule 3 to forecast the enrollments; (5) If $j=i$ and the difference of the differences of the enrollments between years *n-1* and *n-2* and between years *n-2* and *n-3* is positive, then the trend of the forecasting will go up, and we use the following Rule 2 to forecast the enrollments; (6) If $j=i$ and the difference of the differences of the enrollments between years *n-1* and *n-2* and between years *n-2* and *n-3* is negative, then the trend of the forecasting will go down, and we us the following Rule 3 to forecast the enrollments, where Rule 1, Rule 2 and Rule 3 are shown as follows:

Rule 1: When forecasting the enrollment of year 1973, there are no data before the enrollments of year 1970, therefore we are not able to calculate the difference of the enrollments between years 1971 and 1970 and the difference of the differences between years 1972 and 1971 and between years 1971 and 1970. Therefore, if *|(the difference of the enrollments between years 1972 and 1971)|/2* > half of the length of the interval corresponding to the fuzzified enrollment A_j with the membership value equal to 1, then the trend of the forecasting of this interval will be upward, and the forecasting enrollment falls at the *0.75*-point of this interval; if *|(the difference of the enrollments between years 1972 and 1971)|/2* = half of the length of the interval corresponding to the fuzzified enrollment A_j with the membership value equal to 1, then the forecasting enrollment falls at the middle value of this interval; if *|(the difference of the enrollments between years 1972 and 1971)|/2* < half of the length of the interval corresponding to the fuzzified enrollment A_j with the membership value equal to *1*, then the trend of the forecasting of this interval will be downward, and the forecasting enrollment falls at the 0.25-point of the interval.

Rule 2: If *(|the difference of the differences between years n-1 and n-2 and between years (n-2) and (n-3)| × 2 + the enrollments of year (n-1))* or *(the enrollments of year (n-1) - |the difference of the differences between years (n-1) and* *(n-2) and between years (n-2) and (n-3)| × 2)* falls in the interval corresponding to the fuzzified enrollment A_j with the membership value equal to *1*, then the trend of the forecasting of this interval will be upward, and the forecasting enrollment falls at the 0.75-point of the interval of the corresponding fuzzified enrollment A_j with the membership value equal to *1*; if *(|the difference of the differences between years (n-1) and (n-2) and between years (n-2) and n-3|/2 + the enrollments of year (n-1))* or *(the enrollments of year (n-1) - |the difference of the differences between years (n-1) and (n-2) and between years (n-2) and (n-3)|/2)* falls in the interval of the corresponding fuzzified enrollment A_j with the membership value equal to *1*, then the trend of the forecasting of this interval will be downward, and the forecasting value falls at the 0.25-point of the interval of the corresponding fuzzified enrollment A_j with the membership value equal to *1*; if neither is the case, then we let the forecasting enrollment be the middle value of the interval corresponding to the fuzzified enrollment A_j with the membership value equal to *1*.

Rule 3: If *(|the difference of the differences between years (n-1) and (n-2) and between years (n-2) and (n-3)|/2 + the enrollments of year (n-1))* or *(the enrollments of year (n-1) - |the difference of the differences between years (n-1) and (n-2) and between years (n-2) and (n-3)|/2)* falls in the interval of the corresponding fuzzified enrollment A_j with the membership value equal to *1*, then the trend of the forecasting o this interval will be downward, and the forecasting enrollment falls at the 0.25-point of the interval corresponding to the fuzzified enrollment A_j with the membership value equal to *1*; *if (|the difference of the differences between years (n-1) and (n-2) and between years (n-2) and (n-3)| × 2 + the enrollment of year (n-1))* or *(the enrollment of year (n-1) - |the difference of the differences between years (n-1) and (n-2) and between years (n-2) and (n-3)| × 2)* falls in the interval corresponding to the fuzzified enrollment A_j with the membership value equal to *1*, then the trend of the forecasting of this interval will be upward, and the forecasting enrollment falls at the 0.75-point of the interval corresponding to the fuzzified enrollment A_j with the membership value equal to 1; if neither is the case, then we let the forecasting enrollment be the middle value of the interval corresponding to the fuzzified enrollment A_j with the membership value equal to *1*.

4.2.5 Fuzzy Forecasting Enrollments Using High-Order Fuzzy Time Series and Genetic Algorithms

Under this topic a total different method is defined for forecasting enrollments using high-order fuzzy time series and genetic algorithms to forecast the enrollments of the University of Alabama. From Table 4.6 the minimum enrollment D_{min} and the maximum enrollment D_{max} of the University of Alabama are 13,055 and 19337. The universe of discourse $U=[13,000, 20,000]$, and the universe of discourse U can be divided into n intervals $u_1, u_2, \ldots, u_n$, where $u_1=[13,000, x_1]$, $u_2=[x_1, x_2]$, ..., $u_n=[x_{n-1}, 20,000]$. Each chromosome consists of $n-1$ genes shown as follows:

$x_1 \quad x_2 \quad x_3 \quad \ldots \quad x_{n-1}$

The algorithm (called Algorithm-GA) for forecasting enrollments and calculating the Mean Square Error (MSE) based on the above chromosome is presented as follows.

$$MSE = \frac{\sum_{i=1}^{n} \left(\text{Actual Enrollment}_i - \text{Forecasted Enrollment}_i\right)^2}{n} \quad \text{.........................(3)}$$

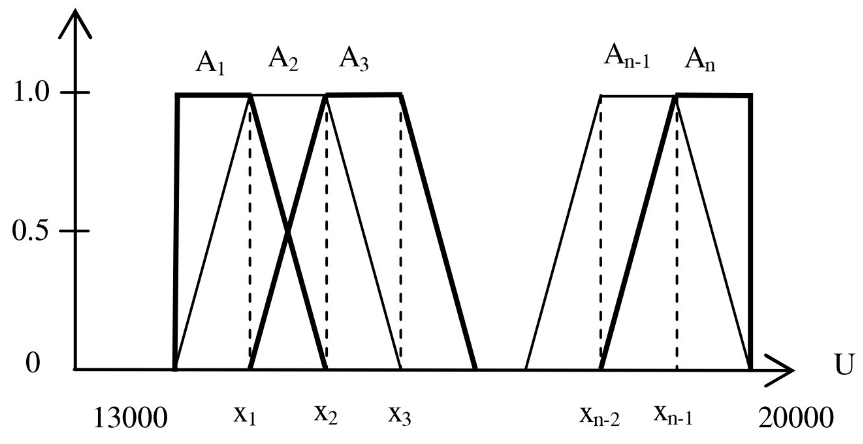


Figure 4.3 Membership functions constructed from genes of a chromosome

Step 1: Construct the membership functions corresponding to the genes $x_1, x_2, ..., x_{n-1}$, of a chromosome, as shown in Figure 4. For example, assume that there is a chromosome shown as follows:

15,29	*15,92*	*15,94*	17,68	*17,73*	*18,25*
1	*6*	*7*	9	*2*	*2*

Then, the membership functions corresponding to the chromosome are as shown in Figure 4.4.

