UNIVARIATE TIME SERIES PREDICTION LEVERAGING LINEAR AND NONLINEAR PATTERNS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ÜMİT ÇAVUŞ BÜYÜKŞAHİN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

SEPTEMBER 2020

Approval of the thesis:

UNIVARIATE TIME SERIES PREDICTION LEVERAGING LINEAR AND NONLINEAR PATTERNS

submitted by **ÜMİT ÇAVUŞ BÜYÜKŞAHİN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Halit Oğuztüzün
Head of Department, **Computer Engineering** _____

Assist. Prof. Dr. Şeyda Ertekin
Supervisor, **Computer Engineering Department, METU** _____

Examining Committee Members:

Prof. Dr. Tolga Can
Computer Engineering Department, METU _____

Assist. Prof. Dr. Şeyda Ertekin
Computer Engineering Department, METU _____

Prof. Dr. Mehmet Reşit Tolun
Computer Engineering Department, KGTÜ _____

Assoc. Prof. Dr. Erhan Eren
Graduate School of Informatics, METU _____

Assoc. Prof. Dr. Ahmet Murat Özbayoğlu
Computer Engineering Department, TOBB ETÜ _____

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: ÜMİT ÇAVUŞ BÜYÜKŞAHİN

Signature :

ABSTRACT

UNIVARIATE TIME SERIES PREDICTION LEVERAGING LINEAR AND NONLINEAR PATTERNS

Büyükşahin, Ümit Çavuş
Ph.D., Department of Computer Engineering
Supervisor: Assist. Prof. Dr. Şeyda Ertekin

September 2020, 123 pages

Making forecasts from time series data has become increasingly important with the increase of data collection capabilities.The forecasting in time series is mostly done in a univariate setting where historical data in the series itself is used for further time steps. Forecasting in univariate time series is a challenging task due to unpredictable variations in the historic data. The objective of this thesis is to develop prediction methods that exhibit better performance results by alleviating the challenges in univariate time series. The methods we developed for this purpose can be divided into three groups. This thesis first proposes a novel Autoregressive Integrated Moving Average (ARIMA)-Artificial Neural Network(ANN) hybrid method that works in a more general framework. This hybrid method is then expanded to feature-based hybrid approach where statistical and structural features of time series are taken into account. Experimental results show that strategies for decomposing the original data and for combining linear and nonlinear models throughout the hybridization process are key factors in the forecasting performance of the methods. Secondly, by incorporating these findings, this thesis proposes two methods that use Empirical Mode Decomposition (EMD) technique which generates more predictable components. These

methods are then enhanced with method selection algorithm which determine the most suitable method for each component. The performed experiments show that our hybrid method with EMD can be an effective way to improve predictive accuracy. These experiments additionally show that having less fluctuations in already stationary time series data leads to more accurate results in forecasting. With the motivation of these consequences, this thesis thirdly proposes another novel method which recursively employs EMD technique for those fast fluctuating components until it achieves more regular and easy-to-predict sub-components. The experiments demonstrate that the recursive algorithm outperforms the previously developed and examined methods. At the end of the thesis, the methods we developed step by step are applied for the prediction of hydropower production data and the results are compared. As a result, the Recursive EMD-based method also produces more accurate predictions for the hydropower data set.

Keywords: Time series prediction, Hybrid Methods, Artificial Neural Network, Autoregressive Integrated Moving Average, Empirical Mode Decomposition

ÖZ

DOĞRUSAL VE DOĞRUSAL OLMAYAN ÖRÜNTÜLERDEN YARARLANARAK TEK DEĞİŞKENLİ ZAMAN SERİSİ TAHMİNİ

Büyükşahin, Ümit Çavuş
Doktora, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Dr. Öğr. Üyesi. Şeyda Ertekin

Eylül 2020 , 123 sayfa

Veri toplama kabiliyetlerinin artmasıyla zaman serisi verilerinden tahmin yapmak giderek daha önemli hale gelmiştir. Zaman serilerinde tahminleme, çoğunlukla serideki geçmiş verilerin gelecek zaman adımları için kullanıldığı tek değişkenli ortamda yapılır. Tek değişkenli zaman serilerinde tahmin yapmak, geçmiş verilerdeki öngörülemeyen varyasyonlar nedeniyle zorlu bir iştir. Bu tezin amacı, tek değişkenli zaman serilerindeki zorlukları azaltarak daha iyi performans sonuçları sergileyen tahmin yöntemleri geliştirmektir. Bu amaçla geliştirdiğimiz yöntemler üç gruba ayrılabilir. İlk olarak bu tez, daha genel bir çerçevede çalışan yeni bir Otoregresif Bütünleşik Hareketli Ortalama (ARIMA) - Yapı Sinir Ağı (YSA) hibrit yöntemi önermektedir. Bu hibrit yöntem daha sonra zaman serilerinin istatistiksel ve yapısal özelliklerinin dikkate alındığı, öznitelik tabanlı hibrit yaklaşıma genişletilir. Deneysel sonuçlar, orijinal verilerin ayrıştırılması ve hibridizasyon işlemi boyunca doğrusal ve doğrusal olmayan modellerin birleştirilmesi için kullanılan stratejilerin tahmin performansında anahtar faktörler olduğunu göstermektedir. Bu bulguları kullanarak, ikinci olarak bu tezde daha öngörülebilir bileşenler üreten Ampirik Kip Ayrıştırma (EMD) tekniğini kulla-

nan iki yöntem önerilmektedir. Bu yöntemler daha sonra her bileşen için en uygun yöntemi belirleyen yöntem seçim algoritması ile geliştirilir. Yapılan deneyler, hibrit yöntemimizin EMD ile kullanılmasının, öngörme doğruluğunu iyileştirmede etkili bir yol olabileceğini göstermektedir. Bu deneyler ayrıca durağan zaman serisi verilerinde daha az dalgalanma olmasının, öngörmede daha doğru sonuçlar verdiğini göstermektedir. Bu sonuçların motivasyonu ile üçüncü olarak bu tez, daha düzenli ve tahmin edilmesi kolay alt bileşenler elde edene kadar, hızlı dalgalanan bileşenler için EMD tekniğini özyinelemeli bir şekilde uygulayan yeni bir yöntem önermektedir. Deneyler, özyinelemeli algoritmanın daha önce geliştirilen ve incelenen yöntemlerden daha iyi performans sergilediğini göstermektedir. Tezin sonunda, adım adım geliştirdiğimiz yöntemler, hidroelektrik üretim verisini tahmini için uygulanır ve sonuçları karşılaştırılmaktadır. Sonuç olarak, Özyinelemeli EMD tabanlı yöntem, hidroelektrik veriseti için de daha doğru tahminler üretmektedir.

Anahtar Kelimeler: Zaman serisi tahminleme, Hibrit yöntemler, Yapay Sinir Ağları, Otoregresif Bütünleşik Hareketli Ortalama, Ampirik Kip Ayrışımı

Zeynep Ece'me

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my supervisor Professor Şeyda Ertekin for her wise guidance, for her invaluable time whenever I needed, and for pushing me always forward with her wisdom. I also benefited greatly from valuable comments and discussions throughout this study with Professor Tolga Can and Professor Erhan Eren, my thesis committee members, to whom I am deeply indebted and thankful.

Next, I would like to express my special thanks and appreciation to my colleagues and managers at Vitus Commodities, for providing me with all kinds of convenience and supports throughout my Ph.D. journey. I owe special thanks to my friends Hayrettin Erdem and Osman Kayataş, for encouragement and true friendship. Thank you for all the fun memories we shared, specially the Catan&Coco nights.

I would like to express my deepest gratitude to my dearest mom Zeynep and my dearest dad Fevzi, who gave me a wonderful life, for their unconditional trust, immense encouragement, and endless love. I am thankful to my brother Ugur and his wife Merve and to my cousin Ahmet for their endless support and motivation, for their continuous and unparalleled love. I am also heartily indebted to my big family, Şahinogulları for their enormously generous and sincere encouragement.

Last but surely not the least, I would like to express my deepest heartfelt appreciation to my beloved wife Latife Gözde for giving me motivation, understanding, and encouragement with her limitless love at every step of this pursuit. Thank you for being, together with our baby, the greatest happiness in my life. Many thanks to my little daughter Zeynep Ece for being a bundle of joy and laughter. You have made me stronger, better and more fulfilled than I could have ever imagined.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xi

LIST OF TABLES xiv

LIST OF FIGURES xvi

LIST OF ABBREVIATIONS xix

CHAPTERS

1 INTRODUCTION 1

1.1 Motivation and Problem Definition 1

1.2 Contributions of the thesis 3

1.3 The Outline of the thesis 5

2 TIME SERIES 7

2.1 Introduction to Time Series 7

2.1.1 Patterns of Time Series 10

2.1.2 Concept of Stationarity 12

2.2 Time Series Prediction 15

2.2.1 Linear Methods 15

2.2.2 Nonlinear Methods 17
2.2.3 Hybrid Methods 20
2.2.4 Prediction Evaluation 26
3 PROPOSED HYBRID METHODS 31
3.1 The ARIMA-ANN Hybrid Method 31
3.1.1 Proposed Method 32
3.1.2 Experimental Results 35
3.2 A feature-based hybrid ARIMA-ANN Method 45
3.2.1 Proposed Method 46
3.2.2 Feature Extraction and Selection 47
3.2.3 Experimental Results 50
3.3 Discussion 56
4 METHODS USING EMPIRICAL MODE DECOMPOSITION 59
4.1 Motivation 59
4.2 Empirical Mode Decomposition (EMD) 60
4.3 Two-level Decomposition Method 65
4.3.1 Proposed Method 65
4.3.2 Experiments and Results 67
4.4 EMD-based Hybrid Method 68
4.4.1 Proposed Method 68
4.4.2 Experiments and Results 69
4.5 Method Selection Approach 71
4.5.1 Method Selection Algorithm 72

4.5.2 Experiments and Results 73

4.6 Discussion 76

5 RECURSIVELY DECOMPOSITION METHOD 79

5.1 Motivation 79

5.2 Proposed Method 80

5.3 Experimental Results 83

5.4 Discussion 92

6 AN APPLICATION: TURKEY HYDROPOWER PRODUCTION PREDICTION 95

6.1 Introduction 95

6.2 Experimental Results 97

7 CONCLUSION AND FUTURE WORK 103

REFERENCES 107

CURRICULUM VITAE 121

LIST OF TABLES

TABLES

Table 2.1 t-ratio and p-value of the ADF-test results of datasets 14

Table 3.1 Performance comparisons of the proposed ARIMA-ANN hybrid method for all datasets 37

Table 3.2 p-value test result comparisons of the proposed ARIMA-ANN hybrid method for all datasets 38

Table 3.3 Summary of the calculated statistical and structural features from time series 48

Table 3.4 Performance comparisons of the proposed Feature-based hybrid method for all datasets 50

Table 3.5 p-value test result comparisons of the proposed Feature-based hybrid method for all datasets 51

Table 4.1 Performance comparisons of the proposed Two-level decomposition method for all datasets 67

Table 4.2 Performance comparisons of methods using EMD for all datasets . . 70

Table 4.3 p-value test result comparisons of the EMD-based hybrid method for all datasets 71

Table 4.4 Performance comparisons of EMD based methods with different approaches for all datasets 75

Table 4.5 Percentage improvement in the methods when EMD is used 76

Table 5.1 Parameters of the methods used in proposed recursive algorithm for all datasets 83

Table 5.2 Datasets setup in the experimental evaluation of the proposed recursive method 84

Table 5.3 Performance comparisons of the proposed recursive EMD-based method for all datasets (*part 1: individual and hybrid methods*) 85

Table 5.4 Performance comparisons of the proposed recursive EMD-based method for all datasets (*part 2: methods using EMD*) 86

Table 6.1 The parameters of the examined methods for forecasting Turkey Hydropower production 99

Table 6.2 The experimental results of the all examined methods for forecasting Turkey Hydropower production 100

Table 6.3 Frequency change rate analysis of the EMD components of Turkey Hydropower production data 101

LIST OF FIGURES

FIGURES

Figure 2.1 Wolf's sunspot series (1700-1987) 8

Figure 2.2 Canadian lynx data series (1821-1934) 9

Figure 2.3 Weekly British pound/US dollar exchange rate series (1980-1993) 9

Figure 2.4 Electricity price of intraday market in Turkey (Jul. 2015 - Dec. 2017) 10

Figure 2.5 Examples of stationary and non-stationary time series created from sine functions 13

Figure 2.6 A sample neural network diagram (multi-layer perceptron) . . . 19

Figure 2.7 The architecture of Zhang's hybrid method 21

Figure 2.8 The architecture of Khashei-Bijari's hybrid method 23

Figure 2.9 The architecture of Babu-Reddy's hybrid method 24

Figure 2.10 Example of hold-out cross-validation process 27

Figure 3.1 The architecture of the proposed ARIMA-ANN hybrid method . 33

Figure 3.2 The Sunspot data forecasts using various methods 39

Figure 3.3 The Lynx data forecasts using various methods 41

Figure 3.4 The GbpUsd data forecasts using various methods 42

Figure 3.5 The Intraday data forecasts using various methods 44

Figure 3.6 The architecture of the proposed feature-based ARIMA-ANN hybrid method 47

Figure 3.7 The sorted feature importance scores of the Sunspot dataset . . . 52

Figure 3.8 The sorted feature importance scores of the Lynx dataset 53

Figure 3.9 The sorted feature importance scores of the GbpUsd dataset . . . 54

Figure 3.10 The sorted feature importance scores of the Intraday dataset . . . 55

Figure 3.11 Comparison of MASE distributions of the ARIMA-ANN hybrid method with ADF test results for all datasets 57

Figure 4.1 Flow Chart of the EMD Algorithm 62

Figure 4.2 A sample EMD decomposition: The Lynx dataset and its components 63

Figure 4.3 The architecture of the proposed Two-level decomposition method 66

Figure 4.4 The architecture of the proposed EMD-based hybrid method . . 69

Figure 4.5 EMD components of the Sunspot dataset with different characteristics 73

Figure 4.6 EMD-based methods with method selection approach 74

Figure 4.7 Average MASE results of the methods for all datasets with/without using EMD 77

Figure 5.1 The architecture of the proposed Recursive EMD-based hybrid method 81

Figure 5.2 The IMFs and residual of the Sunspot series (1700-1987) 87

Figure 5.3 The IMFs and residual of the Consumption Series (2016-2019) . 88

Figure 5.4 The IMFs and residual of the GbpUsd Series (1980-1993) 90

Figure 5.5 The IMFs and residual of the Intraday Series (Jul. 2015 - Dec. 2017) 91

Figure 5.6 Improvement in MAE and MSE for all datasets with the Recursive EMD-method as compared to the EMD-based method 92

Figure 5.7 Improvement of MAE results of the Recursive EMD-based algorithm 93

Figure 6.1 The percentage distribution of resources in electricity generation of Turkey in 2018 96

Figure 6.2 The hydropower production of Turkey between Jan. 2017 and Dec. 2018 98

Figure 6.3 The IMFs and residual of the Turkey Hydropower productin data 102

LIST OF ABBREVIATIONS

ACF	Autocorrelation Function
ADF	Augmented Dickey Fuller
AIC	Akaike's Information Criterion
ANN	Artificial Neural Networks
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
FT	Fourier Transform
GBT	Gradient Boosted Trees
HBV	Hydrologiska Byråns Vattenbalansavdelning
KPSS	Kwiatkowski, Phillips, Schmidt and Shin
LSTM	Long Short-Term Memory
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
MDL	Minimum Description Length
MFE	Mean Forecast Error
MLP	Multi-Layer Perceptron
MPE	Mean Percentage Error
MSE	Mean Squared Error
NMAE	Normalized Mean Absolute Error
NMSE	Normalized Mean Squared Error
PACF	Partial Autocorrelation Function

PP	Phillips-Perron
RMSE	Root Mean Squared Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
SVM	Support Vector Machines
SVR	Support Vector Regression
VC	Vapnik-Chervonenkis Dimension
WPD	Wavelet Packet Decomposition

CHAPTER 1

INTRODUCTION

> Science is the most real guide for civilisation, for life, for success in the world. To search for a guide other than science is absurdity, ignorance and heresy.
>
> *Mustafa Kemal Atatürk*

1.1 Motivation and Problem Definition

A time series is a set of data points listed in time order. There is usually a fixed time interval between these data points.(every minute, hour, day, week, etc.,). With the developing technology, many applications in different domains produce large amount of time series data. For example, weather conditions such as precipitation, temperature, wind speed in the meteorological domain; stock indexes, currency exchange rates and spreads in economy domain; renewable power production and electricity consumption in energy domain; voltage and sensor data in industry domain; heart-rate, pulse and electrocardiogram signals in biomedical domain; search engine logs and user clicks in web technologies domain.

