ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

TIME SERIES FORECASTING VIA COMPUTATIONAL INTELLIGENCE METHODS

M.Sc. THESIS

Atakan ŞAHİN

Department of Control and Automation Engineering

Control and Automation Engineering Programme

MAY 2016

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

TIME SERIES FORECASTING VIA COMPUTATIONAL INTELLIGENCE METHODS

M.Sc. THESIS

Atakan ŞAHİN
(504141102)

Department of Control and Automation Engineering

Control and Automation Engineering Programme

Thesis Advisor: Asst. Prof. Dr. Tufan KUMBASAR

MAY 2016

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

ZAMAN SERİLERİ TAHMİNLEMEDE BİLGİ İŞLEMSEL ZEKA UYGULAMALARI

YÜKSEK LİSANS TEZİ

Atakan ŞAHİN
(504141102)

Kontrol ve Otomasyon Mühendisliği Anabilim Dalı

Kontrol ve Otomasyon Mühendisliği Programı

Tez Danışmanı: Yrd. Doç. Dr. Tufan KUMBASAR

MAYIS 2016

Atakan Şahin, a M.Sc. student of ITU Graduate School of Science Engineering And Technology student ID 504141102, successfully defended the thesis/dissertation entitled "TIME SERIES FORECASTING VIA COMPUTATIONAL INTELLIGENCE METHODS", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Asst. Prof. Dr. Tufan Kumbasar**
Istanbul Technical University

Jury Members : **Prof. Dr. İbrahim Eksin**
Istanbul Technical University

Asst. Prof. Dr. İlker Üstoğlu
Yıldız Technical University

Date of Submission : 02 May 2016
Date of Defense : 06 June 2016

I still have a long way to go but
I am already so far from where I used to be

To the future,

FOREWORD

I would like to thank the persons and organizations listed as respectively below:

- Asst. Prof. Dr. Tufan Kumbasar
- Dr. Engin Yeşil
- M.Sc. Furkan Dodurka
- My friends, especially the members of 002
- Getron
- Istanbul Technical University Scientific Research Projects Unit

Firstly, I would like to thank my advisor Asst. Prof. Dr. Tufan Kumbasar, for all his guidance and encouragements he has given me over the past three years. His vision and sincerity provide ideas to me continue as academic person for rest of my life. His support for my PhD applications is already unbelievable. Without his and Dr. Yesil's encouragements, I would have never given myself the chance to continue my education in abroad.

A special thanks must go to my B.Sc. advisor Dr. Engin Yesil for provide a chance to me involved the IPC workgroup and Getron three years ago. Thanks to his helps and suggestions, I can manage my further life.

I would like to thank both of these incredible persons to provide the opportunities to me develop myself in both academic and thought ways. They have made a great impact on my education and life.

I also need to thank M.Sc. Furkan Dodurka for being helpful to me as always without any question.

Finally, I would like to thank Getron for their support and helps. I would like to thank Istanbul Technical University Scientific Research Projects Unit for their support to my master thesis under the Support Program for Graduate Thesis Research Project.

May 2016 Atakan ŞAHİN

TABLE OF CONTENTS

Page

ABBREVIATIONS

ANFIS : Adaptive Network based Fuzzy Inference System
ANN : Artificial Neural Network
AR : Auto Regressive
ARCH : Autoregressive Conditional Heteroskedastic
ARMA : Auto Regressive – Moving-Average
ARIMA : Auto Regressive Integrated Moving-Average
CI : Confidence Interval
COA : Center of Area
CPCT : Center Point Correction Term
ElinGRA : Enhanced Linguistic Generation and Representation
FIS : Fuzzy Inference System
FL : Fuzzy Logic
FLS : Fuzzy Logic System
FLUBE : Fuzzy Logic based Lower and Upper Bound Estimator
FTS : Fuzzy Time Series
GARCH : Generalized Autoregressive Conditional Heteroskedastic
iid : independent and identically distributed
INFUMO : Interval Fuzzy Modelling
LFLS : Lower Fuzzy Logic System
LinGRA : Linguistic Generation and Representation Approach
LS : Least Squares
LUBE : Lower and Upper Bound Estimation
MA : Moving-Average
MAE : Mean Absolute Error
MAPE : Mean Absolute Percentage Error
MF : Membership Function
MLP : Multilayer Perceptron
MOM : Mean of Maximum
MSE : Mean Square Error
MVE : Mean Variance Estimation
NN : Neural Network
PI : Prediction Interval
PICP : Prediction Interval Coverage Probability
PINAW : Prediction Interval Normalized Average Width
POA : Percentage of Accuracy
RMSE : Root Mean Square Error
RMSPE : Root Mean Square Percentage Error
SA : Selection Algorithm
SARIMA : Seasonal Auto Regressive Integrated Moving-Average
SMAPE : Symmetric Mean Absolute Percentage Error
TFN : Triangular Fuzzy Number
T-S : Takagi Sugeno

UFLS **:** Upper Fuzzy Logic System

LIST OF TABLES

Page

LIST OF FIGURES

Page

TIME SERIES FORECASTING VIA COMPUTATIONAL INTELLIGENCE METHODS

SUMMARY

Information technologies have many improvements about data storage and its usage in the last decades and it will have more according to technological breakthroughs. Data storage and data obtaining will be much easier than now after the Internet of Things revolution. This revolution can call as differently according to governed country. It named as Industry 4.0 in Germany, the Factory of the Future in France and Italy and Catapult in United Kingdom. The data explosion will also make the data analysis techniques especially in forecasting more important.

A forecast is a prediction of some future things or events. Forecasting is an important problem that link many fields such as economics, industry, environmental sciences and much more.

Most of usage, the forecasting involves from the time series data. Many applications of forecasting about the business exploit daily, weekly, monthly or any defined interval of the data. These applications can be listed in the areas such as operations management, marketing, finance and risk management, economics, industrial process control and, demography. These are only a few where forecast is required to make good decisions.

The forecast model always aims to represent best estimate of the future value of the variable of interest. As it might be expected, these forecasts are not always accurate. Therefore, there is a definition between estimated and real data values named as forecast error. Eventually, it is a good practice to accompany a forecast with an estimate of the error bounds to represent the interval about how large a forecast error might be expected. Prediction Interval (PI) and Confidence Interval (CI) are the most used ones for representing the errors.

The CIs handle with the accuracy of the prediction of the regression while the PIs consider the accuracy with the prediction to the targets values. A PI is constructed from interval bound that covers the future unknown value with a prescribed probability called a confidence level. The availability of PIs allows the decision makers to quantify the level of uncertainty associated with the point forecasts. A relatively wide PI indicates the presence of high level of uncertainties in the underlying system operation. On the other hand, narrow PIs give the decision makers the opportunity to decide more confidently with less chance of confronting an unexpected condition in the future. This useful information can guide the decision makers to avoid the selection of risky actions under uncertain conditions. Thus, the construction of PIs has been a subject of much attention. Thus, different methods haven been proposed for the construction of PIs such as delta technique, Bayesian technique, bootstrap, mean-variance estimation, lower and upper bound estimation method.

In this thesis, alternative approach to the error bounds will be represent. Fuzzy linguistic term generation and representation via Fuzzy Logic based Lower and Upper Bound Estimator (FLUBE) based Triangular Fuzzy Number (TFN) is presented to estimate the uncertainty in the forecast. As it can be noticed via the titles, the thesis mainly based on the fuzzy logic. Fuzzy logic has been successfully implemented in various engineering areas including control, robotics, image processing, decision-making, estimation and modelling. Therefore, the proposed representation includes two different methodologies which will give the opportunity to the decision maker to quantify the uncertainty of the point forecasts with linguistic terms which might increase the interpretability. Moreover, the proposed approaches will provide valuable information about the accuracy of the forecast by providing a relative membership degree with respect to the target data. The proposed approaches consist of two main phases, the offline FLUBE design and the online TFN generation part as Linguistic Generation and Representation Approach (LinGRA) and Enhanced Linguistic Generation and Representation Approach (ElinGRA).

In the context of the thesis, firstly, the time series concept and its analysis methodologies are discussed in addition to the basic information of the forecasting. Time series concept is also a basis of forecasting especially in our daily life's events. Therefore, the components and characteristics of the time series are handled to be a light for the time series analysis. The modelling techniques is the key factor of the forecasting models and the error evaluations of the forecasts. They are handled as concisely to give information as much as needed.

Secondly, the error terms obtained from the real data values and the forecast models' outputs are modelled via the fuzzy modelling approach. Thanks to both fuzzy model, the error bounds can shape as nonlinear and nonsymmetrical conversely the classical erro bounds like PI. Furthermore, the forecasting error bounds, fuzzy logic systems, fuzzy time series and fuzzy modelling approaches are introduced as the basis of the proposed linguistic term generation approaches. The methodologies have differences on the determination of the linguistic terms phase that also illustrated as comparatively.

Finally, the linguistic forecast generation approaches are used on the several data sets in comparison with conventional PI bounds to prove the efficiency. The methodologies can also follow at the part named as experimental results.

Thanks to the membership degree of the proposed linguistic terms, the error evaluation of the forecast can be done with fuzzy numbers. Rather than the classical PIs, the proposed bounds utilize the realized value of the project not only bound-in or out consideration but also grading the forecast via the triangular fuzzy numbers. Therefore, the decision can have the opportunities to critize the last forecasts and their methods.

ZAMAN SERİLERİ TAHMİNLEMEDE BİLGİ İŞLEMSEL ZEKA UYGULAMALARI

ÖZET

Bilgi işlem teknolojileri son yıllarda önemli gelişmeler göstermektedir. Özellikle veri kaydetme ve işleme alanındaki gelişmeler gelecekte verilerin nasıl kullanılacağı konusunda da yol haritasını şekillendirmektedir. Günümüzde adı sıkça duyulan Nesnelerin İnterneti devrimi ile herhangi bir alanda elde edilebilecek verilerin miktarı ve içeriği önemli ölçüde değişecektir. Bu devrim farklı ülkelerde farklı adlar ile de anılmaktadır. Almanya'da Endüstri 4.0, Fransa ve İtalya'da Geleceğin Fabrikası, İngiltere'de ise Mancınık olarak adlandırılmaktadır. Temel içerik aynıdır. Bu devrimlerin sonucu olarak karşılaşılacak veri miktarındaki patlama ile veri işlemenin gerekliliği daha da belirgin şekilde hissedilecektir. Data işleme tekniklerinin gelişmesi ile birlikte bu alanın en önemli içeriklerinden biri olan tahminleme işlemlerinin daha da önem kazanacağı düşünülmektedir.

En basit anlamı ile tahminleme gelecekte gerçekleşecek bir olay hakkında çıkarımlar öner sürmektir. Tahminleme etkilediği alanlar düşünüldüğünde en önemli problemlerden biridir. Ekonomi, endüstri, çevresel etkenler sadece belli başlı çalışma alanları olmakla beraber, sadece ufak bir kısmıdır. Birçok kullanımda da görüldüğü gibi tahminleme zaman serileri üzerinden uygulanmaktadır. Birçok alanda günlük, haftalık, aylık veya tanımlanacak olan herhangi bir aralık üzerinde çalışmalar sıkça görülmektedir. Bahsedilen alanların başlıca uygulamaları ise yönetim, pazarlama, risk yönetimi, süreç kontrol gibi dallarda karşılık bulmaktadır. Bahsedilenler iyi tahminlemenin gerektiği alanlardan sadece bazılarıdır.

Tahminlemeyi sağlayan başlıca yapı olan tahminleme modeli, daima ilgilenilen alanın gelecekteki değerlerini en iyi şekilde tahmin etmeye çalışmaktadır. Literatürde tahminleme modelleri üzerine birçok çalışma bulunmaktadır. Tahminleme modellerin başlıcaları durağan gürültüleri modellemek için önerilen Otoregresif hareketli ortalamalar modeli (ARMA) ve durağan olmayan gürültüleri modellemek için önerilen Otoregresif bütünleşik hareketli ortalamalar modelidir (ARIMA). En çok kullanılan yöntemler bunlar olmakla birlikte karmaşık olayları modellemek için önerilen doğrusal olmayan modeller de bulunmaktadır. Tüm bu modellerin temel amacı ise hatasız modelleme sağlayabilmektir. Bu beklenti genelde boş çıkmaktadır çünkü mükemmel model diye bir karşılık neredeyse imkânsızdır. Bu durumu ifade etmek için ise hata tanımı yapılmıştır. Hata, tahmin değeri ile gerçek değer arasındaki farkın karşılığıdır. Genelde iyi bir tahminlemenin yanında sağlanacak olan hata bandı tahmini son kullanıcı için oldukça önemlidir. Tanımlanan hata bantları aynı zamanda tahminlemenin, tahminleme yapılan zamanda ne kadar hata yapılabileceği ifade etmektedir. Hata bantların içinde en yaygın kullanıma sahip olan iki yapı Güven Aralığı ve Kestirim Aralığı'dır.

Güven Aralığı, kestirimlerin parametrik doğruluğu konusunda çıkarımların yapılmasını sağlarken Kestirim Aralığı ise kestirimlerin hedef değerler ile

karşılaştırılarak tutarlılıkları konusunda çıkarımlar yapmaktadır. Bu kapsamda Kestirim Aralığı ve Güven Aralığı, tanımlanan güven seviyesi (confidence level) doğrultusunda gelecek tahminleri etrafına onları çevreleyecek bir aralık değeri üretmektedir. PI, karar verici kişi veya sistemlere tanımlanan güven seviyesi doğrultusunda tahminlerin değerlendirme imkânı sunar. Teoride geniş Kestirim Aralığı, çalışılan sistemin gelecek tahminlerinin yüksek belirsizlik belirttiğini ifade etmektedir. Diğer taraftan dar Kestirim Aralıkları karar verici sisteme gelecek tahminlerin oldukça güven içinde yapıldığını ve belirsizliklerin olmadığını ifade etmektedir. Üretilen bu bantlar karar verici sistemin gereksiz riskli hareketlerden kaçınmasını sağlayarak, belirsizlik içeren durumlarda karar verici sisteme yardımcı olur. Kestirim Aralığı tahminlemeler için hayati olduğu düşünüldüğünde bu konu üzerinde birçok çalışma yapılması gerektiği ve yapıldığı fark edilebilir. Literatürde Kestirim Aralığı bandı üretimi konusunda birçok çalışma vardır bunlardan başlıcaları ise Bayes tekniği, bootstrap, ortalama-varyans kestirimi, alt ve üst bant kestirim metodudur. Kestirim aralığının başlıca sorunlarından biri ise simetrik olarak olmasını sağlayacak tek bir değerin üretilmesi tahminlemenin etrafına artı ve eksi olmak üzere ilave edilmesidir. Birçok durumda tahminlemenin gerçekleşenden fazla olarak devam etmesi (üstte kalma) ve tam tersi (altta kalma) durum uzun süreçler boyunca devam etmektedir. Bu durumda simetrik bantların fayda sağlayıp sağlamadığı tartışmalara açık bir konudur.

Bu tez kapsamında, tahminleme hataları ve belirsizlikleri ifade etmede bahsedilen hata bantlarından yeni ve alternatif bir yöntem önerilmektedir. Önerilen yapı Üçgen Bulanık Sayılar aracılığıyla bulanık mantık tabanlı alt ve üst bant kestirici ile bulanık dilsel terimlerin üretilmesi ve ifade edilmesidir. Farkedileceği üzere önerilen yapı temel olarak Bulanık Mantık üzerine kurulmaktadır. Bulanık mantık tabanlı sistemler kontrol, robotik, görüntü işleme, karar verme, tahmin, modelleme gibi birçok farklı mühendislik uygulanasında başarıyla uygulanmıştır. Bulanık Mantık uygulamaları sunulan alanlar içinde özellikle belirsizlik modelleme özelliği ile öne çıkmaktadır. Bu doğrultuda, karar verici sistemleri desteklemek ve belirsizlikleri daha uygun bir şekilde modellemek adına, 2 farklı dilsel terim üretim yaklaşımı bu tez kapsamında sunulacaktır. Sunulan yaklaşım ile tahminlemeler ve onlar için üretilmiş Bulanık Üçgen Sayılar aracılığıyla kazandıkları bulanık dilsel terimler ve aitlik dereceleri aracılığıyla, tahminleme doğrulukları daha efektif bir şekilde değerlendirilebilmektedir. Önerilen iki yapıda temelde Kestirim Aralığı ile aynı işlevde olan FLUBE'u kullanmaktadır. Bunun yanında Bulanık Üçgen Sayıları üretecek olan kısımda ise ekstradan bir modelleme ile temel sunu olan dilsel terimlerin üretilmesi ve ifadesi ve bu versiyonun geliştirilmiş halidir.

Tez içeriksel olarak öncelikle zaman serileri konsepti ve bunların analizine değinmektedir. Analiz yöntemleri, tahminlemenin temelleri ile birlikte aktarılmaktadır. Günlük hayatta da sıklıkla kullandığımız zaman serisi kavramı, tahminlemenin de temelini oluşturmaktadır. Kavramlarda bu doğrultuda verilmektedir. Zaman serilerinin analizi ise tahminlemenin temelini oluşturmaktadır. Özellikle durağanlık kavramının önemi üzerinde durularak bir sonraki aşama olan modelleme kısmına nasıl geçilebileceği konusunda fikirler de bu giriş bölümünde verilmektedir. Tahminleme esas önemli kısım olan modelleme yöntemleri de bu kavramlarla birlikte ilk bölümlerde anlatılmaktadır. Modelleme yöntemleri literatürde oldukça geniş olarak işlenmesine karşın tez kapsamında ve literatürde en çok kullanılan yöntemlerin aktarılması amaçlanmıştır.

İkincil olarak ise tahminleme ve bağlı alanlarda sıklıkla kullanılan hata tanımları üzerinde durulmaktadır. Hata tanımlamaları bir tahminlerme yöntemini değerlendirmede en önemli birim olmakla beraber birçok çeşidi bulunmaktadır. Bunların arasından kullanımı en yaygın ve güncel yöntemlerin literatürdeki karşılıkları bu bölümde aktarılmaktadır. Hata değerlendirmelerinin yanında tezinde başlıca kapsamında bulunan hata bantları hakkında bilgilendirme ve özellikle Kestirim Aralığı hakkındaki literatür özeti takip eden başlığında yapılmaktadır. Devam edene başlıklarda ise yine sunulan yöntemin gelişmesinde vaşlıca etken olan kavramlar aktarılmaya devam etmektedir. Bulanık mantık bunlardan bir diğeri olmakla beraber tezin gelişimindeki her aşamasında katkısı görülmektedir. Bulanık mantık 1965 yılında önerilmiş ve özellikle belirsizlikleri modellemede oldukça etkili olan matematiksel bir araçtır. 1965'ten günümüze kullanımı oldukça yaygınlaşan bu araç günümüzde özellikle dil ve kelime işleme gibi insan-makine etkilişimindeki belirsizlikleri modellemede kullanılmaktadır. Tez kapsamında yine belirsizlik modelleme özelliği kullanılacak olan Bulanık Mantık, aynı zamanda dilsel terim karşılığı üretmede de yardımcı olacaktır. Önerilen yapıların oluşturulmasında katkıda bulunan bir diğer kavram olan Bulanık Zaman Serileri ise belirsizlikleri ifade etmede zaman serilerini kullanmaktadır. Bu yaklaşım ile zaman serilerinin Bulanık Üçgen Sayıları ile veya farklı dilsel terimler ile de ifade edilebileceği keşfedilebilmektedir. Bulanık modelleme teknikleri ise özellikle belirsizliklerin modellenmesi önemli bir araç olarak tez kapsamında kullanılmıştır. Bu kapsamda iki adet bulanık modelleme tekniği aktarılmaktadır. Tüm bu ön kavramların devamında ise önerilen yöntem olan bulanık mantık tabanlı alt ve üst bant kestirici ve bulanık mantık tabanlı alt ve üst bant kestirici yardımıyla dilsel terim üreteci olarak kullanılan yaklaşımları en son bölümde aktarılmadır.

Son olarak ise sunulan yöntemlerin görsel olarak da gösteriminin sağlandığı deneysel çalışmalar yapılmıştır. Deneysel çalışmalar için farklı veri setleri kullanılmıştır. Sunulan yöntemler, klasik Kestirim Aralığı ile karşılaştırmalı olarak farklı veri setleri üzerine uygulanmıştır. Kestirim Aralığı karşılaştırmasında önerilen yapının ilk kısmı ile karşılaştırma söz konusudur. Bulanık mantık tabanlı alt ve üst bant kestirici hataları modelleyerek Kestirim Aralığının sunduğu bilgiye alternatif olmayı amaçlamakta ve sağlamaktadır. Bunun yanında karar verici sistemlerin dilsel olarak tahminlerini ve aynı zamanda hata bantlarını değerlendirebileceği dilsel terim üreticilerinin çıktıları olan Bulanık Üçgen Sayılar aracılığıyla tahminleme hakkında çıkarımlar yapılabilmektedir. Burada Bulanık Üçgen Sayıların aitlik değerlerinin ortalamaların bir çıkarım olarak kullanılabileceği görülmektedir. Sunulan yöntemlerin tasarı adımları bu bölümden de sırasıyla takip edilebilir.

Tez kapsamında sunulan yapılar aracılığıyla özellikle tahminleme yönteminden bağımsız olarak sadece hataları modelleme üzerine sunulan yöntem ile hataların doğrusal olmayan modelleme tekniği olan bulanık mantık araçları ile modellenmiştir. Modellenen hatalar klasik Güven Aralığının aksine simetrik değilken daha da dar olması sebebiyle karşılaştırma dahilinde klasik Kestirim Aralığına üstünlük kurabilmektedir. Önerilen yapının dilsel terim üretme araçları sayesinde ise son kullanıcıların gerçekleşen veriler aracılığıyla üretilen hata bantlarının kalitesini inceleyebilecekleri, klasik Kestirim Aralığının değerlendirme yöntemi olan içeride dışarıda (0-1) değerlendirmesi aksine yeni bir değerlendirme yöntemi sunulmuştur. Bulanık üçgen sayılar ile ifade edilen yeni aralıklar sayesinde tahminlemeler birer aitlik değerine sahip olurken, bunların toplu olarak değerlendirilmesi sonucunda

tahminleme ve hata bantları birbiri ile aitlik değerleri bağlantısıyla tahminlemenin kalitesi konusunda son kullanıcıya bilgilendirme yapabilmektedir.

1. INTRODUCTION

Information technologies have many improvements about data storage and its usage in the last decades and it will have more according to technological breakthroughs. Data storage and data obtaining will be much easier than now after the Internet of Things revolution. This revolution can call as differently according to governed country. It named as Industry 4.0 in Germany, the Factory of the Future in France and Italy and Catapult in UK (Davies, 2016). The data explosion will also provide that data-acquiring process getting easier day after day, for several sector such informatics and bioinformatics (Ashton, 2009). Wide data scope of the several sectors will make the data analysis techniques more important especially in forecasting.

A forecast is a prediction of some future things or events. Forecasting is an important problem that link many fields such as economics, industry, environmental sciences and much more. Most of usage, the forecasting involves from the time series data. Many applications of forecasting about the business exploit daily, weekly, monthly or any defined interval of the data. These applications can be listed in the areas such as operations management, marketing, finance and risk management, economics, industrial process control and, demography. These are only a few where forecast is required to make good decisions (Montgomery et al, 2015).

The forecast model always aims to represent best estimate of the future value of the variable of interest. As it might be expected, these forecasts are usually wrong; so there is definition between estimated and real ones named as forecast error. It has been stated that there are two main problems with state of art forecasting methods: (i) the models become unreliable in the presence of uncertainty and (ii) no indication of the accuracy of the single point forecasts is provided (Khosravi et al, 2014). The accuracy of the forecast is usually measured with performance indexes such as Mean Absolute Percentage Error (MAPE), Percentage of Error (POA), etc. (Hyndmann & Koehler, 2006). However, since there is always an error margin in the predictions, there is a need to define error bounds in the forecast with its Confidence Interval

(CI), Prediction Interval (PI) or using other novel approaches (Khosravi et al, 2014b). Eventually, it is a good practice to accompany a forecast with an estimate of the error bounds to represent the interval about how large a forecast error might be experienced. Prediction Interval (PI) and Confidence Interval (CI) are the most used ones for representing the errors (Montgomery et al, 2015).

The main motivation for the construction of error bounds is to quantify the likely uncertainty in the point forecasts. Availability of error bounds allows the decision makers to quantify the level of uncertainty associated with the point forecasts. Error bounds that are relatively wide indicate the presence of high level of uncertainties in the underlying system operation. This useful information can guide the decision makers to avoid the selection of risky actions under uncertain conditions. On the other hand, narrow error bounds give the decision makers the opportunity to decide more confidently with less chance of confronting an unexpected condition in the future (Khosravi et al, 2011c).

The applications of the PIs increased in the last decade because of the increasing complexity of the man-made systems. Manufacturing enterprises, industrial plants, transportation systems, and communication networks are the some of the examples of these systems. More complexity adds to high levels of uncertainty in the operation of large systems. Operational planning and scheduling in large systems is often performed based on the point forecasts of the system's future. The reliability of these point forecasts is low and there is no indication of their accuracy. Therefore, the systems meet the interval bounds (Khosravi et al, 2011d).

In literature, several PI construction methods for many applications is proposed. Delta (Hwang and Ding, 1997; De Vieaux et al, 1998), Bayesian (Bishop, 1995; MacKay, 1992), mean-variance estimation (MVE) (Nix and Weigend, 1994), bootstrap (Efron, 1992; Heskes, 1997) and Lower Upper Bound Estimation (LUBE) method (Khosravi et al, 2011b) are proposed for construction of the PIs. PIs have been completed in a variety of disciplines, including transportation (van Hinsbergen et al, 2009; Khosravi et al, 2011c; Khosravi et al, 2011d), energy market (Zhao et al, 2008; Khosravi et al, 2010), manufacturing (Papadopoulos et al, 2001), and financial services (Benoit et al, 2009).