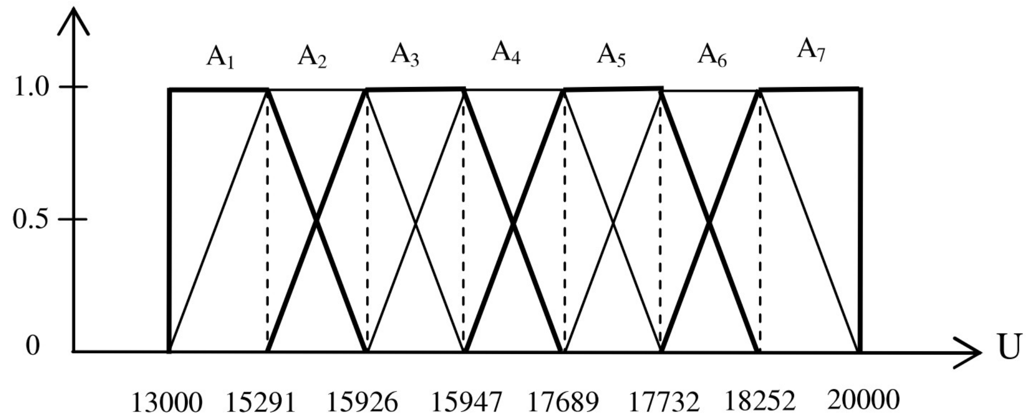


Figure 4.4 Membership functions constructed from genes of a chromosome

Step 2: Based on Figure 4.3 and Chen's (1996) method fuzzify the historical enrollments of the University of Alabama shown in Table 4.6. For example, from Figure 4.4, we can see that u_1= *[13,000, 15,291]*, u_2= *[15,291, 15,926]*, u_3= *[15,926, 15,947]* u_4= *[15,947, 17,689]*, u_5= *[17,689, 17,732]*, u_6= *[17,732, 18,252]*, u_7= *[18,252, 20,000]*. Then, the fuzzified results of the historical enrollments shown in Table 4.6 are shown in Table 4.10.

Table 4.10 Alabama enrollments with fuzzified enrollments

	Actual Enrollments	Fuzzified Enrollments
1971	*13,055*	A_1
1972	*13,563*	A_1
1973	*13,867*	A_1
1974	*14,696*	A_1
1975	*15,460*	A_2
1976	*15,311*	A_2
1977	*15,603*	A_2
1978	*15,861*	A_2
1979	*16,807*	A_4
1980	*16,919*	A_4
1981	*16,388*	A_4
1982	*15,433*	A_2
1983	*15,497*	A_2
1984	*15,145*	A_1
1985	*15,163*	A_1
1986	*15,984*	A_4
1987	*16,859*	A_4
1988	*18,150*	A_6
1989	*18,970*	A_7
1990	*19,328*	A_7
1991	*19,337*	A_7
1992	*18,876*	A_7

Step 3: Generate fuzzy logical relationships based on the j^{th} order fuzzy time series, where $j>1$. For example, from Table 4.10, we can get the third-order fuzzy logical relationships as shown in Table 4.11, where the symbol # denotes an unknown value.

Table 4.11 Alabama enrollments with fuzzified enrollments

$\#, A_1, A_1 \rightarrow A_2$	$A_1, A_1, A_1 \rightarrow A_1$	$A_1, A_1, A_1 \rightarrow A_2$	$A_1, A_1, A_2 \rightarrow A_2$
$A_1, A_2, A_2 \rightarrow A_2$	$A_2, A_2, A_2 \rightarrow A_2$	$A_2, A_2, A_2 \rightarrow A_4$	$A_2, A_2, A_4 \rightarrow A_4$
$A_2, A_4, A_4 \rightarrow A_4$	$A_4, A_4, A_4 \rightarrow A_2$	$A_4, A_4, A_2 \rightarrow A_2$	$A_4, A_2, A_2 \rightarrow A_1$
$A_2, A_2, A_1 \rightarrow A_1$	$A_2, A_1, A_1 \rightarrow A_4$	$A_1, A_1, A_4 \rightarrow A_4$	$A_1, A_4, A_4 \rightarrow A_6$
$A_4, A_4, A_6 \rightarrow A_7$	$A_4, A_6, A_7 \rightarrow A_7$	$A_6, A_7, A_7 \rightarrow A_7$	$A_7, A_7, A_7 \rightarrow A_7$

Step 4: Based on the j^{th} order fuzzy logical relationships, where $j>1$, forecast enrollments using the following principles:

(1) If the j^{th} order fuzzified historical enrollments for year i are $A_{ij}, A_{i(j-1)}, \ldots, A_{i1}$, where $j>1$, and if there is the following fuzzy logical relationship in which the current state is "$A_{ij}, A_{i(j-1)}, \ldots, A_{i1}$", shown as follows:

$$A_{ij}, A_{i(j-1)}, \ldots, A_{i1} \longrightarrow A_j$$

where $A_{ij}, A_{i(j-1)}, \ldots, A_{i1}$ are fuzzy sets, the maximum membership value A_k occurs at interval u_k, and the midpoint of u_k is m_k, then the forecasted enrollment of the i^{th} year is m_k.

(2) If the j^{th} order fuzzified historical enrollments for year i are $A_{ij}, A_{i(j-1)}, \ldots, A_{i1}$, where $j>1$, and if there are the following fuzzy logical relationships in which the current state is "$A_{ij}, A_{i(j-1)}, \ldots, A_{i1}$,", shown as follows:

$$A_{ij}, A_{i(j-1)}, \ldots, A_{i1} \longrightarrow A_{j1}$$
$$A_{ij}, A_{i(j-1)}, \ldots, A_{i1} \longrightarrow A_{j2}$$
$$\vdots$$
$$A_{ij}, A_{i(j-1)}, \ldots, A_{i1} \longrightarrow A_{jp}$$

where A_{ij}, $A_{i(j-1)}$, …, A_{i1}, A_{j1}, A_{j2}, and A_{jp} are fuzzy sets, the maximum membership values of A_{j1}, A_{j2}, …, and A_{jp} occurs at interval u_1, u_2, …, and u_p, respectively, and the midpoint of the intervals u_1, u_2, …, and u_p, and m_1, m_2, …, and m_p, respectively, then the forecasted enrollment of the i^{th} year is $(m_1+m_2+\ldots+m_p)/p$.