Time series forecasting uses models that predict future values based on historically observed values. It is necessary and important to make good prediction models in different domains where the time series is produced. For example, determining the routes of cruise ships requires weather forecasts; deciding whether a coal power station produce electricity considering costs requires the electricity price forecast; de-

termining how much of a good will be produced on a production line requires the forecasts of the consumption of that product; adjusting the capacity of a server for load balancing requires the forecasts of the number of users coming to the website. Moreover, the forecast horizon may vary according to the area where time series is used. For example, it is important to reach forecasts in areas such as telecommunications minutes ago; it is necessary to obtain day-ahead forecasts in electricity trading; it may be enough to have yearly forecasts with large capital. As a result, a good forecasting model in the time series from different areas with different requirements is an effective and efficient way for decision makers.

While it is easy to predict series that follow a particular pattern, such as the sunrise time, it is very difficult to predict series that change rapidly over time, such as the Bitcoin/Dollar exchange rate. There are two important conditions that affect the predictability of an event or a value:

1. sufficient amount of data
2. existence of exogenous explanatory factors

For example, forecasting in solar power production data, which satisfies both these conditions, can provide high accuracy. Solar power production is largely driven by the weather conditions such as solar irradiance and cloud cover. These exogenous series provide explanatory information to yield better approximations of solar power production. In addition, when a sufficient history of data on weather conditions and solar power production are provided, the forecast performance can be significantly high. On the other side, only the first condition, a sufficient amount of data, can be provided for forecasting currency exchange rates. We have limited knowledge of exogenous explanatory factors affecting exchange rates. Therefore, it is difficult to predict high accuracy in this type of time series.

As a result, the existence of explanatory exogenous data plays an important role in predictability of a time series. The prediction methods are classified into two types in literature according to the existence of these data. The first one is multivariate method in which predictions of a given time series depend on one or more exogenous data variables. The second one is univariate method in which predictions depend only

on historical observed values of series being predicted.

In this thesis, we focus on univariate time series prediction which is a dynamic research area where many studies are conducted in the literature. These methods can be classified into three groups. The first one is linear methods assuming a linear relationship between past data and future data. These methods might be insufficient to explain real-world series, which have a rather complex structure. The second one is nonlinear methods assuming a nonlinear relationship between past data and future data. These methods might be ineffective and inefficient for forecasting simple time series, as compared to linear methods. Finally, hybrid methods that combine those two methods. Hybrid methods decompose the given time series into linear and nonlinear components which are then modelled by linear and nonlinear methods, respectively. The strong assumptions that the hybrid method make at the processes of data decomposition and combining the obtained predictions might decrease the accuracy results. In order to have better forecasting performance results, our successively proposed methods avoid limiting the relationship between the past and future data, and we develop the models with effective decomposition methods to make this relationship as stationary as possible.

1.2 Contributions of the thesis

Our contributions are listed below, as follows:

- We propose a hybrid method that models the linear and nonlinear components, which are properly decomposed from the univariate time series, by ANN and ARIMA methods, respectively, and combines the model results effectively. The proposed method, called *ARIMA-ANN hybrid method*, eliminates the important assumptions that reduce the performance results of the traditional hybrid methods. Thus, our hybrid method outperforms both individual methods and the well-accepted hybrid methods in forecasting for the benchmark and publicly available datasets.
- We enhance our ARIMA-ANN hybrid method to improve its effectiveness in univariate time series forecasting, by using the feature-based representation of

series. The method selects the most explanatory features and presents them to our hybrid method, thereby capturing the underlying structural characteristics of the observed values to improve average accuracies.

- We propose a method that employs a linear method ARIMA on each linear component obtained from time series by applying decomposition process at two levels with Moving Average (MA) filter and Empirical Mode Decomposition (EMD) technique. The high performance results obtained with this method, called *Two-level decomposition method*, show the effectiveness of using EMD in time series.

- We propose an improvement to our hybrid method by combining with EMD. This method applies our hybrid method, that is able to produce accurate predictions, for each EMD components which are relatively stationary and shows more regularity in themselves. Taking advantage of both EMD and our hybrid method in this algorithm, called *EMD-based hybrid method*, we observe that accuracy performance gets better with the increasing level of linearity in time series.

- We introduce the *method selection* algorithm, which suggests determining the appropriate method according to the different characteristics of EMD components such as fluctuation rate. Modeling fast fluctuating components with our hybrid method and slow fluctuating components with ARIMA, this algorithm is implemented for both Two-level decomposition method and EMD-based hybrid method. Given that EMD components are stationary, experimental results show the importance of considering different fluctuation rates belonging to those components.

- We propose a novel approach, called *Recursive EMD-based hybrid method*, that recursively decomposes the fast fluctuating EMD components until obtaining slow fluctuating sub-components which have more regular structure in themselves and less variation. Numerous sub-components obtained from this recursive process, which are relatively easy-to-predict, are modeled by our hybrid method, thereby resulting in a remarkable improvement in the forecasting results.

1.3 The Outline of the thesis

The organization of this thesis is as follows:

- In Chapter 2, we introduce the basic concepts and the patterns of time series. It presents the definition of stationarity in time series and explains how to measure it.In addition, the chapter discusses the time series prediction methods and reviews the related literature. Furthermore, the model development and evaluation processes are examined. The error metrics used in these processes are explained and compared.

- In Chapter 3, we propose a novel hybrid method for time series forecasting which combines the individual ARIMA and ANN methods. Then, we extend the hybrid method with a feature-based approach. This section also details experimental setup used in the next chapters. Finally, the experimental results of the hybrid method, which outperforms the well-known methods in the literature, are investigated.

- In Chapter 4, we examine Empirical Mode Decomposition (EMD) multiscale decomposition technique and explain how it decomposes a given series into components. In addition, we mention why EMD is important in time series analysis and review the time series prediction literature. Moreover, in this chapter we present two more methods using EMD, called Two-level decomposition method and EMD-based hybrid method, respectively. Furthermore, we introduce the method selection algorithm which is implemented on these methods. Finally, we discuss the remarkable performance increase of prediction results in experiments by taking advantage of using EMD.

- In Chapter 5, we propose another novel algorithm, called Recursive EMD-based method. We explain how this algorithm recursively uses EMD with our hybrid method for forecasting univariate time series. The experimental evaluation results report that the recursive algorithm outperforms the previously examined and developed methods. Finally, we comprehensively discuss these improved prediction results together with demonstrating the other significant findings.

- In Chapter 6, we apply the algorithms developed step by step in previous chapters for forecasting Turkey Hydropower production and compare performance results.
- In Chapter 7, we briefly summarize our methods and obtained findings and conclude the thesis by presenting future research directions.

CHAPTER 2

TIME SERIES

> Don't never prophesy: If you prophesies right, ain't nobody going to remember and if you prophesies wrong, ain't nobody going to let you forget.
>
> *Mark Twain*

In this chapter, we introduce time series, prediction methods, and prediction evaluations in three sections. In the first section, we provide a fundamental knowledge about time series and its patterns. The importance of stationarity in time series is examined and the most widely used stationarity testing technique Augmented Dickey Fuller (ADF) is investigated. In the second section, the time series prediction methods are examined, in particular Autoregressive Integrated Moving Average Method (ARIMA) as a linear method and Artificial Neural Networks (ANN) as a nonlinear method. In addition, we describe the well-known hybrid methods which are the combination of linear and nonlinear methods such as Zhang, Khashei - Bijari, and Babu and Reddy methods. In addition, at the end of the this section, we summarize the other efforts on the literature on time series prediction. In the last section, evaluation techniques of a time series modelling are examined.

2.1 Introduction to Time Series

Time series include data points listed in time order. It is generally a sequence of discrete-time data which consists of points equally spaced in time. In general, time series are represented like y_t, where the subscript t symbolizes time, ranged between

$t = 1$ being the first observation and $t = T$ being the last observation. The complete of the series $y_1, y_2, \ldots, y_T$ is often called observation period. The measurements of the observations are generally at equally spaced intervals, such as every hour, day, or week, etc., therefore, the order of observations is crucial in order to preserve the temporal relationship.

Time series are mainly used in any domain of engineering and applied science which involves temporal measurements. For example, econometric, finance, statistics, astronomy, weather forecasting, control engineering, pattern recognition, signal processing and etc.

Figure 2.1 shows the Wolf's sunspot dataset, a time series from the science of astronomy. The time series contains annual records of incidence of spots on the sun surface from 1700 to 1987, giving a total of 288 observations. Sunspots are darker spots that appear temporarily in the photosphere of the Sun. Since it is directly related with the solar activity, sunspot occurrence is mainly used in predicting weather conditions and global warming. Therefore the dataset has been extensively used in numerous time series prediction models [1, 2, 3, 4, 5].

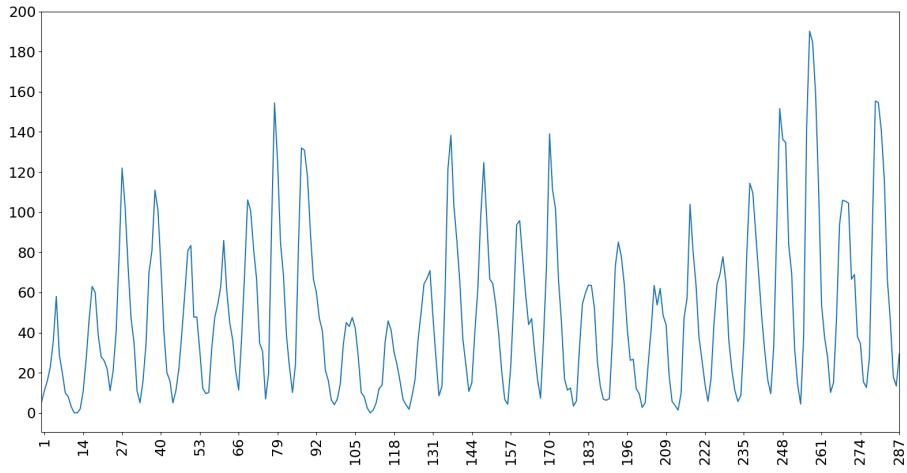

Figure 2.1: Wolf's sunspot series (1700-1987)

Figure 2.2 shows the Canadian lynx dataset, a time series from the social sciences. The Lynx dataset contains the annual number of trapped lynx in the Mackenzie River

district of Northern Canada between 1821–1934, giving a total of 114 observations. The Lynx time series is another extensively analyzed time series data in the literature [1, 2, 4, 5].

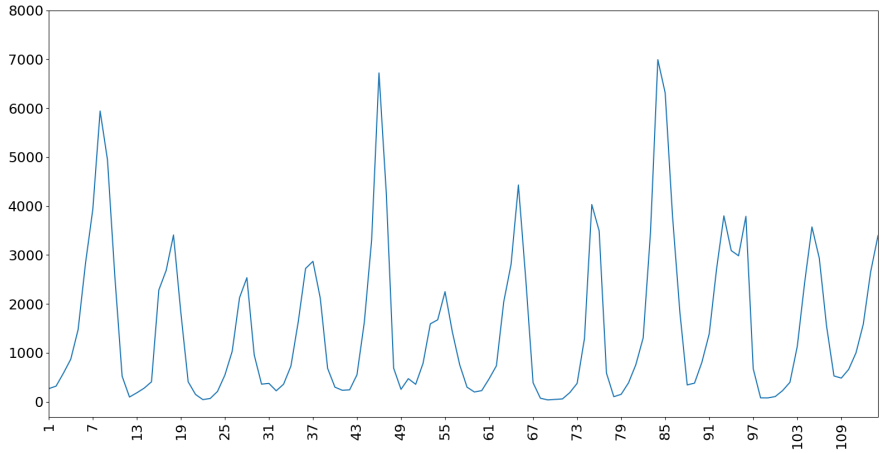

Figure 2.2: Canadian lynx data series (1821-1934)

Figure 2.3 shows the British pound/US dollar (Gbp/Usd) exchange rate dataset, from the field of economics and finance . The time series contains weekly observations from 1980 to 1993, giving 731 data points. Like The Sunspot and the Lynx dataset, the GbpUsd is another benchmark dataset widely used in time series analysis literature [1, 2, 4, 5].

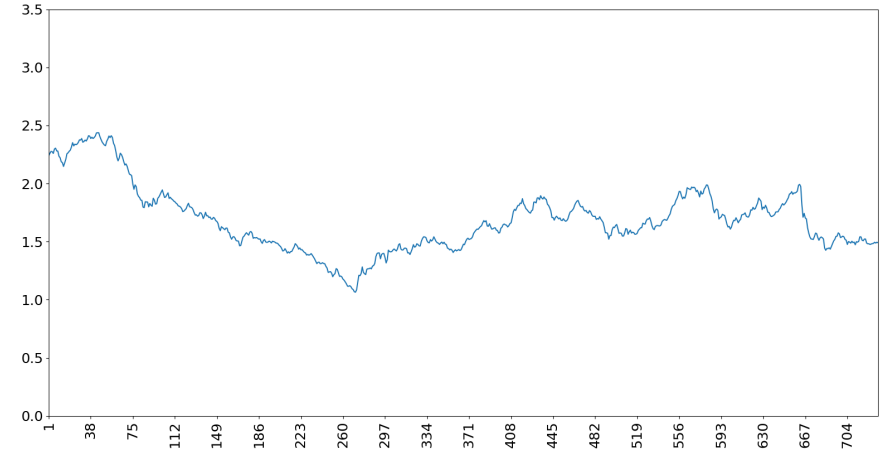

Figure 2.3: Weekly British pound/US dollar exchange rate series (1980-1993)

Figure 2.4 shows Turkey intraday electricity market price dataset from the field of energy. The dataset is publicly available [6]. The Intraday dataset includes 581 observations which consist of daily averaged prices from July 2015 to December 2017. As compared to datasets such as Sunspot and Lynx, the data pattern in this dataset is also highly fluctuating .

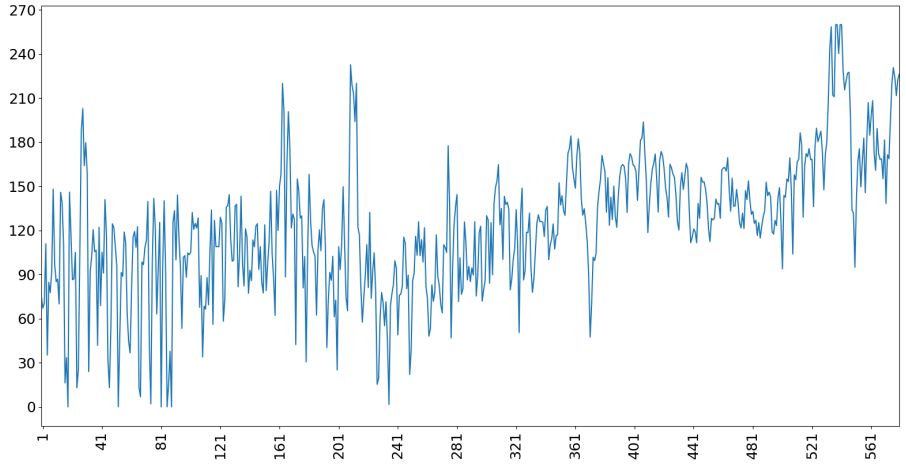

Figure 2.4: Electricity price of intraday market in Turkey (Jul. 2015 - Dec. 2017)

2.1.1 Patterns of Time Series

There are various patterns that affect the characteristics of time series [7]. These components are : Trend, Seasonal, Cyclic, and Irregular.

- **Trend**: When a given time series has a long-term increase or decrease, this phenomenon is termed as trend. Trend may not always be in a time series or it may be partial. For example, while there is not an explicit trend in the Sunspot time series shown in Figure 2.1, there is a decreasing trend in the first quarter of the GbpUsd dataset shown in Figure 2.3. Besides, it can be quite difficult to determine whether a time series contains a trend. In the Intraday dataset shown in Figure 2.4, although there is definitely a general upward movement throughout the observation period, it is never monotonic and it even shows decreases at some points. Nevertheless, this upward movement in Turkey intraday

electricity market price is also a sign of an implicit trend.