In this thesis, alternative approach to the error bounds will be represent. Fuzzy linguistic term generation and representation via Fuzzy logic based Lower and Upper Bound Estimator (FLUBE) based Triangular Fuzzy Number (TFN) is presented to estimate the uncertainty in the forecast.

Zadeh firstly introduced the fuzzy logic concept in 1965 (Zadeh, 1965). Mamdani accomplished the first industrial application example of the fuzzy logic in 1974 in the area of control theory by constructing to the linguistic control rules of a skilled human operator (Mamdani, 1974). Nowadays, Fuzzy Logic Systems (FLSs) have been successfully implemented in various engineering areas and their applications including control, robotics, image processing, decision-making, estimation and modelling. The FLSs become more and more popular and take part in many studies from that date. One of the most well known structure fuzzy logic systems is Takagi-Sugeno-Kang fuzzy logic system that was first proposed in 1985. That system is widely used in control applications (Takagi and Sugeno, 1985). Moreover, since most of the systems and processes are characterized by uncertainties and nonlinearities, various fuzzy modeling and fuzzy control strategies have been successfully implemented in many engineering problems (Babuška et al., 2012).

The proposed method and its approaches will give the opportunity to the decision maker to quantify the uncertainty of the point forecasts with linguistic terms that might increase the interpretability with using Fuzzy Logic and its modelling tools. Fuzzy Logic modelling provides not only nonlineer modelling but also non symmetrical bounds for errors while classical PIs only concludes the one value for the creation of the error bounds. Moreover, the proposed approaches will provide valuable information about the accuracy of the forecast by providing a relative membership degree with respect to the target data. The proposed approaches consist of two main phases, the offline FLUBE design and the online TFN generation part as Linguistic Generation and Representation Approach (LinGRA) and Enhanced Linguistic Generation and Representation Approach (ElinGRA). By the help of mentioned linguistic terms, the decision makers will have oppurtunities to compare the forecasting methods. Classical PIs only ensure the binary conclusion that determined from the realized value of the forecast, is it in the interval or not, while the mentioned methodologies will provide the membership degrees of the realized

values of the forecasts that can be determined from the Triangular Fuzzy Numbers of the created linguistic terms.

The thesis organized in five sections including conclusion. In Section 2, the general information about the time series is represented in terms of the components of the time series and its modelling approaches especially the basic ones like AutoRegressive (AR), Moving Average (MA), and ARMA. The basic forecasting concept is also mentioned at this part of the thesis. Section 3 is the main part of the thesis. The proposed method is represented at the Section 3 with its concepts that provide to comprising the structure of the method. Evaluation of the forecast errors in special interest with forecast error bounds, Fuzzy Logic, Fuzzy Modelling, Fuzzy Time Series and the proposed method is presented respectively. In section 4, the experimental results of the proposed method is illustrated on several data sets. Consequently, the results are concluded and discussed in the conclusion section.

2. TIME SERIES FORECASTING

In the last decades of the information technologies includes many improvements about data storage and its usage. Getting data from anything will be much easier than now after the Internet of Things revolution named as Industry 4.0 in Germany, the Factory of the Future in France and Italy and Catapult in UK (Ashton, 2009). Obtaining the data from the cheap way also stimulates the analysis of the data especially in forecasting.

In this chapter, the data analysis from the point of the view of the time series will be discussed. First subsection is representing the time series concept and its inferior parts. Second subsection tries to explain the forecasting and forecasting process briefly. The studies will be present at the next chapter construct on the forecasting and tries to improving the forecasting approaches. Third and Fourth subsection will present the time series analysis via the residual concept and its modelling approaches. The modelling approaches are selected according the most used ones according selective references.

2.1 Time Series

A time series is a sequence of observations x_t, each one taken sequentially in time t. In other word, an ordered sequence of values of a variable at equally spaced time intervals like hourly, daily, weekly or yearly. Examples occur in a diversity of fields, ranging from engineering to economics, and methods of analyzing time series constitute an important area of statistics. A monthly sequence of the quantity of goods shipped from a factory, a weekly series of the number of road accidents, hourly observations made on the yield of a chemical process, and so on can be listed (Box et al, 2015) . In the following chapters, some representative time series are illustrated with their statements (Chatfield, 1995).

2.1.1 Economic time series

Time series have always been used in the field of econometrics. Tinbergen is firstly constructed the first econometric model for the United States and thus started the scientific research programme of empirical econometrics (Tinbergen et al, 1939). After that first application, many time series arise in economics. Figure 2.1 shows the Beveridge wheat price index time series that consist of the average wheat price in nearly 50 places in various countries measured from 1500 to 1869 as yearly. The complete time series is presented by Anderson as also published as online at Url-1 (Anderson, 2011; Url-1, 2016).

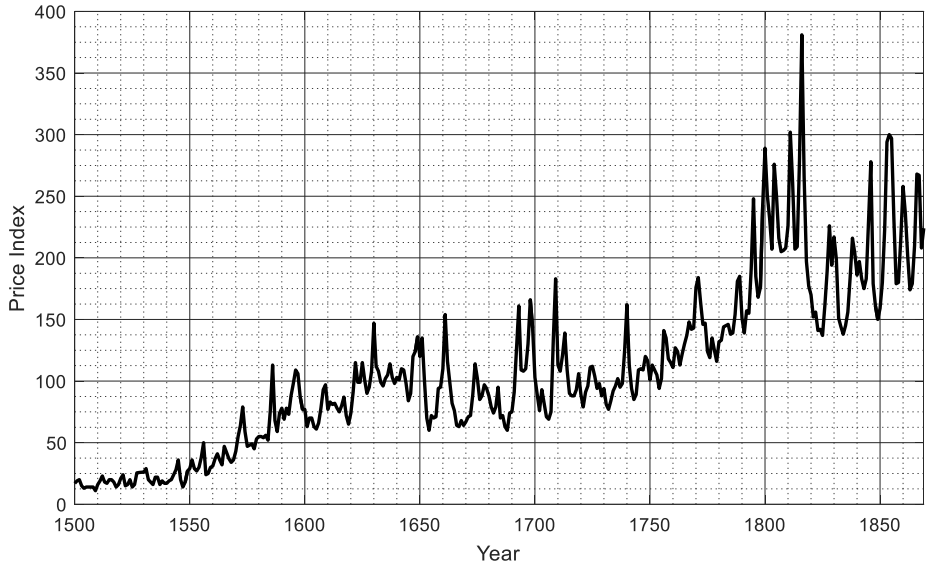


Figure 2.1 : Beveridge monthly wheat price index time series.

Figure 2.2 shows weekly closings indexes of the Dow-Jones industrial average also called DJIA from July 1970 to August 1974 (Hsu, 1979). DJIA is a stock market index, and one of several indices created by *Wall Street Journal* editor and Dow Jones & Company co-founder Charles Dow. The industrial average was first calculated on May 26, 1896.

2.1.2 Physical time series

Various types of the time series occur in the physical sciences, especially in meteorology, marine science and geophysics. Rainfall on successive days, air temperature in successive hours, days or months, water level of the lakes are examples about the physical time series. Figure 2.3 shows the surface temperature at Edirne, in Turkey, averaged over successive months. The data comes from

HadCRUT3, a gridded dataset of global historical surface temperature anomalies produced by the Met Office Hadley Centre and the Climatic Research Unit at the University of East Anglia (Url-2, 2016).

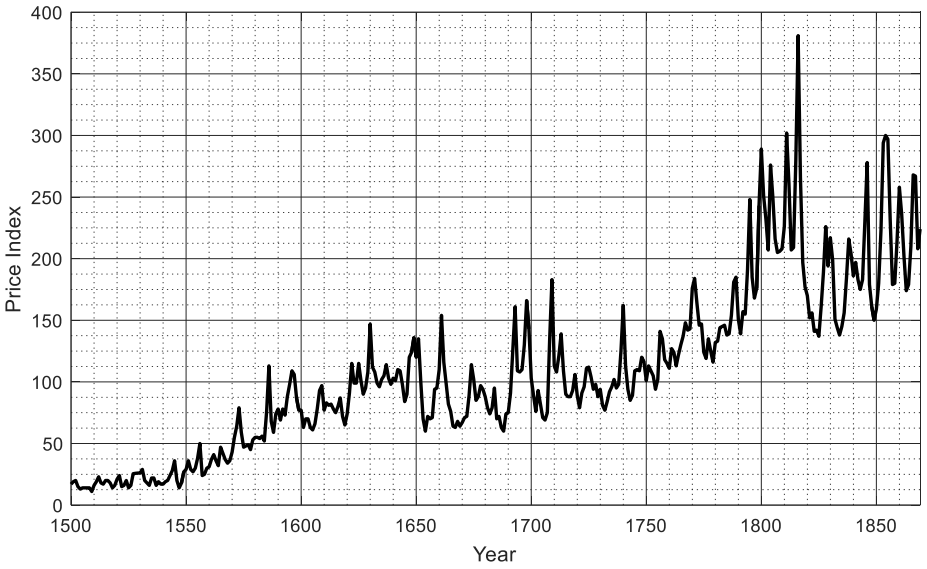


Figure 2.2 : Dow Jones weekly industrial average index.

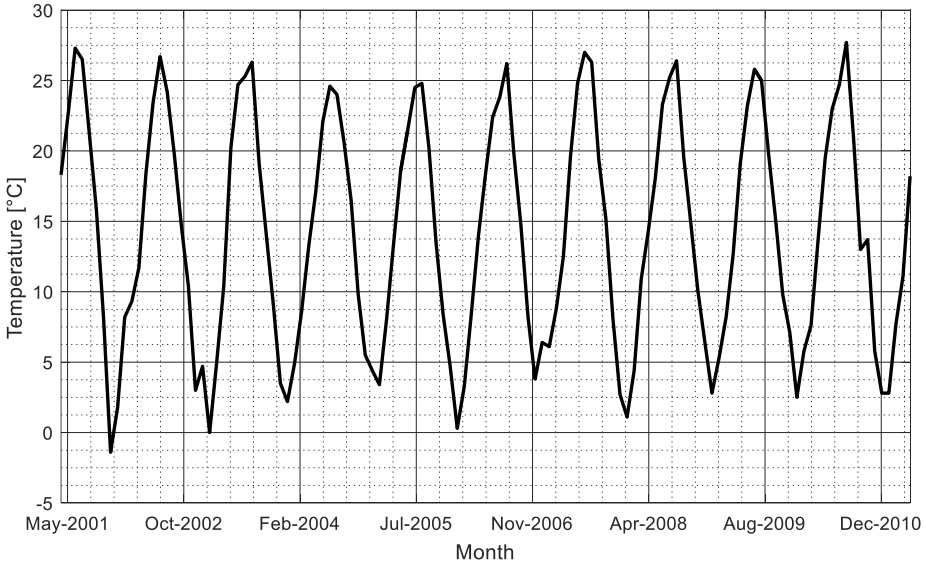


Figure 2.3 : Monthly surface temperature at Edirne in Turkey.

Some mechanical recorders take measurements continuously and produce a continuous trace. In process control, the problem is to detect changes in the performance of a manufacturing process can be evaluate as physical problems and includes clarified categories. Figure 2.4 shows time series of chemical process concentration observes from readings every 2 hours with 197 observations (Box &

Jenkins, 2015). According to Box & Jenkins, the unit of the concentration is currently unknown.

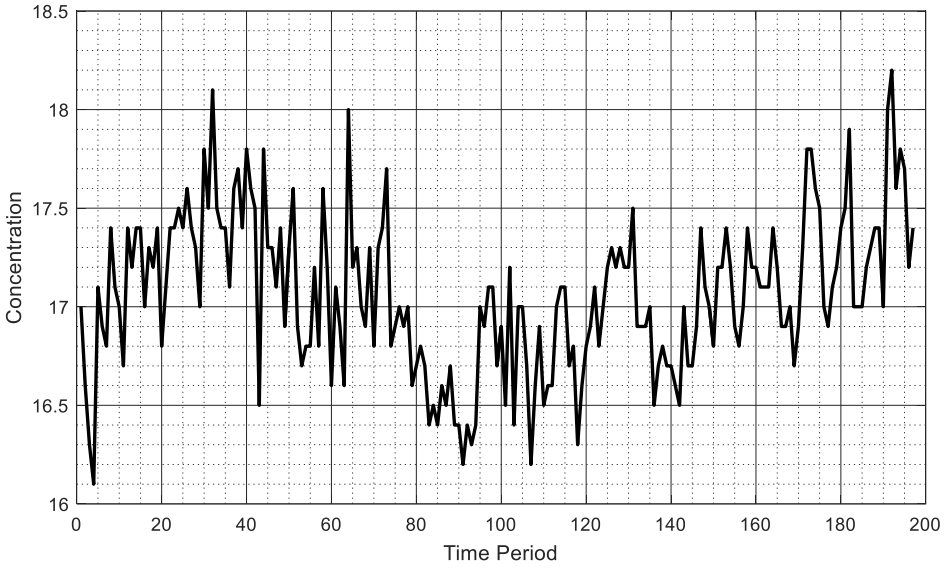


Figure 2.4 : Chemical process concentration readings every 2 hours.

2.1.3 Marketing time series

The analysis of sales and stock figures in successive days, weeks or months is a critical management problem. It may also be of interest examine the relationship between sales and other time series such as advertising expenditure. Figure 2.5 shows the monthly sales (in thousands of liters) of red wine by Australian winemakers from January 1980 through July 1995 (Brockwell and Davis, 2002). The sales data clarify that sales performance of the red wine. Like the preceding one, Figure 2.6 states the quarterly clay brick production from March 1956 to September 1994 (Makridakis et al, 1998).

2.1.4 Demographic time series

Demographic time series occur from the study of population, generally. Most of times, changes are tried to predict for as long as ten or twenty into the future (Brass, 1974). Figure 2.7 shows the change of the population of the U.S.A., measured at ten-year intervals (Brockwell and Davis, 2002).

2.1.5 Other time series

The observations can take only two values mostly denoted with 0 or 1. Nevertheless, in some cases it could be changes as -1 or 1. These types of time series called binary

processes. All-star baseball games' results are instance for binary processes (Brockwell and Davis, 2002).

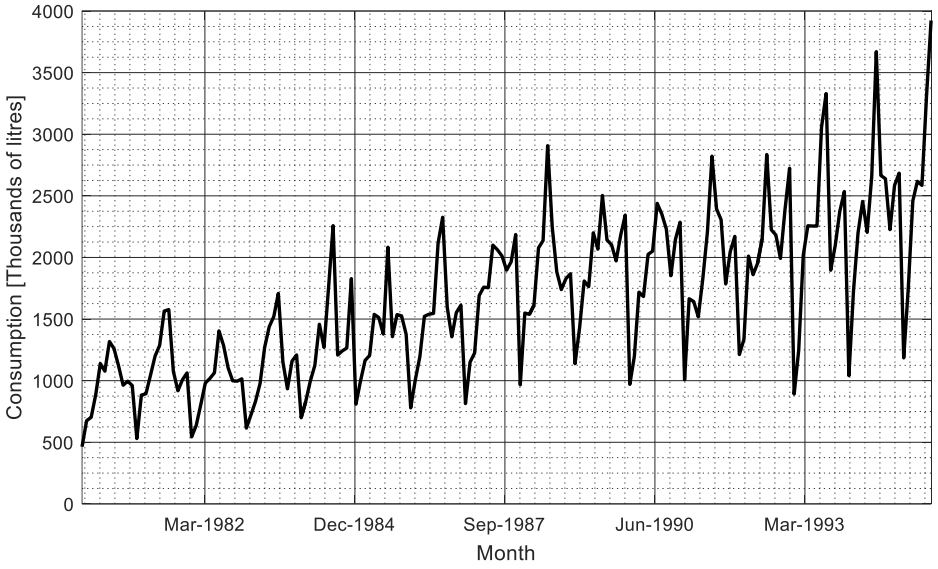


Figure 2.5 : Australian monthly red wine sales time series.

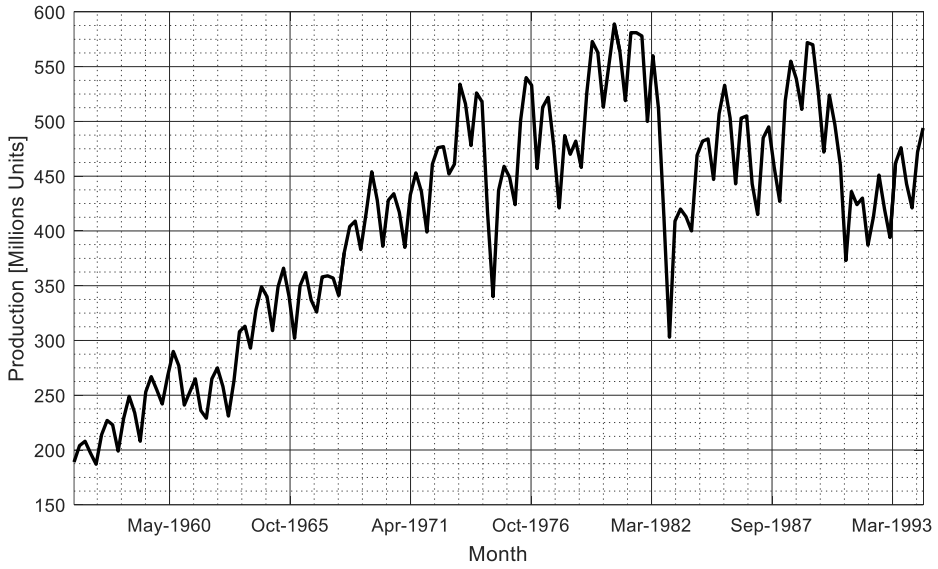


Figure 2.6 : Australian quarterly clay brick productions time series.

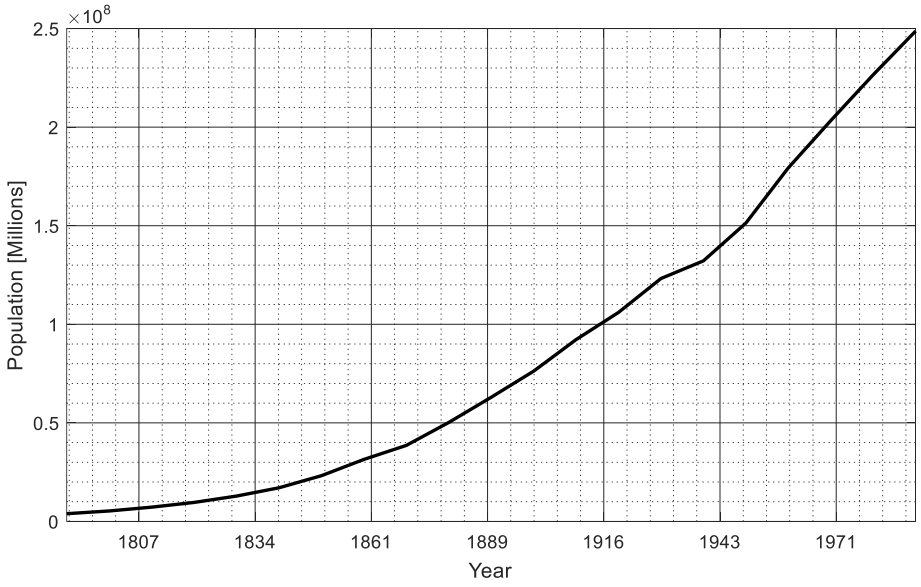


Figure 2.7 : Population of the U.S.A at ten-year intervals: 1790-1990.

2.2 Forecasting

A forecast is a prediction of some future things or events. According to Neils Bohr, it is not always easy. In book *Bad Predictions* (Lee, 2000), there are several forecasts that can be evaluated as 'bad' forecasts as the most famous ones (Montgomery et al, 2015):

- "The population is constant in size and will remain so right up to the end of mankind." (Diderot & d'Alembert, 1756)
- "1930 will be a splendid employment year." U.S Department of Labor, *New Year's Forecast* in 1929, just before the Dow Jones crash on October 29 (Mannermaa, 2004).
- "Computers are multiplying at a rapid rate. By the turn of the century there will be 220,000 in the U.S." *Wall Street Journal*, 1966 (Montgomery et al, 2015).

In consideration of the given examples, it can be said that forecasting is an important problem which link many fields such as economics, industry, environmental sciences and much more. Forecasting can be classified according to their scopes such as short and long-term forecasts. As it can clearly have noticed that short-term forecasting problems involve predicting a few times of the event into the future. Because of the

forecasts obtaining from the historical data sets especially time series, the sort-term forecast are mostly more reliable than the long ones.

Most of usage, the forecasting involve from the time series data. Many applications of forecasting about the business exploit daily, weekly, monthly or any defined interval of the data. These applications can be listed in the areas such as operations management, marketing, finance and risk management, economics, industrial process control and, demography. These are only a few where forecast are required to make good decisions. Although the wide range of the problems, there are two extensive types of the forecasting techniques – qualitative and quantitate forecasting techniques (Montgomery et al, 2015).

Qualitative forecasting techniques depends on the judgment of the experts. It is useful where there is not much more historical data for forecasting. It depends not only experts' decision but also the surveys of potential customers, and experiences about the similar products or services.

Quantitave forecasting techniques is under the scope of this thesis that use the historical data and forecasting models. Most of times, the model formally summarizes patterns and express statistical relations between observations. Therefore, the model use for estimating past and current behavior into the future.

The forecast model always aims to represent best estimate of the future value of the variable of interest. As it might be expected, these forecasts are usually wrong; so there is definition between estimated and real ones named as forecast error. Eventually, it is a good practice to accompany a forecast with an estimate of the error bounds that represent the interval about how large a forecast error might be experienced. Prediction Interval (PI) and Confidence Interval (CI) are the most used ones for representing the errors. In the context of the thesis, alternative approach to the error bounds will be represent (Montgomery et al, 2015).

All of the mentioned process can be illustrated as a flow chart given at Figure 2.8. First two part is mentioned before chapter as time series data and third part will be discussed at Time Series Analysis chapter. Model selection part of the process will be discussed briefly at the next chapters. Model validation, deployment and monitoring will be the part of the implementation section of the thesis.

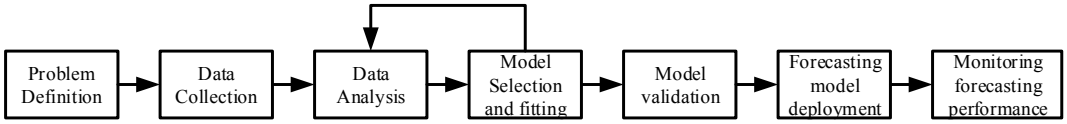


Figure 2.8 : The forecasting process.

2.3 Time Series Analysis

In this chapter, time series analysis will be discussed by presenting the components of the time series that are effects the characteristics of the time series. Most of times, time series occurs from the sum of a specified trend, and seasonal and random terms. By the means of definitions, the trend is evolutionary movement, either upward or downward. It will be long-term or relativity short duration with its dynamics. Seasonality is the component that repeats on regular basis, such as each month. The remove process of components named as seasonal adjustment is important to recognizing of the time series or data for not to confuse basic model of the time series.

The selection of a suitable probability model or class is the most important part of the time series analysis. The fitting of the time series models can be an ambitious undertaking at literature. There are many methods of model fitting including, AutoRegressive Integrated Moving Average (ARIMA), Multivariate, and Holt-Winters Exponential Smoothing. In this chapter, the basic ones and at the same time the easiest ones according to given methods are discussed.

2.3.1 Random walk

The simplest model for a time series is one is which is there is no trend or seasonal component. The observations of the time series are independent and identically distributed (iid) random variables with zero mean. The random walk is obtained by cumulatively integrating iid random variables. According to Brockwell, it can be obtained by defining its starting value S_0 as 0 and

$$S_t = X_1 + X_2 + \cdots X_t, \qquad for\ t = 1,2 \ldots \tag{2.1}$$

where $\{X_t\}$ is iid noise. As another aspect, it can be defined for each time period that going from left to right, the value of the variable takes an independent random step up or down, a so-called random walk. If up and down movements are equally likely

at each intersection, then every possible left-to-right path through the grid is equally likely a priori. In addition to zero mean version, random walk with drift provides addition with nonzero member and it can be represent as below:

$$S_t = S_{t-1} + X_t \tag{2.2}$$

The daily change of the exchange rates can be model with random walk according to its characteristics like many peaks and valleys.

2.3.2 Models with trend and seasonality

In several of the time series examples of Section 2.1 contains a clear trend in the data. The Australian electricity consumption and the air passenger data sets also includes trend components that are represented at Figure 2.9 and Figure 2.10 (Makridakis et al, 1998). First data set involves the electricity consumption values from January 1956 to August 1995, thus the data set has 476 samples. Second data set involves monthly counts of the international airline passengers, measured in thousands, for the period January 1949 through December 1960, thus the data set has 144 samples.