Table 4.12 Alabama enrollments with fuzzified enrollments

Year	Actual Enrollments	Forecasted Enrollments	Year	Actual Enrollments	Forecasted Enrollments
1971	*13,055*		*1982*	*15,433*	*15,609*
1972	*13,563*		*1983*	*15,497*	*15,609*
1973	*13,867*	*14,146*	*1984*	*15,145*	*14,146*
1974	*14,696*	*14,878*	*1985*	*15,163*	*14,146*
1975	*15,460*	*14,878*	*1986*	*15,984*	*16,818*
1976	*15,311*	*15,609*	*1987*	*16,859*	*16,818*
1977	*15,603*	*15,609*	*1988*	*18,150*	*17,992*
1978	*15,861*	*16,214*	*1989*	*18,970*	*19,126*
1979	*16,807*	*16,214*	*1990*	*19,328*	*19,126*
1980	*16,919*	*16,818*	*1991*	*19,337*	*19,126*
1981	*16,388*	*16,818*	*1992*	*18,876*	*19,126*

For example, when using the third-order forecasting model to forecast the enrollments of the University of Alabama, then, according to the third-order fuzzy logical relationships shown in Table 4.11, the forecasting results are as shown in Table 4.12, where the process to forecast the enrollment of 1973 is described as follows. When forecasting the enrollment of 1973, we must look at the enrollments of the previous three years before 1973. From Table 4.11, we can see that the enrollment of 1970 is unknown. Thus, the fuzzified enrollment of 1970 is denoted as

#, the fuzzified enrollment of 1971 is A_1, and the fuzzified enrollment of 1972 is A_1, where the symbol # denotes an unknown value.

Then, we look at the fuzzy logical relationships shown in Table 4.11 to find any fuzzy logical relationship whose current state is *"#, A_1, A_1"*. In Table 4.11, we can find the fuzzy logical relationship *"#, A_1, $A_1 \rightarrow A_1$"* whose current state is *"#, A_1, A_1"*. Because the maximum membership value of A_1 occurs at interval u_1, u_1=[13,000, 15,291], and the midpoint of u_1 occurs at 14,146, the forecasted enrollment of 1973 is 14,146. In the same way, the enrollments of the University of Alabama from 1974 to 1992 can be forecasted, as shown in Table 4.12. The mean square error (MSE) of a chromosome can be calculated, where

$$MSE = \frac{\sum_{i=1}^{n}\left(\text{Actual Enrollment}_{\text{i}} - \text{Forecasted Enrollment}_{\text{i}}\right)^2}{n} \quad \text{.........................(3)}$$

Let the value of MSE be the fitness value of the genetic algorithm, where *m* denotes the number of historical data. The system will choose chromosomes with smaller MSE values for evolution. For example, from the experimental result, we can see that the MSE value of the chromosome

15,291 *15,926* *15,947* 17,689 *17,732* *18,252*

is 209,003, as shown in Table 4.13.

Table 4.13 MSE values of the chromosome

	Actual Enrollments (A_i)	*Forecasted Enrollments (F_i)*	*(A_i - F_i)*	*$(A_i - F_i)^2$*
1971	*13,055*			
1972	*13,563*			
1973	*13,867*	*14,146*	*279*	*77,841*
1974	*14,696*	*14,878*	*182*	*33,124*
1975	*15,460*	*14,878*	*-582*	*33,8724*
1976	*15,311*	*15,609*	*298*	*88,804*
1977	*15,603*	*15,609*	*6*	*36*
1978	*15,861*	*16,214*	*353*	*124,609*
1979	*16,807*	*16,214*	*-593*	*351,649*
1980	*16,919*	*16,818*	*-101*	*10,201*
1981	*16,388*	*16,818*	*430*	*184,900*
1982	*15,433*	*15,609*	*176*	*30,976*
1983	*15,497*	*15,609*	*112*	*12,544*
1984	*15,145*	*14,146*	*-999*	*998,001*
1985	*15,163*	*14,146*	*-1017*	*1,034,289*
1986	*15,984*	*16,818*	*834*	*695,556*
1987	*16,859*	*16,818*	*-41*	*1,681*
1988	*18,150*	*17,992*	*-158*	*24,964*
1989	*18,970*	*19,126*	*156*	*24,336*
1990	*19,328*	*19,126*	*-202*	*40,804*
1991	*19,337*	*19,126*	*-211*	*44,521*
1992	*18,876*	*19,126*	*250*	*62,500*

The algorithm for forecasting enrollments using high-order fuzzy time series and genetic algorithm is presented as follows. Assume that the system randomly generates 30 chromosomes as the initial population as shown in Table 4.14.

Step 1: Randomly choose two chromosomes from the population for performing crossover operations and randomly choose a crossover point. For example, if the crossover point selected by the system is "2", then it performs the crossover operations after the crossover point, as shown in Figure 4.5.

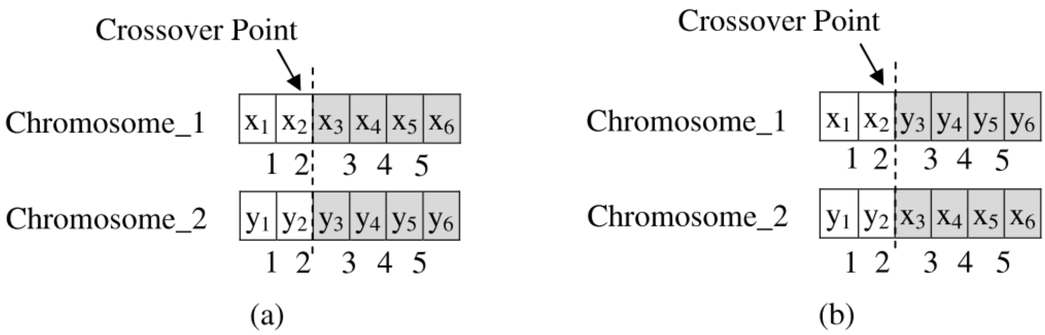


Figure 4.5 (a) Before the crossover operation. (b) After the crossover operation.