- **Seasonal**: Some time series has seasonal pattern in which variations are specific to a particular time range, e.g coat and boot sales increase in winter season while decrease in summer season. Seasonal factors such as the quarter of the year or the week of the month are always known frequency. Therefore, seasonal patterns provide very important information about time series for making proper future estimations. Due to the high volatility in exchange currency rates in the financial markets, it is very difficult to find a seasonal pattern. For example, as seen in Figure 2.3, GbpUsd currency rate dataset does not have any seasonal pattern.

- **Cyclic**: A cyclic pattern exists when a time series exhibits incline and decline which are not of a fixed frequency. Unlike seasonal patterns, the fluctuations are not of a fixed frequency. The term "business cycles" in economics such as depression, prosperity are the best known examples of these cyclic fluctuations. In Figure 2.1, sunspots dataset shows a periodic movement of approximately 11 years (the elapsed time from one maximum to next maximum). Since this movement does not occur in a fixed frequency, it is termed as a cycle rather than seasonal.

- **Irregular**: Given that the unexpected influences, time series might include irregular or random patterns. These patterns might be caused by either extreme events which can not be defined by any statistical technique. For example, as seen in Figure 2.4, daily Turkey intraday electric market price has excessively exceeded the average price three times in the given time period and reached 220 Turkish Lira. In addition, high volatility is also the cause of irregular patterns. Volatility indicates the variation of a time series over time. Therefore, high variance or equivalently high volatility means that there is a high irregular pattern. For example, a highly volatile dataset, the GbpUsd shown in Figure 2.3 which contains numerous changing turning points and irregular patterns.

The effect of these four patterns on time series can be shown as follows:

$$Y(t) = f(T(t), S(t), C(t), I(t)) \tag{2.1}$$

where $Y(t)$ is the time series observed; f is the additive or multiplicative model that characterizes the particular time series; $T(t)$, $S(t)$, $C(t)$, and $I(t)$ are trend, seasonal, cycle and irregularity, respectively. Since all time series include irregular patterns, pure trend, seasonal and cyclical patterns or combinations of them are seen together with the irregular patterns.

2.1.2 Concept of Stationarity

Stationarity is a very important feature in time series. Because it is a necessary condition to create high accuracy predictions [8]. The concept of stationarity in time series occurs when there are not any irregular movements in the series (e.g. sustained increase or decrease, high fluctuations) and statistical properties such as mean and variance do not depend on the time which the time series is observed. Therefore, time series including the patterns which affect the time series values at different times such as trend and seasonality are not regarded as stationary series.

In literature there are two different type of stationary process, *strongly* and *weakly* stationarity [9], respectively.

A given time series y_t is called as strongly stationary if the joint distribution of $(y_1, y_2, ..., y_t)$ and $(y_{1+k}, y_{2+k}, ..., y_{t+k})$ are identical for all t, where k is an arbitrary positive integer and $(1, 2, ..., t)$ is time period. Therefore, the joint distribution of any possible set of variables of a strongly stationary time series does not depend on the time [10]. Since it is empirically hard to verify this condition, weakly stationary is considered in practical applications. A weakly stationary time series has constant mean and variance and covariance values do not depend on time. More specifically, a given time series y_t is called weakly stationary if (1) $E(y_t) = \mu$ which is a constant, (2) $VAR(y_t) = \sigma^2$ which is a constant, and (3) $Cov(y_t, y_{t-l}) = \gamma_l$ which does not depend on t, only depends on l. As a result, weakly stationarity in time series indicates that data fluctuates with a constant variation around a certain level. With this property of the weakly stationary, future data is easily inferred, which makes it very important to make predictions in the time series [11].

Figure 2.5 shows four different kinds of time series charts obtained using the sine

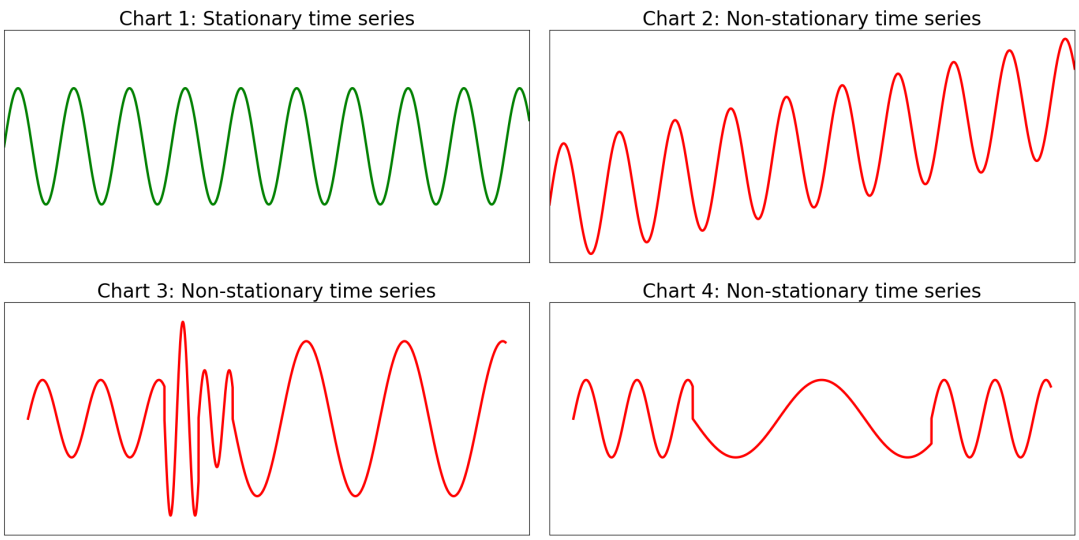

Figure 2.5: Examples of stationary and non-stationary time series created from sine functions

function. Chart 1 is a stationary time series as it satisfies all the conditions of the weakly stationary. On the other hand, Chart 2 and Chart 3 violate the properties of having a constant mean and variance, respectively. Therefore these two time series are not stationary. Similarly; since the covariance value of Chart 4 varies with time, it is non-stationary.

In order to determine whether a given time series is stationary or not, unit root test techniques are applied [10]. The existence of a unit root on a given dataset indicates that there is an unpredictable systematic pattern, which implies that the corresponding data is non-stationary. There are many such kinds of techniques in statistical theory, such as augmented Dickey-Fuller (ADF) [12], Kwiatkowski, Phillips, Schmidt and Shin (KPSS) [13], Phillips-Perron (PP) [14], etc.

ADF is the most widely used and well-known unit root test [10]. It uses the autoregressive model of order p, AR(p) as follows:

$$y_t = c_t + \beta y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + e_t \tag{2.2}$$

where c_t is the constant and $\Delta y_j = y_j - y_{j-1}$ is the differenced series of y_t. Then,

ADF , t-ratio can be shown as following:

$$ADF - test = t - ratio = \frac{\hat{\beta} - 1}{std(\hat{\beta})} \tag{2.3}$$

where $\hat{\beta}$ is the least squares estimate of β. The first differencing of the Equation 2.2 can be rewritten as follows:

$$\Delta y_t = c_t + \beta_c y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + e_t \tag{2.4}$$

where $\beta_c = \beta - 1$. The null hypothesis $H_o : \beta_c = 0$ implies that there is a unit root and the corresponding process is non-stationary. The alternative hypothesis $H_a : \beta_c < 0$, implies that there is not a unit root and the corresponding process is stationary.

Table 2.1 indicates t-ratio and p-value of ADF test results of datasets; Sunspot, Lynx, GbpUsd, and Intraday which are shown in Figure 2.1, Figure 2.2, Figure 2.3, and Figure 2.4, respectively. The more negative p-value of the ADF test result means the stronger rejection of the existence of unit root for a given time series. Therefore, a negative ADF result implies that the given dataset is stationary. The well accepted threshold is 0.05, datasets with p-value results lower than this value are considered stationary.

Table 2.1: t-ratio and p-value of the ADF-test results of datasets

Datasets	**t-ratio**	**p-value**
Sunspot	-5.116	0.083
Lynx	-5.871	0.056
GbpUsd	-2.056	0.581
Intraday	-3.674	0.273

ADF stationarity test result of the Sunspot dataset is 0.083 which is greater than the threshold 0.05. This implies that there is a unit root on the dataset, thus the dataset can be regarded as a non-stationary time series. Similarly, p-values of the GbpUsd and the Intraday datasets are 0.581 and 0.273, respectively, which are highly greater than the threshold which implies that they are highly non-stationary datasets. This non-stationarity can be even seen in their plots, given in Figure 2.3 and Figure 2.4, which show numerous changing turning points in their corresponding series. On the

other hand, the p-value of the Lynx dataset is 0.056 which implies that the dataset is almost stationary.

2.2 Time Series Prediction

In literature, various forecasting methods have been proposed which use linear and nonlinear models separately or combination of both [4]. Linear methods interpret the pattern of a time series through linear correlation. These methods expect that a given time series is quite regular with not including sudden movements, in order to have high accuracy results in prediction [15]. However, there are sudden movements in variation and extreme values in many real world time series such as financial data, renewable energy data, etc. This nonlinear behaviour could not be efficiently modelled by a linear time series [16, 17, 18]. In order to capture this nonlinearity, numerous nonlinear time series prediction methods have been proposed in literature [19]. Nonlinear methods are able to model a complex structure in time series which includes sudden erratic bursts with random variables. Although, nonlinear time series methods are able to achieve more consistent and accurate prediction results as compared to linear methods on nonlinear datasets, there are several studies report that statistical and conventional linear methods are capable of having better approximation under the ideal conditions which do not include high volatility and multicollinearity [20, 21, 22]. In short, each type of method outperforms in different domains. It is apparent that there is no universal model which is suitable for all circumstances. In order to overcome this limitation, various hybrid techniques have been proposed which aim to take advantage of the unique strength of each different type of models [4].

2.2.1 Linear Methods

Linear methods assume that the future data values are linearly dependent on the current and past data values. In the simplest case, linear regression model defines the linear relationship between the prediction variable y and a single predictor x, as follows:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{2.5}$$

where β_0 is the intercept value, β_1 is the regression slope which shows the change in prediction result y when the predictor x changes one unit, and ε_t denotes the random error which captures the effect on prediction results of all factors other than the predictor. As seen, the model assumes that there is a linear relation between y and x by this linear equation. However, it is extremely difficult to have linearly linked variables in real world problems.

Autoregressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) is widely used linear time series forecasting method that is used in numerous applications including finance [23], engineering [24], social sciences [25], agriculture [26], and energy [27]. ARIMA models are integration of Autoregressive models (AR) and Moving Average models (MA). The model building process of ARIMA depends on Box-Jenkins methodology [28] where it provides a step-by-step procedure for AR and MA analysis.

The method is composed of three main steps in order: Identification, estimation parameters and forecasting steps. In the identification step, stationarity check is performed on given time series data. A stationary time series' statistical characteristics such as mean and autocorrelation patterns constant over time. When a trend or heteroscedasticity is observed on a time series data, differencing or power transformation methods are generally applied to remove the trend and make variance stabilize. If stationary property is not satisfied after the first attempt, differencing (or power transformation) method is performed continuously until non-stationary property is disposed. If this operation is performed d times, the integration order of ARIMA model is set to be d. Thereafter, an autoregressive moving average(ARMA) is applied on the resultant data as follows:

Let the actual data value is y_t and random error ϵ_t at any given time t. This actual value y_t is considered as a linear function of the past p observation values, say $y_{t-1}, y_{t-2}, ..., y_{t-p}$ and q random errors, say $\epsilon_t, \epsilon_{t-1}, ..., \epsilon_{t-q}$. The corresponding

ARMA equation is given in the following equation:

$$\begin{aligned} y_t &= \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + \alpha_p y_{t-p} \\ &+ \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - ... - \theta_q \epsilon_{t-q} \end{aligned} \tag{2.6}$$

In Equation 2.6, the coefficients from α_1 to α_p are Autoregression coefficients, θ_1 to θ_q are Moving Average coefficients. Note that random errors ϵ_t are identically distributed with a mean of zero and a constant variance.

Similar to the d parameter, p and q coefficients are referred to as the orders of the model. When q equals to zero, the model is reduced to AR model of order p. If p is equal to zero, the model becomes MA model of order q. The main issue in ARIMA modeling is to determine the appropriate model orders (p, d, q). In order to estimate the order of the ARIMA model, Box and Jenkins [28] proposed to use correlation analysis tools, such as the autocorrelation function (ACF) and the partial autocorrelation function (PACF). In literature, there are various alternative order selection methods, for sample, Gaussian maximum likelihood estimation (GMLE) [29] which is computationally more complex, information-theoretic approaches such as the Akaike's information criterion (AIC)[30] and the minimum description length (MDL) [31]. When model coefficient estimation is finalized, the future values of the time series data are forecasted using available past data values and estimated model coefficients.

As a result, ARIMA is a linear method which means the future value of a variable to be forecasted is assumed to be linear function of past observations. As a consequence, time series data that is fed to ARIMA is expected to be linear and stationary. However, real world problems are often nonlinear [32], thus ARIMA could be inadequate for complex nonlinear real world problems.

2.2.2 Nonlinear Methods

Most of the real world time series are generated by nonlinear dynamic equations and features such as asymmetric cycles, time-changing momentums, heteroscedastic-

ity. As a result of it, they can not be efficiently and effectively modelled by a linear method [33]. In the simplest case, nonlinear regression model defines the nonlinear relationship between the prediction variable y and a single predictor x with a nonlinear function logarithm log, as follows:

$$y_t = \beta_0 + \beta_1 log x_t + \varepsilon_t \tag{2.7}$$

where β_0 is the intercept value, β_1 is the regression coefficient, and ε_t denotes the random error. A one unit change in predictor x_t results in nonlinear change in the prediction varibel y.

Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are one of the most widely used algorithms [34] in many fields, such as finance [35], energy [36], hydrology [37], and network communications [38]. ANNs have several advantages over ARIMA and other forecasting models. Firstly, ANNs are capable of fitting a complex nonlinear function, therefore it does not need to make data stationary as ARIMA does. This ability helps ANNs to approximate any continuous measurable function with arbitrarily desired accuracy [39, 40]. Moreover, ANNs are adaptively data-driven in nature which means ANN models can be adaptively formed based on the features of time series data [1]. Finally, ANNs provides simplicity as compared to the traditional polynomial, spline, and trigonometric expansions in terms of used parameter. While only linear parameters are enough for ANNs, other expansions requires many parameters in order to achieve the same approximation rate [2, 41].

The architecture of the most widely used ANN model in time series forecasting, which is also called as multilayer perceptrons, contains three-layers (see Figure 2.6). The neurons of the processing units are acyclically linked. In order to model time series data using such a network, nonlinear function f of y_t sequence from y_{t-1} to y_{t-N} is constructed as shown in the following equation:

$$y_t = w_0 + \sum_{j=1}^{H} w_j f(w_{0j} + \sum_{i=1}^{N} w_{ij} y_{t-i}) + e_t \tag{2.8}$$

where, at any given time t, w_{ij} and w_j are model weights and H and N are the number of hidden and input nodes, respectively. In this equation, e_t corresponds to a noise

or error term. The transfer function of the hidden layers f in ANN architecture is generally a sigmoid function.