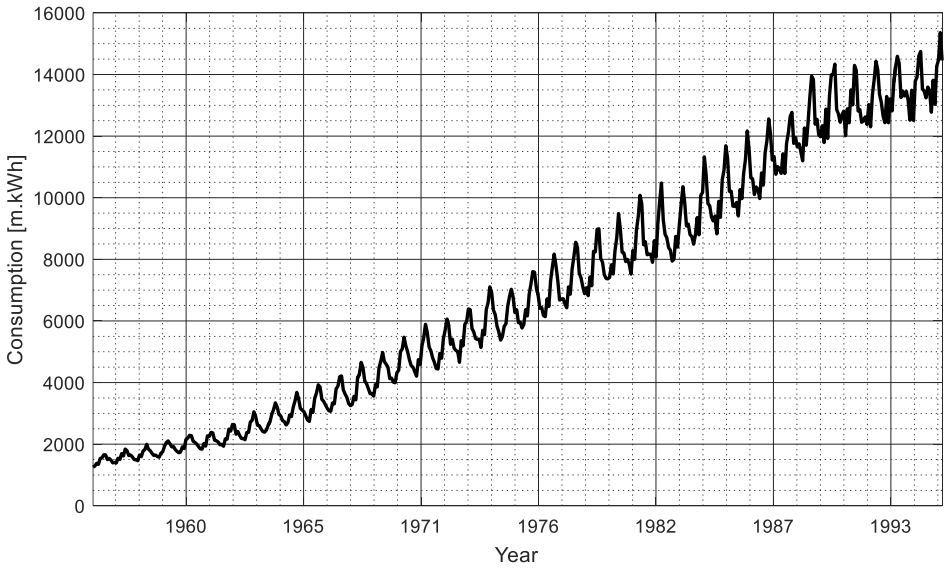


Figure 2.9 : The Australian monthly electricity consumption data set.

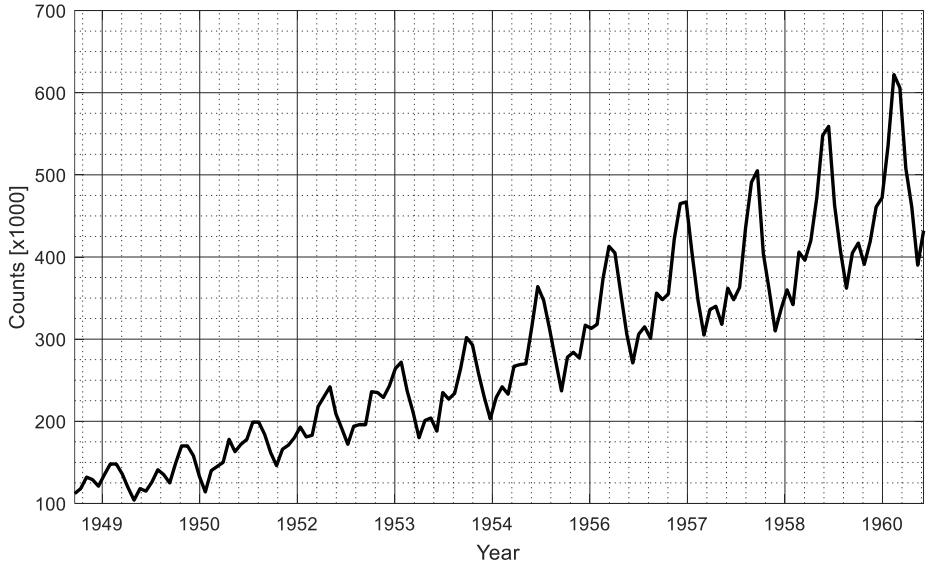


Figure 2.10 : The air passenger data set.

The zero mean models, as random walk is inappropriate for these type data. Therefore, the trend component of the time series is expressed as follows:

$$X_t = m_t + Y_t \tag{2.3}$$

where m_t is an evoluting function kown as the trend component and Y_t has zero mean. Most general method for the estimating m_t is the least squares (LS) method. The LS method is a parametric method like given as follow:

$$m_t = l_0 + l_1 t + l_2 t^2 \tag{2.4}$$

where $l_0, l_1,$ and l_3 are the weight of the time vector. In addition to the trend characteristics, many time series are influenced by seasonally varying factors such as the weather that can be modeled by a periodic component with fixed known period. For example, the Australian electricity consumption data set (Figure 2.9) shows repeating annual patterns with peaks and trend in July and troughs in February, strongly suggesting a seasonal factor with period 12. In order to represent the seasonal effect with assumption of there is no trend; the simple model can be defined as follow:

$$X_t = s_t + Y_t \tag{2.5}$$

where s_t is a periodic function of t with period d ($s_{t-d} = s_t$). s_t is a sum of harmonics which is given by:

$$s_t = a_0 + \sum_{j=1}^{k} \left(a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)\right) \tag{2.6}$$

where $a_0, a_1, \dots, a_k$ and $b_1, \dots, b_k$ are unknown parameters and $\lambda_1, \dots, \lambda_k$ are fixed frequencies, each being some interger multiple of $2\pi/d$ (Brockwell and Davis, 2002).

The illustrative examples about the trend and seasonality modelling, and the component decompositions will be discuss at the next section.

2.3.3 General approach to time series modelling

The overview of the time series modelling is gave by Brockwell (Brockwell and Davis, 2002). According to this flow, the time series can be model with applying items given below:

- Plot the time series and then checking any outlying observations and apparent sharp changes,
- After the remove process of the outliers, seasonal and trend components, the stationary residuals obtain,
- Choose a model to fit the residuals,
- Forecasting will be achieved by forecasting the residuals,
- Inverting the transformations described at the previous steps to arrive at forecast of the original series.

According to the given items, the stationary term and the model decomposition process must be define. Both two concepts will be discussed at the next subsections.

2.3.4 Stationary concept

The stationary of a time series is related to its statistical properties in time. In general manner, a time series $\{X_t, t = 0, \pm 1, \dots\}$ will be stationary if it has statistical properties similar to its time shifted version $\{X_{t+h}, t = 0, \pm 1, \dots\}$ for each time shift integer(h). If these statistical properties depends only on the first $\left(E(X_t)\right)$and second

order $E(X_t^2)$ moments of the time series (X_t) and also provides the following definitions, it can be said that the time series is stationary (Brockwell & Davis, 2002):

$\{X_t\}$ is weakly stationary while $E(X_t^2) < \infty$ for all t:

- If the mean function $\mathrm{M}_X(t)$ is independent of t
- If the covariance function $\gamma_X(t+h,t)$ is independent of t for each h

where the mean and covariance function defined as below:

$$\mathrm{M}_X(t) = E(X_t) \tag{2.7}$$

$$\gamma_X(r,s) = Cov(X_r, X_s) = E\big[\big(X_r - \mathrm{M}_X(r)\big)\big(X_s - \mu_X(s)\big)\big] \tag{2.8}$$

For example, if $\{X_t\}$ is iid noise and $E(X_t^2) = \sigma^2 < \infty$, then the first requirement of the conditions is obvious satisfied, since $E(X_t) = 0$ for all t. By assumed independence:

$$\gamma_X(t+h,t) = \begin{cases} \sigma^2, & if\ h = 0 \\ 0, & if\ h \neq 0 \end{cases} \tag{2.9}$$

which does not depend on t. Hence iid noise with finite second moment is stationary.

2.3.5 Estimation and elimination of trend and seasonal components

Elimination of trend and seasonal components also named as decomposition that is used after the outliers' elimination to the time series data. Decomposition process apply for getting stationary time series to modelling by the forecast models. In this section, the decomposition process will be discussed and it applied on Australian monthly electricity consumption data set given at Figure 2.9.

The classical decomposition model can be define as:

$$X_t = m_t + s_t + Y_t \tag{2.10}$$

where m_t is a slowly changing function known as a trend component, s_t is a periodic function with d period known as a seasonal component, and lastly Y_t is a random noise component which provide stationary condition given at previous section (Brockwell and Davis, 2002). In other world, the purpose of the decomposition is

determination of the deterministic components (m_t and s_t) and after the elimination of these components, getting the stationary residuals (Y_t).

Another decomposition approach is developed by Box and Jenkins that apply differencing operators repeatedly to the time series $\{X_t\}$ until the differenced observations. Therefore the model can be construct with remain residuals (Box et al, 2015).

The first step of the decomposition is the trend elimination stage. According to Brockwell and Davis, the trend elimination can be done with four different method: Smoothing with a finite moving average (MA), Exponential smoothing, Smoothing by elimination of high-frequency components, polynomial fitting.

The moving average based filter applied to get raw trend components given as follow. It illustrated as $\widehat{m}_t$ because of that is the estimated one.

$$\widehat{m}_t = 0.5X_{t-q} + x_{t-q+1} + \cdots + x_{t-q-1} + 0.5x_{t+q}/d \tag{2.11}$$

where q is shifting operator and d is the window length of the moving average. While $q < t \leq n - q$, the shifting operator is the half of the period ($q = d/2$) if the period is even. Therefore, the raw trend can be obtained according to simple version of the equation:

$$\widehat{m}_t = \frac{\sum_{j=-q}^{q}(x_{t-j})}{d} \tag{2.12}$$

The time series definable at $0 \leq t$ and $t < n$. Therefore the observations with $t \leq q$ or $t > n - q$ are not numerable. The estimated raw trend $\widehat{m}_t$ is illustrated according to the Australian monthly electricity consumption data set at Figure 2.11 with the knowledge of the seasonality term as 12 month.

Following the estimated raw trend taking out from the raw data, the next step will be determining the seasonal component of remain data. Therefore, an average deviation (w_k) is determining according to period (d) and complete season count of the data (r) where $k = 1, \ldots, d$ and $j = 0,1, \ldots, r$

$$w_k = \frac{x_{k+jd} - \widehat{m}_{k+jd}}{r+1} \tag{2.13}$$

under circumstance of $q < k + jd \leq n - q$. By the help of average deviation, the seasonal component determines as without any condition about the zero mean:

$$\hat{s}_k = w_k - \frac{1}{d}\sum_{i=1}^{d} w_i \tag{2.14}$$

where $k = 1, \ldots, d$. The seasonal component $\hat{s}_t$ is illustrated according to the Australian monthly electricity consumption data set at Figure 2.12 with the knowledge of the seasonality term as 12 month. As it can be noticed, the consumption is maximized at the month July and minimized at the month February. Therefore, the seasonal pattern of the data set can be defined by the help of average deviations.

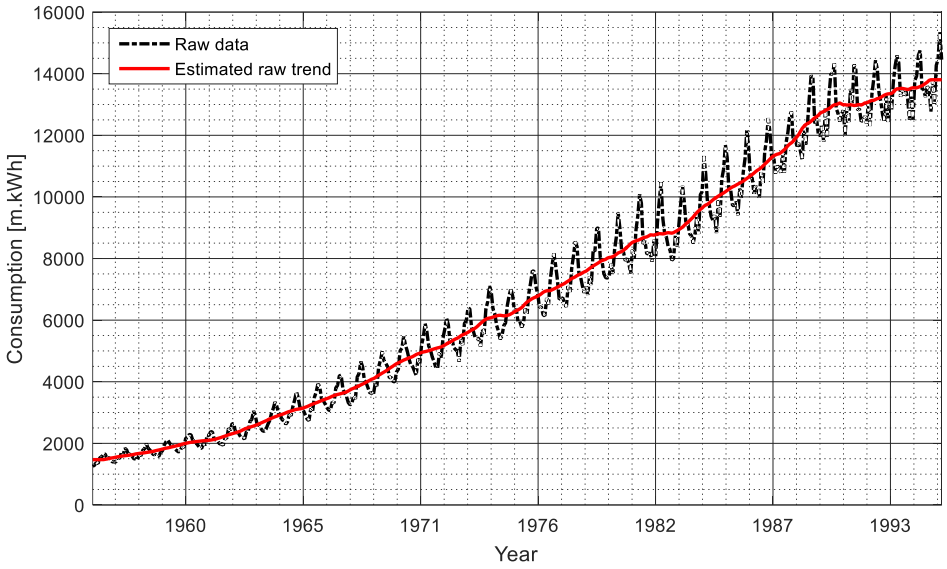


Figure 2.11 : Estimated raw trend ($\widehat{m}_t$) illustration on Australian monthly electricity consumption data set.

After the determination of the seasonal component, the rest of the part of the time series will include only the residual and trend components. The trend component determination will define again without any seasonal components at the next stage. Thus, the remainder components define as follow:

$$d_t = x_t - s_t \tag{2.15}$$

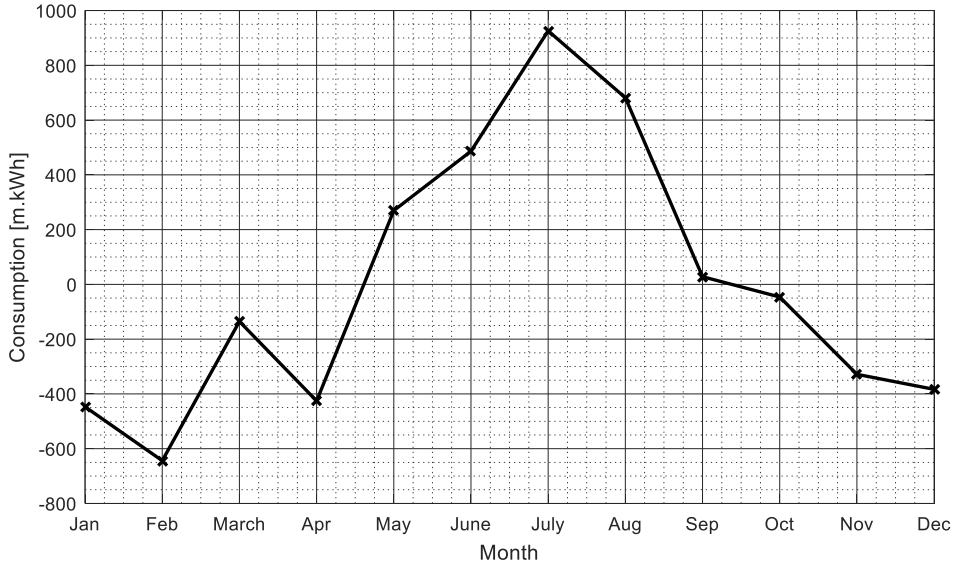


Figure 2.12 : Seasonal component ($\hat{s}_t$) illustration on Australian monthly electricity consumption data set.

Deseasonalized time series of the example data is illustrated at Figure 2.13 on the Australian monthly electricity consumption data set. As it can be noticed on deseasonalized data, the seasonal component is determined according to the general seasonal behavior of the time series. Therefore, sometimes the seasonal component decomposition cause creating new seasonal effect on the beginning and end of the time series.

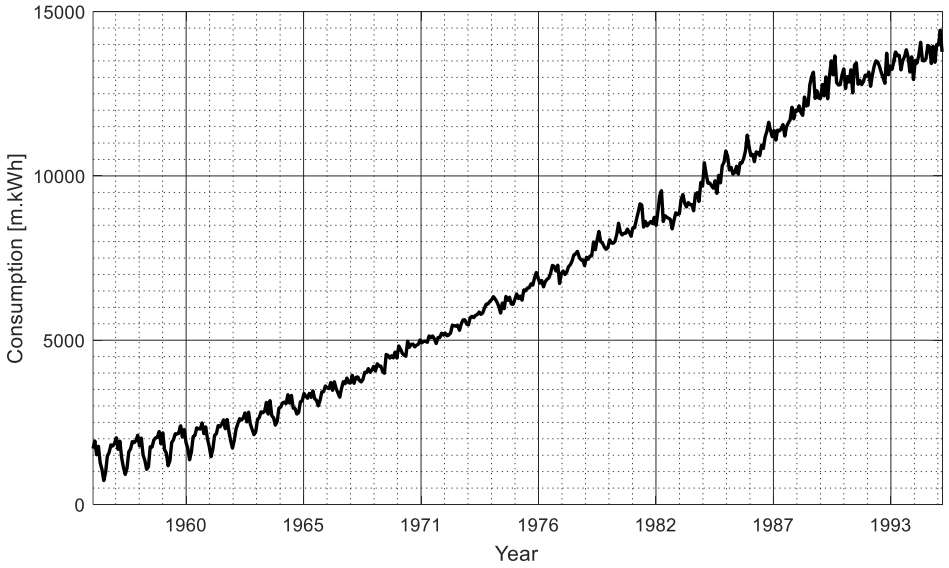


Figure 2.13 : Deseasonalized data (d_t) illustration on Australian monthly electricity consumption data set.

The trend component determination of the decomposition process is occurred after the getting deseasonalized data (d_t). As it can clearly seen at Figure 2.13 deseasonalized data has trend. The trend component can be defined with polynomial function by the help of the Least Squares (LS) algorithm. The model can be selected as first or second order to help the create simple model as possible (Brockwell and Davis, 2002). Thus second order polynomial is obtained from the example data by the help of LS algorithm given as follow equation. The second order polynomial regulation can be seen at Figure 2.14:

$$m_t = 0.0266t^2 + 16.1250t + 1042.5 \tag{2.16}$$

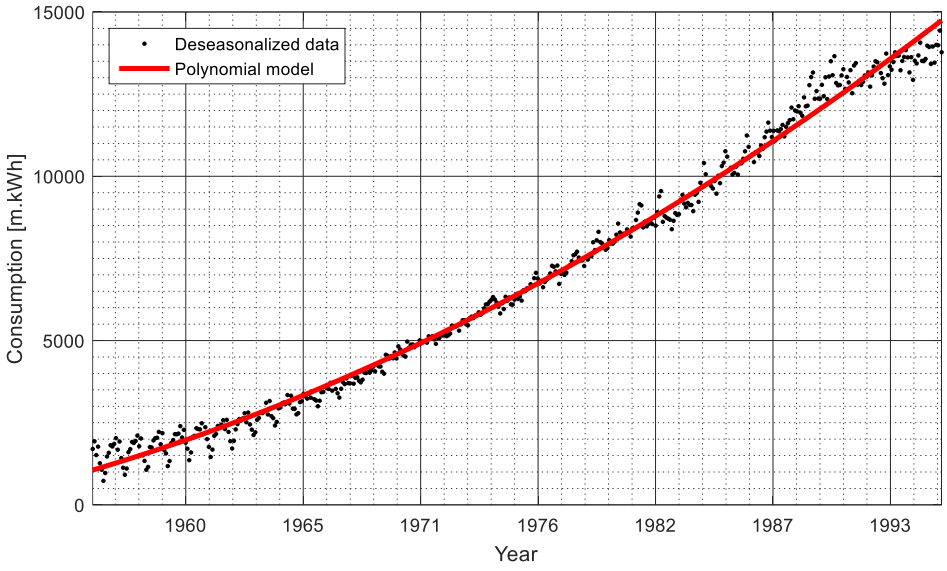


Figure 2.14 : Estimated trend function for Australian monthly electricity consumption data set.

The residual data (y_t) of the time series is determined after decomposing the trend components on deseasonalized data according to equation given as follow:

$$y_t = d_t - m_t \tag{2.17}$$

The residual data of the example data given at Figure 2.15. The residual modelling is discussed at the next sections with its modelling approaches.

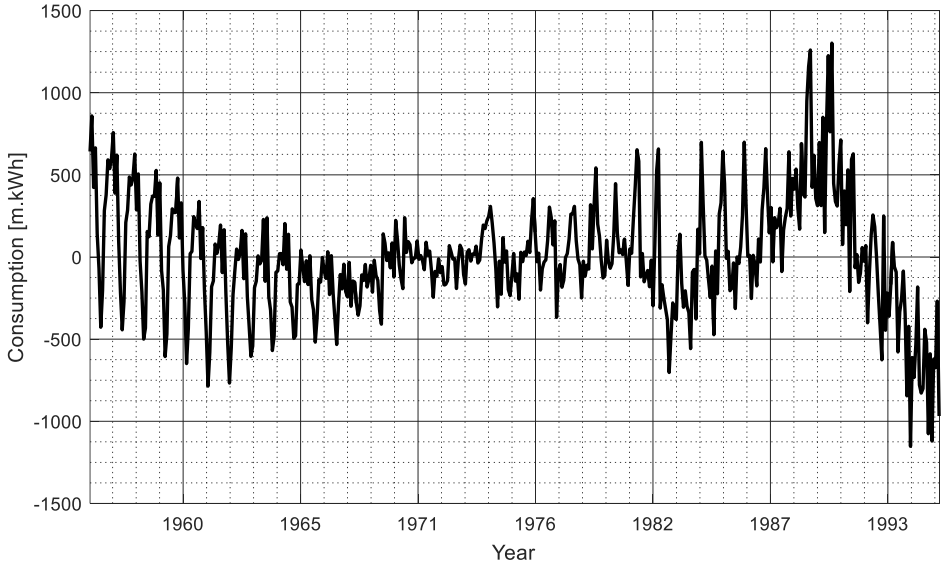


Figure 2.15 : Residuals of the Australian monthly electricity consumption data set.

2.4 Forecasting Methods

In time series modelling, the selection of the forecasting methods is the most influential factor of fitting to the data. The fitting process can be done with model and its parameters selection. In this chapter, several types of time series models are discussed.

2.4.1 Stationary models

The stationary or weak stationary condition is discussed at previous chapter. In this chapter, the forecasting methods for stationary time series will be discussed.

2.4.1.1 Moving Average (MA) models

The Moving Average (MA) model is a common approach for modelling univariate time series. The notation MA(q) refers to the moving average model of order q:

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \tag{2.18}$$

where μ is the mean of the series, $\theta_1, \ldots, \theta_q$ are the parameters of the model which aim to weighting the previous error terms, and $\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ are the white noise error terms which obtains from the subtracting the estimated value and the real observation given as follow:

$$\varepsilon_t = y_t - \hat{y}_t \tag{2.19}$$

q^{th} order MA model state as MA(q) and includes q parameter which are the weights of previous error terms.

2.4.1.2 Autoregressive (AR) models

Autoregressive (AR) model specifies that the output variable depend linearly on its own previous values and on a stochastic term. Therefore, the autoregressive model can be formulized depend on the model order p:

$$\begin{aligned} y_t &= c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \\ y_t &= c + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \end{aligned} \tag{2.20}$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are the parameters of the model which aim to weighting the previous observations, ε_t is white noise, and c is the model constant. p^{th} order AR model state as AR(p) and includes $p + 1$ parameter which are the weights of previous observations and the constant.

2.4.1.3 Autoregressive – Moving Average (ARMA) models

Autoregressive moving average (ARMA) model is the combination of the two polynomials, one for the auto-regression and the second for the moving average. The general ARMA model was described by Whittle in 1951 (Whittle, 1951). Box and Jenkins popularized it in the book in 1971 (Box and Jenkins, 2015). The ARMA model can be formulized depend on the model order p and q:

$$y_t = c + \varepsilon_t + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{2.21}$$

ARMA(p, q) includes $p + q + 1$ parameter which are the weights of previous observations, error terms and the constant.

2.4.2 Nonstationary models

It is often the case that while the processes may not have a constant level, they exhibit homogeneous behavior over time. Consider, for example, the linear processes. It can be seen that different snapshots taken in similar behavior.

Therefore, there are several modelling approaches for these type data. In this chapter, they will discussed as respectively.

2.4.2.1 Autoregressive–Integrated–Moving Average (ARIMA) models

The autoregression integrated moving average (ARIMA) model is a generalization of the ARMA model. The model is applied in some cases where data show evidence of nonstationary. ARIMA models form an important part of the Box-Jenkins approach to time-series modelling (Box and Jenkins, 2015). ARIMA models generally denoted ARIMA (p, d, q) where parameters $p, d,$ and q are non-negative integers, p is order of the AR model, d is the degree of differencing, q is order of the MA model. The ARIMA model can be formulized as:

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right)(1 - B)^d y_t = c + \left(1 + \sum_{i=1}^{q} \theta_i B_i\right)\varepsilon_t \tag{2.22}$$

where L is the lag operator, where ϕ_i are the parameters of the AR model which aim to weighting the previous observations, c is the model constant, θ_i are the MA parameters of the model which aim to weighting the previous error terms, and ε_t are the white noise error terms. The random walk process, ARIMA$(0,1,0)$ is the simplest nonstationary model. It is suggesting that first differencing eliminates all serial dependence and yields a white noise processes.

2.4.2.2 Seasonal Autoregressive–Integrated–Moving Average (SARIMA) models

The Seasonal ARIMA (SARIMA) models are usually denoted ARIMA(p, d, q) $(P, D, Q)_m$ where m refers to the number of periods in each season. They are enhancements of ARMA class in order to include more dynamics, respectively, non-stationary in mean and seasonal behaviors. Distinctly from the ARIMA models, the uppercase P, D, Q refer to the autoregressive, differencing, and moving average terms for the seasonal part of the ARIMA model. The SARIMA model can be formulized as:

$$\begin{aligned}\left(1 - \sum_{i=1}^{p} \phi_i B^i\right)(1 - B)^d \left(1 - \sum_{i=1}^{P} \Phi_i L^i\right)(1 - B^m)^D y_t \\ = c + \left(1 + \sum_{i=1}^{q} \theta_i B_i\right)\left(1 + \sum_{i=1}^{Q} \Theta_i B_i^m\right)\varepsilon_t\end{aligned} \tag{2.23}$$

where B is the lag operator, where ϕ_i and Φ are the parameters of the AR and SAR model which aim to weighting the previous observations, c is the model constant, θ_i and Θ are the MA and SMA parameters of the model which aim to weighting the previous error terms, and ε_t are the white noise error terms.

2.4.3 Other models

In addition to ARMA or ARIMA models, there are several forecasting methods in literature. Vector ARIMA models are used for forecasting on multivariate time series. Multivariate time series models involve several variables that are not only serially but also cross-correlated. The state space model and its Kalman recursions applications used as forecasting techniques in different approaches (Brockwell and Davis, 2002).

The constant variance assumption is often taken as a given for most of the models. In many practical cases and particularly in finance, it is common to observe the violation of this assumption. Autoregressive Conditional Heteroskedastic (ARCH) models proposed by Engle for dealing heteroskedastic data (Engle, 1982). The notation also enhanced as Generalized ARCH (GARCH) by Bollerslev (Bollerslev, 1986).