After performing the crossover operations, if the values of the genes of a chromosome are not in an ascending sequence, then the system sorts the values of the genes in the chromosomes in an ascending sequence, as shown in Table 4.15.

Table 4.14 The initial population generated from the system.

Chromosome Number	Gene Number					
	Gene 1	*Gene 2*	*Gene 3*	*Gene 4*	*Gene 5*	*Gene 6*
Chrom.1	*15,291*	*15,926*	*15,947*	*17,689*	*17,732*	*18,252*
Chrom.2	*13,162*	*13,381*	*13,538*	*17,312*	*18,313*	*18,777*
Chrom.3	*14,239*	*15,111*	*17,219*	*17,289*	*18,409*	*18,497*
Chrom.4	*13,212*	*17,194*	*17,939*	*18,265*	*18,847*	*19,101*
Chrom.5	*14,283*	*16,820*	*18,627*	*18,709*	*18,745*	*19,158*
Chrom.6	*14,500*	*16,008*	*16,227*	*18,469*	*19,074*	*19,834*
Chrom.7	*13,886*	*16,106*	*16,290*	*16,668*	*17,821*	*18,473*
Chrom.8	*14,516*	*15,510*	*15,627*	*18,030*	*18,126*	*19,482*
Chrom.9	*14,641*	*14,729*	*15,417*	*19,023*	*19,621*	*19,884*
Chrom.10	*13,150*	*13,757*	*15,307*	*19,275*	*19,356*	*19,960*
Chrom.11	*14,124*	*14,535*	*15,876*	*16,895*	*17,848*	*18,457*

Table 4.14 The initial population generated from the system (continues)

Chromosome Number	Gene Number					
	Gene 1	*Gene 2*	*Gene 3*	*Gene 4*	*Gene 5*	*Gene 6*
Chrom.12	*14,539*	*14,621*	*17,582*	*18,355*	*18,542*	*19,203*
Chrom.13	*13,169*	*15,081*	*15,883*	*18,012*	*19,586*	*19,943*
Chrom.14	*14,186*	*16,709*	*17,376*	*18,108*	*19,525*	*19,569*
Chrom.15	*15,093*	*15,795*	*16,684*	*17,812*	*17,951*	*18,116*
Chrom.16	*13,510*	*14,715*	*17,545*	*18,646*	*19,157*	*19,975*
Chrom.17	*13,445*	*13,925*	*16,653*	*17,672*	*19,225*	*19,264*
Chrom.18	*13,038*	*13,326*	*14,133*	*15,612*	*17,003*	*18,005*
Chrom.19	*13,121*	*14,175*	*15,601*	*16,464*	*16,579*	*18,011*
Chrom.20	*13,732*	*16,025*	*16,127*	*16,798*	*19,795*	*19,858*
Chrom.21	*13,987*	*16,052*	*16,069*	*16,959*	*17,578*	*19,805*
Chrom.22	*13,708*	*14,143*	*15,269*	*15,390*	*16,602*	*19,733*
Chrom.23	*14,754*	*15,648*	*15,769*	*17,237*	*17,880*	*18,854*
Chrom.24	*14,194*	*14,396*	*17,748*	*18,193*	*18,734*	*19,456*
Chrom.25	*13,187*	*13,209*	*13,367*	*16,787*	*16,819*	*19,552*
Chrom.26	*14,109*	*14,432*	*15,421*	*17,501*	*18,327*	*19,052*
Chrom.27	*15,339*	*15,583*	*15,895*	*16,540*	*19,304*	*19,607*
Chrom.28	*14,228*	*14,399*	*14,439*	*14,514*	*15,869*	*18,372*
Chrom.29	*13,709*	*14,022*	*16,110*	*16,724*	*18,830*	*19,502*
Chrom.30	*13,636*	*15,723*	*16,215*	*17,796*	*17,831*	*19,195*

Step 2: Randomly select a chromosome from the population and randomly select a gene from the selected chromosome to perform the mutation operation. In this article, we set the mutation rate to 0.05. If the random number generated by the system is smaller than or equal to the mutation rate (0.05) the mutation operation mutates the selected gene of the selected chromosome. Assume that gene x_3 is randomly selected by the system to perform the mutation operation, and then the value of x_3 will be replaced by a random value between x_2 and x_4. For example, let us consider the chromosomes shown in Figure 4.6.

Chromosome	14124	14535	15876	16895	19525	19569

(a)

Chromosome	14124	14535	15489	16895	19525	19569

(b)

Figure 4.6 (a) Chromosome before the mutation. (b) Chromosome after the mutation.

Assume that the system selects the third gene of the chromosome to perform the mutation operation; then the value of the third gene (i.e., 15,876) will be replaced by a random number between 14,535 and 16,895 generated by the system. Assume that the random number generated by the system is 15,489, and then the value "15,876" of the third gene of the chromosome is replaced by "15,489", as shown in Figure 4.6.

Table 4.15 After crossover and sorting the index of genes in the chromosomes

Chromosome Number	Gene Number *Gene 1*	*Gene 2*	*Gene 3*	*Gene 4*	*Gene 5*	*Gene 6*
Chrom.1	*14,124*	*14,535*	*15,876*	*16,895*	*19,525*	*19,569*
Chrom.2	*15,093*	*15,795*	*16,684*	*17,812*	*17,951*	*18,473*
Chrom.3	*14,516*	*15,510*	*15,627*	*18,030*	*18,126*	*19,733*
Chrom.4	*14,754*	*15,648*	*15,769*	*18,193*	*18,734*	*19,456*
Chrom.5	*13,038*	*13,326*	*14,133*	*15,612*	*15,869*	*18,372*
Chrom.6	*13,121*	*14,175*	*15,601*	*16,464*	*19,304*	*19,607*
Chrom.7	*15,291*	*15,926*	*15,947*	*17,689*	*17,732*	*19,101*
Chrom.8	*13,886*	*16,106*	*16,290*	*16,668*	*17,821*	*18,116*
Chrom.9	*13,169*	*15,081*	*15,883*	*16,724*	*18,830*	*19,502*
Chrom.10	*13,636*	*15,723*	*17,219*	*17,289*	*18,409*	*18,497*
Chrom.11	*13,367*	*13,987*	*16,052*	*16,787*	*16,819*	*19,552*
Chrom.12	*14,109*	*14,432*	*15,421*	*17,501*	*18,313*	*18,777*
Chrom.13	*13,445*	*13,925*	*16,653*	*16,798*	*19,795*	*19,858*