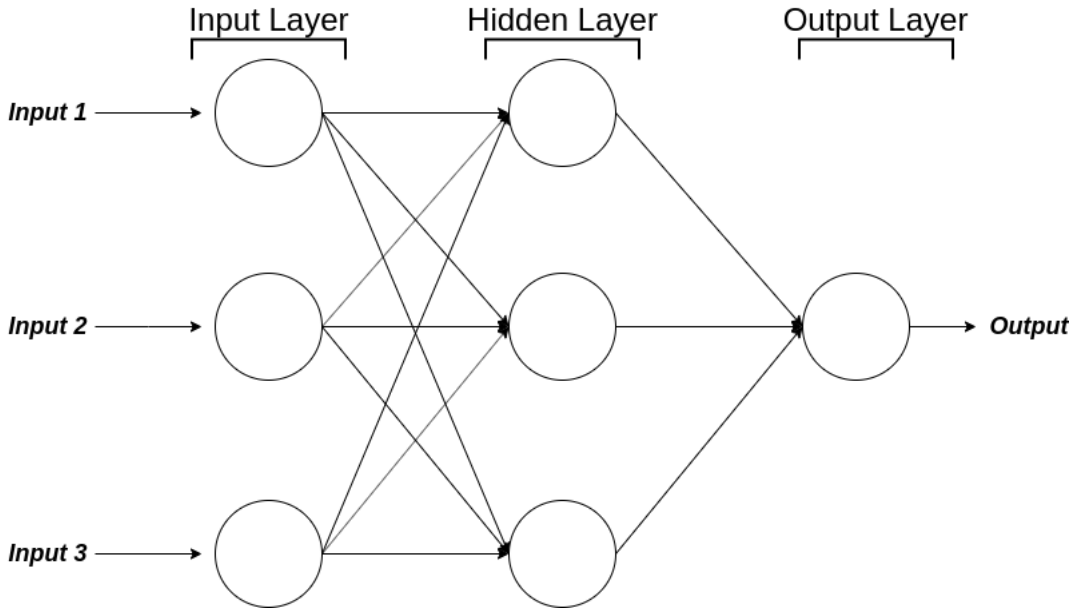

Figure 2.6: A sample neural network diagram (multi-layer perceptron)

The power of ANN comes from its flexibility to approximate any continuous function by changing the number of layers N and hidden nodes H. The choice of number of layers and the nodes at each of them play important role in ANNs' forecasting performance. Large numbers of N and H can give very high training accuracies, but since it tends to memorize the training data, it suffers from overfitting. On the other hand, a too simple network of ANN leads to poor generalization. Unfortunately, there is no systematic set of rules to decide the value of these parameters. Thus, extensive number of experiments are required to tune functions and the parameters.

As a result, ANN provides a flexible computation framework for nonlinear modeling in a wide range of applications. Due to its flexible architecture, the number of layers and the neurons at each layer can be easily varied. In addition, ANN does not require any prior assumption, such as data stationarity, in model building process. Therefore, the network model is largely determined by the characteristics of the data. In literature, numerous studies state that ANNs are remarkably and consistently more accurate than the traditional linear methods. However, some other studies report that the accuracy performance of ANN could be inconsistent in some specific situa-

tions [2]. For example, Foster et al. [20] figures out that simple linear regression and exponential smoothing methods are better than ANNs in short and noisy time series. Similarly, Brace et al. [21] finds that ANNs do not produce better forecasting results as compared to many other statistical methods for power load prediction problems. Aras et al. [22] explains the selection of the final ANN architecture is very difficult and the inappropriate architecture causes overfiting/underfitting problems.

2.2.3 Hybrid Methods

It is apparent that there is no universal model which is suitable for all circumstances. While linear models such as ARIMA give better forecasting accuracy with stationary time series data, nonlinear methods such as ANN are more appropriate for non-stationary datasets. In order to take advantage of the unique strength of each different type of methods in a more general setting, hybrid methods are proposed [42].

In time series, it is difficult to completely identify data generation process [43] and characteristics of the data cannot be easily determined [44]. Therefore, a single linear or nonlinear prediction method might not sufficiently capture the time series pattern with high accuracy. By using hybrid methods, combining several models reduces the risk of using an insufficient or inappropriate model and achieves more accurate prediction results [4].

The common practice in hybrid techniques is to decompose time series data into its linear and nonlinear forms, then use appropriate types of models on them separately. Various hybrid methods have been proposed in literature. The best known and well accepted ones; Zhang's Method [1], Khashei-Bijari's Method [2], Babu-Reddy's Method [3] are examined in detail below. These three successive hybrid methods where each one has taken the previous one(s) as reference study. Since benchmark datasets are used in the evaluation of these hybrid methods, researchers are able to repeat their studies for their comparisons.

Zhang's Hybrid Method

For time series forecasting, Zhang proposed a hybrid ARIMA-ANN model[1]. The Figure 2.7 simply summaries the Zhang's method. The method is able to benefit from both linear and nonlinear patterns separately, by way of using both linear and nonlinear models. Then, these models are combined to improve overall forecasting performance.

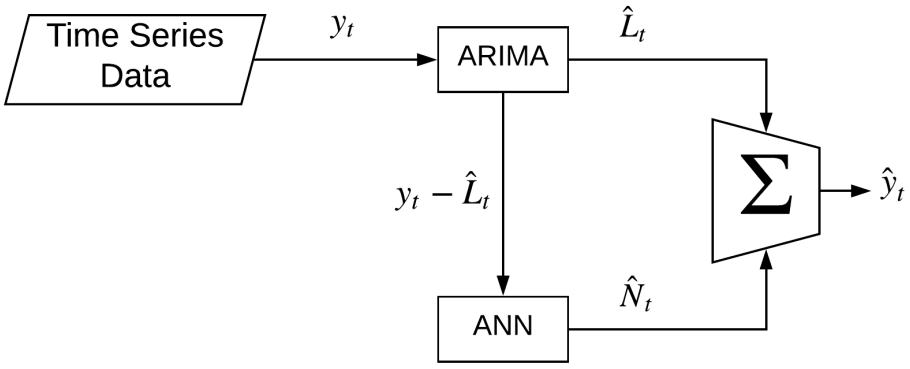

Figure 2.7: The architecture of Zhang's hybrid method

According to this model, it is assumed that time series data is a sum of linear and nonlinear components, given in the form of:

$$y_t = L_t + N_t \tag{2.9}$$

where L_t denotes the linear and N_t denotes the nonlinear component.Firstly, ARIMA is used with the given time series data and linear forecasts are obtained. Residuals from linear component is assumed to contain only nonlinear relationship. This method uses ARIMA to make forecast from the linear component and ANN from the nonlinear component. Then, these models are combined to improve overall forecasting performance. This method gives better foreceasting accuracy than using ARIMA and ANN methods individually, as seen in the experimental results in three well-known real datasets - the Wolf's sunspot data, the Canadian lynx data, and the British pound/US dollar exchange rate data which are shown in Figure 2.1, 2.2, 2.3, respectively.

The Zhang's hybrid method is applied in various applications in literature such as water quality [45], wind speed [46], selling prices of the fuelwood [47], and currency exchange rate [48].

Khashei-Bijari's Hybrid Method

For time series forecasting, in 2010 Khashei and Bijari proposed another hybrid ARIMA-ANN [2]. Similar to Zhang's model, this models assumes that any time series data composed of linear and nonlinear components. Likewise, linear component is modelled by using ARIMA and residuals are assumed to contain nonlinear relationship which can not be modelled by a linear approach. When stepped to next stage in which linear forecast and the computed nonlinear residuals are combined, this model is differentiated from the Zhang's model. This model claims that when the relationship between linear and nonlinear components is assumed to be additive, this relationship might be underestimated and degrade the performance [49]. In addition, one may not guarantee that the residuals of the linear component may comprise valid nonlinear patterns. Therefore, such assumptions are likely to lead to unwanted degeneration of performance if the opposite situation occurs [42].

Because of these reasons, a time series can be considered as a function of a linear and a nonlinear component, as shown in the Equation 2.10:

$$y_t = f(\hat{L}_t, \hat{N}_t) \tag{2.10}$$

where $\hat{L}_t$ is the linear component which is modelled by ARIMA, $\hat{N}_t$ is the nonlinear one modelled by ANN. For modelling nonlinear component, multilayer component is used by introducing past original values, present forecast of linear component, and past error data, that is residuals as inputs to ANN, as seen in the Figure 2.8 which simply summaries the Khashei-Bijari's method. ANN does not only get residuals as input but also takes past linear forecasts and actual values. Therefore, it is able to undermines the strict assumption that is additive association between linear and nonlinear components. The performed experiments support this approach by having better performance than Zhang's method in various application in terms of forecasting accuracy.

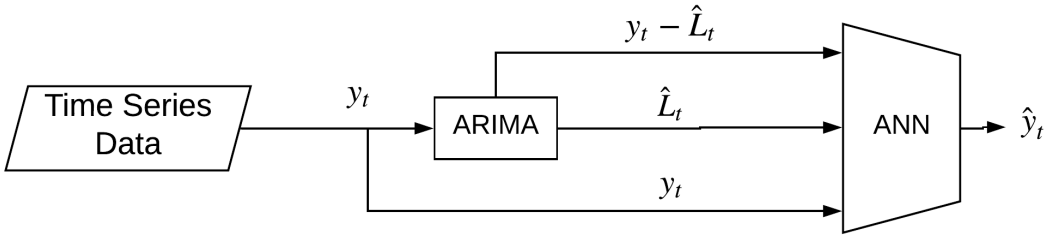

Figure 2.8: The architecture of Khashei-Bijari's hybrid method

Khashei-Bijari's hybrid approach is applied in various applications in literature such as machine industry for prediction of production values [50] , inspection facilities for predicting the number of goods subject to inspection [51], finance for predicting carbon price [52] .

Babu-Reddy's Hybrid Method

The hybrid ARIMA-ANN method proposed by Babu and Reddy [3] integrates moving average filter into hybrid ANN-ARIMA model. Like other methods, this model also assumes that any time series data is composed of linear and nonlinear components. However, this study emphasizes that neither Zhang nor Khashei and Bijari's methods decompose original time series data into its linear and nonlinear components; instead, they use a linear ARIMA model to extract the linear component and the error sequences is assumed to be nonlinear component. On the other hand, this study separates the linear and nonlinear components, then feed them into appropriate methods.

The architecture of the method is indicated in Figure 2.9. This method tries to fix Moving Average (MA) filter length until *kurtosis* value of the data becomes approximately 3. Kurtosis is a shape of a probability distribution which measure thickness or heaviness of the tails of a distribution. The kurtosis value is 3 if the data has normal distribution [53]. Shortly, the method aims to find out normal distributed component, which shows low volatility, in time series data by using kurtosis value. When the low-volatile component is separated from the original data, high-volatile compo-

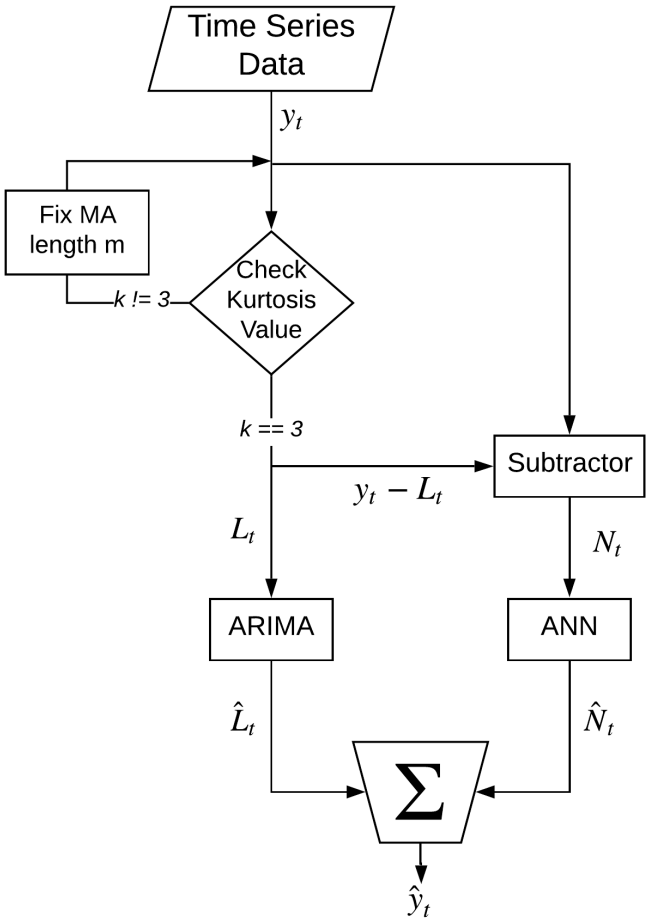

Figure 2.9: The architecture of Babu-Reddy's hybrid method

nent, which is assumed to be nonlinear, is achieved. In the final step, like in Zhang's method, the decomposed components are fed into ARIMA and ANN accordingly and forecast results are summed up to achieve final forecast.

In addition to these three well-known hybrid methods, there are various hybrid methods which combine linear and nonlinear methods and achieve remarkable results in terms of accuracy in different fields of application.

Firmino et al. [54] introduce an approach to recursively employ ARIMA and ANN methods until the remaining residuals behave as a white noise. In this approach, each model is supposed to correct a predetermined model's error. Their experimental results, which are performed on various finance datasets such as, S&P500 Index, Google Stock Value, Nasdaq Index and benchmark datasets such as Wolf's Sunspot,

and Canadian Lynx, shows that the proposed hybrid method is able to achieve better forecasting results as compared to single usage of ANN and ARIMA.

Panigrahi et al. [55] perform extensive experiments on sixteen different time series datasets by using combined nonlinear method ANN with linear methods ARIMA and Exponential Smoothing. The experimental results indicate that the hybrid methods statistically give promising results.

Domingos et al. [5] focus on searching for a proper combination function for linear and nonlinear models in hybrid methods. While linear component is modelled by ARIMA, nonlinear component is separately modelled by Multi-Layer Perceptron (MLP) and Support Vector Regression (SVR) which are also used in combination phase. The paper reports that the proposed hybrid system attains superior performance as compared with single models.

Durdu et al. [45] propose to use a ARIMA-ANN hybrid method to predict water quality in a river. The experiments, which are performed on 108-month observations, show that the hybrid model provides more robust and accurate results. Moreover, Koutroumanidis et al. [47] perform a similar approach for predicting fuelwood prices in Greece forest farms. Due to the high accuracy results, the proposed ARIMA-ANN hybrid method can be used by decision-makers to present more rational fuelwood market planning.

Zhou et al. [56] deploy a hybrid method combining ARIMA and neural network methods for predicting the prevalence of schistosomiasis in the future four years. The paper states that since the hybrid method is able to produce high accuracy results, it can be used for future monitoring in the prevalence of schistosomiasis.

Ruiz-Aguilar et al. [51] aim to predict the number of goods subject to inspection at European Border Inspections Post using a hybrid method combining Seasonal Autoregressive Integrated Moving Average (SARIMA) and ANN. In this method, the time series is firstly modelled by SARIMA, then the output of SARIMA and historical observations are fed into an ANN model to generate final prediction. In the experiments, the hybrid method outperforms the single methods used separately. The same approach is applied by Jeong et al. [57] for determining the annual energy cost

budget in educational facilities. Similarly, they obtain better accuracies as compared to single usage of SARIMA and ANN.

Wongsathan et al. [58] combine ARIMA and ANN methods to predict the pollution in Chiang Mai city moat area. The given time series is firstly modelled by ARIMA, then the residuals are predicted by ANN to discover non-stationary pattern. The authors assume that the data is a composition of stationary and non-stationary components which are linearly related. Therefore, the predictions of both methods are summed and presented as final result. Their experiments demonstrate that the hybrid method outperforms over single usage of ANNs and ARIMA respectively.

Cadenas et al. [46] propose to predict hourly wind speed with a hybrid method. Since wind speed data is highly volatile, a linear time series prediction method ARIMA is not sufficient to capture nonlinear tendencies. Thus, the authors deploy an ANN method for modelling obtained errors that ARIMA could not identify. In the experiments they show that this hybrid method gives higher accuracy performance results for wind speed prediction.

Nie et al. [59] uses a combination of ARIMA and Support Vector Machines (SVMs) for short-term load prediction. Firstly, the daily load data is predicted by ARIMA, then the results are subtracted from the original time series. The remaining component is considered as deviation which shows non-stationary patterns. Therefore, this deviation component is modelled by ANN and the final result is obtained. The experiments show that the hybrid method is much better than the two separate models of themselves.

2.2.4 Prediction Evaluation

Model evaluation process is as important as model development process. According to accuracy performance results, the process of model development including selection of a proper method, hyperparameter optimization, etc. could be reevaluated until obtaining the most appropriate model. The accuracy performance of predictions can only be determined by considering how accurate a model performs on a new dataset which is not used while developing the model [60, 61].

While applying a particular model on a time series, it is common practise to divide the available data into two non-overlapping parts, training and test datasets. Train dataset, is also called in-sample set, includes the observations of a time series which is used to develop the desired model. On the other side, test dataset, is also called out-of-sample set, is used to evaluate the established model. The observations in test dataset are never used while constructing the model for evaluating how accurate the fitted model performed on predicting these values [62].