3. FUZZY LINGUISTIC TERM GENERATION AND REPRESENTATION

In the recent years, computational intelligence methods have been widely employed as prediction and estimation approaches (Wang et al, 2009). It has been stated that there are two main problems with state of art forecasting methods: (i) the models become unreliable in the presence of uncertainty and (ii) no indication of the accuracy of the single point forecasts is provided (Khosravi et al, 2014). The accuracy of the forecast is usually measured with performance indexes such as Mean Absolute Percentage Error (MAPE), Percentage of Error (POA), etc. (Hyndmann and Koehler, 2006; Makridakis, 2003; Armstrong, 2001). However, since there is always an error margin in the predictions, there is a need to define error bounds in the forecast with its Confidence Interval (CI), Prediction Interval (PI) or using other novel approaches (Khosravi et al, 2014b).

The main motivation for the construction of error bounds is to quantify the likely uncertainty in the point forecasts. Availability of error bounds allows the decision makers to quantify the level of uncertainty associated with the point forecasts. Error bounds that are relatively wide indicate the presence of high level of uncertainties in the underlying system operation. This useful information can guide the decision makers to avoid the selection of risky actions under uncertain conditions. On the other hand, narrow error bounds give the decision makers the opportunity to decide more confidently with less chance of confronting an unexpected condition in the future (Tsay, 2005).

The construction of PIs has been a subject of much attention and has been implemented in temperature prediction, travel time prediction in baggage handling system, watershed simulation; solder paste deposition process, and time series forecasting (Heskes, 1997). In literature, different methods exist for the construction of PIs as delta technique (Chryssolouris et al, 1996; Hwang and Ding, 1997), Bayesian technique (MacKay, 1992), bootstrap method (Dybowski and Roberts,

2001), mean-variance estimation (Nix and Weigend, 2001), Lower and Upper Bound Estimation (LUBE) method (Khosravi et al, 2011b).

In this study, fuzzy linguistic term generation and representation via Fuzzy logic based Lower and Upper Bound Estimator (FLUBE) based Triangular Fuzzy Number (TFN) is presented to estimate the uncertainty in the forecast. The representation includes two different methodologies that will give the opportunity to the decision maker to quantify the uncertainty of the point forecasts with linguistic terms which might increase the interpretability. Moreover, the proposed approaches will provide valuable information about the accuracy of the forecast by providing a relative membership degree with respect to the target data. The proposed approaches consist of two main phases, the offline FLUBE design and the online TFN generation part as Linguistic Generation and Representation Approach (LinGRA) and Enhanced Linguistic Generation and Representation Approach (ElinGRA).

Thus, the chapter includes methodologies and its basics. The chapter will start with the forecast error evaluation to clarify that what are the approaches mainly model. The first subsection also includes the literature survey of the error modelling via PI. Second subsection gives brief information about the fuzzy logic to use it as tool by the presented methodologies to modelling errors. Third subsection presented Fuzzy Time Series methodology to use for the development of the presented approaches from the beginning. Fourth subsection is also related with the fuzzy logic. It clarifies the tools with its mathematical background. The last subsection presented the approaches as two part as FLUBE and Linguistic Generation and Representation Approaches.

3.1 Forecast Errors Evaluation

Box wrote, "Essentially, all models are wrong, but some are useful" in his book on response surface methodology with Norman R. Draper (Box et al, 2015). In other words, all models include modelling errors. Thus, several assessments presented to evaluate the forecast models via their errors according to original data and forecasted ones. Conventionally forecasting error (e) can be state as:

$$e = f - y \tag{3.1}$$

where f is forecasted value and y is the real data.

In this chapter, the measures of forecast accuracy and forecast error bounds will be discussed. In general definition, forecast accuracy measures provide the evaluation of the forecast model while the error bounds derive the foresight for the future to the decision maker or the system.

3.1.1 Measures of forecast accuracy

In statistics, the accuracy of forecast is the degree of closeness of the statement of quantity to that quantity's actual (true) value. The actual value usually cannot be measured at the time the forecast is made because the statement concerns the future. For most businesses, forecasts with more accuracy increase their effectiveness to serve their aims.

According to Hyndmann and Koehler, there are four type error measures for determining the accuracy of the forecast. Most commonly used two ones are listed as respectively below. In addition to these, relative error based measures and relative measures for forecasting can be found at different sources (Hyndmann and Koehler, 2006; Chatfield 1988; Gardner, 1990).

3.1.1.1 Scale dependent measures

Some commonly used accuracy metrics can be defined as this type of measures like Mean Square Error (MSE) or Mean Absolute Error (MAE) (Thompson, 1990). Scale-dependent measures are useful when comparing different methods on the same set of the same data but should not be used while there are different scales across the data.

The most commonly used scale-dependent measures are based on absolute or squared errors listed as below:

- Mean Square Error (MSE)

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(e_t)^2 \tag{3.2}$$

- Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(e_t)^2} \quad (3.3)$$

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t| \quad (3.4)$$

where n is the sample size, e is the error term**.**

3.1.1.2 Percentage errors based measures

Percentage errors have the advantage of being scale-independent, and frequently used to compare forecast performance across different data sets (Armstrong and Collopy, 1992). The most commonly used measures are listed below:

- Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{e_t}{y_t}\right| \quad (3.5)$$

- Root Mean Square Percentage Error (RMSPE)

$$\text{RMSPE} = \sqrt{\frac{100}{n}\sum_{t=1}^{n}\left(\frac{e_t}{y_t}\right)^2} \quad (3.6)$$

- Percentage of Accuracy (POA)

$$\text{POA} = \sum_{t=1}^{n} f_t \,/\, \sum_{t=1}^{n} y_t \quad (3.7)$$

where n is the sample size, e is the error term, f is the forecast term and y is the real data value.

These measures have the disadvantage of being infinite or undefined if $y_t = 0$ for any t in the period of interest, and having an extremely skewed distribution when any y_t is close to zero. A further disadvantage of methods based on percentage errors is that they assume a meaningful zero. For instance, measures for temperatures on the Fahrenheit or Celsius scales are meaningful.

The MAPE also have the disadvantage that it puts a heavier penalty on positive errors than on negative errors. This observation led to the use of the so-called "symmetric" measures defined by Makridakis (Makridakis, 1993; Goodwin and Lawton, 1999).

- Symmetric Mean Absolute Percentage Error (SMAPE)

$$SMAPE = \frac{200}{n}\sum_{t=1}^{n}\left|\frac{e_t}{y_t+f_t}\right| \tag{3.8}$$

where n is the sample size, e is the error term, y is the real data value, and f is the forecast value.

3.1.2 Forecast error bounds

Forecast accuracy measure can only evaluate the point forecast quality. Since there is always an error margin in the forecasts, there is a need to define error bounds in the forecast with its Confidence Interval (CI), Prediction Interval (PI) or using other novel approaches (Nakagawa and Cuthill, 2007; Hosmer and Lemeshow, 1992; Briggs et al, 1997). A point estimate is a single value given as the estimate of a population parameter that is of interest, for example the mean of some quantity. An interval estimate specifies instead a range within which the parameter is estimated to lie. The main motivation for the construction of the error bounds is to quantify the likely uncertainty in the forecasts. CIs are commonly reported in tables or graphs along with point estimates of the same parameters, to show the reliability of the estimates.

A confidence interval is ideal for quantifying the degree of uncertainty around common parameters of interest such as the center of a sampled population, or its spread. For normally distributed data, the CI for the mean looks like (Ramírez, 2009):

$$x_{n+1} \pm t_{\left(1-\frac{\alpha}{2},n-1\right)}\sigma\sqrt{\frac{1}{n}} \tag{3.9}$$

where x_{n+1} is the sample which is needing to construct the interval, $t_{\left(1-\frac{\alpha}{2},n-1\right)}$ is the t distribution quantile with degrees of freedom equal to the number of observations

minus one $(n-1)$ and its confidence level α, σ is the standart deviation with the sample size n. Similarly, the CI for the standard deviation gives lower and upper bounds for the variation of the standard deviation estimate.

CIs are constructed at a confidence level, such as 95 %, selected by the user. It means that if the same population is sampled on numerous occasions and interval estimates are made on each occasion, the resulting intervals would bracket the true population parameter in approximately 95 % of the cases. A confidence stated at a $1-\alpha$ level can be thought of as the inverse of a significance level, α.

While a CI describes the uncertainty in the prediction of an unknown but fixed value, a PI deals with the uncertainty in the prediction of a future realization of a random variable. The PI is an interval associated with a random variable yet to be observed, with a specified probability of the random variable lying within the interval. It is wider the CI because it takes into account the prediction noise by adding a 1 to the expression inside the square root (Ramírez, 2009):

$$x_{n+1} \pm t_{\left(1-\frac{\alpha}{2},n-1\right)}\sigma\sqrt{1+\frac{1}{n}} \tag{3.10}$$

It is known that the intervals are widely used at point forecasting and its related works. The PIs and its utility helps the decision makers and operational planners to provide uncertainties bounds around the point forecasts. It also provides multiple of solutions/scenarios for the best and worst conditions. Wide PIs are an indication of presence of a high level of uncertainty in the operation of the underlying system. In the circumstances, the decision makers avoid the selection of risky actions under uncertain conditions. Narrow PIs mean that decisions can be decide more confidently with less chance of threatening an unexpected condition in the future (Khosravi et al, 2011a).

The applications of the PIs increased in the last decade because of the increasing complexity of the man-made system. Manufacturing enterprises, industrial plants, transportation systems, and communication networks are the some of the examples of these systems. More complexity adds to high levels of uncertainty in the operation of large systems. Operational planning and scheduling in large systems is often performed based on the point forecasts of the system's future. The reliability of these

point forecasts is low and there is no indication of their accuracy. Therefore, the systems meet the interval bounds.

In literature, several PI construction method for many application is proposed. Delta (Hwang and Ding, 1997; De Veaux et al, 1998), Bayesian (Bishop, 1995; MacKay, 1992), mean-variance estimation (MVE) (Nix and Weigend, 1994), bootstrap (Efron, 1992; Heskes, 1997) and Lower Upper Bound Estimation (LUBE) method (Khosravi, 2011b, Quan et all, 2014) are proposed for construction of the PIs. PIs have been completed in a variety of disciplines, including transportation (Hinsbergen et al, 2009; Khosravi et al, 2011c; Khosravi et al, 2011d), energy market (Zhao et al, 2008; Khosravi et al, 2010), manufacturing (Papadopoulos et al, 2001), and financial services (Benoit et al, 2009). These represented studies are the few of the all studies to give the general information about the PI construction.

In addition to that, the new approach to representing the forecast errors will be clarified in this thesis. Therefore, there will be need of the assessments measurement for error bounds like PI and CI to compare the classical ones and the new one.

3.1.2.1 Quantitave measures for error bound assessments

Khosravi proposes PI assessment methods at the last decade in several studies (Khosravi et al, 2010, Khosravi et al, 2011a). PI and CI can be evaluated as the same for the comparison. Therefore, the error bounds mostly named as PI for this study. The most important characteristic of PIs is their coverage probability. PI coverage probability (PICP) is measured by counting the number of target values covered by the constructed PIs:

$$PICP = \frac{1}{n}\sum_{i=1}^{n} p_i \tag{3.11}$$

where n is the number of the samples and p_i is:

$$p_i = \begin{cases} 1, & if\ y_i \in [L_i\ U_i] \\ 0, & if\ y_i \notin [L_i\ U_i] \end{cases} \tag{3.12}$$

where y_i is the i^{th} real data value, L_i and U_i are the lower and upper value of the i^{th} interval, respectively. According to Khosravi, PICP should be very close or larger than the nominal as ideally (Khosravi et al, 2011c). Otherwise, PIs are invalid and

should be discarded. The ideal case for PICP is the case where PICP = 100%, which means all the targets are covered by PIs.

If the upper and lower bounds of PIs are chosen as extreme values of the targets (maximum and minimum values), a high PICP (even 100% PICP) can be easily achieved. Too wide PIs carry a little information and not useful for decision-making. Width of PIs determines their informativeness. In the literature, a quantitative measure of the width is defined as PI normalized average width (PINAW) (Khosravi et al, 2011e, Quan et al, 2014):

$$PINAW = \frac{100}{nR}\sum_{i=1}^{n}(U_i - L_i) \tag{3.13}$$

where R is the range of the targets in other words universe of discourse of the targets (maximum value of the targets minus minimum value of the targets). R is normalized the PI average width in percentage. The format of PINAW is similar to the MAPE used for point forecasts. It gives equal weights to each width of PIs.

3.2 Basic of Fuzzy Logic

Fuzzy Logic (FL) is a powerful tool for complexity modelling especially on the things that includes uncertainty, vagueness and imprecision while classical methodologies has inadequate models for them. L. A. Zadeh (Zadeh, 1965) introduced the fuzzy logic theory in 1965. In the last 50 years, fuzzy sets and fuzzy logic theory have found a great variety of applications in control engineering, power systems, telecommunication, pattern recognition, machine intelligence, qualitative modeling, motor industry, robotics, and so on. This chapter introduces the basic concepts and components of the Fuzzy Logic Systems (FLSs) that will be needed in the following chapters. In what follows, firstly introduce the basic concepts of fuzzy sets, and membership functions (MFs). Then the typical structure of a fuzzy inference system with its four components are introduced.

3.2.1 Fuzzy sets and membership functions

Fuzzy logic starts with the concept of a fuzzy set (Klir and Yuan, 1995; Mendel, 2001; Zadeh, 1996). A fuzzy set is a set that does not have a crisp boundary. It can

contain elements with only a partial degree of member-ship. The main advantage of fuzzy logic is that words or sentences can be used as expressions instead of numeric values. The inverse of this process is also invalid for the fuzzy sets. Associative expressions are involved our daily life. The velocity example can be selected to express the concept of the fuzzy logic. Velocity can be stated with three linguistic variables: slow, medium, and fast and they can be represented as Figure 3.1. Each linguistic variable is represented by membership function in the universe of discourse.

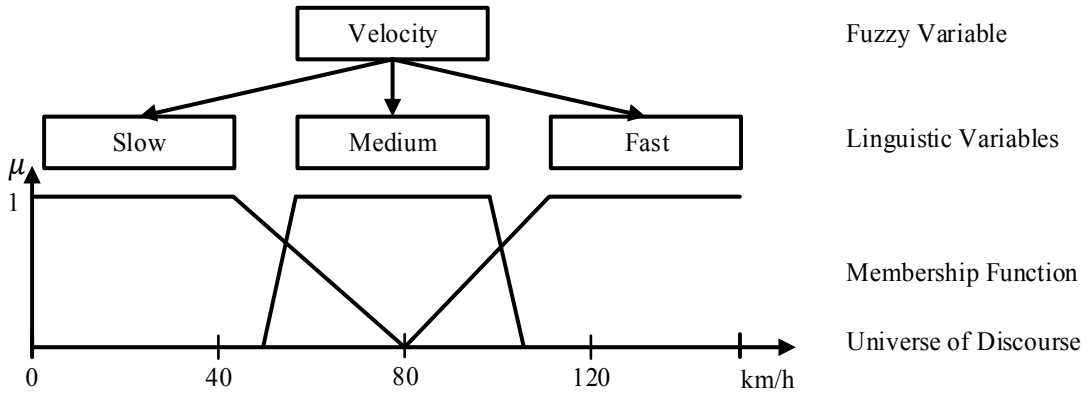


Figure 3.1 : Terms of Fuzzy Logic.

Consider a universe of discourse $\mathcal{U}$, whose elements are denoted as x. A fuzzy set A in $\mathcal{U}$ may be defined as follows:

$$A = \{(x, \mu_A(x)) \mid x \in \mathcal{U}\} \tag{3.14}$$

where $\mu_A(x)$ is the MF of x in A, which represents the degree of the x belongs to A. The MF ($\mu_A(x)$) maps each element in $\mathcal{U}$ to a continuous unit interval $[0, 1]$. If $\mu_A(x) = 0$, it means that x is definitely not an element of fuzzy set A and if $0 < \mu_A(x) < 1$, it indicates that x falls in the fuzzy boundary of fuzzy set A. It is obvious that a fuzzy set is an extension of a classical crisp set by generalizing the range of the characteristic function from the crisp numbers 0, 1 to the unit interval $[0, 1]$.

A MF is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The input space is some-times referred to as the universe of discourse (Hanss, 2005). The only condition a MF must really satisfy is that it must vary between 0 and 1 and the

function itself can be an arbitrary curve. Figure 3.2 depicts a variety of MFs to use in FLSs. The mathematical characterization of the triangular MF and Gaussian MF use extensively in defining MFs in FLSs which are defined as below (Hanss, 2005).

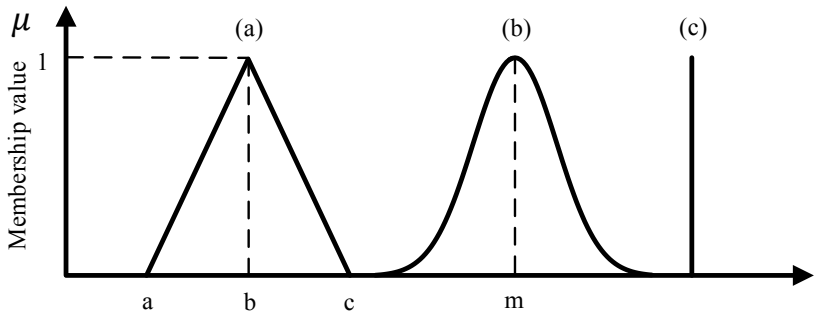


Figure 3.2 : Different shapes of membership functions. (a) triangular, (b) gaussian, (c) singleton.

- Triangular Membership Function: A triangular MF is illustrated at Figure 3.2 which is specified by three parameters a, b, c, which determine the x coordinates of three corners as (Dodurka et al, 2015; Yesil et al, 2014):

$$\mu(x) = \begin{cases} 0, & x \leq 0 \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \\ \dfrac{c-x}{c-b}, & b \leq x \leq c \\ 0, & c \leq x \end{cases} \tag{3.15}$$

where the parameters a and c locate the x coordinates of the feet of the triangle and the parameter b locates the x coordinate of the peak of the triangle.

The triangular membership function also defines with different names like Triangular Fuzzy Number (TFN) or Linear Fuzzy Number (Hanss, 2005). In this work, the TFN will use as extremely.

- Gaussian Membership Function: A Gaussian MF is illustrated at Figure 3.2 which is determined completely by m and σ:

$$\mu(x) = \exp\left[\frac{-(x-c)^2}{2\sigma^2}\right] \tag{3.16}$$

where c represents the MFs center and σ determines the MFs width.

- Fuzzy Singleton: Crisp numbers can be considered as a special case of fuzzy numbers, for they all their properties. The membership function of the fuzzy singleton can be defined as:

$$\mu(x) = \begin{cases} 0, & x < \bar{x} \\ 1, & x = \bar{x} \\ 0, & x > \bar{x} \end{cases} \tag{3.17}$$

where $\bar{x}$ is a crisp number expressed by the membership function.

3.2.2 Fuzzy reasoning

Fuzzy reasoning is performed using if-then rules (Tanaka, 1997; Zadeh, 1978). The fuzzy rules state the connection between input and output fuzzy variables. An antecedent and a consequence part are the two parts of the rule. A classical fuzzy rule can be defined as follow:

$$\begin{gathered} IF\ input1\ is\ big\ AND\ input2\ \ positive\ small \\ THEN\ output\ is\ negative\ small \end{gathered} \tag{3.18}$$

First part is the antecedent part and the second part is the consequence part. *Big, positive small, negative small* are the linguistic variables which are stated as fuzzy sets. AND is a connective operation which aggregates the results within the premise part.

The Fuzzy Inference System (FIS) performs fuzzy reasoning which was first introduced by Zadeh (Zadeh, 1965). A typical structure of a FIS consists of four components: The "fuzzification" block converts the crisp inputs to fuzzy sets, the "rule base" contains a selection of fuzzy rules, the "inference mechanism" uses the fuzzy rules in the rule base to produce fuzzy conclusions and the "defuzzification" block converts the fuzzy conclusions into the crisp outputs. The basic structure of a FLS can be seen in Figure 3.3.

- Fuzzification: The fuzzification block comprises the process of transforming crisp values (e.g. x_0) into grades of membership for linguistic terms of fuzzy sets. The MF use to associate a grade to each linguistic term.

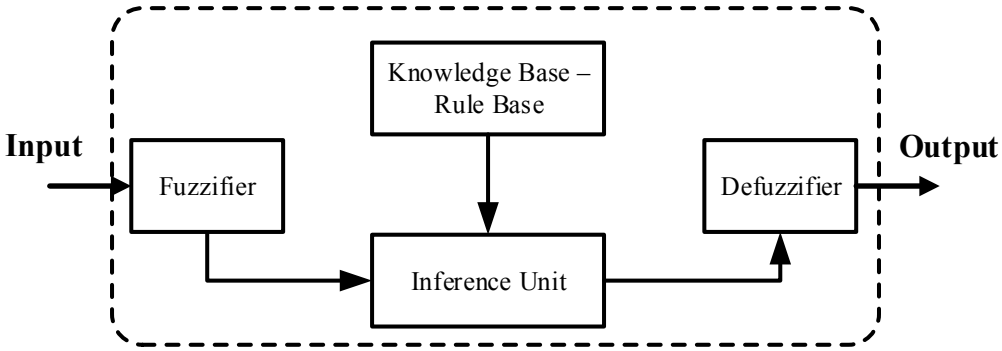


Figure 3.3 : The structure of the fuzzy logic inference system.

- Rule Base: The classical if-then rule is already defined before. The rule structure of FLS with p inputs ($x_1 \in X_1, \ldots, x_p \in X_p$) and one output $y \in Y$ is as follows:.

$$Rule\ l^{th}:\ IF\ x_1\ is\ A_{1l}\ and \ldots and\ x_p\ is\ A_{pl}\ Then\ y\ is\ B \tag{3.19}$$

where $l = 1, \ldots, M$, M is the number of rules and $A_1, \ldots, A_p$ are the values for each input linguistic variables in the universes of discourse. This rule represents a relation between the input space $X_1 \times \cdots \times X_p$, and the output space, Y, of the fuzzy logic system.

- Inference Mechanism: After inputs fuzzified, we know the degree to which each part of the antecedent is satisfied for each rule. If the antecedent of a given rule has more than one part, the fuzzy operator is applied to obtain one number that represents the result of the antecedent for that rule. The input for the implication process is a single number given by the antecedent, and the output is a fuzzy set. Implication is implemented for each rule. Since decisions are based on the testing of all of the rules in a FIS, the rules must be combined in some manner in order to make a decision. Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set.

The operators can be defined in several ways, but the best-established ones are presented as follows (Yager and Zadeh, 2012; Ross, 2009):

Intersection

$$\begin{aligned} \boldsymbol{AND}\ (\mu_A, \mu_B) &= \min\{\mu_A, \mu_B\} \\ \boldsymbol{AND}\ (\mu_A, \mu_B) &= \mu_A \mu_B \end{aligned} \tag{3.20}$$

Union

$$\boldsymbol{OR}\ (\mu_A, \mu_B) = \max\{\mu_A, \mu_B\} \tag{3.21}$$

Complement

$$\boldsymbol{NOT}\ (\mu_A) = 1 - \mu_A \tag{3.22}$$

where μ_A and μ_B are membership values which are combined by operators. The firing angles are calculated according to these rules.

- Defuzzification: Defuzzifier refers to the way a crisp value is extracted from a fuzzy set as a representative value, which is a necessary step as in often case a crisp number is required for real application. In general, there are five methods for defuzzifying a fuzzy set but the center of area (COA) and the mean of maximum (MOM) are the two most commonly used methods in generating the crisp system output. A brief explanation of these defuzzification methods are presented as follow:

COA method produces the center of the fuzzy output area. The crisp system out-put for discrete universe of discourse will be calculated as

$$y = \frac{\sum_{i=0}^{m} x_i \mu_A(x_i)}{\sum_{i=0}^{m} \mu_A(x_i)} \tag{3.23}$$

where m is the number of the discrete elements in the universe of discourse, x_i is the value of the discrete element and $\mu_A(x_i)$ represents corresponding MF value at the x_i.

The MOM defuzzification calculates the mean value of all the m output points. In the case of a discrete universe, the crisp output is expressed as:

$$y = \frac{\sum x_i}{m} \tag{3.24}$$

where x_i is the support value at these points, whose MF reaches the maximum value $\mu_A(x_i)$. The MOM method does not consider the shape of the fuzzy output, however the defuzzification calculation is simplified.

The fuzzy inference process discussed so far is Mamdani's fuzzy inference method (Mamdani, 1974), the most common methodology. The Takagi-Sugeno (TS) Type Fuzzy Inference System was proposed by Takagi and Sugeno in an effort to develop a systematic approach to generating fuzzy rules from a given input-output data set (Takagi and Sugeno, 1985). The first two parts of the fuzzy inference process, fuzzifying the inputs and applying the fuzzy operator, are exactly the same as Mamdani method. The main difference that the TS output membership functions are either linear or constant. A typical rule in a TS fuzzy model has the form as:

$$\begin{aligned} &Rule\ l^{th}:\ IF\ x_1\ is\ A_{1l}\ AND\ \dots AND\ x_p\ is\ A_{pl} \\ &THEN\ y = f_1(x_1, \dots, x_p) \end{aligned} \tag{3.25}$$

where $l = 1, \dots, M$, M is the number of rules, $A_1, \dots, A_p$ are the values for each input linguistic variables in the universes of discourse and $y = f(x_1, \dots, x_p)$ is a predefined function of the input variables for output. A simple expression is the linear and affine functions as:

$$\begin{aligned} &Rule\ l^{th}:\ IF\ x_1\ is\ A_{1l}\ AND\ \dots AND\ x_p\ is\ A_{pl} \\ &THEN\ y = a_{1l}x_1\ \cdots\ + a_{0l} \end{aligned} \tag{3.26}$$

If $a_{0l} = 0$, then the f_l mapping is a linear mapping and if $a_{0l} \neq 0$, then the mapping is called affine. When f_l is constant, then the mapping is called "zero-order TS", which can be viewed as a special case of the Mamdani inference system, in which each rule's consequent is specified by a fuzzy singleton.