Table 4.15 After crossover and sorting the index of genes in the chromosomes (continues)

Chromosome Number	Gene Number					
	Gene 1	*Gene 2*	*Gene 3*	*Gene 4*	*Gene 5*	*Gene 6*
Chrom.14	*14,186*	*16,709*	*17,376*	*17,848*	*18,108*	*18,457*
Chrom.15	*14,500*	*16,008*	*16,227*	*18,469*	*19,074*	*19,975*
Chrom.16	*13,709*	*14,022*	*16,110*	*18,012*	*19,586*	*19,943*
Chrom.17	*15,339*	*15,583*	*15,895*	*16,540*	*16,579*	*18,011*
Chrom.18	*14,28*	*16,820*	*18,627*	*18,709*	*18,745*	*19,203*
Chrom.19	*14,228*	*14,399*	*14,439*	*14,514*	*17,003*	*18,005*
Chrom.20	*13,708*	*14,143*	*15,269*	*15,390*	*16,602*	*19,482*
Chrom.21	*13,732*	*16,025*	*16,127*	*17,672*	*19,225*	*19,264*
Chrom.22	*13,212*	*17,194*	*17,939*	*18,252*	*18,265*	*18,847*
Chrom.23	*14,239*	*15,111*	*16,215*	*17,796*	*17,831*	*19,195*
Chrom.24	*13,162*	*13,381*	*13,538*	*17,312*	*18,327*	*19,052*
Chrom.25	*13,187*	*13,209*	*16,069*	*16,959*	*17,578*	*19,805*
Chrom.26	*13,510*	*14,715*	*17,545*	*18,646*	*19,157*	*19,834*
Chrom.27	*14,194*	*14,396*	*17,237*	*17,748*	*17,880*	*18,854*
Chrom.28	*14,539*	*14,621*	*17,582*	*18,355*	*18,542*	*19,158*
Chrom.29	*14,641*	*14,729*	*15,417*	*19,023*	*19,621*	*19,960*
Chrom.30	*13,150*	*13,757*	*15,307*	*19,275*	*19,356*	*19,88*

Step 3: Calculate the MSE value of each derived chromosome obtained from Step 2 based on the algorithm Algorithm-GA* presented previously. The results are shown in Table 4.16.

Table 4.16 The MSE values of the chromosomes after crossover and mutation operations.

Chromosome Number	MSE	Chromosome Number	MSE
Chromosome 1	*297,470*	*Chromosome 16*	*253,133*
Chromosome 2	*129,169*	*Chromosome 17*	*270,647*
Chromosome 3	*173,649*	*Chromosome 18*	*501,718*
Chromosome 4	*520,723*	*Chromosome 19*	*1,438,292*
Chromosome 5	*498,592*	*Chromosome 20*	*693,002*
Chromosome 6	*523,545*	*Chromosome 21*	*329,334*
Chromosome 7	*233,348*	*Chromosome 22*	*883,698*
Chromosome 8	*252,649*	*Chromosome 23*	*225,963*
Chromosome 9	*255,308*	*Chromosome 24*	*890,302*
Chromosome 10	*242,733*	*Chromosome 25*	*379,907*
Chromosome 11	*537,351*	*Chromosome 26*	*980,279*
Chromosome 12	*337,461*	*Chromosome 27*	*1,088,842*
Chromosome 13	*904,830*	*Chromosome 28*	*1,073,041*
Chromosome 14	*312,478*	*Chromosome 29*	*1,586,739*
Chromosome 15	*382,629*	*Chromosome 30*	*1,778,588*

Step 4: Select 20 chromosomes from the population shown in Table 4.16 that have lower MSE values, and put 10 chromosomes randomly generated by the system into them to form the new population of the next generation, as shown in Table 4.17. In the same way, after repeatedly performing Step 1 to Step 4 to evolve 1000 generations, the chromosome that has the lowest value of MSE is the best solution to be used to forecast the enrollments of the University of Alabama.

Table 4.17 The chromosomes of the next generation.

	Chrom. Number	Gene Number						MSE
		Gene 1	*Gene 2*	*Gene 3*	*Gene 4*	*Gene 5*	*Gene 6*	
20 Chromosome having lower MSE values selected from the population	*Chrom.1*	*14,124*	*14,535*	*15,876*	*16,895*	*17,848*	*18,457*	*114,129*
	Chrom.2	*15,093*	*15,795*	*16,684*	*17,812*	*17,951*	*18,473*	*129,169*
	Chrom.3	*15,093*	*15,795*	*16,684*	*17,812*	*17,951*	*18,116*	*162,231*
	Chrom.4	*14,516*	*15,510*	*15,627*	*18,030*	*18,126*	*19,733*	*173,649*
	Chrom.5	*14,516*	*15,510*	*15,627*	*18,030*	*18,126*	*19,482*	*177,756*
	Chrom.6	*14,754*	*15,648*	*15,769*	*17,237*	*17,880*	*18,854*	*180,976*
	Chrom.7	*13,038*	*13,326*	*14,133*	*15,612*	*17,003*	*18,005*	*185,036*
	Chrom.8	*13,121*	*14,175*	*15,601*	*16,464*	*16,579*	*18,011*	*193,039*
	Chrom.9	*15,291*	*15,926*	*15,947*	*17,689*	*17,732*	*18,252*	*209,008*
	Chrom.10	*13,886*	*16,106*	*16,290*	*16,668*	*17,821*	*18,473*	*218,605*
	Chrom.11	*14,239*	*15,111*	*16,215*	*17,796*	*17,831*	*19,195*	*225,963*
	Chrom.12	*15,291*	*15,926*	*15,947*	*17,689*	*17,732*	*19,101*	*233,348*
	Chrom.13	*13,636*	*15,723*	*17,219*	*17,289*	*18,409*	*18,497*	*242,733*
	Chrom.14	*13,886*	*16,106*	*16,290*	*16,668*	*17,821*	*18,116*	*252,649*
	Chrom.15	*13,709*	*14,022*	*16,110*	*18,012*	*19,586*	*19,943*	*253,133*
	Chrom.16	*13,169*	*15,081*	*15,883*	*18,012*	*19,586*	*19,943*	*254,270*
	Chrom.17	*13,169*	*15,081*	*15,883*	*16,724*	*18,830*	*19,502*	*255,308*
	Chrom.18	*13,636*	*15,723*	*16,215*	*17,796*	*17,831*	*19,195*	*266,532*
	Chrom.19	*13,987*	*16,052*	*16,069*	*16,959*	*17,578*	*19,805*	*268,621*
	Chrom.20	*15,339*	*15,583*	*15,895*	*16,540*	*16,579*	*18,011*	*270,65*

The loop will continue till the best 20 generations are created and the cycle is ending. The process provides great forecasts with lowest mean square errors comparing to various methods introduced in this master thesis. That means this method with artificial computer technologies might be the most accurate forecasting method for time series which are time-variant or time-invariant.