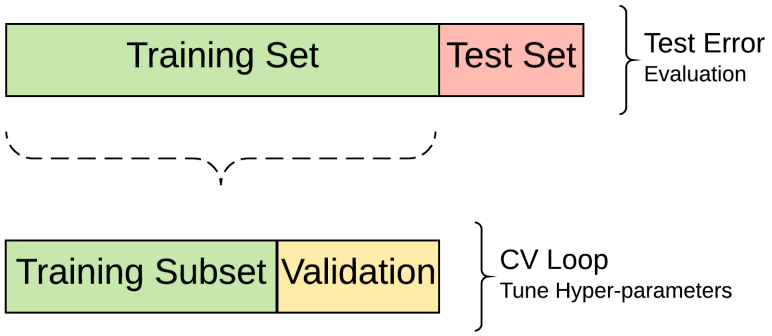

Figure 2.10: Example of hold-out cross-validation process

A small subpart of the training set (20% of the set, but there is not a theoretical foundation for this percentage [9]) is separated for validation process and is called validation set. This set is used to tune hyperparameters and produce robust performance measurements in a given time series prediction method. The most common use case of validation set is plotted in Figure 2.10: Different models are created with different parameters by using the observations in the training set. These models then perform predictions with values in the validation set. By comparing the performances of these predictions, the most proper model is chosen which produces better accuracy than other models. This process is called hold-out Cross-Validation (CV) [63].

It is important to find the best prediction method that produces the most accurate results in the evaluation process. Various error metrics have been developed in the literature to measure *"the most accurate"* one [9].

The error metrics define *"error"* e_t as the difference between actual observed value y_t and its prediction $\hat{y}_t$ at time t. This difference refers to the unpredictable part of the

corresponding observation. It can be written as:

$$e_t = y_t - \hat{y}_t \tag{2.11}$$

The well-known error metrics are described below in the literature [64].

- **Mean Forecast Error (MFE)** is defined as follows [65]:

$$\text{MFE} = \frac{1}{n}\sum_{t=1}^{n} e_t \tag{2.12}$$

  MFE measures the average difference between predicted values and observed ones. Since the error values are directly counted, Forecast Bias could be easily determined. For example, very low negative MFE values indicate that the model tends to underestimate bias, and very high positive values indicate that it tends to overestimate bias. However, it should be noted that zero MFE value does not indicate that the model performs without errors, rather; it demonstrates that the error distribution is balanced with minimum bias. Moreover, since the metric equally weights all values in time series, extreme errors are not penalized.

- **Mean Absolute Error (MAE)** is defined as follows [66]:

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} |e_t| \tag{2.13}$$

  MAE measures the average absolute difference between predicted values and observed ones. MAE is also called Mean Absolute Deviation (MAD). Unlike MFE, MAE does not give any information about the direction of errors. Instead, it measures the amount of overall error of the prediction. Therefore, the smaller the MAE results means that the higher the model success. On the other hand, like MFE it does not penalize the extreme errors.

- **Mean Absolute Scaled Error (MASE)** is defined as follows [67]:

$$\text{MASE} = \frac{n-1}{n}\frac{\sum_{t=1}^{n} |e_t|}{\sum_{t=2}^{n} |y_t - y_{t-1}|} \tag{2.14}$$

  MASE measures the average absolute difference between predicted values and observed ones, divided by the mean absolute error of the in-sample one-step

naïve forecast. Therefore it is scale-free, namely; MASE results are independent of the scale of the data. Thanks to this property, this metric could be used to compare predictions across datasets with different scales. MASE result means that the model is successful as it approaches zero. On the other hand, when it is greater than one, it means that the corresponding model is even worse than a naïve method. Similar to MAE and MFE, it equally penalizes positive and negative prediction errors.

- **Mean Percentage Error (MPE)** is defined as follows [68]:

$$\text{MPE} = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{e_t}{y_t}\right) * 100 \tag{2.15}$$

MPE measures the percentage of average difference between predicted values and observed ones. Thus, it becomes independent of the scale of measurement. Like MFE, it indicates the direction of prediction error. Similarly, MPE values close to zero only implies that the model does not have a specific bias; it does not provide information about the accuracy performance of the model. Like the previous ones, it does not punish extreme errors, it weighs all errors equally.

- **Mean Absolute Percentage Error (MAPE)** is defined as follows [69]:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{e_t}{y_t}\right| * 100 \tag{2.16}$$

MAPE measures the percentage of average absolute difference between predicted values and observed ones. Thus, it is a scale-free error metric like MPE and MSE. Since it takes the absolute value into account, it does not give information about the direction of the prediction error. Like MAE, MAPE value should be as small as possible for a good prediction performance. It does not penalize the extreme errors, either.

- **Mean Squared Error (MSE)** is defined as follows [70]:

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}{e_t}^2 \tag{2.17}$$

MSE measures the average squared prediction error. Because the error metric squares the errors, the knowledge of the direction of the overall error is lost.

Because of the same reason, MSE emphasizes the large errors, so that it penalizes extreme prediction errors unlike previous error metrics. MSE is not independent of scale, it is even very sensitive to small scale differences. Although it could be difficult to understand and interpret MSE results, it is more useful when large errors are particularly undesirable in the prediction model.

- **Root Mean Square Error (RMSE)** is defined as follows [66]:

$$\text{RMSE} = sqrt(MSE) = \sqrt{\frac{1}{n}\sum_{t=1}^{n} {e_t}^2} \tag{2.18}$$

Taking the square root of MSE yields RMSE. Therefore, all features specified for MSE also apply to RMSE.

In order to evaluate the prediction performance results of a model, the most widely used error metrics with different properties are discussed above. Each of these metrics is different from others and separately explains the amount, direction, and magnitude of the prediction error. Therefore, it is better to take more than one error metrics into account to capture characteristics of the model and the prediction errors.

CHAPTER 3

PROPOSED HYBRID METHODS

> Prediction is very difficult, especially if it's about the future.
>
> *Nils Bohr*

In this chapter, it is proposed a novel hybrid method for univariate time series forecasting which aims to overcome the limitations of the traditional hybrid methods by eliminating making strong assumptions [4]. In this method, the nature of nonlinearity is firstly characterized by the help of moving-average filter, then ARIMA is applied to the linear component. In the final step, ANN is used to combine the output of ARIMA, the nonlinear component, and the original data. In addition, this hybrid method is enhanced into a feature-based hybrid ARIMA-ANN model [71]. The features computed by statistical methods are presented to the hybrid method as variables that explain the given time series. The performed experiments show the effectiveness of the proposed methods in univariate time series forecasting.

3.1 The ARIMA-ANN Hybrid Method

Many decision processes need high forecasting accuracies in time series applications. Although there are numerous available time series models, none of them consistently gives the best results in various situation. There are two main challenges for making an accurate forecast. The first challenge is that underlying data generating process of time series cannot easily identified [44]. The second one is that non-hybrid individual models are generally insufficient to determine all the characteristics of the time series [1]. Many researches in time series forecasting literature show that hybrid

models improve the forecasting performances [42]. By taking the advantage of each individual method in a combined model, error risk of using an inappropriate method is reduced and more accurate results are obtained.

Each hybrid method mentioned in Section 2.2.3 brings different perspectives into time series forecasting problem. However, strong assumptions these methods make might degenerate their performances in certain circumstances. For example, Zhang's and Babu-Reddy's methods assume that the relationship between linear and nonlinear components is additive. If linear and nonlinear components are not additively associated and the relation is different (i.e., it can be multiplicative), the possible complex relationship between components might be overlooked and the forecasting performance might be degenerated. Another assumption that might not always hold is that residuals might not show nonlinear pattern property. Additionally, Zhang's and Khashei-Bijari's methods do not actually decompose data into linear and nonlinear components, but they assume that linear component can be extracted by ARIMA and error sequences shows nonlinear pattern. As a result, such assumptions may lead to low forecasting performances when unexpected scenarios take place.

3.1.1 Proposed Method

In this study, we propose a novel hybrid method for time series forecasting which aims to overcome the limitations of the traditional hybrid methods by eliminating the need to make strong assumptions [4]. In this method, nature of nonlinearity is first characterized by the help of moving-average (MA) filter, then ARIMA is applied to the linear component. In the final step, ANN is used to combine the output of ARIMA, the nonlinear component, and the original data. By this means, we do not make any assumptions in the extraction of components and in the modeling of components. The architecture of the proposed hybrid method is shown in Figure 3.1

The algorithm starts with data decomposition. In this method time series data y_t is considered as a function of linear L_t and nonlinear N_t components in the same way as given in Equation 3.1.

$$y_t = f(L_t, N_t) \tag{3.1}$$

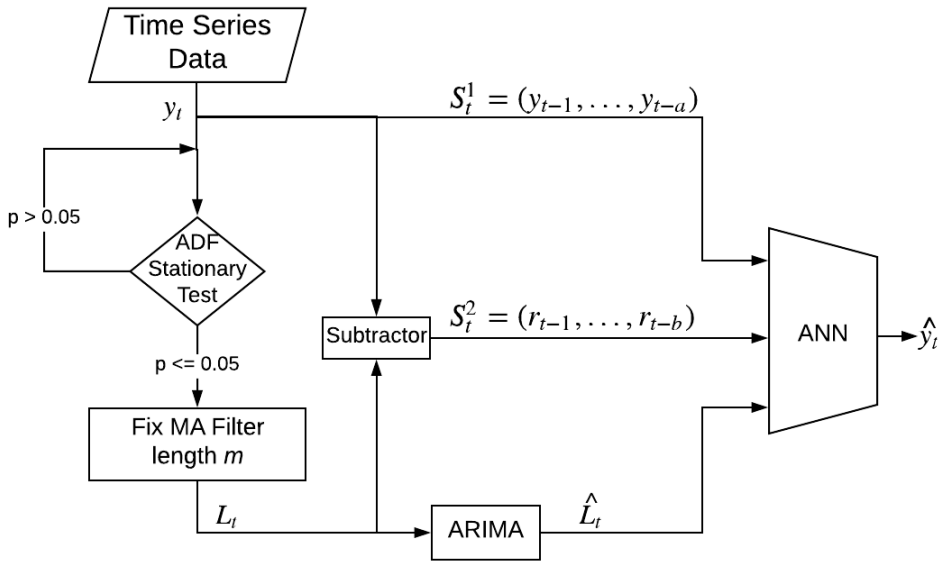

Figure 3.1: The architecture of the proposed ARIMA-ANN hybrid method

where $\hat{L}_t$ and $\hat{N}_t$ are the linear and nonlinear components, respectively. These two components are separated from the original data by using moving average (MA) filter with the length of m, as given in Equation 3.2.

$$l_t = \frac{1}{m} \sum_{i=t-m+1}^{t} y_i \tag{3.2}$$

While the linear component l_t has low volatility, the residual r_t, which is the difference between the original data and the decomposed linear data in Equation 3.3, shows high fluctuation.

$$r_t = y_t - l_t \tag{3.3}$$

In order for a proper decomposition, the length of the MA filter m has to be adjusted. Augmented Dickey Fuller (ADF) test which is unit root test (see Section 2.1.2) can be performed to determine whether a given data series is stationary or not. The existence of a unit root on a given dataset indicates that there is an unpredictable systematic pattern. The more negative ADF test result means the stronger rejection of the existence of unit root for a given time series. Therefore, a negative ADF result implies that the given dataset is stationary. The well accepted threshold is 0.05 which is also used in this study to adjust MA filter length.

After the linear component is achieved with MA filter, a linear model is constructed as

shown in Equation 3.4. The stationary component l is modeled as a linear function of past values of the data series $l_{t-1}, l_{t-2}, ..., l_{t-p}$ and random error series $\epsilon_{t-1}, \epsilon_{t-2}, ..., \epsilon_{t-q}$ in Equation 2.6 using ARIMA model.

$$\hat{L_t} = g(l_{t-1}, l_{t-2}, ..., l_{t-p}, \epsilon_{t-1}, \epsilon_{t-2}, ..., \epsilon_{t-q}) \tag{3.4}$$

where g is a linear function of ARIMA. The forecast values achieved from ARIMA model $\hat{L_t}$ and residual r_t are used in the next step.

Finally, nonlinear modeling ANN is used to implement functional relationship between components as indicated in Equation 3.1. The past observed data $y_{t-1}, y_{t-2}, ..., y_{t-a}$, present ARIMA forecast result of the decomposed stationary data $\hat{L_t}$, and residuals of the data decomposition $r_{t-1}, r_{t-2}, ..., r_{t-b}$ are fed to ANN as indicated in Equation 3.5:

$$\begin{aligned} S_t^1 &= (y_{t-1}, y_{t-2}, ..., y_{t-a}) \\ S_t^2 &= (r_{t-1}, r_{t-2}, ..., r_{t-b}) \\ \hat{y_t} &= f(S_t^1, \hat{L_t}, S_t^2) \\ \hat{y_t} &= f(y_{t-1}, y_{t-2}, \ldots, y_{t-a}, \hat{L_t}, r_{t-1}, r_{t-2}, ..., r_{t-b}) \end{aligned} \tag{3.5}$$

where f is the nonlinear function of ANN, a and b are parameters of the model which show how much we will go back in time to use as features to ANN. Time series data determines how many of those features in the residual path and observed data path are going to be used in the nonlinear model. For example, if the given data does not show volatility, then the residual variable b in Equation 3.5 might come out even as zero in tuning process. Likewise, a variable in Equation 3.5 is also empirically determined in the tuning process.

The proposed model does not only exploit the unique strength of single models, but also eliminates the three strong assumptions performed by other hybrid methods. Therefore, risk of low forecasting performance in unexpected situations is highly avoided. The competitive performance of our proposed algorithm is shown in our experimental results by using various type of datasets.

3.1.2 Experimental Results

The performance results of the proposed hybrid method along with the other methods discussed in this paper are evaluated on four different datasets. Three of them are well-known benchmark datasets - the Wolf's sunspot data, the Canadian lynx data, and the British pound/US dollar exchange rate data - which are indicated in Figure 2.1, Figure 2.2, and Figure 2.3, respectively. The other dataset is publicly available electricity price of Turkey Intraday Market, indicated in Figure 2.4. These four different time series data are originated from different disciplines and show different characteristics. While one dataset (e.g. GbpUsd) is highly non-stationary and volatile, another one (e.g. Lynx) shows linear behaviors. Similarly, while one of them includes seasonality (e.g. Sunspot), the remaining dataset(e.g. Intraday) does not. This diversity has enabled us to make our experiments with a wide spectrum of datasets.

In the experiments, only one-step-ahead forecasting is considered. In order to compare accuracy performances, three evaluation metrics are used: Mean Absolute Error (MAE), Mean Squared Error (MSE) and Mean Absolute Scaled Error (MASE) of which their equations are showed in Equation 2.13, Equation 2.17, and Equation 2.14, respectively. As metioned in Section 2.2.4, while MAE specifies the average of the absolute errors over the performed prediction, MSE measures the average of the squared error. Since prediction errors are squared in MSE metric, it gives relatively greater influence to larger errors. Hence, this property makes MSE useful when large errors are not desired. On the other hand, MAE does not penalize outliers like MSE, since it equally weights prediction errors in the average. Both metrics are good when comparing algorithms for a dataset. However, when comparing an algorithm for time series which have different scales, they might not be reliable, since both MAE and MSE results depend on the scale of the given data. Therefore, a scale-free error metric MASE can be used to compare forecast accuracy between scaled series.

To assess the forecasting performance of the different methods, each dataset is divided into training and testing sets. While the training data is used for model development, the test data is used to evaluate the established model. In order to tune the hyperparameters the methods, the last 20% of training dataset is used as validation set.

While choosing the right ANN model, since there is not an accepted method for network configuration, an empirical approach is performed. In this approach, ANN is methodically build for each combination of algorithm parameters and evaluated on the validation set. The achieved best configuration is used as a final model.

In addition, due to the fact that ANN performs random initialization and produces different results at each run, the methods which include ANN algorithm are executed 50 times and average results are reported. This number is not a hyperparameter to tune but high enough to get reliable and robust results. Table 3.1 gives the forecasting results of all examined methods on these all four datasets. The table also presents the results of the *naïve* method in order to have a baseline comparison. This method $(y_t = y_{t-1})$ uses the value at the previous time step $t-1$ to predict the expected outcome at the current time step t.