3.3 Fuzzy Time Series

Fuzzy Time Series (FTS) is one of the unique research fields, and it contributes to the conventional time series analysis by data clustering and rule-based features. The FTS can be stated as the union of the fuzzy sets according to their time properties. Basically, it used for the modelling uncertainties. The illustration and more detail about the fuzzy time series also clarified by Möller and Reuter. (Möller and Reuter, 2007). The classical illustration of the FTS can be seen at Figure 3.4 on the Air Passenger data set as 2D. 3D illustration of the FTS via Triangular Fuzzy Numbers can be seen at Figure 3.5.

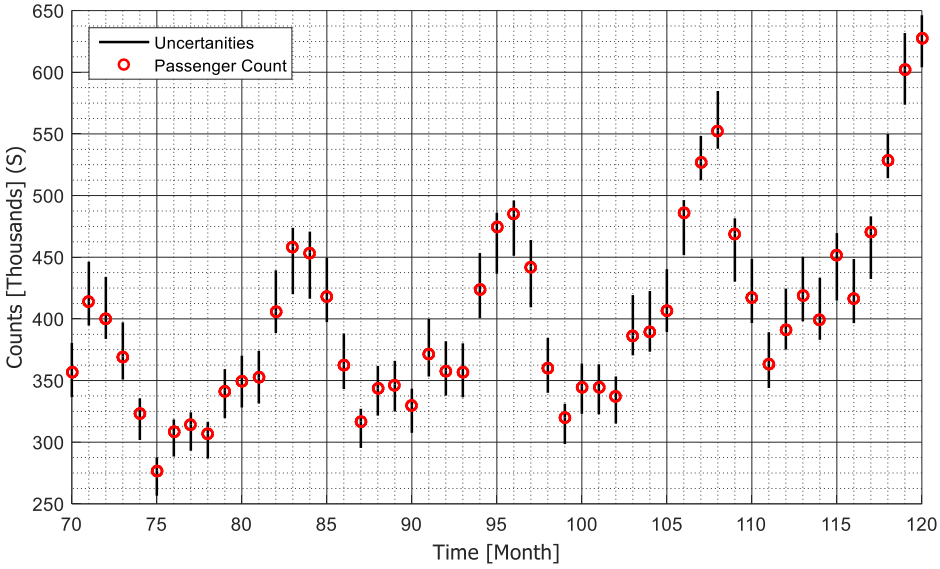


Figure 3.4 : The illustration of the Fuzzy Time Series in 2 Dimension.

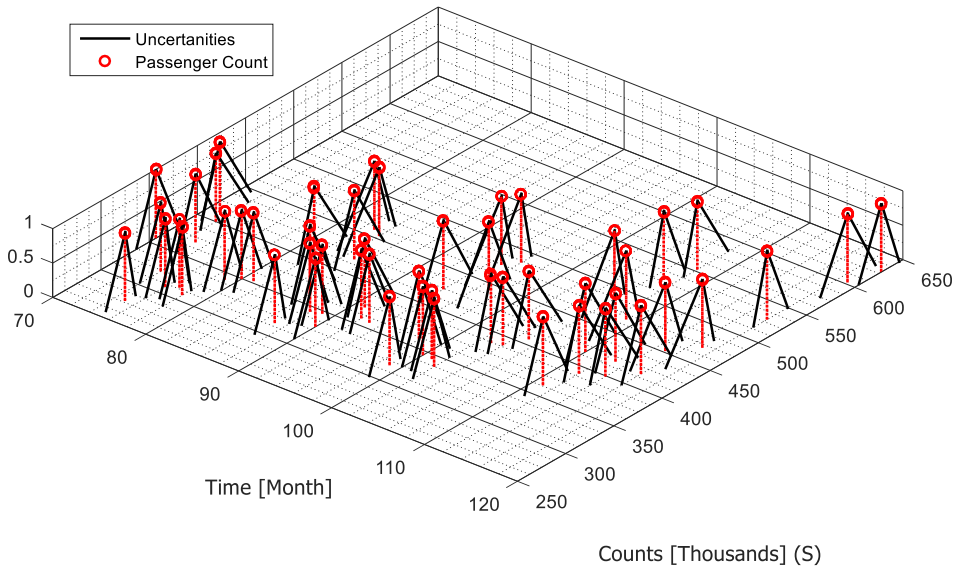


Figure 3.5 : The illustration of the Fuzzy Time Series in 3 Dimension.

In respect of forecasting activities in daily life, it can be said that it has an important role, obviously. Traditional forecasting methods cannot deal with the problems where they are mainly related with the linguistic terms as the historical data sets. With the aim of to solve the drawback of the traditional forecasting methods, Song and Chissom first defined the FTS forecasting method to predict university enrollment levels in 1993 (Song and Chissom, 1993; Song and Chissom, 1994). However, this model used max–min composition operations that is significantly

complicate the computational process. It was improved by Chen which is generally accepted by researchers and is the common form of FTS (Chen, 1996).

As mentioned precedings declaration, FTS and its modelling approaches are based on rule-based structures. Therefore, there is need to declare the relations between the observations. First, FTS samples states with their intervals in other name its universe of discourse. If the historical time series data set is distributed among n intervals (i.e $a_1, a_2, \dots, a_n$) then n linguistic variables $\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_n$ which can be represented by fuzzy sets , as shown below (Chen, 1996):

$$\begin{aligned} \tilde{A}_1 &= 1/a_1 + 0.5/a_2 + 0/a_3 + \cdots + 0/a_{n-1} + 0/a_n \\ \tilde{A}_2 &= 0.5/a_1 + 1/a_2 + 0.5/a_3 + \cdots + 0/a_{n-1} + 0/a_n \\ &\dots \\ \tilde{A}_n &= 0/a_1 + 0/a_2 + 0/a_3 + \cdots + 0.5/a_{n-1} + 1/a_n \end{aligned} \tag{3.27}$$

The degree of membership of each time series value belonging to each $\tilde{A}_i$ are obtain after the determination of the interval of the intervals. The maximum degree of membership of fuzzy set occurs at interval a_i, and $1 \leq i \leq n$. Then, each historical time series value is fuzzified. For ease of computation, the degree of membership values of $\tilde{A}_j (j = 1,2, \dots, n)$ are considered as either $0, 0.5$ or 1, and $1 \leq j \leq n$ (Chen and Hwang, 2000; Huarng, 2001)

If $F(t)$ consists of linguistic variables $f_i(t)(i = 1,2, \dots)$, $F(t)$ is defined as a fuzzy time series on $Y(t)(t = \cdots, 0,1,2, \dots)$ where $Y(t)$ is a subset of real line of the observations with time parameter t. The relationship between $F(t)$ and $F(t-1)$ can be symbolized by:

$$F(t-1) \rightarrow F(t) \tag{3.28}$$

The relation can also define with relation matrix:

$$F(t) = F(t-1) o R(t, t-1) \tag{3.29}$$

where the symbol o is the Max-Min composition operator and $R(t, t-1)$ can be define as a fuzzy relationship if $F(t)$ is caused by $F(t-1)$ (Chen and Tanuwijaya, 2002).

The relation of the observations can be extend with group and multivariate version of the FTS. More details can be found at different sources (Sahin et al, 2015; Singh, 2014). FTS modelling is the one of the inspiration of the thesis with its uncertainty modelling techniques and its rule-based approaches. Therefore, using the time series time series modelling with fuzzy modelling is decided after the review study about the FTS approaches (Sahin et al, 2015).

3.4 Fuzzy Modelling

Fuzzy Logic is the powerful tool for modelling the uncertainties and nonlinearities where will be several cases like noises at signal processing, error distributions forecasting processes, parameter uncertainties at several estimation processes, and nonlinearities at several cases like bioprocess. Fuzzy modelling is the solution which use the Fuzzy Logic to overcome the uncertainties and nonlinearities (Pedrycz, 2012; Yager and Filev, 1994). In preceding chapter, the Fuzzy Inference System (FIS) is already discussed which is the basis of the fuzzy modelling. Within the context of the thesis, two fuzzy modelling approaches will use for modelling the forecast error uncertainties: the Interval Fuzzy Modelling (INFUMO) and Adaptive Network based Fuzzy Inference System (ANFIS).

3.4.1 Interval fuzzy modelling (INFUMO)

The interval fuzzy model was introduced by Skrjanc as a means of robust system identification, and it was used in the work to provide a functional description of a static nonlinear area approximation (Skrjanc and Blazic, 2005; Oblak and Blazic, 2006; Oblak et al, 2007).

The modelling of the upper and lower bound of the observations ($\overline{f}(x,t)$ and $\underline{f}(x,t)$) will be carried out using an interval fuzzy model. The procedure of obtaining the exact lower and upper bounds for the family of functions and assigning the INFUMO to them will be briefly reviewed. Let $C \subset \mathbb{R}$ be a compact set and $\mathcal{G} = \{g(y): C \rightarrow \mathbb{R}\}$ be a class of nonlinear functions. Let us assume that there exist the exact upper bound $\overline{g}$ and the exact lower bound $\underline{g}$ that satisfy the following conditions for an arbitrary $\varepsilon > 0$ and for each observation $y \in C$:

$$\begin{aligned} \overline{g}(y) &\geq \max_{g \in \mathcal{G}} g(y), \qquad \exists g \in \mathcal{G}: \overline{g}(y) < g(y) + \varepsilon \\ \underline{g}(y) &\leq \max_{g \in \mathcal{G}} g(y), \qquad \exists g \in \mathcal{G}: \underline{g}(y) > g(y) + \varepsilon \end{aligned} \tag{3.30}$$

The exact upper and lower boundary functions will be approximated by fuzzy functions, and the Stone–Weierstrass theorem is extended to prove that there exist fuzzy systems $\overline{f}$ and $\underline{f}$ such that the bounds of an arbitrary nonlinear area can be approximated with any precision, i.e.,

$$\begin{aligned} 0 < \overline{f}(y_i) - g(y_i) < \varepsilon \quad \forall i \\ -\varepsilon < \underline{f}(y_i) - g(y_i) < 0 \quad \forall i \end{aligned} \tag{3.31}$$

The INFUMO is the Tagaki Sugeno (TS) type fuzzy model with one antecedent variable and it can be defining as a set of rules:

$$\begin{aligned} R_j : IF\ x_p\ is\ A_j\ THEN\ \overline{f}(y_i) = \overline{\theta}_j^T x_c, j = 1, \dots, m \\ \underline{f}(y_i) = \underline{\theta}_j^T x_c, j = 1, \dots, m \end{aligned} \tag{3.32}$$

The antecedent variable $x_p \in \mathbb{R}$ denotes the input or variable in premise, and variables $\overline{f}, \underline{f} \in \mathbb{R}$ are the outputs of the interval fuzzy model that provide the upper and lower boundary function. The antecedent variable is connected with m fuzzy sets A_j, and each fuzzy set $A_j (j = 1, \dots, m)$ is associated with a real-valued function $\mu_{A_j}(x_p): \mathbb{R} \to [0, 1]$, that produces a membership grade of the variable x_p with respect to the fuzzy set A_j. The consequent vector is denoted $x_c^T = [u_f\ 1] \in \mathbb{R}^2$. As the output functions are in affine form, 1 was added to the vector x_c, The system outputs are linear combinations of the consequent states, and $\overline{\theta}, \underline{\theta} \in \mathbb{R}^2$ are vectors of fuzzy parameters. The system can be described in closed form

$$\begin{aligned} \overline{y} &= \beta^T(x_p)\overline{\Theta}x_c \\ \underline{y} &= \beta^T(x_p)\underline{\Theta}x_c, \end{aligned} \tag{3.33}$$

where $\overline{\Theta}^T = [\overline{\Theta}_1, \dots, \overline{\Theta}_m]$ and $\underline{\Theta}^T = [\underline{\Theta}_1, \dots, \underline{\Theta}_m]$ denote the upper and lower coefficient matrices for the complete set of rules, and $\beta^T(x_p) =$

$[\beta_1(x_p), \dots, \beta_m(x_p)]$ is a vector of normalized membership functions with elements that indicate the degree of fulfilment of the respective rule. Functions $\beta_j(x_p)$ can be defined as

$$\beta_j(x_p) = \frac{\mu_{A_j}(x_p)}{\sum_{j=1}^{m} \mu_{A_j}(x_p)}, \quad j = 1, \dots, m \tag{3.34}$$

if the partition of unity is assumed. Note that $\beta_j(x_p)$ and x_c are equal in both the upper and lower boundary-function calculations at before.

Illustration of the presented fuzzy modelling can be seen at Figure 3.6 according to Skrjanc and Blazic for nonlinear input-output signals with Interval Fuzzy Model (INFUMO) (Skrjanc and Blazic, 2005).

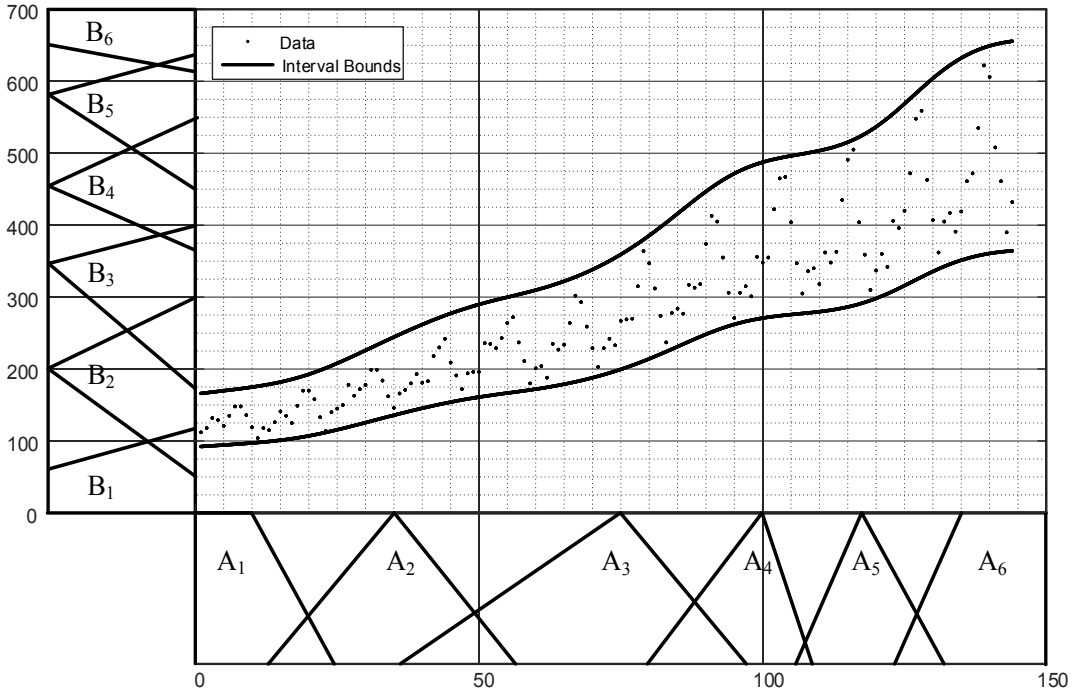


Figure 3.6 : Nonlinear static curve and membership functions of INFUMO.

The membership functions $\mu_{A_j}(x_p)$ and the fuzzy parameters in $\overline{\Theta}, \underline{\Theta}$ of the INFUMO have to be determined for optimal approximation of the bounds. While the proper arrangement and shape of the membership functions will not be considered-the mentioned can be approached by different clustering methods for the optimal

fuzzy parameters the minimization of the maximum modelling error ε over the completely input set X is performed.

According to Oblak, the minimization of the modelling errors have done with the implementation of the min–max optimization method, and l_∞-norm. In the context of the forecasting error modelling which will presented at the next chapter about the proposed approach, the membership functions and fuzzy parameters of the nonlineer bound are defined by MATLAB Adaptive Network based Fuzzy Inference System (ANFIS) Toolbox with selected boundary points as training data set. Therefore, the ANFIS structure will be discussed at the next chapter.

3.4.2 Adaptive Network based Fuzzy Inference System (ANFIS)

Fuzzy Logic (FL) and Neural Network (NN) or Artificial Neural Networks (ANN) are capable of learning nonlinear mapping of numerical data. However, the operation of NN have also many weaknesses. The classical most popular form of the NN, Multilayer Perceptron (MLP) has some complication about the uncertainties. Fuzzy Logic uses human understandable linguistic terms to express the knowledge of the system.

The origin of the ANFIS is to incorporate neural concepts, such as learning and parallelism, into fuzzy logic systems (Jang, 1993; Jang et al, 1997). ANFIS can be implemented as a five layer neural network as illustrated at Figure 3.7.

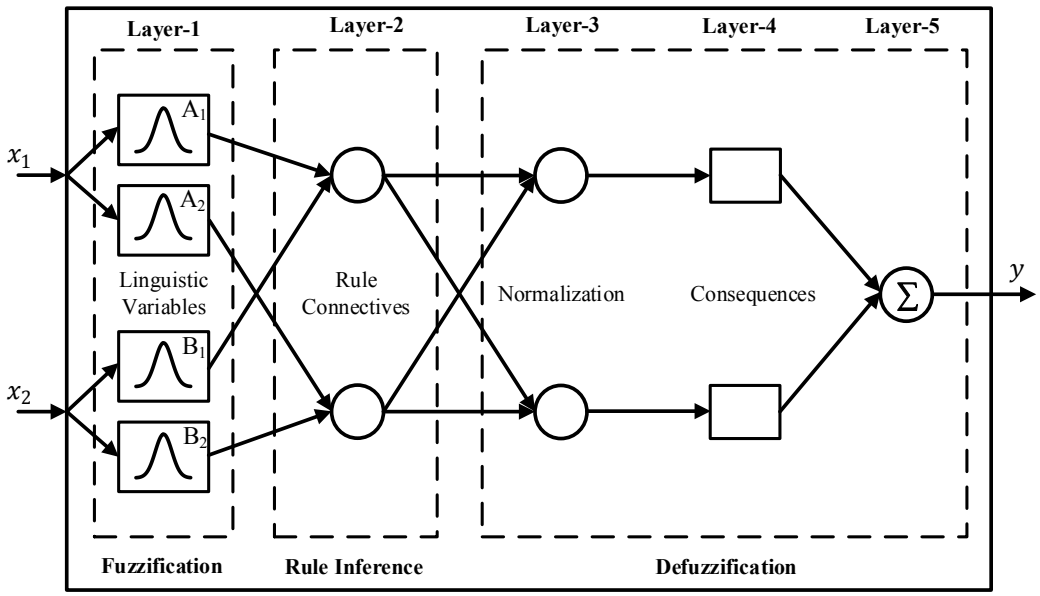


Figure 3.7 : Illustration of the Adaptive Network based Fuzzy Inference System.

The illustrated network is the most common among network based fuzzy inference systems. It uses Takagi Sugeno type fuzzy rule defined at the preceding chapter. The operation of the ANFIS can be described layer by layer as follows:

- **Layer-1:** Fuzzification

The linguistic variables which can also be named as fuzzy sets are found at this layer. The crisp inputs are fuzzified by using membership functions. The fuzzification process is clarified at the FIS chapter.

- **Layer-2:** Rule Nodes

The layer includes one node for each fuzzy if-then rule. Each rule node performs connective operation between rule antecedents. The minimum or the dot product is used as intersection AND as usually. The union OR is done using maximum operation as usually.

- **Layer-3:** Normalization

The fuzzy rules are normalized according to:

$$\mu^* = \frac{\mu_r}{\sum_{i=0}^{m} \mu_i} \tag{3.35}$$

- **Layer-4:** Consequence Layer

According to T-S type, the values of the singletons are multiplied by normalized firing strengths:

$$y_r = \sum_{i=0}^{m} z_i \mu^* \tag{3.36}$$

- **Layer-5:** Summation

It determines the overall output:

$$y = \sum_{i=0}^{m} y_r \tag{3.37}$$

The parameters and structure identification of the ANFIS is the main concern of the modelling via the ANFIS. Suitable number of fuzzy rules and a proper partitioning of

the input and output space related with the structure identification. Most of times it depends on the user of the ANFIS.

The backpropagation is probably the most popular neural network learning technique. It can be used for the adjustment of the system parameters, such as membership functions and other parameters (Sun, 1994; Takagi et al, 1992). It is an application of the gradient descent algorithm originally for MLP.

Jang combines the LS method and gradient descent methods for tuning the ANFIS as a hybrid approach. It is initialized with a number of membership functions and a fixed rule base. The learning process can be done with two part: the forward pass and the backward pass. Forward pass provides the determination of the consequent parameters via the LS method. The antecedents are updated at backward pass via the gradient descent method.

In the context of the thesis, the ANFIS is used at the modelling of the forecasting errors via MATLAB Fuzzy Logic toolbox. The usage process of the ANFIS will be represented at the next chapters.

3.5 Forecast Representation via Triangular Fuzzy Numbers

In recent years, the research works on point forecast and prediction approaches using computational intelligence methods have been widely increased. It has been reported by Wang that there are two main problems with forecasting methods presented in literature: (i) the models become unreliable in the presence of uncertainty and (ii) no indication of accuracy of point forecasts is provided (Wang et al, 2009). The accuracy of the forecast is measured with performance indexes such as Mean Absolute Percentage Error (MAPE), Percentage of Error (POA), etc. (Khosravi et al, 2014c). However, since there is always an error margin in predictions, there is a need to define error bounds in forecast like Confidence Interval (CI) and Prediction Interval (PI).

The main motivation for the construction of PIs is to quantify the likely uncertainty in the point forecasts. Availability of PIs allows the decision makers to quantify the level of uncertainty associated with the point forecasts. PIs that are relatively wide indicate the presence of high level of uncertainties in the underlying system operation. This useful information can guide the decision makers to avoid the

selection of risky actions under uncertain conditions. On the other hand, narrow PIs give the decision makers the opportunity to decide more confidently with less chance of confronting an unexpected condition in the future.

Fuzzy logic based Lower and Upper Bound Estimator (FLUBE) provides the intervals from the forecast errors which are generated according to the forecast models without any dependency of the specific one. Thus, instead of a classical statistical approaches, the level of uncertainty associated with the point forecasts will be defined within the FLUBE bounds and these bound can be used for defining fuzzy linguistic terms for the forecasts. Then, the FLUBE is used for the Linguistic Term Generation and Representation Approach (LinGRA) so that the forecasts are represented with the linguistic terms, which are defined with Triangular Fuzzy Numbers (TFNs). In the LinGRA, the support of the TFN is constructed on the output interval generated by the FLUBE while its center is directly assigned to the forecast value.

Enhanced LinGRA (ElinGRA) is enhanced version of the FLUBE based LinGRA which use an extra FLS to generate a Center Point Correction Term (CPCT) for the TFN. In comparison with the LinGRA, ElinGRA will increase the information about the accuracy and success of the single point forecast by providing a relative membership degree (μ).

In the context of the forecast representation, both methodologies will give the opportunity to the decision maker to quantify the uncertainty of the point forecasts with the linguistic terms that might increase the interpretability for further assessments. The proposed approaches consist of two main parts, the FLUBE design and the linguistic forecast generation with the ElinGRA and LinGRA via TFNs. Thus, the chapter will start by presenting the internal structure and the design steps of the FLUBE. Then the LinGRA and the ElinGRA will present in comparison with each other. To illustrate the methodology steps, the Australian monthly electricity consumption data set is used as example data.

3.5.1 Fuzzy Logic based Lower and Upper Bound Estimator (FLUBE)

The FLUBE consists of two Fuzzy Logic Systems (FLSs) which will define the uncertainty bounds of the point forecast error. It tries to determine two bounds for the determination of the interval as nonsymmetrical conversely PI bounds. By the

looking of that point, the FLUBE approach intercept with the Interval Fuzzy Modelling (INFUMO) (Skrjanc and Blazic, 2005). According to error modelling idea, the FLUBE is constructed by choosing the input to be the target data (x) and the output as the forecast (f) error terms $(e = f - x)$. It will be known that the decomposition and modelling approaches of the time series are realized by MATLAB as automatically.

The required data sets are chosen by the Selection Algorithm (SA) to inhibit the general approximation and the creating a bound around the mean value of the forecasting error given at Table 3.1.

Table 3.1 : Selection Algorithm.

1: Choose design parameter P.
2: Divide the universe of discourse of the target data (x) for P subpiece with its ranges.
3: for each sub piece
4: for each member of the sub piece
5: if $e < 0$
6: select minimum error term in response of the sub piece target data range.
7: save the points for training data set of the LFLS $[x_{min}, e_{min}]$
8: if $e > 0$
9: select with the same procedure for the maximum one.
10: save the points for training data set of the UFLS $[x_{max}, e_{max}]$.

After the determination of the training data sets via the SA, FLUBE has two unique data sets which are symbolized as $[x_{min}, e_{min}]$ and $[x_{max}, e_{max}]$ for the next notations. $[x_{min}, e_{min}]$ for the training of the Lower FLS (LFLS) and $[x_{max}, e_{max}]$ for the training of the Upper FLS (UFLS). The ANFIS toolbox/MATLAB will be used to generate the LFLS and UFLS for the training of the FLUBE in the context of the all study. The fuzzy rule base structures of the LFLS and UFLS are as follows:

LFLS: $$R_{\omega}^{L}: IF\ x\ is\ A_{\omega}^{L}, THEN\ e^{L}\ is\ D_{\omega}^{L} \tag{3.38}$$

UFLS: $$R_{\omega}^{U}: \text{IF } x \text{ is } A_{\omega}^{U}, \text{THEN } e^{U} \text{ is } D_{\omega}^{U} \tag{3.39}$$

where A_ω^L and A_ω^U, are the antecedent membership functions (MFs), D_ω^L and D_ω^U are the consequent crisp sets and W $(\omega = 1, \dots W)$ is the total number of rules. e^L and e^U represent the lower and upper error forecast values.