Table 4.17 The chromosomes of the next generation. (Continues)

Chromosomes randomly generated by the system

Chrom. Number	Gene Number						MSE
	Gene	*Gene*	*Gene*	*Gene*	*Gene*	*Gene*	
Chrom.21	*13,832*	*18,015*	*18,035*	*19,356*	*19,523*	*19,590*	*1,795,597*
Chrom.22	*13,953*	*14,470*	*15,475*	*16,799*	*16,830*	*17,786*	*174,952*
Chrom.23	*13,462*	*13,473*	*14,128*	*14,473*	*16,485*	*19,184*	*750,216*
Chrom.24	*13,494*	*14,261*	*16,077*	*17,787*	*19,600*	*19,925*	*244,710*
Chrom.25	*13,853*	*14,282*	*16,067*	*17,791*	*18,454*	*18,767*	*214,065*
Chrom.26	*13,352*	*13,800*	*17,240*	*17,722*	*18,543*	*19,574*	*1,014,534*
Chrom.27	*13,306*	*14,349*	*17,524*	*18,055*	*18,303*	*18,954*	*1,065,016*
Chrom.28	*17,100*	*17,334*	*17,554*	*17,752*	*17,963*	*19,925*	*1,485,862*
Chrom.29	*15,470*	*16,752*	*17,065*	*17,386*	*18,961*	*19,481*	*268,896*
Chrom.30	*14,039*	*14,200*	*14,680*	*15,675*	*17,279*	*18,61*	*152,876*

4.2.6 Fuzzy Metric Approach for Fuzzy Time Series

Since first appearance of Fuzzy Time Series, nearly all the researchers present their method to forecast the enrollments of the University of Alabama based on fuzzy time series based on Jilani & Burney (2007) and Jilani, Burney & Ardil (2007). The historical enrollments of the University of Alabama are shown in Table 4.5.

Firstly, we defined the partition the universe of discourse into equal length intervals. Then based on frequency density portioning, we redefine the intervals. After this, define some membership function for each interval of the historical enrollment data to obtain fuzzy enrollments to form a fuzzy time series. Then, it establishes fuzzy logical relationships (FLRs) based on the fuzzified enrollments in Table IV. Finally, it uses our proposed method. The proposed method bases on Hsu & Chen (2004) approach of partitioning universe of discourse are as follows:

Step 1: Define the universe of discourse U and partition it into several even and equal length intervals $u_1, u_2... u_n$. For example, assume that the universe of discourse U= [13000,20000] is partitioned into seven even and equal length intervals.

Step 2: Get a weighted aggregation Zimmermann (2001) of the fuzzy distribution of the historical enrollments in each interval. Sort the intervals based on the number of historical enrollment data in each interval from the highest to the lowest. Find the interval having the largest number of historical enrollment data and divide it into four sub-intervals of equal length. Find the interval having the second largest number of historical enrollment data and divide it into three sub-intervals of equal length. Find the interval having the third largest number of historical enrollment data and divide it into two sub-intervals of equal length. Find the interval with the fourth largest number of historical enrollment data and let the length of this interval remain unchanged. If there are no data distributed in an interval then discard this interval. For example, the distributions of the historical enrollment data in different intervals are summarized as shown in Table 4.18 from Hsu & Chen (2004).

Table 4.18 The intervals for enrollments of the university of alabama

Intervals	Number of historical enrollment data
[13,000, 14,000]	*3*
[14,000, 15,000]	*1*
[15,000, 16,000]	*9*
[16,000, 17,000]	*4*
[17,000, 18,000]	*0*
[18,000, 19,000]	*3*
[19,000, 20,000]	*2*

After executing this step, the universe of discourse [13,000, 20,000] is re-divided into the following intervals Hsu & Chen (2004), see Table 4.19.

Table 4.19 The fuzzy intervals using frequency density based partitioning

Linguistic	Intervals
U_1	*[13,000, 13,500]*
U_2	*[13,500, 14,000]*
U_3	*[14,000, 15,000]*
U_4	*[15,000, 15,250]*
U_5	*[15,250, 15,500]*
U_6	*[15,500, 15,750]*
U_7	*[15,750, 16,000]*
U_8	*[16,000, 16,333]*
U_9	*[16,333, 13,500]*
U_{10}	*[16,667, 17,000]*
U_{11}	*[18,000, 18,500]*
U_{12}	*[18,500, 19,000]*
U_{13}	*[19,000, 20,000]*

Step 3: Define each fuzzy set A_i based on the re-divided intervals and fuzzify the historical enrollments shown in Table 4.6, where fuzzy set A_i denotes a linguistic value of the enrollments represented by a fuzzy set. We have used triangular membership function to define the fuzzy sets A_i. The reason for fuzzify the historical enrollments into fuzzified enrollments is to translate crisp values into fuzzy sets to get a fuzzy time series.

Table 4.20 Third-order fuzzy logical relationships

$A_2, A_2, A_3 \rightarrow A_5$	$A_7, A_{10}, A_{10} \rightarrow A_9$	$A_4, A_4, A_7 \rightarrow A_{10}$
$A_2, A_3, A_5 \rightarrow A_5$	$A_{10}, A_{10}, A_9 \rightarrow A_5$	$A_4, A_7, A_{10} \rightarrow A_{11}$
$A_3, A_5, A_5 \rightarrow A_6$	$A_{10}, A_9, A_5 \rightarrow A_5$	$A_7, A_{10}, A_{11} \rightarrow A_{12}$
$A_5, A_5, A_6 \rightarrow A_7$	$A_9, A_5, A_5 \rightarrow A_4$	$A_{10}, A_{11}, A_{12} \rightarrow A_{13}$
$A_5, A_6, A_7 \rightarrow A_{10}$	$A_5, A_5, A_4 \rightarrow A_4$	$A_{11}, A1_2, A_{13} \rightarrow A_{13}$
$A_6, A_7, A_{10} \rightarrow A_{10}$	$A_5, A_4, A_4 \rightarrow A_7$	$A_{12}, A_{13}, A_{13} \rightarrow A_{12}$

Step 4: Establish fuzzy logical relationships based on the fuzzified enrollments where the fuzzy logical relationship "A_p, A_q, $A_r \rightarrow A_s$" denotes that "if the fuzzified enrollments of year p, q and r are A_p ,A_q and A_r respectively, then the fuzzified enrollments of year *(r)* is A_r".