In order to assess whether the average results of the methods are statistically different from each other, t-test analysis is used [72]. As a result of the test, p-values are obtained for pairwise comparison between our method's and the other methods' error metrics. If p-value is less than 0.05, it can be accepted as our method is statistically different from the compared method in the test [72]. Table 3.2 presents the p-values for four pairwise comparison for all datasets. Since MASE is the scaled up of MAE, the test results in two metrics are same. To avoid the repetition instead of two of them, the results for only MAE are reported.

Table 3.2 shows that the proposed method is significantly different from all examined methods for all experimented datasets in MSE metric. In MAE metric, the proposed method is statistically significant than at least three of the examined methods.

Forecasts for Sunspot dataset

The Wolf's sunspot series (see Figure 2.1) has 0.083 ADF stationarity test (see Table 2.1) which is greater than the threshold 0.05. This implies that there is a unit root on the dataset, thus the dataset can be regarded as non-stationary time series. 288 observations in the dataset is divided into two samples: 221 observations between 1700-1920 years are considered as training data to develop the model, the last 67 ob-

Table 3.1: Performance comparisons of the proposed ARIMA-ANN hybrid method for all datasets

Datasets	**Methods** / ***Metrics***	Naïve Method	ANN	ARIMA	Zhang's Method	K-B's Method	B-R's Method	Proposed Method
Sunspot	*MAE*	22.96	14.23 ±0.40	13.37	13.14 ±0.34	10.62 ±0.26	11.39 ±0.22	**10.48** ±0.27
	MSE	920.73	353.12 ±24.26	306.97	289.31 ±23.01	205.08 ±15.06	239.90 ±22.89	**194.29** ±13.36
	MASE	1.01	0.62 ±0.01	0.59	0.58 ±0.01	0.47 ±0.01	0.50 ±0.01	**0.46** ±0.01
Lynx	*MAE*	0.2308	0.1249 ±0.0029	0.1198	**0.1003** ±0.0023	0.1025 ±0.0025	0.1102 ±0.0028	0.1013 ±0.0031
	MSE	0.0687	0.0241 ±0.0019	0.0231	0.0173 ±0.0016	0.0175 ±0.0016	0.0189 ±0.0019	**0.0162** ±0.0018
	MASE	1.1425	0.6185 ±0.0144	0.5932	**0.4966** ±0.0113	0.50757 ±0.0127	0.5457 ±0.0138	0.5016 ±0.0157
GbpUsd	*MAE*	512.41	428.55 ±60.45	435.72	429.52 ±48.54	406.22 ±24.65	436.34 ±36.87	**404.90** ±23.43
	MSE	3.97	3.47 ±0.37	3.52	3.45 ±0.30	3.10 ±0.22	3.50 ±0.33	**2.95** ±0.26
	MASE	1.297	1.085 ±0.153	1.103	1.087 ±0.122	1.028 ±0.062	1.104 ±0.145	**1.025** ±0.059
Intraday	*MAE*	21.54	20.10 ±1.242	20.22	19.16 ±0.925	19.79 ±1.708	19.50 ±0.981	**18.81** ±0.972
	MSE	797.03	617.46 ±37.17	652.72	594.09 ±27.10	600.93 ±46.75	619.67 ±37.07	**581.38** ±34.77
	MASE	1.12	1.04 ±0.06	1.05	0.99 ±0.04	1.03 ±0.08	1.01 ±0.05	**0.98** ±0.05

* Khashei-Bijari's method is shown as K-B's method.

** Babu-Reddy's method is shown as B-R's method.

*** MAE and MSE results are multiplied with 10^5 in GbpUsd dataset

**** The values shown with $\pm$ give the standard deviation values.

Table 3.2: p-value test result comparisons of the proposed ARIMA-ANN hybrid method for all datasets

Datasets	**Methods** / ***Metrics***	ANN	Zhang's Method	Khashei-Bijari's Method	Babu-Reddy's Method
Sunspot	*MAE*	0.001	0.001	0.007	0.001
	MSE	0.001	0.001	0.001	0.001
Lynx	*MAE*	0.001	0.135	0.038	0.001
	MSE	0.001	0.002	0.011	0.001
Gbp/Usd	*MAE*	0.017	0.022	0.124	0.037
	MSE	0.001	0.001	0.036	0.001
Intraday	*MAE*	0.001	0.012	0.001	0.001
	MSE	0.001	0.049	0.017	0.001

servations between 1921-1987 years are considered as test data and used to evaluate the model performance.

In Sunspot dataset, when ARIMA is individually used as a forecasting method, we also set the order of ARIMA to 9 (AR(9)) as same as the other many studies [1, 2, 3]. When ANN is individually used as a forecasting method, similar to these studies, three layered 4 x 4 x 1 ANN architecture is used which is composed of four input nodes, four hidden nodes, and one output node.

In the proposed method, the linear component comes out when the MA filter length is 15. After using the filter, the achieved linear component has 0.006 stationary test result which indicates its stationarity, since it is a value less than the threshold, 0.05. The best fitted neural network in the final step of the proposed hybrid method has 7 nodes in the input layer where 4 of them are observed values, 2 of them are residuals, and one node is assigned for the result of linear component forecast. According to our experiments, when the number of hidden nodes are adjusted to same number of nodes as in the input layer, the best fitted ANN model is achieved.

When the numerical results of the Sunspot dataset given in Table 3.1 are analyzed, individual methods such as ARIMA, ANN have apparently lowest performance as compared to hybrid methods. This suggests that either ARIMA or ANN, when individually used, do not capture all patterns in the data series. Therefore, combining two methods by taking advantage of each of them can be an effective way to overcome this limitation. We indeed observe in Table 3.1 that the hybrid methods Zhang, Khashei-Bijari, and Babu-Reddy methods produce better results as compared to individual ones. However, they produce lower forecasting performance than our proposed hybrid method. The assumptions those hybrid methods make can be restricting in many situations. Our proposed hybrid method eliminates those assumptions and yields better generalization performance. Figure 3.2 compares the actual and forecast values for all examined methods.

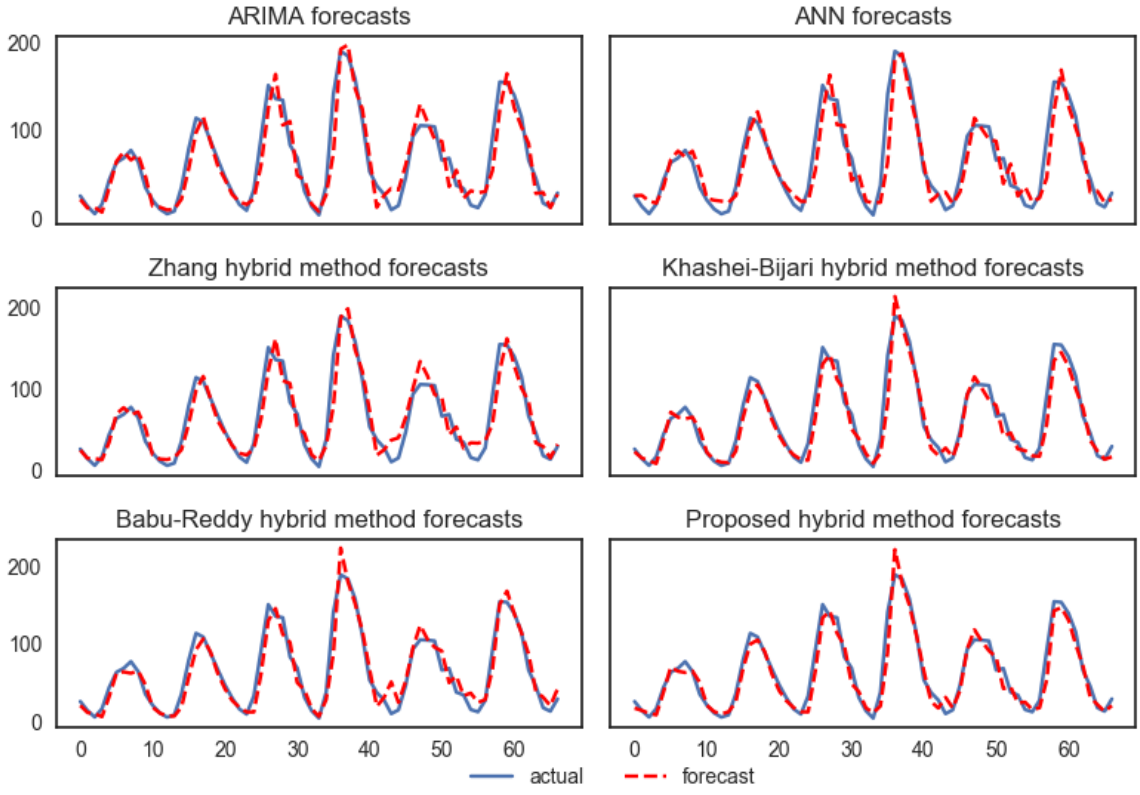

Figure 3.2: The Sunspot data forecasts using various methods

Forecasts for Lynx dataset

The Lynx series (see Figure 2.2) has 0.056 ADF stationarity test result (see Table 2.1) which implies that the dataset is almost stationary. There are 114 observations in the data, corresponding to the period of 1821-1934. The first 100 observations between 1821-1920 years are considered as training data to develop the model, the last 14 observations between 1921-1934 years are used as test data to evaluate the model performance. In addition, like in other studies [1, 2], the logarithms (to the base 10) of the data are used in the analysis.

When ARIMA is used as an individual model, we used the AR model of order 12 (AR(12)) for Lynx dataset which is also used by [1, 2]. Similar to these studies, three layered 7 X 5 X 1 ANN architecture is used when ANN is individually used as a forecasting method.

In the proposed method, the linear component is extracted from the Lynx dataset when the MA filter length is 5. The relatively short MA filter length was expected, since the ADF test result shows a certain level of stationarity in the data. As a result of MA filter, the achieved linear component has 0.006 stationary test result which indicates even more stationarity to be properly modeled by ARIMA. The best fitted neural network in the final step of the proposed hybrid method has 9 nodes in the input layer where 5 of them are observed values, 3 of them are residuals, and one node is assigned for the result of linear component forecast. According to our tuning experiments, when the number of hidden nodes are adjusted to the same number of nodes as in the input layer, the best fitted ANN model is achieved.

Figure 3.3 compares the actual and forecast values for all examined methods. In this dataset, among the individually used methods, ARIMA gives better accuracy as compared to ANN in contrast to the Sunspot dataset (see in Table 3.1). This is most likely due to the fact that Lynx dataset is more stationary dataset compared to Sunspot dataset. Due to this relative stationarity, the effect of hybrid methods might not be easily observed. Although we have circumstances which do not necessarily favor the data decomposition and model combination, hybrid methods do not give a lower performance than the individual linear models, and they even provide better

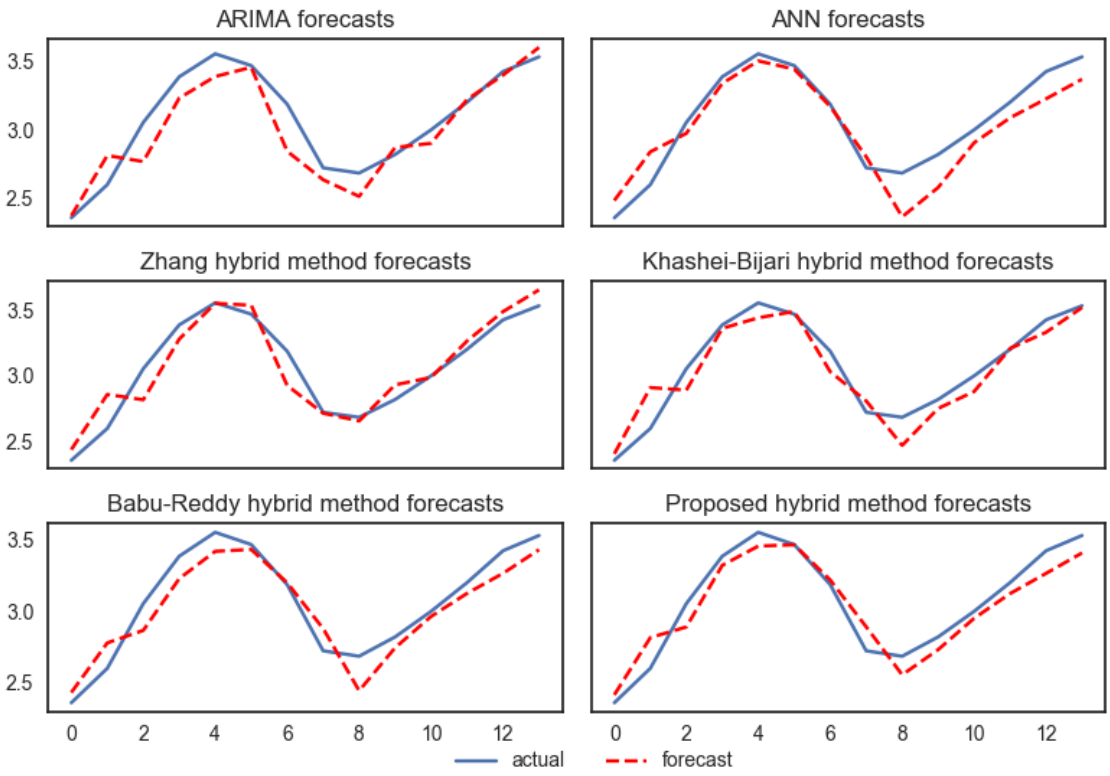

Figure 3.3: The Lynx data forecasts using various methods

results. The highest performance is mostly achieved by our proposed hybrid method.

Forecasts for GbpUsd dataset

Predicting exchange rate is an important yet difficult task due to high volatility. ADF stationary test result of the dataset is 0.58 (see Table 2.1) which is highly greater than the threshold 0.05. This implies that the dataset is highly volatile and non-stationary. This non-stationarity can be even seen in the plot, given in Figure 2.3, which shows numerous changing turning points in the series. Similar to other datasets, the experimental setup is same as in previous hybrid studies [1, 2] where data is transformed using natural logarithmic function and separated into two samples. The first 679 observations from 1980-1992 years are considered as training data to develop the model, the last 52 observations between 1992-1993 years are used as test data to evaluate the model performance.

In this dataset, when ARIMA is individually used, rather than using regression type

of model in the ARIMA itself, random walk model is chosen as best-fitted ARIMA model. This approach has been used by Zhang [1] and also been suggested by many studies in the exchange rate literature [73]. In this model, the most recent observation is the best guide for the next forecast. When ANN is individually used as a forecasting method, the best fitted ANN is set as three layered 7 x 6 x 1 architecture.

In the proposed method, in order to decompose this highly volatile data, MA filter length comes out to be 40. As a result of this decomposition, ADF test result of the obtained component is 0.007 which indicates the stationarity of the component. To compute the best final forecast in the proposed model, the ANN is constructed as three layered 9 x 9 x 1 architecture. In this architecture, input layer is composed of the last 5 of observed values, the last 3 of residuals, and the result of linear component forecast.