In Figure 3.8, the flow chart of the training procedure of the FLUBE is presented which consists of two main steps. The first step is to construct a conventional forecast model while in the second step the FLUBE is designed. The presented analysis will apply on Australian monthly electricity consumption data set for illustrative purposes.

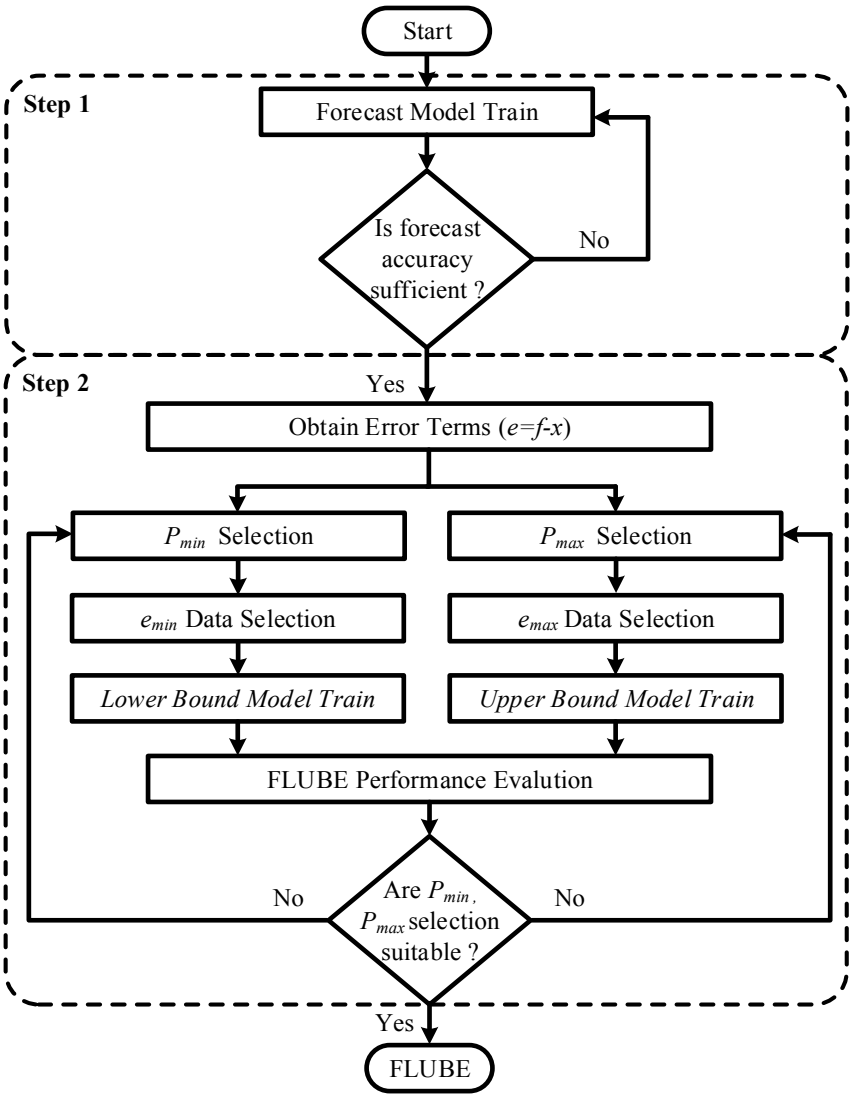


Figure 3.8 : The flow chart of the training procedure of the FLUBE.

- **Step-1:** Forecast phase

Here, the time series forecast model is constructed, by using the target data (x) to predict the forecast value (f). The forecast model can be constructed from different

structures such as time series regression, conditional means/variance and multivariate models (Box and Jenkins, 2015). In general, the forecast model should provide a satisfactory performance to result with an acceptable FLUBE training. In this study, a SARIMA forecast model would prefer since it results with an acceptable forecast performance for the handled electrical consumption data set. Note that the FLUBE can be also easily employed if other forecast models are preferred.

- **Step-2:** FLUBE training

In this step, we will collect the required training data sets which are defined as follows:

$$e_k = f_k - x_k \tag{3.40}$$

where f_k is the forecast value at the k^{th} sample. In Figure 2.5, the training data set is illustrated for the electricity consumption data set. Then, the SA is employed to collect the required the data sets which will be used at the training of FLUBE. The data set for LFLS will be constructed with negative error terms and will be labeled as $[x_{min}, e_{min}]$ which are illustrate with red circles in Figure 3.9. In a similar manner, the dataset for the UFLS will be labeled as $[x_{max}, e_{max}]$ which are illustrate with blue circles in Figure 3.9.

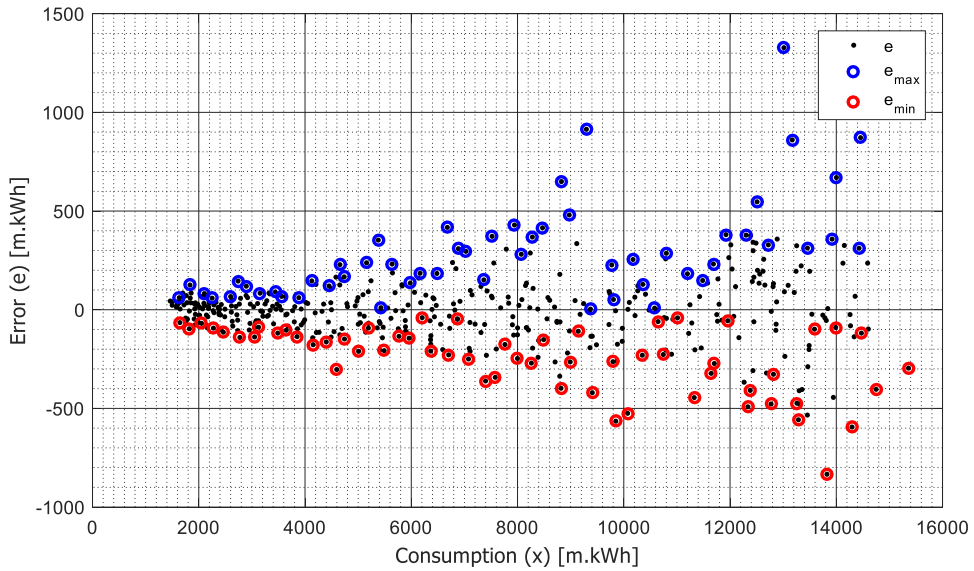


Figure 3.9 : Selected error terms by Selection Algorithm according to target data of the Australian monthly electricity consumption data set.

As it has been asserted, these two data sets will be chosen by the SA with a tuning parameter P. The design parameter (P) of the SA decomposes the minimum and maximum target data with equal and fixed subintervals. The SA picks the maximum/minimum error values for each decomposed subinterval. Thus, it can be concluded that the parameter P defines the maximum size of the training data set.

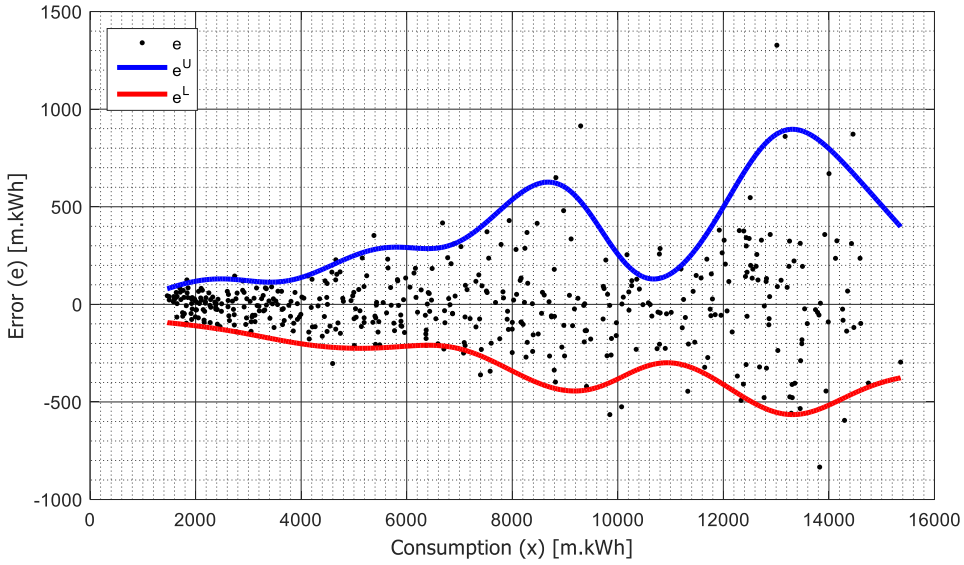
For the collection of training data set of the UFLS $[x_{max}\ e_{max}]$, the SA will perform as follows:

- If $e_k > 0$, then the SA will select and insert the corresponding data into training data set of the UFLS $[x_{max},\ e_{max}]$.
- If $e_k < 0$, then the SA will not update the training data set for this subinterval.

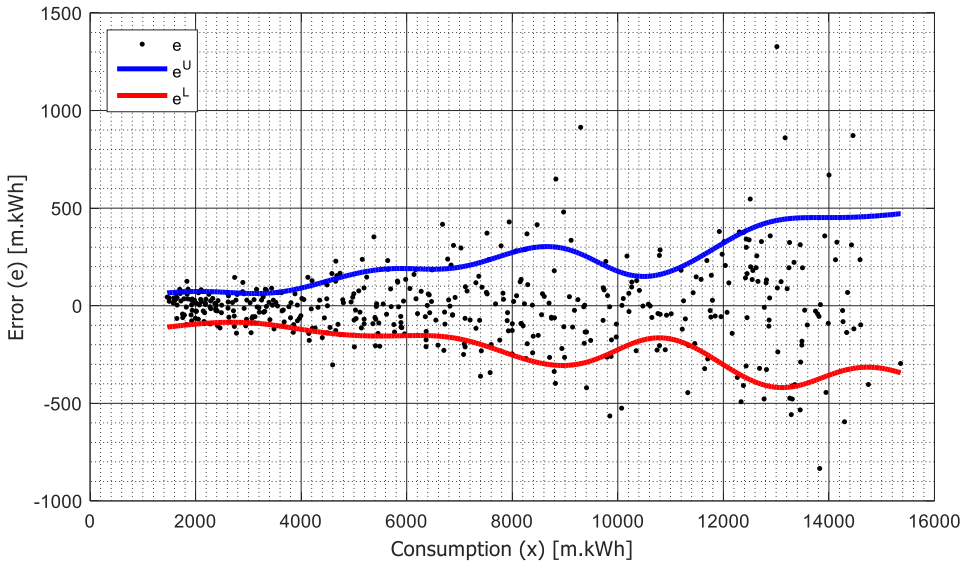
In a similar manner, the $[x_{min}, e_{min}]$ data set will be collected.

It can be observed that selection of the SA parameter *P* affects size of the train data sets. The SA will collect more training data points for relatively big *P* values. Thus, the generated FLUBE might result with a general approximation while neglecting the characteristics of the error values. Therefore, *P* is an important design parameter and needs to be tuned for each handled data set. To illustrate this concept, for the Data Set-1, the performance of the FLUBE is illustrated in Figure 3.10a for the *P* values $P = 60$ and $P = 320$. (Note that, we have employed $P = P_{min} = P_{max}$ throughout the paper). It can be clearly seen that the relatively small value of P $(P = 60)$ provides bounds which cover almost all error terms while the relatively big value of P $(P = 320)$ results a smoother bound characteristic as van be seen at Figure 3.10b. Hence, it can be concluded that the *P* parameter must be selected such that to provide a tradeoff between coverage performance and bound characteristic.

Remark: The mean value of the error distribution must be shifted from the error terms (e_{min}, e_{max}) to make the expected value of the error as zero. Thus, the shifted mean value, at the end, has to be added to the output of the FLUBE.



(a)



(b)

Figure 3.10 : The selection parameter effect on the FLUBE bounds determination on the Data Set-1 for a) *P*=60, b) *P*=320.

3.5.2 Linguistic Forecast Generation

Linguistic forecast generation is a data analysis approach based on the forecast error terms. The main purpose of the method helping the last decisions and the decision makers. Classical error bounds like PIs are only interested the bounds convergence performance Therefore, they cannot help the decisions makers when the decision makers need to evaluate the forecasting method because the evaluation measurement like PICP cannot be helpful with bound-in or out counts. Linguistic forecast

generation become a part of the evaluation at this stage while providing the evaluation of the forecasting methods with TFNs' membership degrees.

In this subsection, the ElinGRA, which is the improved version of the online LinGRA and ElinGRA will be presented. In this context, two novel approaches that try to explain how the target values (x) are similar to the forecast values (f) by using linguistic terms are presented which cannot be accomplished by the conventional PI representations. As it has been asserted, the FLUBE has been constructed where the target data (x_k) is the input. However, since the x_k value will not be available for evaluating FLUBE at the k^{th} sample, the single point forecast value f_k will used as the input of the FLUBE by assuming the expected value (mean) of the error terms is zero. This will give the opportunity to the FLUBE to generate the e_k^L and e_k^U. Thus, firstly a brief overview of LinGRA and then the ElinGRA will be give.

3.5.2.1 Linguistic Generation and Representation Approach (LinGRA)

In this subsection, the novel online TFN representation and generation of the uncertainty in the forecast will be explained. This presentation tries to explain how the target values are similar to the forecast values by using linguistic terms, instead of the conventional PI representations in literature. Here, at each sample, the forecast uncertainty of the next sample will be approximately defined by TFN. Consequently, the TFN generation will be accomplished in an online manner as shown in Figure 3.11a.

In this context, we will prefer and employ TFNs which are represented with triplet (L, C, U) (Yeşil et all, 2014) which can be seen at Figure 3.11b. Thus, at each sample k, the TFN parameters will be assigned in an online manner as follows:

$$\begin{aligned} L_k &= f_k + e_k^L \\ C_k &= f_k \\ U_k &= f_k + e_k^U \end{aligned} \tag{3.41}$$

where e_k^L and e_k^U are the lower and upper error forecasts generated from the FLUBE. Thus, the decision maker has the opportunity to evaluate the success of the forecast by using linguistic terms. This can be accomplished by calculating the membership degree of the TFN. Thus, for a crisp target value (x'), the success of the forecast can be defined as follows:

$$\mu_k^{TFN}(x') = \begin{cases} 0, & \text{if } x' \notin [L_k\ U_k] \\ \dfrac{x' - L_k}{|e_k^L|}, & \text{if } x' \in [L_k\ C_k] \\ \dfrac{U_k - x'}{|e_k^U|}, & \text{if } x' \in [C_k\ U_k] \end{cases} \tag{3.42}$$

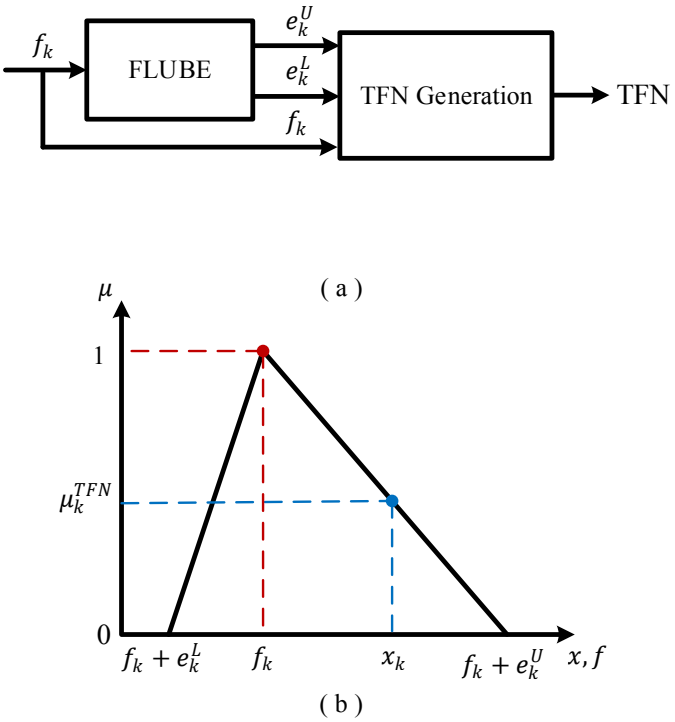


Figure 3.11 : Illustration of (a) the TFN generation method and (b) generated TFN according to LinGRA

The membership degree will represent the success of the forecast that is represented by TFNs. Thus, for a μ value of the TFNs that is close to 1, the success of the forecast will be relatively high.

3.5.2.2 Enhanced Linguistic Generation and Representation Approach (ElinGRA)

As it has been asserted in LinGRA, the support of the TFN will be generated by using the outputs of the FLUBE. Furthermore, the center of the TFN is determined by the forecast value, which is generated by forecast model, since the forecast value is expected to be equal to the Target value at k^{th} sample. However, the error of the forecast is not always equal to zero since forecasting model cannot overcome with uncertainness and nonlinearities in the data to be forecasted. Likewise, this fact is not considered in the LinGRA, so a correcting mechanism for the center of the TFN

is needed. To overcome the above mentioned problem, the proposed ElinGRA inherits an extra fuzzy logic system called CPCT-FLS (Center Point Correction Term - FLS) to define the center of the TFN with a correction term (e^C) as shown in Figure 3.12a.

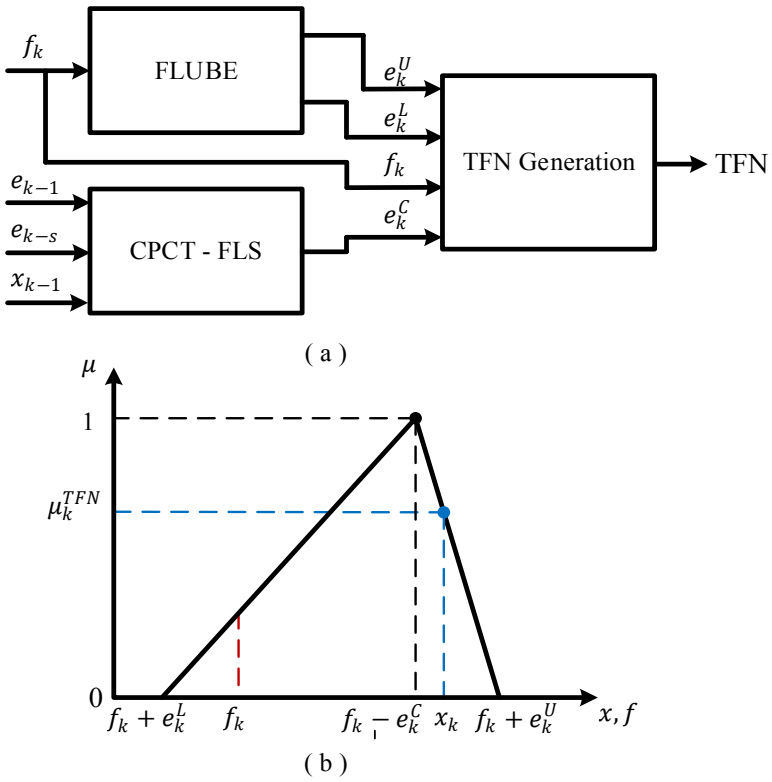


Figure 3.12 : Illustration of (a) the TFN generation method and (b) generated TFN according to ElinGRA

The CPCT-FLS is constructed by choosing the inputs to be target data (x_{k-1}) and the error term (e_{k-1}) at the $(k-1)^{th}$ sample and the error term e_{k-s} which is defined with respect to the seasonality of the target data. (e_k^C), which is the output of the CPCT-FLS, will be the extra input to the TFN generation block as shown in Figure 3.12a. The fuzzy rule base of the CPCT-FLS is as follows:

$$R_\omega^C: \text{IF } x_{k-1} \text{ is } A_\omega^C, e_{k-1} \text{ is } B_\omega^C, e_{k-s} \text{ is } C_\omega^C \text{ THEN } e^C \text{ is } D_\omega^C \tag{3.43}$$

where A_ω^C, B_ω^C and C_ω^C, are the antecedent MFs, D_ω^C is the consequent singleton and W $(\omega = 1, \dots W)$ is the total number of rules. The CPCT-FLS will be designed via the ANFIS toolbox/MATLAB in an offline manner.

The extra information about the forecast error will be used to evaluate the linguistic term to represent the forthcoming target value. In this context, we will use the output of the Extra FLS e^C to redefine the center of the TFN (C_k) while keeping the L_k and U_k as given in Equation 3.41 in order to provide an identical PINAW and PICP values of the LinGRA. Thus, the triplet of the TFN will be defined as follows:

$$C_k = \begin{cases} L_k = f_k + e_k^L \\ f_k + e_k^L > f_k - e_k^C, & f_k + e_k^L \\ f_k + e_k^U < f_k - e_k^C, & f_k + e_k^U \\ else, & f_k - e_k^C \\ U_k = f_k + e_k^U \end{cases} \tag{3.44}$$

where e_k^C is the error forecast value at the k^{th} sample generated from the extra FLS. Note that, since the CPCT-FLS trained with all error terms distinctively from other FLS ($e_{max} > 0, e_{min} < 0$), the signal of the error forecast terms are different from each other at Equation 3.44.

In a similar manner, the success of the forecasts will be evaluated with the membership degrees for LinGRA. Thus, for a crisp target value (x'), the success of the forecast can be defined as follows:

$$\mu_k^{TFN}(x') = \begin{cases} 0, & \text{if } x' \notin [L_k\ U_k] \\ \dfrac{x' - L_k}{|e_k^L|}, & \text{if } x' \in [L_k\ C_k] \\ \dfrac{U_k - x'}{|e_k^U|}, & \text{if } x' \in [C_k\ U_k] \end{cases} \tag{3.45}$$

It can be observed that at each sample (k) the interval of the forecast uncertainty (i.e. the support of the TFN) varies with respect to the current forecast value and thus nonsymmetrical TFNs will are generated for both the LinGRA and ElinGRA. However, since their center definition is not identical, both approaches resulted with unique TFN representation for the same forecast uncertainty. In order to clearly illustrate this concept, we have presented the generated TFNs of the LinGRA and ElinGRA for certain samples.

To compare two approaches, two samples are illustrated for two approaches that are obtained from different data sets at Figure 3.13. First one generated for the 357^{th}

sample on the Australian monthly electricity consumption data set and second one generated for the 120^{th} sample on the airlines passenger data set.

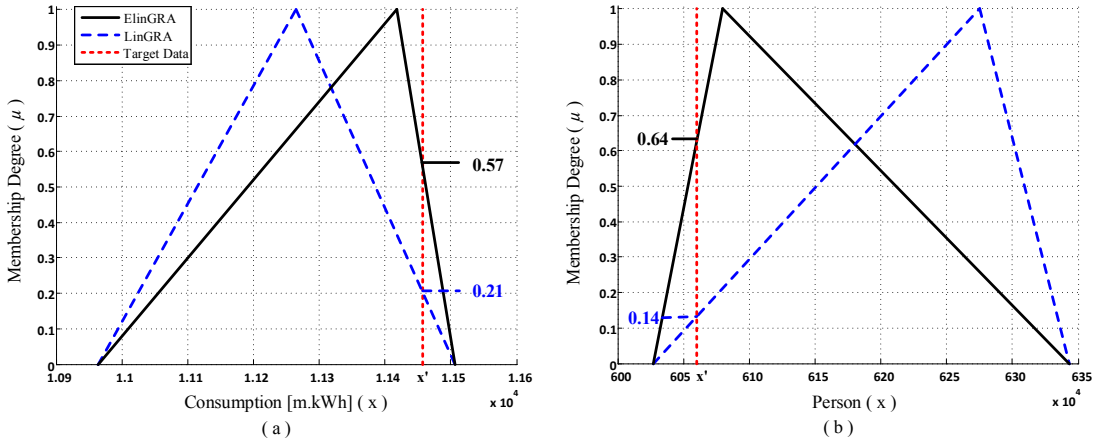


Figure 3.13 : Enhancement of the membership degree with ElinGRA: (a) 357^{th} sample on the Australian monthly electricity consumption data set (b) 120^{th} sample on the airlines passenger data set.

For the 357^{th} sample of the Australian monthly electricity consumption data set, the current forecast and consumption values are $f_k = 11265$ and $x_k = 1145$, respectively. Since the success of the forecast is measured with respect to the μ value, and as mentioned previously a value close to one shows the success of the forecast which is represented with TFN. For the studied samples, the μ values of the TFNs generated from the LinGRA and ElinGRA are calculated as $\mu_{LinGRA} = 0.21$ and $\mu_{ElinGRA} = 0.57$, respectively. It can be concluded that ElinGRA was able to represent the success of the forecast about 2.8 times better in comparison to the LinGRA. A similar analysis can be also done for the 120^{th} sample of the airlines passenger data set given in Figure 3.13b where the ElinGRA resulted with a 4.5 times better representation of forecast success in comparison with the LinGRA.

Thanks to the membership degree of the proposed linguistic terms, the error evaluation of the forecast can be done with fuzzy numbers. Rather than the classical PIs, the proposed bounds utilize the realized value of the project not only bound-in or out consideration but also grading the forecast via the triangular fuzzy numbers. Therefore, the decision can have the opportunities to critize the last forecasts and their methods.

The further usage of the proposed methodologies will be discussed at the next section of the thesis. Proposed methodologies will used on different data sets in comparison with the conventional PIs.

4. EXPERIMENTAL RESULTS

In this section, the proposed approaches, fuzzy linguistic term generation and representation, will apply on two data sets. The data set can be evaluated as benchmark data sets that have been widely used in several forecasting studies (Makridakis et al, 1998; Sahin et al, 2014, Sahin et al 2015b). The proposed methodology is applied and illustrated as respectively both on two data sets. Applied methodology has two part, FLUBE and linguistic term generators. FLUBE is proposed for the same purpose as Prediction Intervals (PIs). Therefore, FLUBE will compare with the conventional PI on data sets. Then the linguistic term generators and their inferences are also evaluated in comparison with each other.