CHAPTER FIVE
CONCLUSION

5.1 Conclusion of the Results

In various methods there has been proven that since the beginning of the idea solving Time Series data sets by fuzzify them is going to improve very fast. The actual forecasting error rate is significantly low, comparing to previous methodologies. Also comparing the Mean Square Error (MSE) during years shown in Table 5.1 indicates the improvement of Fuzzy Time Series Analysis.

This table also includes the Average Forecasting Error Rate (AFER) which is calculated as

$$AFER = \frac{\left(A_i - F_i\right) / A_i}{n} \times 100\%$$

this formula shows that how far the forecasting falls away from the actual data, and how accurate its forecasting’s could project the real time values. Observing the error rate is a significant advantage to decide, if the method or model does it’s job, or not. But also, it should give the same results as a classical Time Series Analysis process produces. Rather using the new one as a fast growing child or the more experienced mature methodology.

There are sure some other benefits using one of the Fuzzy Time Series Methods. In my opinion two of them are strictly strong: against assumptions, small data set effectiveness.

As typically in classical time series, an effective fitting of Box & Jenkins models requires at least a moderately long series, which consists at least of 50 observations (Chatfield, 1996). Many other would recommend at least 100 observations. In Fuzzy Time Series Models there are no upper/lower limits, which stricken the analyst to carve dozens of information or data. While using soft computing technologies in Fuzzy Time Series, competing with enormous data sets, is also an easy task to accomplish.

It's been briefly explained in second chapter that, during processing the Time Series Set with Box & Jenkins Methodology, a lot of assumptions, hypothesis should be proven, to run the forecasting steps. In Fuzzy Time Series no assumptions needed to build a model and forecast a future value. The meaning of using no assumptions is deep; because that greatly reduces the analysis time, lesser complications in theoretical iterations, more flexibility to improve the model, smarter basis for soft computing algorithms.

The mathematical language used in Fuzzy Time Series is based on Fuzzy Logic, which makes it more understandable. The great ability comes with Fuzzy Logic: Computing with linguistic terms or definitions. Working with qualitative historical data by Fuzzy Time Series, has no issue to convert them in real numbers. Also the dynamics of fuzzy logic improves, the understanding of trend and other components affecting the historical data.

In summary, lots of different approaches have been made and there are going to be more. The best parts of Fuzzy Time Series have been tried to define in this Master Degree Thesis. Fuzzy Logic integration in Time Series Analysis, made a lot of improvement in this field of research. As a strong, accurate and fast processing Time Series forecasting tool, Fuzzy Time Series Methodology's doing his job way over. Since, comparing the invention time of Fuzzy Time Series theory, there is a giant leap of improvement. In a short period of time this theory is going to pass all the prejudice, fortify itself in science history.

Table 5. 1 MSE results for methods used in this chapter.

Year	Actual Enroll-ments	Song & Chissom (1993)	Chen (1996)	Chen (2002)	Chen (2004)	*Jilani, Burney & Ardil (2007)*	Chung & Chen (2009)
1971	*13,055*					*13,579*	
1972	*13,563*	*14,000*	*14,000*		*13,750*	*13,798*	
1973	*13,867*	*14,000*	*14,000*		*13,875*	*13,798*	
1974	*14,696*	*14,000*	*14,000*	*14,500*	*14,750*	*14,452*	
1975	*15,460*	*15,500*	*15,500*	*15,500*	*15,375*	*15,373*	
1976	*15,311*	*16,000*	*16,000*	*15,500*	*15,312.5*	*15,373*	
1977	*15,603*	*16,000*	*16,000*	*15,500*	*15,625*	*15,623*	
1978	*15,861*	*16,000*	*16,000*	*15,500*	*15,812.5*	*15,883*	
1979	*16,807*	*16,000*	*16,000*	*16,500*	*16,833.5*	*17,079*	*16,846*
1980	*16,919*	*16,813*	*16,833*	*16,500*	*16,833.5*	*17,079*	*16,846*
1981	*16,388*	*16,813*	*16,833*	*16,500*	*16,42*	*16,497*	*16,420*
1982	*15,433*	*16,789*	*16,833*	*15,500*	*15,375*	*15,373*	*15,462*
1983	*15,497*	*16,000*	*16,000*	*15,500*	*15,375*	*15,373*	*15,462*
1984	*15,145*	*16,000*	*16,000*	*15,500*	*15,125*	*15,024*	*15,153*
1985	*15,163*	*16,000*	*16,000*	*15,500*	*15,125*	*15,024*	*15,153*
1986	*15,984*	*16,000*	*16,000*	*15,500*	*15,937.5*	*15,883*	*15,977*
1987	*16,859*	*16,000*	*16,000*	*16,500*	*16,833.5*	*17,079*	*16,846*
1988	*18,150*	*16,813*	*16,833*	*18,500*	*18,250*	*17,991*	*18,133*
1989	*18,970*	*19,000*	*19,000*	*18,500*	*18,875*	*18,802*	*18,910*
1990	*19,328*	*19,000*	*19,000*	*19,500*	*19,250*	*18,994*	*19,334*
1991	*19,337*	*19,000*	*19,000*	*19,500*	*19,250*	*18,994*	*19,334*
1992	*18,876*	*19,000*	*19,000*	*18,500*	*18,875*	*18,916*	*18,910*
MSE		*423,027*	*775,687*	*86,694*	*5,353*	*41,426*	*1,101*
AFER		*% 4.38*	*% 3.11*	*%2.4452*	*%1.5294*	*%2.3865*	*%1.0242*

REFERENCES

Batyrshin, I. & Kaynak, O. & Rudas, I. (2002). Fuzzy Modelling Based on Generalized Conjunction Operations, *IEEE Trans. on Fuzzy Systems, 105,* 678-683.

Box, G.E.P. & Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

Chen, S.M. (1996). Forecasting enrollments based on fuzzy time series. *Fuzzy Sets and Systems, 81*, 311–319.