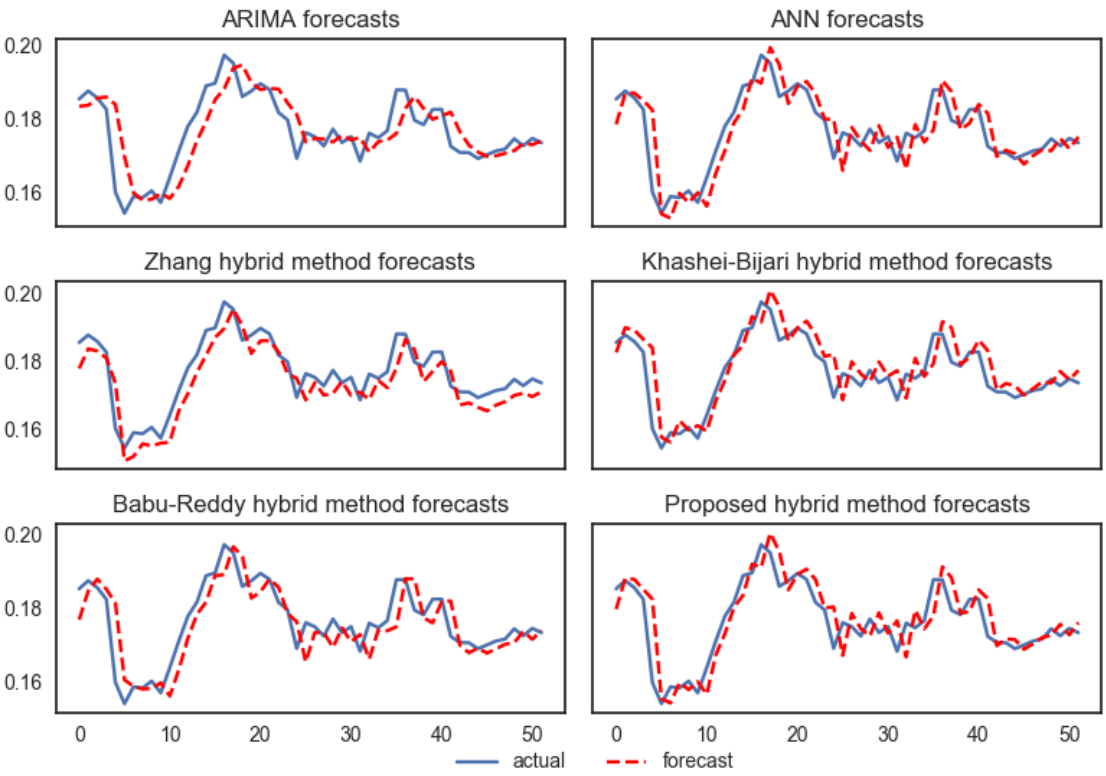

Figure 3.4: The GbpUsd data forecasts using various methods

Results of the GbpUsd dataset forecasts are compared in Figure 3.4. Both ANN and hybrid methods have much better performance than the individual ARIMA method for a highly fluctuating forecast horizon. The proposed hybrid method is able to

capture this volatile pattern much better and outperforms the other methods in all error metrics as seen in Table 3.1.

Forecasts for Intaday dataset

The last analyzed dataset is Turkey intraday electricity market price data. As compared to other datasets such as Sunspot and Lynx, data pattern in this dataset is also highly fluctuating (see Figure 2.4). The ADF stationary test result of the dataset shows 0.27 (see Table 2.1) value which is highly greater than the threshold of 0.05. Natural logarithmic transformation is applied on the dataset for scaling purposes.

In this dataset, when ARIMA is individually used as a forecasting method, we found autoregressive model of order 9 (AR(9)) to be the most parsimonious one among all ARIMA models. The best fitted individual ANN model is achieved in three layered 3 X 6 X 1 architecture, after our tuning process.

In this dataset, the length of MA filter which we use in the proposed hybrid method comes out as 6. The obtained linear component after the MA filter has 0.004 ADF stationary test result; this implies that the component can be properly modeled by ARIMA method in our architecture. The best fitted ANN in the final step of the proposed method has 17 x 17 x 1 architecture. Input layer is composed of the last 8 of original data, the last 8 of residuals and the result of linear component forecast.

Similar to GbpUsd dataset, Intraday dataset is highly volatile and non-stationary which cannot be effectively modeled by using only a linear model. As can be seen in Table 3.1, ANN and hybrid methods significantly outperform the individual linear model of ARIMA. Furthermore, the proposed method gives remarkably superior accuracy as compared to other hybrid methods in all error metrics. The comparison of the actual and forecast values for all examined method are given in Figure 3.5.

Conclusions

Time series forecasting is an important yet often a challenging task used in many different application domains. The studies in the literature mainly focus on either linear

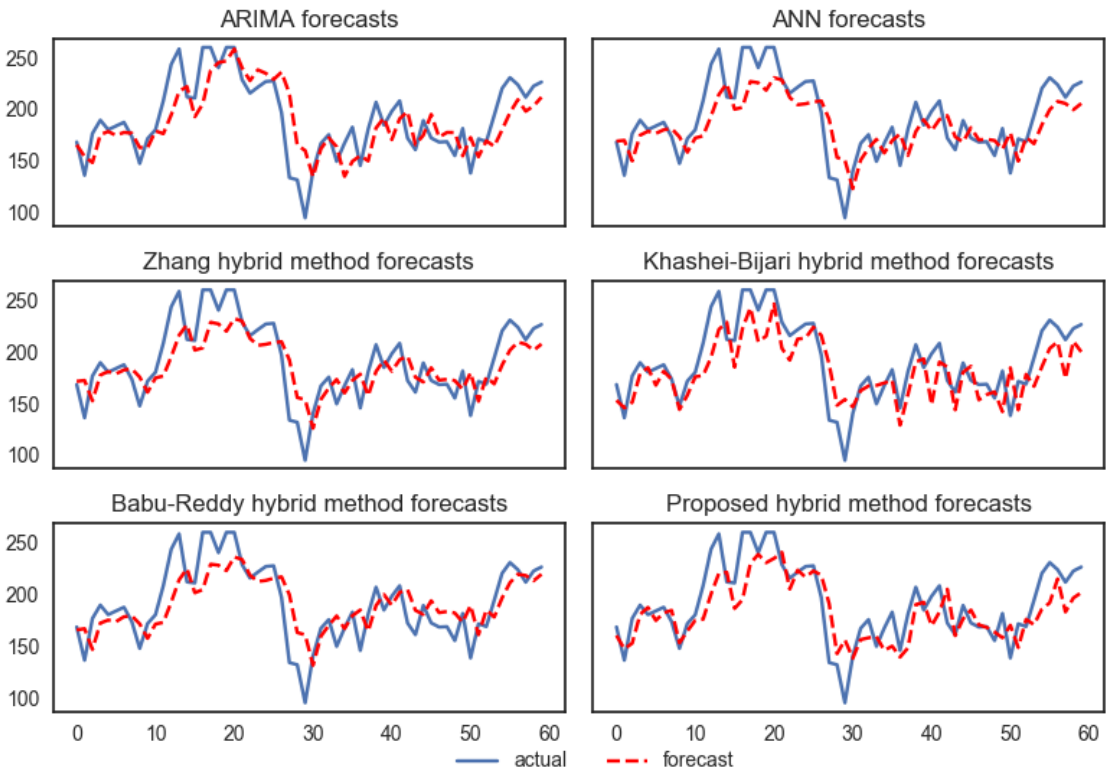

Figure 3.5: The Intraday data forecasts using various methods

or nonlinear modelings individually or a combination of them. While linear models such as ARIMA gives better forecasting accuracy with stationary time series data, nonlinear methods such as ANN is more appropriate for non-stationary datasets. In order to take advantage of the unique strength of each different type of methods in a more general setting, hybrid methods are proposed. Hybrid methods basically use linear and nonlinear modeling, ARIMA and ANN respectively on the corresponding decomposed components and then combine the results. Hybrid ARIMA-ANN methods give better results in general as compared to the cases where they are individually used. However, they generally suffer from the assumptions they make while constructing their model. These assumptions lead to produce inconsistent results and give low accuracies in overall if unexpected situations occur.

In this section, a new hybrid ARIMA-ANN model based forecasting method is proposed to overcome three main assumptions made by traditional hybrid ARIMA-ANN models. Firstly, the proposed method removes the assumption that the linear component is the ARIMA model output of the given data. Rather, it extracts the linear

component by using MA filter. It is known that data showing linear characteristics can be more accurately modeled by linear methods. Therefore, properly decomposed data yields more accurate linear forecasting and consequently more accurate final results in the hybrid methods. Secondly, the proposed method does not directly model residuals via a nonlinear method ANN, since the assumption of that residuals might comprise valid nonlinear patterns, does not always hold. Thirdly, the proposed method does not restrict linear and nonlinear component modeling and also combining the results of them. Rather, it can capture structures of the linear and nonlinear components in a better way, and produces more general models than those existing hybrid models.

3.2 A feature-based hybrid ARIMA-ANN Method

The methods described in Section 2.2.3 and the ARIMA-ANN hybrid method introduced in Section 3.1 of this Chapter use only values from previous time steps to predict the next value. On the other hand, capturing the underlying structural features of the historical values can provide more meaningful summarized information that improves the forecasting performance. In this section, we propose to develop the highly accurate hybrid ARIMA-ANN model examined in Section 3.1 of this Chapter by converting it into a feature-based model [71]. The features calculated by statistical methods are presented to the hybrid method as input variables that explain the given time series. Since the most relevant set of features might change according to the given time series, the proposed feature-based method deploys a feature selection preprocessing step. In this step, Gradient Boosted Trees (GBT) are constructed from the features to intelligently obtain the feature importance scores. The importance scores of the features are calculated through GBTs and the most informative ones are selected. These selected attributes are then used as explanatory variables in the hybrid method, which allows the improvement of the time series prediction accuracy results.

The features of the time series consist of statistical calculations of the values in certain time intervals. In literature, there are several studies on the idea of modeling using features in time series. For example, there are studies targeting clustering in time series [74, 75, 76] and studies for detection of abnormalities [77, 78]. The study proposed for time series clustering in [75] aims to define the time series with the features

obtained by statistical methods applied on time series. Similarly, the study proposed in [74] extracts the global features of the time series, rather than looking at historical data for clustering in the time series. By applying the dimension reduction method for extracted features, it shows that time series can be clustered more effectively and successfully with a small number of relevant features. The study proposed in [76] extracts various time series features, such as correlation structures, distributions, entropy, stationarity and scaling from a large amount of time series databases. After selecting the most informative features among them, it automatically learns the differences between the classes in the time series with feature-based machine learning classifiers. The study proposed in [77] develops a feature-based method to detect unusual situations over time on Yahoo email servers. This method calculates a feature vector from the time series data from each server, which measures the characteristics of the series (correlation, seasonal power, entropy, etc.). The study then detects unusual movements from this feature vector using the outlier detection method. The study proposed in [78] aims to develop a feature-based method to detect unusual situations occurring on holidays or other special occasions in the time series datasets of Uber. For this purpose, the study defines each point in the time series with a feature vector. Then, the constructed feature-based datasets are modeled using the Long Short-Term Memory (LSTM) neural network method. In this way, they are able to achieve up to 18% performance increase in detecting unusual situations.

3.2.1 Proposed Method

The proposed feature-based hybrid method is designed for univariate time series and is built on the ARIMA-ANN hybrid method presented in Section 3.1. The proposed feature-based method converts the univariate time series into a multivariate format by using extracted features (see Figure 3.6). In this method, time series are considered as a function of the linear component L_t, nonlinear component N_t and selected feature set F_t as given in Equation 3.6 which extends Equation 3.1 with the set of features.

$$y_t = f(L_t, N_t, F_t) \tag{3.6}$$

In addition to the ARIMA-ANN hybrid method, the feature-based hybrid method calculates the classical and advanced statistical features which explain the characteristics

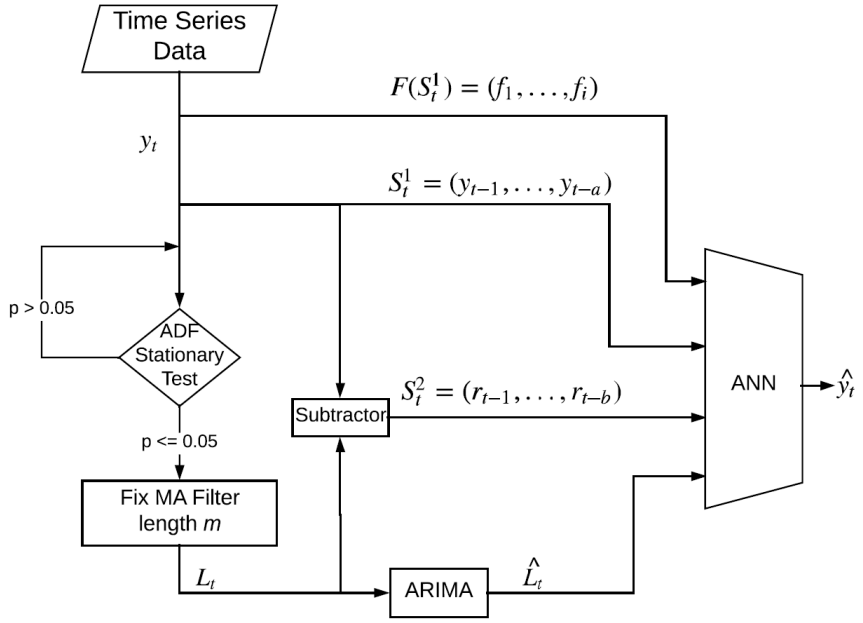

Figure 3.6: The architecture of the proposed feature-based ARIMA-ANN hybrid method

of a given time series such as mean, variance, skewness, kurtosis, etc. These features are calculated based on historical data which are defined as S_t^1 in Equation 3.5. After that, the importance scores of these features are computed by using GBT. The computed importance scores show how much each feature describes the given time series. Those with the highest importance score, that is, the most descriptive k features $F_t = f_1, f_2, ..., f_k$, are selected to use in the hybrid method.

3.2.2 Feature Extraction and Selection

Univariate time series forecasting uses only observations from the past of the series itself. However, the character of the time data given by this approach might not be well understood. Therefore, our method converts the time series as a derived set of features, making the dynamic variability in the series more static. In other words, univariate time series are converted to multivariate form by creating feature based representation.

The features are calculated from a frame that shifts a certain length on the observed historical data. These features consist of classical and advanced statistical calcula-

tions that define the characteristics of a particular time series. The calculated attributes and their descriptions are listed in Table 3.3.

In addition, the feature set is enriched by calculating the rate of change of each feature. At a given time t, the change rate of a given feature f_t is calculated by Equation 3.7.

$$ch_rate(f_t) = \frac{f_t - f_{t-1}}{f_{t-1}} \tag{3.7}$$

Table 3.3: Summary of the calculated statistical and structural features from time series

Features	**Description**
Mean	mean
Variance	variance
Mchange	change in mean
Index of Dispersion (IoD)	dispersion of a probability distribution
Skewness	asymmetry in a statistical distribution
Kurtosis	tailedness of the probability distribution
Autocorrelation	serial correlation
Entropy	spectral entropy
Trend	strength of trend
Linearity	strength of linearity
Crossing points (Cpoints)	number of crossing points

The calculated and enriched feature set increases the size of the dataset. Despite the fact that high dimensionality enriches the content of the dataset, it might memorize the data in the training set of the model, leading to the problem of overfitting to training dataset. Therefore, reducing the size by extracting and selecting the most informative features in the dataset is important for the forecasting process. Feature selection not only reduces the size of the dataset, but also eliminates unnecessary features that do not contribute to the predictive performance or cause the accuracy results to drop. Moreover, having less features reduces the complexity of the dataset, resulting in a faster and more cost-effective model. In many data mining and machine learning studies in the literature, the importance of feature extraction and selection has been

emphasized [44, 76, 77, 78, 79].

The feature selection methods in the literature are divided into three groups. 1) The Filter-based methods reveal the features with the strongest correlation with the output value with the help of statistical tests. Methods such as Variance threshold [80] falls into this category. The disadvantage of this approach is that it captures only linear relationships between features and output value. 2) The wrapped-based methods aim to find the most descriptive features in a heuristic way with the iterative search method. The most well-known method in this category is Genetic algorithms [81]. This approach is time consuming and tends to overfitting problem due to the use of machine learning algorithms during iterative search. 3) The Embedded methods incorporate feature selection into the process of creating a particular learning model. In this way, the selected features are obtained by optimizing the learning model. There are tree-based methods, such as Gradient Boosted Tree (GBT) [82] in this group. The fact that it also captures nonlinear relationships between the output value and the features and has a low risk of overfitting problem, makes the embedded methods more preferable as compared to other methods for feature selection.

GBT method [82] is used for feature selection of the proposed feature-based hybrid method. This method basically creates weak learning decision tree models and constructs a combined model from them. While creating decision trees with gradient boosting, the importance score of each feature is calculated. The most descriptive and informative features have more importance scores. There are three main methods used to calculate the significance score [83]. The first one is *gain* which calculates the relative contribution of a given feature for each tree to the model. The second method is *frequency* which calculates the number of a particular feature that occurs in all trees and is represented in percent. The third method is *cover* which is the relative number of observations based on the given feature. In this study, the *gain* method of GBT is used to determine the importance point values for each feature.