4.1 Data Set-1: The Australian Monthly Electricity Consumption Data Set

The Australian monthly electricity consumption data set (Makridakis et al, 1998) will be used to illustrate the proposed approach. This data set involves the electricity consumption values from January 1956 to August 1995, thus the data set has a total of 476 samples. As it can be clearly seen in Figure 2.9, the consumption of the electricity has always increased with respect the time which shows the trend property of the data. Moreover, the data has a seasonality characteristic of 12 months described with parameter s which can be clearly seen from subplots presented in Figure 2.9. Thus, the data has trend and seasonality characteristics which are commonly encountered in time series analysis (Box et al, 2015; Brockwell and Davis, 2002).

The design step starts with the choosing the forecast model to getting error terms. As mentioned in the FLUBE design, the bound characteristics are determined via the error terms e_k which are calculated from forecast model. Accordingly, since the handled data sets inherit trend and seasonality characteristics, the model have designed as $SARIMA\ (2,1,2)(15,1,2)_{12}$ for Data Set-1.

The performance values of the SARIMA models are can be listed as 1.95 % and 99.01 % as MAPE and POA. It can be observed that the performances of the SARIMA models are satisfactory since they resulted with relatively low MAPE values while providing high POA values (Frechtling, 2001).

Consequently, to design the FLUBE, the error terms e_k are calculated via SARIMA forecast model as given at Figure 4.1. The error terms (e_k) with respect to the consumption values (x_k) will be used for the error model training. As it has been mentioned in the FLUBE subsection as remark, there is need to shift the error terms by their mean values which are calculated as -3.04.

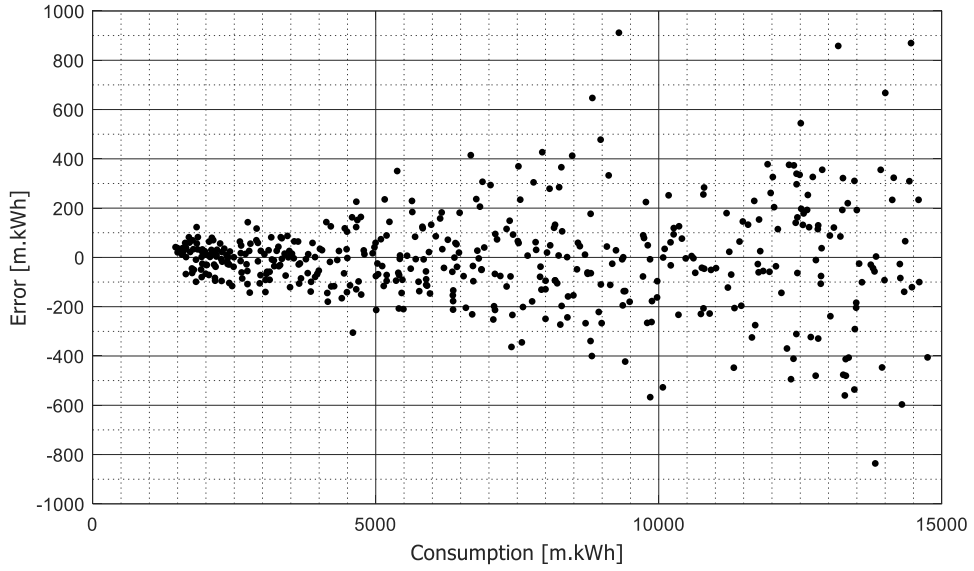


Figure 4.1 : Error terms of Data Set-1 according to SARIMA model.

The Selection Algorithm has been employed on data set with the tuning parameter of the SA (P) as 60 to collect the training data sets $[x_{min}, e_{min}]$ and $[x_{max}, e_{max}]$ for LFLS and UFLS, respectively. The selected points of the data set are also illustrated at Figure 4.2.

The LFLS and UFLS are constructed with 5 Gaussian antecedent MFs and 5 linear consequent MFs ($W = 5$) to provide an approximate bounds on the uncertainty (E^L and E^U). The outputs of the FLUBE are given in Figure 4.3.

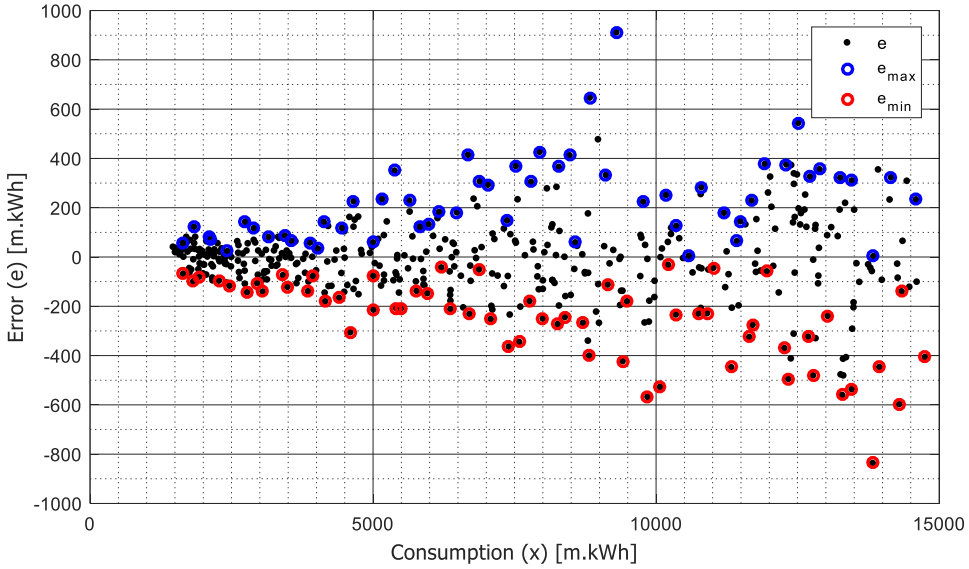


Figure 4.2 : Selected error terms of Data Set-1 via Selection Algorithm.

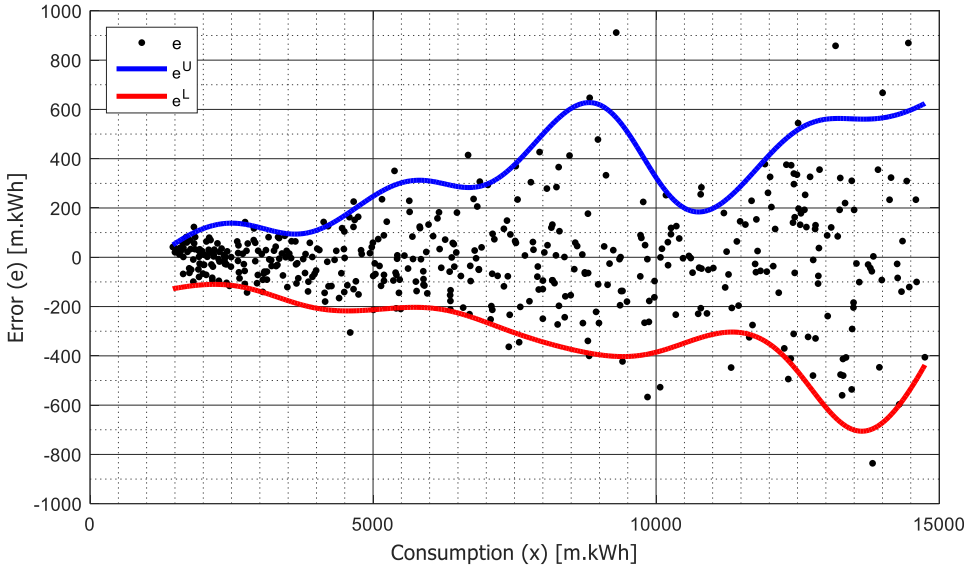


Figure 4.3 : Illustration of the FLUBE bounds on the Data Set-1's error terms.

Here the classical PI bounds which is already formulated at Section 3.1.2 can be compared with the proposed FLUBE bounds. Therefore, PI bounds construct with the parameter α as 0.95. The constructed bounds and their coverage properties are illustrated at Figure 4.4.

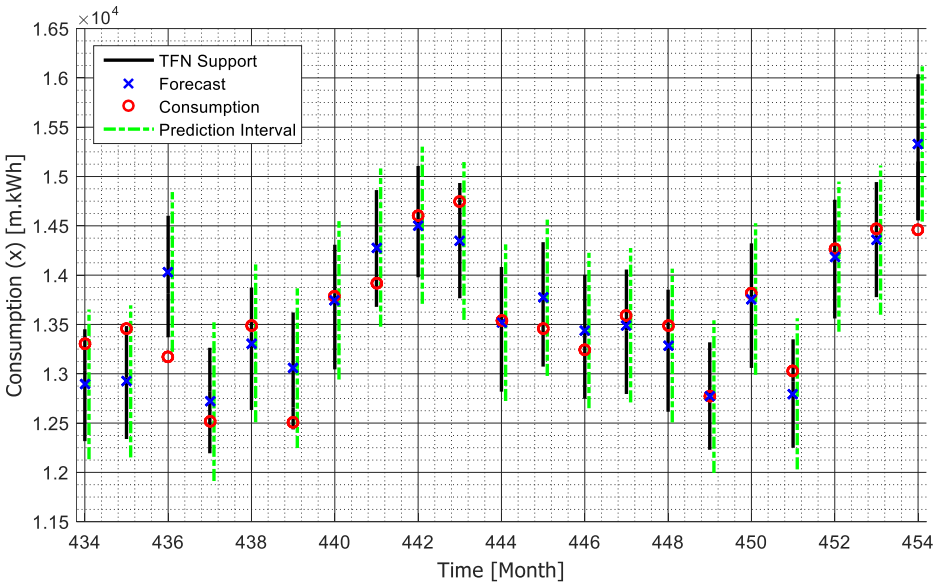


Figure 4.4 : The error bounds on the Data Set-1's last 20 samples.

The forecast coverage performances of the FLUBE and conventional PI are evaluated with respect to the performance indexes; PICP and PINAW which are also presented in Table 4.1. It can be concluded that the FLUBE provided a satisfactory performance since they resulted with a relatively small PINAW value and a high PICP value (Khosravi et al, 2010).

Table 4.1 : Performance values of the FLUBE and Conventional PI on Data Set-1.

Modelling Tool	Performance values	
	PICP	PINAW
FLUBE	92.72%	22.04%
PI	94.54%	57.56%

After the construction of the FLUBE, to generating TFN representation can be done for both approaches (LinGRA and ElinGRA) with outputs of the FLUBE. The main difference between these two approaches is the representation of the center point of the TFN. In the LinGRA, the center point will be assigned directly to the forecast value, while in the ElinGRA the center point of the TFN is redefined with the center point correction term (e^C) which generated from the Extra FLS. For the handled data set, the Extra FLS of the ElinGRA is constructed by defining the antecedent parts of its fuzzy rules with 4 Gaussian antecedent MFs while their consequent part with 4 linear MFs.

In the TFN generation of LinGRA, the outputs of the FLUBE at the k^{th} sample (e_k^L and e_k^U) to define the uncertainty bounds of the forecast value via Equation 3.41 will be used. In a similar manner, are used for ElinGRA with Equation 3.44 instead of Equation 3.41 to generate the TFNs. For instance, the uncertainty bounds of the forecast (L_k and U_k) is shown in Figure 4.4 for the Data Set-1. Now, the TFN can be generated by using Equation 3.41. Since the TFN will naturally results with an extra dimension (μ), the forecast is illustrated in a 3-D plot as shown in Figure 4.5 for Data Set-1 where the generated TFNs for each sample can be clearly seen for both approaches. It must be said that conventional PI bounds did not illustrate at the figures due to the extra dimensions which PIs did not have.

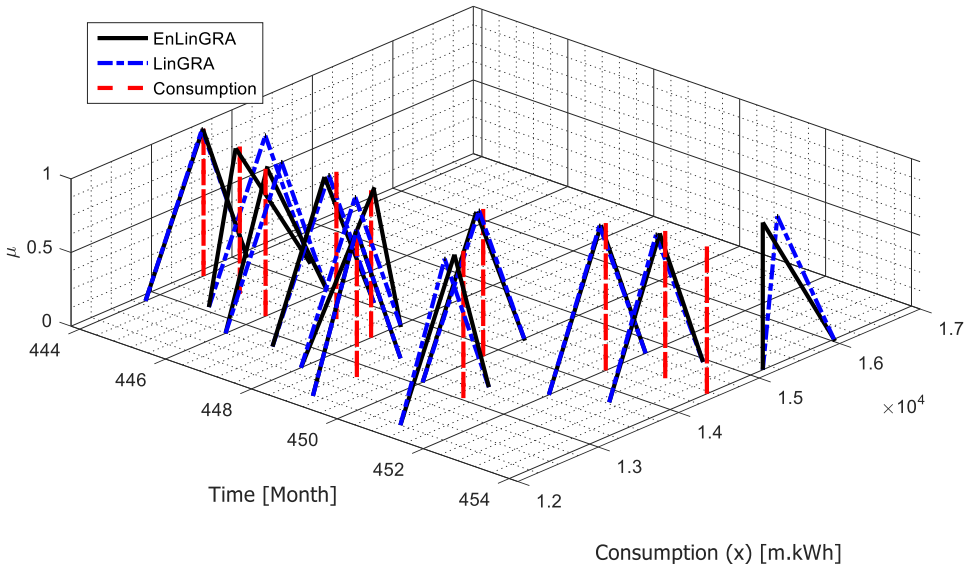


Figure 4.5 : 3-D representation of the LinGRA and ElinGRA for the Data Set-1.

Moreover, to analyze the overall forecast quality of the LinGRA and ElinGRA, the distribution of the membership grades (μ) of the studied data set is illustrated in Figure 4.6. It can be clearly observed that the membership grades of both approaches are mostly distributed around the maximum membership grade value 1. However, in comparison to the LinGRA, the count numbers of the μ values of the ElinGRA around 1 are relatively higher. In order to make a fair comparison, the mean values of the μ_{LinGRA} and $\mu_{ElinGRA}$ have been calculated and presented in Table 4.2. It can be clearly observed that the ElinGRA approach was able to represent overall forecast quality almost by 9% better in comparison to LinGRA.

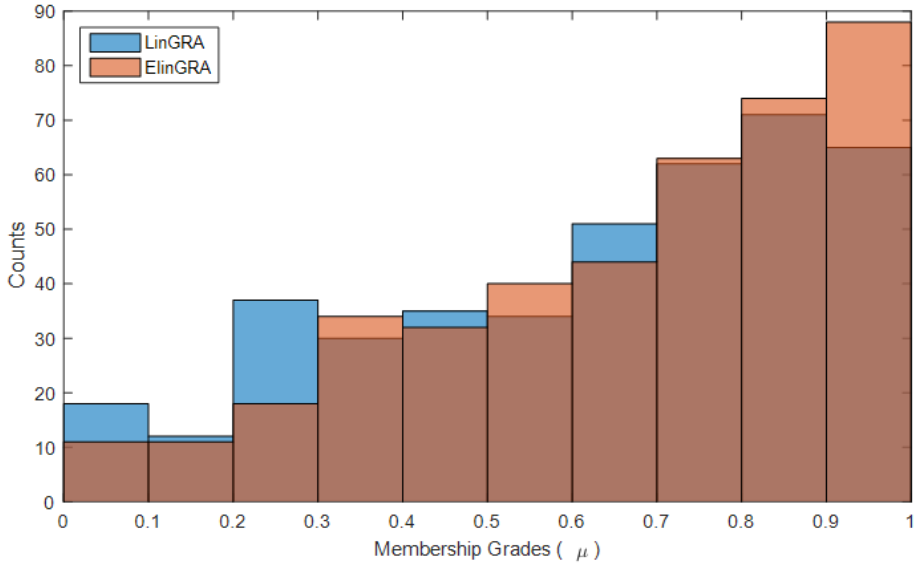


Figure 4.6 : The histogram of the membership degrees generated from the LinGRA and ElinGRA for the Data Set-1.

Table 4.2 : Success of the LinGRA and ElinGRA on Data Set-1.

Data Set	Mean μ values	
	LinGRA	ElinGRA
Data Set-1	0.57	0.62

4.2 Data Set-2: The Air Passenger Sata Set

The air passenger data set will also be used to illustrate the proposed approach. This data set involves monthly counts of the international airline passengers, measured in thousands, for the period January 1949 through December 1960, thus the data set has a total of 144 samples (Makridakis et al, 1998). Just as electricity consumption data set, air passenger data set has same characteristic as it can be seen in Figure 2.10.

The design steps are the same that also mentioned for the first data set. Therefore, the error term determination is done via the forecasting model which is design as $SARIMA\ (0,1,2)(13,1,14)_{12}$ for Data Set-2.

The performance values of the SARIMA models are can be listed as 3.13 % and 99.32 % as MAPE and POA. It can be observed that the performances of the SARIMA models are satisfactory since they resulted with relatively low MAPE values while providing high POA values (Frechtling, 2001).

Consequently, to design the FLUBE, the error terms e_k are calculated via SARIMA forecast model as given at Figure 4.7. The error terms (e_k) with respect to the consumption values (x_k) will be used for the error model training. As it has been mentioned in the FLUBE subsection as remark, there is need to shift the error terms by their mean values which are calculated as -0.26.

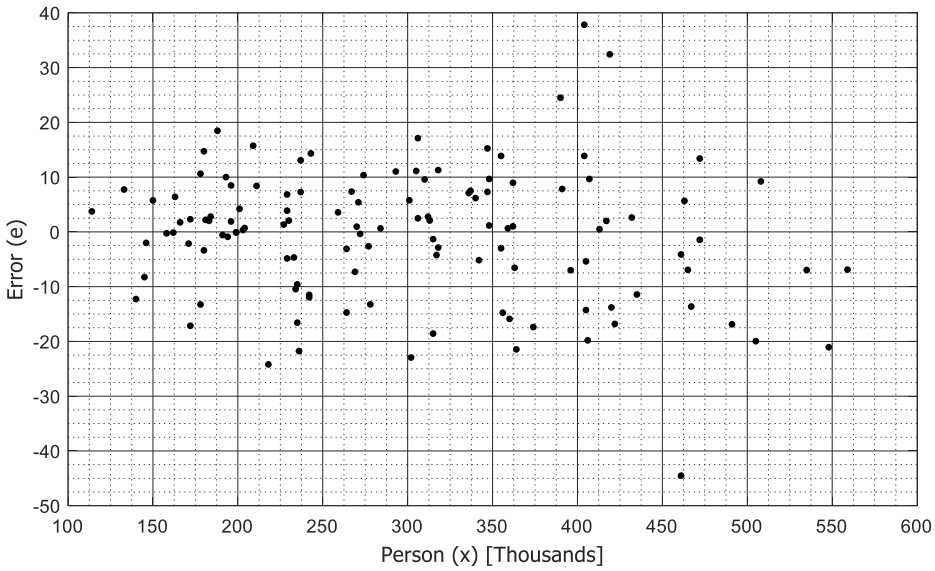


Figure 4.7 : Error terms of Data Set-2 according to SARIMA model.

The Selection Algorithm has been employed on data set with the tuning parameter of the SA (P) as 60 to collect the training data sets $[x_{min}, e_{min}]$ and $[x_{max}, e_{max}]$ for LFLS and UFLS, respectively. They are also illustrated at Figure 4.8.

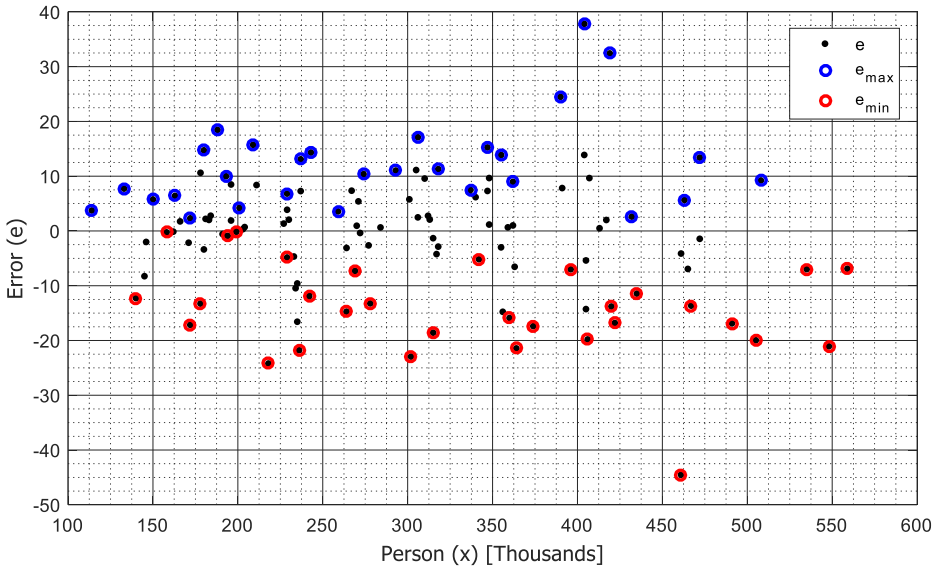


Figure 4.8 : Selected error terms of Data Set-2 via Selection Algorithm.

The LFLS and UFLS are constructed with 5 Gaussian antecedent MFs and 5 linear consequent MFs ($W = 5$) to provide an approximate bounds on the uncertainty (E^L and E^U). The outputs of the FLUBE are given in Figure 4.9.

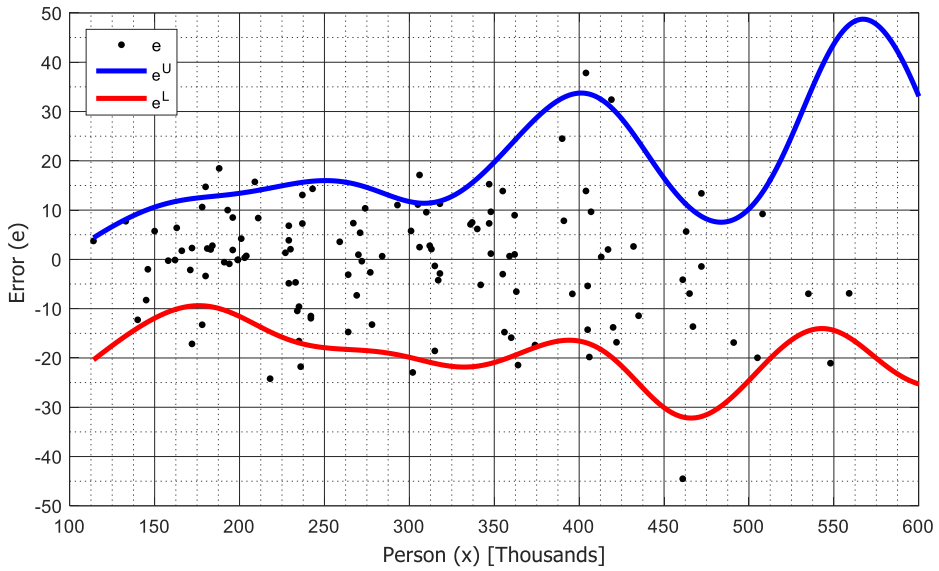


Figure 4.9 : Illustration of the FLUBE bounds on the Data Set-2's error terms.

Here the classical PI bounds which is already formulated at Section 3.1.2. It can be compared with the proposed FLUBE bounds. Therefore, PI bounds construct with the parameter α as 0.95. The constructed bounds and their coverage properties are illustrated at Figure 4.10.

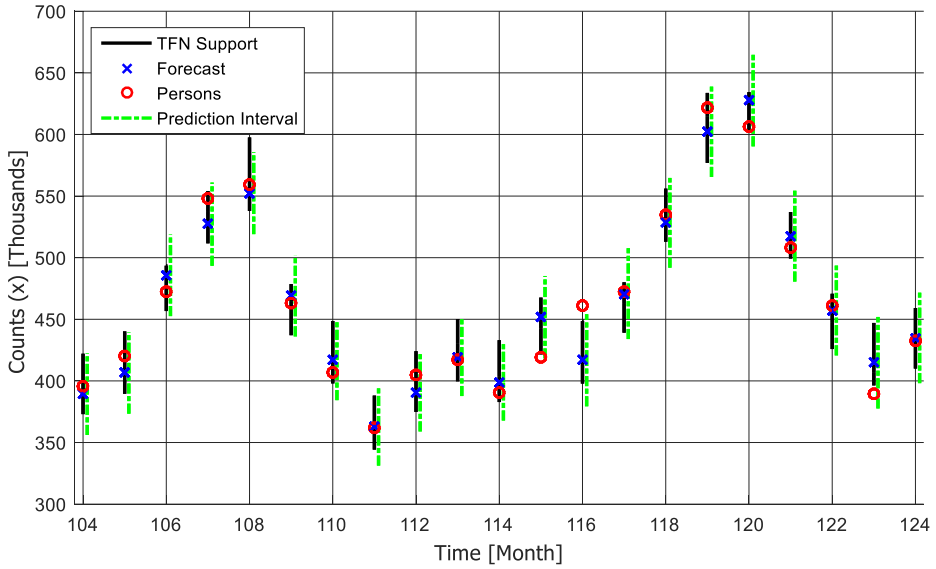


Figure 4.10 : The error bounds on the Data Set-2's last 20 samples.

The forecast coverage performances of the FLUBE and conventional PI are evaluated with respect to the performance indexes; PICP and PINAW which are presented in Table 4.3. Same as the first example, the FLUBE provided a satisfactory performance since they resulted with a relatively small PINAW value and a high PICP value (Khosravi et al, 2010).

Table 4.3 : Performance values of the FLUBE and Conventional PI on Data Set-2.