Chen, S.M. (2002). Forecasting enrollments based on high-order fuzzy time-series. *Cybernetics and Systems, 33,* 1–16.

Chen, S.M. & Hsu, C.C. (2004). A new method for forecasting enrollments based on fuzzy time series, *International Journal of Applied Science and Engineering 2004. 2, 3:* 234-244.

Chen, S.M., & Hwang, J.R. (2000). Temperature prediction using fuzzy time series. *IEEE Transaction on Systems, Man, & Cybernetics, 30,* 263–275.

Dubois, D. & Prade, H. (1982). Towards fuzzy differential calculus Part 1: Integration of fuzzy mappings, *Fuzzy Sets and Systems 8,* 1-17.

Huarng, K.H. (2001a). Heuristic models of fuzzy time series for forecasting. *Fuzzy Sets and Systems, 123,* 369–386.

Huarng, K.H. (2001b). Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy Sets and Systems, 123*, 387–394.

Hwang, J.R., Chen, S.M., & Lee, C.H. (1998). Handling forecasting problems using fuzzy time series. *Fuzzy Sets and System, 100,* 217–228.

Jang, J.S.R. & Sun, C. T. & Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Upper Saddle River, NJ: Prentice Hall.

Jilani, T.A. & Burney, S.M.A. (2007). M-factor high order fuzzy time series forecasting for road accident data, *World Congress, Cancun, Mexico, June 18-21,* Forthcoming in Book series Advances in Soft Computing, Springer-Verlag.

Jilani, T.A. & Burney, S.M.A. & Ardil, C. (2007). Multivariate high order fuzzy time series forecasting for car road accidents, *International Journal of Computational Intelligence, 4, 1,* 15-20.

John, R. I. (1996). Type-2 Fuzzy Sets for Community Transport Scheduling, *In Proceedings of the Fourth European Congress on Intelligent Techniques and Soft Computing EUFIT'96, 2,* 1369-1372

John, R. I. (1998) Type-2 Fuzzy Sets for Knowledge Representation and Inferencing, *Research Monograph 10, School of Computing Sciences, De Montfort University*.

John, R. J. & Coupland, S. (2007). Type-2 Fuzzy Logic: A Historical View, *IEEE Computational Intelligence Magazine, 2, 1,* 57-62.

Karnik, N. N. & Mendel J. M. (1998a). Introduction to Type-2 Fuzzy Logic Systems, In *Proc. IEEE World Congress on Computational Intelligence,* 915-920, Anchorage, Alaska, USA.

Karnik, N. N. & Mendel J. M. (1998b). Type-2 Fuzzy Logic Systems: Type Reduction, *In Proceedings IEEE Systems, Man and Cybernatics,* 2046-2051.

Karnik, N. N. & Mendel, J.M. (1999). Application of Type-2 Fuzzy Logic System to Forecasting of Time Series, *Information Sciences, 120,* 89-111.

Karnik, N. N. & Mendel J. M. (2001). Centroid of A Type-2 Fuzzy Set, *Information Sciences, 132,* 195-220.

Khinchin, A.Y. (1934) Korrelations Theorie der Stationaren Stochastischen Prozesse, *Mathematical Annals 109,* 604.

Klement, E.P. (1980) Fuzzy α-algebras and fuzzy measurable functions, *Fuzzy Sets and Systems, 4,* 83-93.

Kolmogorov, A.N. (1931) Uber die Analytischen Methoden in der Wahrscheinlich-keitsrechnung, *Mathematical Annals 104,* 604.

Lee, H.S. & Chou, M.T. (2004). Fuzzy forecasting based on fuzzy time series. *International Journal of Computer Mathematics, 81,* 781–789.

MATLAB *Fuzzy Logic Toolbox Tutorial,* 1994-2005 The MathWorks, Inc.

Mendel, J. M. (2001a). *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions,* Upper Saddle River, NJ: Prentice Hall.

Mendel, J. M. (2001b). The Perceptual Computer: An Architecture for Computing with Words, *In Proc. FUZZ-IEEE 2001, Melbbourne, Australia.*

Mendel, J. M. (2003). Fuzzy Sets for Words: A New Beginning, *In Proc. FUZZY,* 37-42, St. Louis, MO, USA.

Mizumoto, M. & Tanaka, K. (1981). Fuzzy Sets of Type 2 Under Algebraic Product and Algebraic Sum, *Fuzzy Sets and Systems, 5,* 277-290.

National Science Council (1995–2004). Indicator of Science & Technology Republic of China. *R.O.C.: National Science Council.*

Ralescu, D.A. & Nikodym, R. (1996). Theorem for fuzzy set valued measures, *Fuzzy Sets Theory and Applications* (Reidel Publishing Company).

Stojakovic, M. (1994). Fuzzy valued measure, *Fuzzy Sets and Systems 65,* 95-104.

Song, Q. & Chissom, B.S. (1993a). Fuzzy forecasting enrollments with fuzzy time series - Part 1. *Fuzzy Sets and Systems, 54,* 1–9.

Song, Q. & Chissom, B.S. (1993b). Fuzzy Time Series and Its Models. *Fuzzy Sets and Systems, 54,* 269–277.

Song, Q. & Chissom, B.S. (1994). Fuzzy forecasting enrollments with fuzzy time series-Part 2. *Fuzzy Sets and Systems, 62*, 1–8.

Wang G. & Zhang, Y. (1992) The theory of fuzzy stochastic processes, *Fuzzy Sets and Systems 51,* 161-178.

Yule, G.U. (1927). On the method of investigating periodicities in disturbed series, *with special reference to Wolfer's sunspot numbers, Philosophical Transactions A226,* 267–298.

Zadeh, L. A. (1994). Soft Computing and Fuzzy Logic, *IEEE Software, 11, 6,* 48-56

Zadeh, L.A. (1968). Probability measures of fuzzy events, *J. Math. Anal. Appl. 23* 421-427

Zadeh, L.A. (1975). The concept of a linguistic variable and its application to approximate reasoning, *Information Science Parts 1-3, 8* , 199-249

Zadeh, L. A. (1996). "Fuzzy Logic: Computing with Words", *IEEE Transactions on Fuzzy Systems, 4,* 103-111

Zhang, W.X. & Li, T. & Ma J.F. & Li, A.J. (1990). Set-valued measure and fuzzy set-valued measure, *Fuzzy Sets and Systems 36,* 181-188

Zimmermann, H.J. (2001). *Fuzzy Set Theory and Its Applications,* Kluwer Publishers, Boston, USA.