Modelling Tool	Performance values	
	PICP	PINAW
FLUBE	92%	12.28%
PI	96%	25.77%

After the construction of the FLUBE, generating TFN representation can be done for both approaches (LinGRA and ElinGRA) with outputs of the FLUBE. The Extra FLS of the ElinGRA constructed by the antecedent parts of its fuzzy rules with 4 Gaussian antecedent MFs while their consequent part with 4 linear MFs.

In the TFN generation of LinGRA and ElinGRA, the outputs of the FLUBE at the k^{th} sample (e_k^L and e_k^U) to define the uncertainty bounds of the forecast value via Equation 3.41 and Equation 3.44 will be used. The uncertainty bounds of the forecast (L_k and U_k) is shown in Figure 4.10 for of the handled data set. Since the TFN will naturally results with an extra dimension (μ), the forecast is illustrated in a 3-D plot as shown in Figure 4.11 for Data Set-2. The PIs did not represented for the same reason which mentioned at the first example.

Moreover, to analyze the overall forecast quality of the LinGRA and ElinGRA, the distribution of the membership grades (μ) of the studied data set is illustrated in Figure 4.12. Same as the first example, the mean values of the μ_{LinGRA} and $\mu_{ElinGRA}$ have been calculated and presented in Table 4.4. It can be observed that the ElinGRA approach was able to represent overall forecast quality almost by 28% better in comparison to LinGRA.

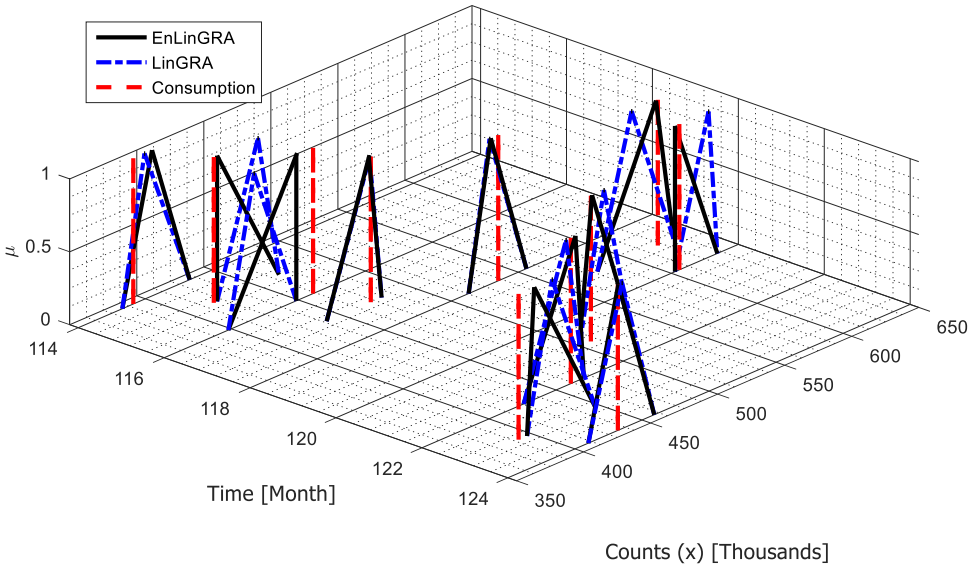


Figure 4.11 : 3-D representation of the LinGRA and ElinGRA for the Data Set-2.

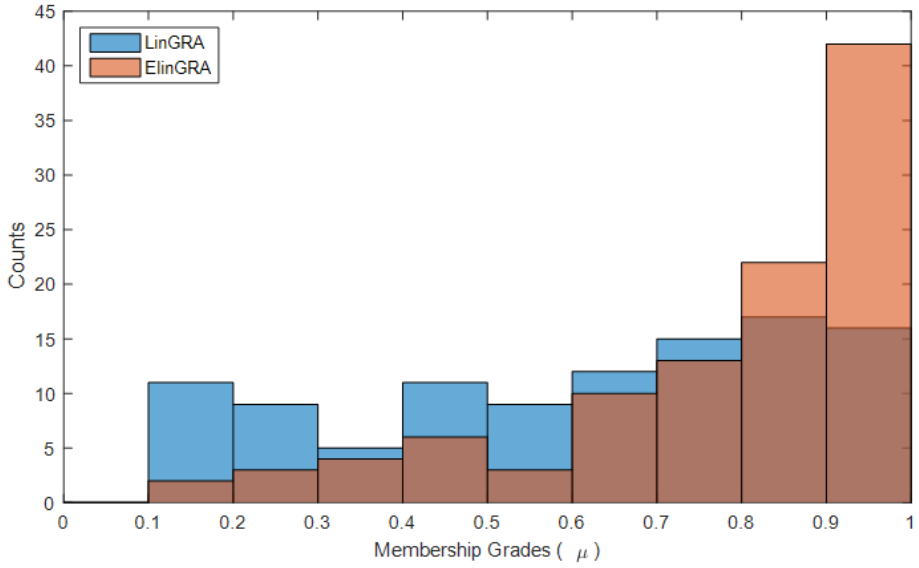


Figure 4.12 : The histogram of the membership degrees generated from the LinGRA and ElinGRA for the Data Set-2.

Table 4.4 : Success of the LinGRA and ElinGRA on Data Set-2.

Data Set	Mean μ values	
	LinGRA	ElinGRA
Data Set-2	0.52	0.67

After two illustrative examples, the proposed approaches' efficiency has been shown in comparison between the conventional PIs. PIs have compared with the proposed

error bounds constructed by the FLUBE. According to coverage results named as PICP and PINAW, FLUBE can provide better coverage probability as much as PIs can do while FLUBE has narrower intervals according to the PIs' width. Therefore, it can be said that the first part of the proposed method, FLUBE, can be more effective approach to modelling the errors via the fuzzy modelling tools

Basically, the error bounds are provided by FLUBE and the extra degree via the TFNs constructed by Linguistic term generator LinGRA and ElinGRA. By the help of TFN and its membership degree according to the forecasted value, the decision maker will be decide the forecast methods' efficiency on the data set. The decision can be made from not only the classical error terms like MSE and MAE but also the membership degree or the mean of the membership degrees. Linguistic term generation can provide the opportunities to the decision makers at the evaluation phase of the forecast accuracy. For example, the decision maker can evaluate the forecast via the membership degree. If the constructed TFN via the linguistic term generator such as ElinGRA provides a closer prediction to the real data value, it will be scale on the interval around 0.8 to 1.0 as the membership degree. Then the decision maker can deduct from the membership degree to evaluate as the linguistic terms like "Very Good" for the interval around 0.8 to 1.0, "Well" for the interval around 0.6 to 0.8, etc. It can be vary with such a wide terms.

In addition to these, ElinGRA also provides more accuracy via the extra FLS on the error terms in comparison with LinGRA. The enhancement can be seen via the tables about the mean of the memberships.

5. CONCLUSION AND DISCUSSION

In this Master thesis, two alternative linguistic term generation approaches to the representation of the forecast values are proposed, the fuzzy approaches named as FLUBE is also proposed to the modelling of the forecasting errors with the same purpose of the classical error bound named as PI. For this purpose, first, the times series and forecasting concept are investigated to clarify that what are approaches proposed for, then the general structures of the linguistic terms generation approaches and FLUBE are examined in detail, and the experimental studies are performed in order to show the effectiveness of the proposed approaches.

Firstly, the time series concept are examined as briefly to give information about the data which is used by the proposed approaches. The significant of the time series concepts especially in the future of the forecasting approaches are another topics which are highly related to why this thesis is proposed. The components of the time series are examined for the modelling of the time series. It will be known that the decomposition and modelling approaches of the time series are realized by MATLAB as automatically. Represented models are a few of the models when considering the all models in literature but they are selected as extremely concise to examine

Secondly, the main part of the thesis, fuzzy linguistic term generation and representation, is examined in throughout the section from the beginning of the method to the linguistic terms. The concept of the forecasting errors and evaluation assessments are presented in detail. These examined as widely when considering for what the method was proposed. The convenient error bound approaches like PI and CI with their literature surveys are also investigated. The FL and fuzzy modelling approaches are also presented according to its usage in FLUBE and linguistic term generation. Then the FLUBE and linguistic term generation are examined in detail.

The experimental studies are performed in order to show the effectiveness of the proposed approaches on several data sets. The first part of the experiments includes the comparison of the convenient PIs and FLUBE as the first part of the proposed

method. PIs will be have more coverage probability according to the PICP ındex but the width of the interval is the another important measurement for the comparison. FLUBE bounds always has the more narrow intervals according to the PIs. In according to this comparison, the first part of the proposed method has good coverage probability as well as convenient PIs and also narrow intervals.

The second part of the proposed method provides to the decision makers to deduce linguistic terms for their daily life usage. Linguistic term generation can provide the opportunities to the decision makers at the evaluation phase of the forecast accuracy. For example, the decision maker can evaluate the forecast via the membership degree. If the constructed TFN via the linguistic term generator such as ElinGRA provides a closer prediction to the real data value, it will be scale on the interval around 0.8 to 1.0 as the membership degree. Then the decision maker can deduct from the membership degree to evaluate as the linguistic terms like “Very Good” for the interval around 0.8 to 1.0, “Well” for the interval around 0.6 to 0.8, etc. For both approaches, LinGRA and ElinGRA, the deduction can be done. By the help extra FLS, ElinGRA can be have more accuracy with its membership degree in comparison with LinGRA. Correction terms will be have more effective performance to fix the top point of the TFN, in other meaning the forecasted point. The membership degree can also be provide the comparison for forecasting method. The mean value of the TFNs gives information about the forecast method quality.

Furthermore, the proposed method depends on the performance of the forecasting method. If the forecasting method has the nonsense errors, the fuzzy modelling tools will be limited in side of the performance. Therefore, it should be known that the methodology could be use on the consistent forecasting projects. The Selection Parameter, P is the important key for the interval width and coverage probability. It will be arrange by the user according to the circumstances. The fuzzy modelling tools (Upper and Lower FLS, CPCT FLS) have different parameters like the type of the membership functions and the count of its. It should also be arrange by the user. Therefore, the performance of the proposed methodology is also depend on the user decisions.

REFERENCES

Anderson, T. W. (2011). The statistical analysis of time series.

Armstrong, J. S., and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International journal of forecasting*, *8*(1), 69-80.

Armstrong, J. S. (2001). Evaluating forecasting methods. In *Principles of forecasting* (pp. 443-472). Springer US.

Ashton, K. (2009). That 'internet of things' thing, *RFiD Journal, 22*, no.7, 97–114.

Babuška, R. (2012). *Fuzzy modeling for control* (Vol. 12). Springer Science & Business Media.

Benoit, D. F., and Van den Poel, D. (2009). Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services. *Expert Systems with Applications*, *36*(7), 10475-10484.

Bishop, C. M. (1995). Neural networks for pattern recognition.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307-327.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brass, W. (1974). Perspectives in population prediction: Illustrated by the statistics of England and Wales. *Journal of the Royal Statistical Society. Series A (General)*, 532-583.

Briggs, A. H., Wonderling, D. E., and Mooney, C. Z. (1997). Pulling cost-effectiveness analysis up by its bootstraps: A non-parametric approach to confidence interval estimation. *Health economics*, *6*(4), 327-340.

Brockwell, P. J., and Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer Science & Business Media.

Chatfield, C. (1988). Apples, oranges and mean square error. *International Journal of Forecasting*, *4*(4), 515-518.

Chatfield, C. (2013). *The analysis of time series: an introduction*. CRC press.

Chen, S. M. (1996). Forecasting enrollments based on fuzzy time series.*Fuzzy sets and systems*, *81*(3), 311-319.

Chen, S. M., and Hwang, J. R. (2000). Temperature prediction using fuzzy time series. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *30*(2), 263-275.

Chen, S. M., and Tanuwijaya, K. (2011). Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering techniques. *Expert Systems with Applications*, *38*(8), 10594-10605.

Chryssolouris, G., Lee, M., and Ramsey, A. (1996). Confidence interval prediction for neural network models. *Neural Networks, IEEE Transactions on*, *7*(1), 229-232.

De Vieaux, R. D., Schumi, J., Schweinsberg, J., and Ungar, L. H. (1998). Prediction intervals for neural networks via nonlinear regression. *Technometrics*, *40*(4), 273-282.

Diderot, D., and d'Alembert, J. L. R. (1756). L'Encyclopédie, ou Dictionnaire raisonné des arts et sciences.

Dodurka, M. F., Sahin, A., Yesil, E., and Urbas, L. (2015, August). Learning of FCMs with causal links represented via fuzzy triangular numbers. In *Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on* (pp. 1-8). IEEE.

Dybowski, R., and Roberts, S. (2001). Confidence intervals and prediction intervals for feed-forward neural networks. *Clinical applications of artificial neural networks*, 298-326.

Efron, B. (1992). Bootstrap methods: another look at the jackknife.*Breakthroughs in Statistics*, 569-593.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Frechtling, D. (2012). *Forecasting tourism demand*. Routledge.

Gardner, E. S. (1990). Evaluating forecast performance in an inventory control system. *Management Science*, *36*(4), 490-499.

Goodwin, P., and Lawton, R. (1999). On the asymmetry of the symmetric MAPE. *International journal of forecasting*, *15*(4), 405-408.

Hanss, M. (2005). *Applied fuzzy arithmetic*. Springer-Verlag Berlin Heidelberg.

Heskes, T. (1997). Practical confidence and prediction. In *Advances in Neural Information Processing Systems 9: Proceedings of the 1996 Conference* (Vol. 9, p. 176). MIT press.

Hosmer, D. W., and Lemeshow, S. (1992). Confidence interval estimation of interaction. *Epidemiology*, *3*(5), 452-456.

Hsu, D. A. (1979). Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis. *Journal of the American Statistical Association*, *74*(365), 31-40.

Huarng, K. (2001). Heuristic models of fuzzy time series for forecasting.*Fuzzy sets and systems*, *123*(3), 369-386.

Hyndman, R. J., and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, *22*(4), 679-688.

Hwang, J. G., and Ding, A. A. (1997). Prediction intervals for artificial neural networks. *Journal of the American Statistical Association*, *92*(438), 748-757.

Jang, J. S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on*, *23*(3), 665-685.

Jang, J. S. R., Sun, C. T., and Mizutani, E. (1997). Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence.

Khosravi, A., Nahavandi, S., and Creighton, D. (2010). Construction of optimal prediction intervals for load forecasting problems. *Power Systems, IEEE Transactions on*, *25*(3), 1496-1503.

Khosravi, A., Nahavandi, S., Creighton, D., and Atiya, A. F. (2011). Comprehensive review of neural network-based prediction intervals and new advances. *Neural Networks, IEEE Transactions on*, *22*(9), 1341-1356.

Khosravi, A., Nahavandi, S., Creighton, D., and Atiya, A. F. (2011). Lower upper bound estimation method for construction of neural network-based prediction intervals. *Neural Networks, IEEE Transactions on*, *22*(3), 337-346.

Khosravi, A., Mazloumi, E., Nahavandi, S., Creighton, D., and Van Lint, J. W. C. (2011). Prediction intervals to account for uncertainties in travel time prediction. *Intelligent Transportation Systems, IEEE Transactions on*, *12*(2), 537-547.

Khosravi, A., Mazloumi, E., Nahavandi, S., Creighton, D., and Van Lint, J. W. C. (2011). A genetic algorithm-based method for improving quality of travel time prediction intervals. *Transportation Research Part C: Emerging Technologies*, *19*(6), 1364-1376.

Khosravi, A., Nahavandi, S., and Creighton, D. (2011). Prediction interval construction and optimization for adaptive neurofuzzy inference systems.*Fuzzy Systems, IEEE Transactions on*, *19*(5), 983-988.

Khosravi, A., and Nahavandi, S. (2014). An interval type-2 fuzzy logic system-based method for prediction interval construction. *Applied Soft Computing*,*24*, 222-231.

Klir, G., and Yuan, B. (1995). *Fuzzy sets and fuzzy logic* (Vol. 4). New Jersey: Prentice hall.

Lee, L. (2000). *Bad predictions*. Elsewhere Press (MI).

MacKay, D. J. (1992). The evidence framework applied to classification networks. *Neural computation*, *4*(5), 720-736.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., ... and Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of forecasting*, *1*(2), 111-153.

Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, *9*(4), 527-529.

Makridakis, S., Wheelwright, S. C., and Hyndman, R. J. (2008). *Forecasting methods and applications*. John Wiley & Sons.

Mamdani, E. H. (1974). Application of fuzzy algorithms for control of simple dynamic plant. In *Proceedings of the Institution of Electrical Engineers* (Vol. 121, No. 12, pp. 1585-1588). IET Digital Library.

Mannermaa, M. (2004). Traps in futures thinking–and how to overcome them. In *Thinking Creatively in Turbulent Times. Proceedings from the Annual Conference of the World Future Society* (Vol. 31, pp. 41-53).

Mendel, J. M. (2001). Uncertain rule-based fuzzy logic system: introduction and new direction

Montgomery, D. C., Jennings, C. L., and Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.

Möller, B., and Reuter, U. (2007). *Uncertainty forecasting in engineering.* Berlin: Springer.

Nakagawa, S., and Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*,*82*(4), 591-605.

Nix, D. A., and Weigend, A. S. (1994). Estimating the mean and variance of the target probability distribution. In *Neural Networks, 1994. IEEE World Congress on Computational Intelligence.* (Vol. 1, pp. 55-60). IEEE.

Oblak, S., and Blažič, S. (2006). If approximating nonlinear areas, then consider fuzzy systems. *Potentials, IEEE*, *25*(6), 18-23.

Oblak, S., Škrjanc, I., and Blažič, S. (2007). Fault detection for nonlinear systems with uncertain parameters based on the interval fuzzy model.*Engineering Applications of Artificial Intelligence*, *20*(4), 503-510.

Papadopoulos, G., Edwards, P. J., and Murray, A. F. (2001). Confidence estimation methods for neural networks: A practical comparison. *Neural Networks, IEEE Transactions on*, *12*(6), 1278-1287.

Pedrycz, W. (Ed.). (2012). *Fuzzy modelling: paradigms and practice* (Vol. 7). Springer Science & Business Media.

Ramírez, J. G. (2009). Statistical intervals: confidence, prediction, enclosure. *SAS Institute Inc., white paper*.

Ross, T. J. (2009). *Fuzzy logic with engineering applications*. John Wiley & Sons.

Sahin, A., Kumbasar, T., Yesil, E., Dodurka, M. F., and Karasakal, O. (2014, November). An Approach to Represent Time Series Forecasting via Fuzzy Numbers. In *Artificial Intelligence, Modelling and Simulation (AIMS), 2014 2nd International Conference on* (pp. 51-56). IEEE.

Sahin, A., Dodurka, M. F., Kumbasar T., Yesil, E., Siradag, S., Review Study on Fuzzy Time Series and Their Applications in the Last Fifteen Years,

FUZZYSS 2015 - International Fuzzy Systems Symposium, İstanbul, Turkey, November 5-6.

Sahin, A., Kumbasar, T., Yesil, E., Dodurka, M. F., Karasakal, O., and Siradag, S. (2015, August). An Enhanced Fuzzy Linguistic Term Generation and Representation for time series forecasting. In *Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on* (pp. 1-8). IEEE.

Singh, P. (2015). A brief review of modeling approaches based on fuzzy time series. *International Journal of Machine Learning and Cybernetics*, 1-24.

Sun, C. T. (1994). Rule-base structure identification in an adaptive-network-based fuzzy inference system. *Fuzzy Systems, IEEE Transactions on*, *2*(1), 64-73.

Song, Q., and Chissom, B. S. (1993). Forecasting enrollments with fuzzy time series—part I. *Fuzzy sets and systems*, *54*(1), 1-9.

Song, Q., and Chissom, B. S. (1994). Forecasting enrollments with fuzzy time series—part II. *Fuzzy sets and systems*, *62*(1), 1-8.

Škrjanc, I., Blažič, S., and Agamennoni, O. (2005). Identification of dynamical systems with a robust interval fuzzy model. *Automatica*, *41*(2), 327-332.

Tanaka, K. (1997). An introduction to fuzzy logic for practical applications.

Takagi, H., Suzuki, N., Koda, T., and Kojima, Y. (1992). Neural networks designed on approximate reasoning architecture and their applications.*Neural Networks, IEEE Transactions on*, *3*(5), 752-760.

Takagi, T., and Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *Systems, Man and Cybernetics, IEEE Transactions on*, (1), 116-132.

Thompson, P. A. (1990). An MSE statistic for comparing forecast accuracy across series. *International Journal of Forecasting*, *6*(2), 219-227.

Tinbergen, J. (1939). *Statistical Testing of Business-Cycle Theories: I. A Method and Its Application to Investment Activity* (Vol. 1). League of Nations, Economic Intelligence Service.

Tsay, R. S. (2005). *Analysis of financial time series* (Vol. 543). John Wiley & Sons.

Quan, H., Srinivasan, D., and Khosravi, A. (2014). Short-term load and wind power forecasting using neural network-based prediction intervals. *Neural Networks and Learning Systems, IEEE Transactions on*, *25*(2), 303-315.

Wang, W. C., Chau, K. W., Cheng, C. T., and Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of hydrology*, *374*(3), 294-306.

Whittle, P. Hypothesis Testing in Time Series Analysis, 1951. *Almquist and Wiksell, Upssala*.

van Hinsbergen, C. I., Van Lint, J. W. C., and Van Zuylen, H. J. (2009). Bayesian committee of neural networks to predict travel times with confidence intervals. *Transportation Research Part C: Emerging Technologies*, *17*(5), 498-509.

Yager, R. R., and Filev, D. P. (1994). Essentials of fuzzy modeling and control.*New York*.

Yager, R. R., and Zadeh, L. A. (Eds.). (2012). *An introduction to fuzzy logic applications in intelligent systems* (Vol. 165). Springer Science & Business Media.

Yesil, E., Dodurka, M. F., and Urbas, L. (2014, July). Triangular fuzzy number representation of relations in Fuzzy Cognitive Maps. In *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on* (pp. 1021-1028). IEEE.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, *8*(3), 338-353.

Zadeh, L. A. (1996). Fuzzy logic = computing with words. *Fuzzy Systems, IEEE Transactions on*, *4*(2), 103-111.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, *1*(1), 3-28.

Zhao, J. H., Dong, Z. Y., Xu, Z., and Wong, K. P. (2008). A statistical approach for interval forecasting of the electricity price. *Power Systems, IEEE Transactions on*, *23*(2), 267-276.

Davies, R., Knapp, S., & Doe, J. (2001). Industry 4.0 Digitalisation for productivity and growth. Date retrieved: 14.02.2016, adress: http://www.europarl.europa.eu/RegData/etudes/BRIE/2015/568337/EPRS_BRI(2015)568337_EN.pdf

Url-1 <*https://datamarket.com/data/set/22ts/beveridge-wheat-price-index-1500-1869#!ds=22ts*>, date retrieved 25.02.2016.

Url-2 < *http://www.metoffice.gov.uk/research/climate/climate-monitoring/land-and-atmosphere/surface-station-records*>, date retrieved 25.02.2016.

CURRICULUM VITAE

Name Surname **:** Atakan Şahin

Place and Date of Birth **:** Merkez/Kırklareli, 09.08.1991

E-Mail **:** atakansah@gmail.com

EDUCATION **:**

- **B.Sc.** **:** 2014, Istanbul Technical Engineering, Control and Automation Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2014-Present: Getron Bilişim Hizmetleri A.Ş.
- 2014-Present: Istanbul Technical University ABB Process Control and Automation Laboratory

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **A. Sahin**, T. Kumbasar, E. Yesil, M.F. Dodurka, O. Karasakal, S. Siradag, 2015: An enhanced fuzzy linguistic term generation and representation for time series forecasting, In FUZZ-IEEE 2015 - International Conference on Fuzzy Systems, August 2-5, 2015, Istanbul, Turkey.
- **A. Sahin**, M.F. Dodurka, T. Kumbasar, E. Yesil, S. Siradag, 2015: Review Study on Fuzzy Time Series and Their Applications in the Last Fifteen Years, In FUZZYSS 2015 - International Fuzzy Systems Symposium, November 5-6, 2015, Istanbul, Turkey.

- M.F. Dodurka, **A. Sahin**, T. Kumbasar, E. Yesil, S. Siradag, 2015: A Systematic Type-2 Fuzzy Modelling Methodology for Time Series Forecasting, In FUZZYSS 2015 - International Fuzzy Systems Symposium, November 5-6, 2015, Istanbul, Turkey.
- **A. Sahin**, T. Kumbasar, E. Yesil, M.F. Dodurka, O. Karasakal, 2014: An approach to represent time series forecasting via fuzzy numbers, In AIMS 2014 - Artificial Intelligence, Modelling and Simulation, November 18-20, 2014, Madrid, Spain.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- **A. Sahin**, E. Atici, T. Kumbasar, 2016: Type-2 Fuzzified Flappy Bird Control System, In IEEE WCCI 2016 - IEEE World Congress on Computational Intelligence, July 25-26, 2016, Vancouver, Canada.
- M.F. Dodurka, A. Sakallı, A. Beke, A. I. Savran, **A. Sahin**, T. Kumbasar, E. Yesil, H. Hagras, 2015: A New Online Self-Tuning Approach for Type-2 Fuzzy PID Controllers," TOK 2015 – Turkish National Meeting on Automatic Control, September, 2015, Denizli, Turkey, (in Turkish).
- M.F. Dodurka, **A. Sahin**, E. Yesil, L. Urbas, 2015: Learning of FCMs with Causal Links Represented via Fuzzy Triangular Numbers, In FUZZ-IEEE 2015 - International Conference on Fuzzy Systems, August 2-5, 2015, Istanbul, Turkey.
- E. Yesil, C. Ozturk, M.F. Dodurka, **A. Sahin**, 2013: Control engineering education critical success factors modeling via Fuzzy Cognitive Maps, In ITHET 2013 - Information Technology Based Higher Education and Training, October 10-12, 2013, Antalya, Turkey.