Breiman et al. [84] computes the importance score of a given feature f_i for a single decision tree T as follows:

$$w_t^2(T) = \sum_{t=1}^{j-1} \hat{\tau_t^2} \tag{3.8}$$

The decision tree splits a node into sub-nodes with the f_i feature, thereby obtaining $j-1$ internal nodes. The features on split points provide maximal estimated improvement ${\hat{\tau}_t}^2$. As a result, the squared importance score of the feature f_i for a given tree T is the sum of squared estimated improvement over the $j-1$ nodes.

3.2.3 Experimental Results

The evaluation of the proposed feature-based hybrid method is performed on the same experimental setup of the ARIMA-ANN hybrid method presented in Section 3.1. In the experiments, the performance results of the proposed feature-based method is compared with the performance results of the hybrid method built on it and individual methods such as ARIMA and ANN for Sunspot, Lynx, GbpUsd, and Intraday datasets. Two error metrics are used to compare the accuracy results of the experiments: Mean Absolute Error (MAE) and Mean Squares Error (MSE). The results are presented in Table 3.4. Moreover, due to the fact that ANN performs random initialization and produces different results at each run, the methods which include ANN algorithm are executed 50 times and average results are reported.

Table 3.4: Performance comparisons of the proposed Feature-based hybrid method for all datasets

Datasets	**Methods** / ***Metrics***	**ANN**	**ARIMA**	**ARIMA-ANN Hybrid Method**	**Feature-based Hybrid Method**	**Improvement (%)**
Sunspot	*MAE*	14,23	13,37	10,48	9,82	5,7
	MSE	353,12	306,97	194,29	16,.40	16,5
Lynx	*MAE*	0,1249	0,1198	0,1013	0,0891	11,8
	MSE	0,1249	0,1198	0,1013	0,0891	11,8
GbpUsd	*MAE*	428,55	435,72	404,90	382,57	5,5
	MSE	3,4681	3,5272	2,9538	2,8426	3,7
Intraday	*MAE*	20,10	20,22	18,81	17,20	8,1
	MSE	617,46	652,72	581,38	487,21	16,1

* MAE and MSE results are multiplied with 10^5 in GbpUsd dataset

Table 3.5 compares the p-value results of t-test analysis of the proposed feature-based method with the other examined methods. The proposed feature-based method is

significantly different from the individual methods, such as ANN and ARIMA for all experimented datasets in both MAE and MSE error metrics. When the p-value results are compared with the ARIMA-ANN hybrid method, built on it, the obtained results are significantly different for Intraday and GbpUsd datasets for both error metrics. On the other side, the feature-based hybrid method gives statistically significant results for at least one error metric for Sunspot and Lynx datasets.

Table 3.5: p-value test result comparisons of the proposed Feature-based hybrid method for all datasets

Datasets	**Methods** / ***Metrics***	**ANN**	**ARIMA**	**ARIMA-ANN Hybrid Method**
Sunspot	*MAE*	0.001	0.001	0.108
	MSE	0.001	0.001	0.029
Lynx	*MAE*	0.001	0.001	0.042
	MSE	0.001	0.001	0.133
GbpUsd	*MAE*	0.001	0.001	0.012
	MSE	0.001	0.001	0.053
Intraday	*MAE*	0.001	0.001	0.001
	MSE	0.001	0.001	0.001

Forecasts for Sunspot dataset

Figure 3.7 shows the features calculated for the Sunspot dataset (see Figure 2.1), sorted by their importance scores. While the Mean attribute provides the highest contribution to forecasting performance, Cpoints makes the lowest contribution. Our experimental observations show that the features with a relative importance score of less than 0.04 cannot sufficiently explain the time series. Since the features with low importance scores do not contribute effectively to the forecasting performance, only the top 12 attributes (see Figure 3.7) are included in the final ANN model. The final ANN model additionally takes the last 4 of observed values, the last 2 of residuals, and the result of linear component forecast from the input layer. (see Figure 1). The

used ANN model for this experiment has 1 hidden layer containing 39 nodes. In the training of this model, 0.01 learning rate, 0.001 bias value (bias), ReLU activation function and Nadam optimization algorithm are used.

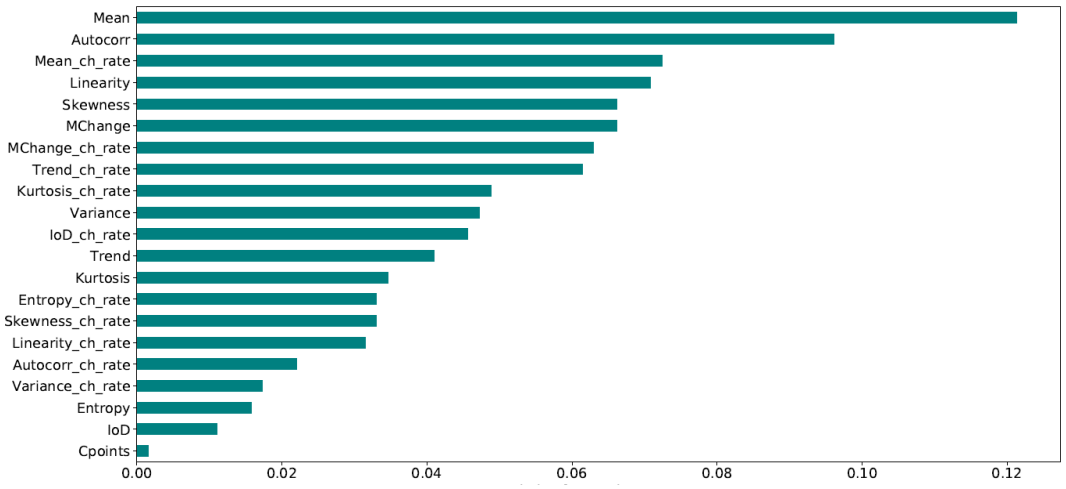

Figure 3.7: The sorted feature importance scores of the Sunspot dataset

When the numerical results of the Sunspot dataset are analyzed (see Table 3.4), it appears that the hybrid methods provide significantly better performance than the ARIMA and ANN singular methods. This is due to the fact that ARIMA and ANN cannot individually capture general patterns in the data series. However, hybrid methods that combine the unique strengths of both methods are an effective way to overcome this shortcoming. Moreover, while the feature-based method proposed in this section uses the advantages of the ARIMA-ANN hybrid method, it also provides feature-based representation of the dataset, giving higher forecasting performance. As a result, the proposed method improved the MAE and MSE error metrics of the hybrid method, built on it, 5.7% and 16.5%, respectively.

Forecasts for Lynx dataset

The Lynx dataset (see Figure 2.2) is sensitive to statistical features such as Mean, Variance, Skewnewss, Linearity, when the calculated feature importance scores are compared in Figure 3.8. As a result of our experimental observations, in this dataset, the most important first 11 features (see Figure 3.8) are selected for using in the final

ANN model. In addition to them, there are 9 more nodes in the input layer of the final ANN model including 5 observed values, 3 residual values and 1 linear component forecast. Moreover, the final ANN model has 1 hidden layer with 41 nodes. In the training of this model, 0.01 learning rate, 0.001 deviation value, ReLU activation function and Nadam optimization algorithm are used.

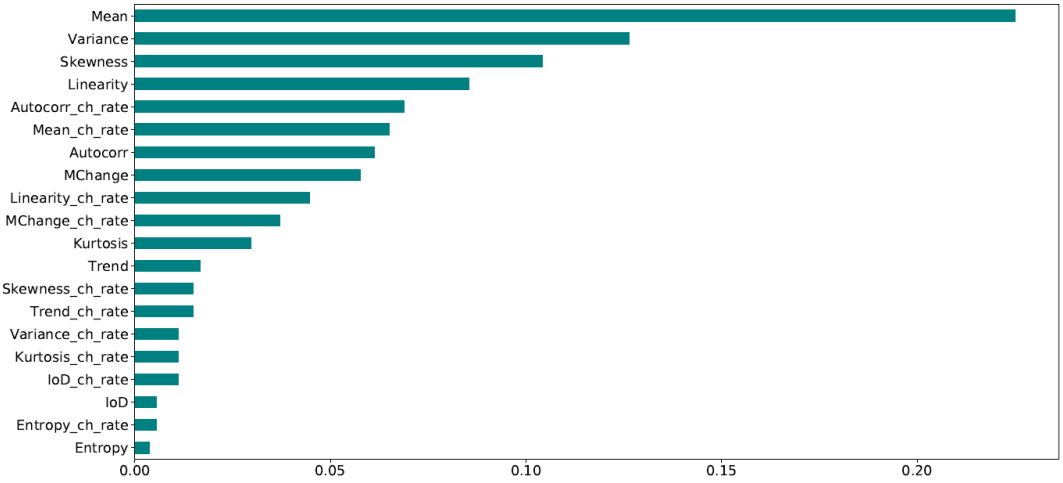

Figure 3.8: The sorted feature importance scores of the Lynx dataset

When the individual methods are compared in Lynx dataset, ARIMA is able to provide a better accuracy result compared to ANN, unlike the Sunspot dataset (see Table 3.4). This is because the Lynx dataset is more stationary and linear than the Sunspot dataset, so that a linear model ARIMA is better than ANN. Due to this stationarity, the extracted features show more regularity among themselves and represent the time series characteristic more accurately. In this way, our method increased the ARIMA-ANN hybrid method accuracy results by more than 10% in both error metrics.

Forecasts for GbpUsd dataset

When the calculated feature importance scores of the GbpUsd dataset are analyzed, it is seen that the Mean feature makes the most important contribution to the prediction performance as in the other datasets (see Figure 3.9). As a result of the experiments we conduct in this dataset, the first 7 features with the highest score are selected to be

used in the final ANN model. In addition, the final ANN model is additionally fed by the last 5 of observed values, the last 3 of residuals, and the result of linear component prediction. The final ANN model has 1 hidden layer containing 33 nodes. In the training of this model, 0.01 learning rate, 0.001 deviation value, ReLU activation function and Nadam optimization algorithm are used.

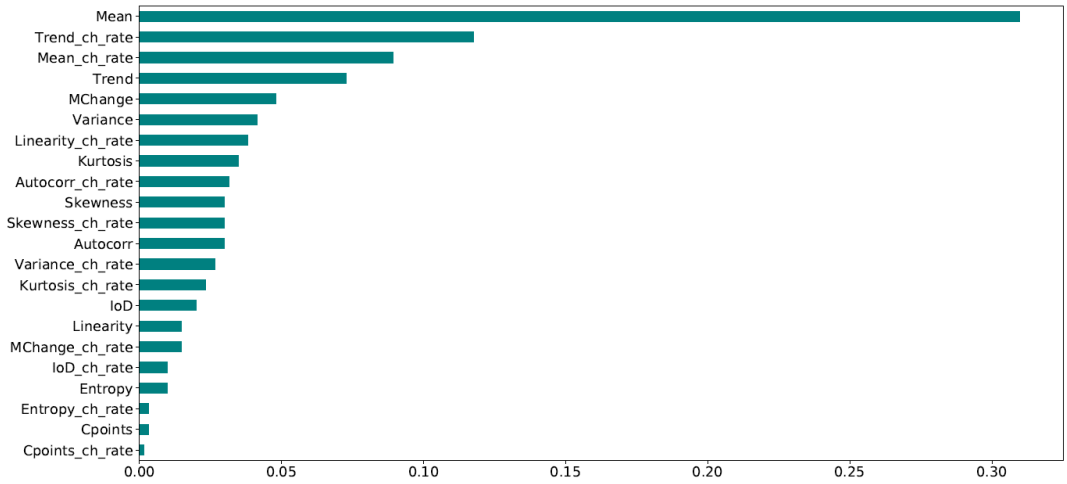

Figure 3.9: The sorted feature importance scores of the GbpUsd dataset

When the experimental results of the GbpUsd dataset are analyzed (see Table 3.4), the ANN method performs better than a linear method ARIMA for this highly volatile dataset. Although the hybrid method provides better accuracy performance compared to the individual methods, the proposed feature-based hybrid method is able to capture the high volatility pattern better than other methods by using the selected features. The proposed feature-based hybrid method succeeds in improving the hybrid method results by 5.5% and 3.7% respectively in MAE and MSE error metrics.

Forecasts for Intraday dataset

Figure 3.10 shows the calculated features for this dataset, sorted by their importance scores. While the Mean feature provides the highest contribution to the prediction performance, the Cpoints feature makes the lowest contribution. As a result of the experiments carried out in this dataset, the first 9 features with the highest importance score are selected to be used in the final ANN model. In addition, the final ANN

model takes the last 8 values of the given time series, the last 8 values of the nonlinear residual series and the prediction of the linear component as input. The final ANN model has 1 hidden layer with 53 nodes. In the training of this model, 0.01 learning rate, 0.001 deviation value, ReLU activation function and Nadam optimization algorithm are used.

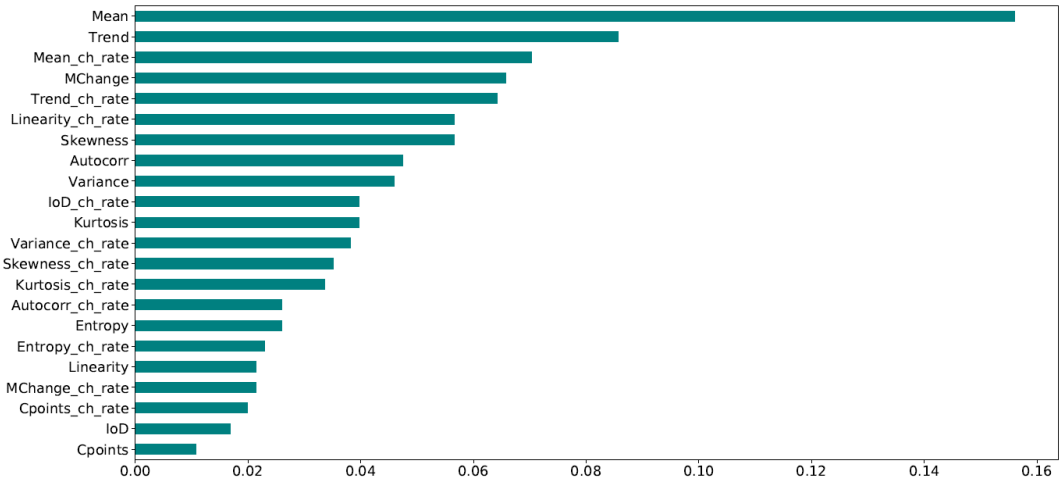

Figure 3.10: The sorted feature importance scores of the Intraday dataset

When the experimental results of the Intraday dataset (see Table 3.4) are analyzed, it is seen that ARIMA, which is a linear method, is not sufficient to capture the nonlinear movements of this high volatility dataset. The proposed method uses the advantages of selected features which calculate the characteristics of the dataset. In this way, the MAE and MSE error metrics of the ARIMA-ANN hybrid method improve 8.1% and 16.1%, respectively.

Conclusions

Time series data, derived from various applications in practice, generally include linear and nonlinear variations. Various hybrid methods that combine the strengths of linear and nonlinear methods are proposed in the literature to capture these different variations. Hybrid methods are better than single methods as they can use the advantages of both types of methods at the same time. However, hybrid models are designed as univariate methods that use values from previous time steps to predict

the next value in the series. They do not take into account the underlying structural features of the series. In this study, a feature-based hybrid method is proposed to increase the predictive performance in time series. This method calculates the features that present a meaningful summary of the time series and presents the most descriptive ones as an input to the ARIMA-ANN hybrid method, proposed in Section 2. To find the most descriptive features of each series, the importance score of the features is calculated separately using Gradient Boosted Trees. After that, the most informative features are selected in order of importance score and presented with other inputs to the ARIMA-ANN hybrid method. The experimental results show that our feature-based hybrid method is able to improve average prediction results as compared to the studied methods.

3.3 Discussion

There are several important results obtained in the experiments of Chapter 3. Firstly, when individual methods' results are compared among themselves, we see that ARIMA outperforms ANN for the datasets which present more linearity and vice versa (see Table 3.1). For example, while ARIMA produces more accurate result in a stationary datasets such as Lynx, ANN obtains the better approximation in a non-stationary datasets such as Intraday. Moreover, hybrid methods have better performance as compared to individual ones especially in more fluctuating datasets. Furthermore, the assumptions made by other hybrid methods degenerate the forecasting performance when unexpected situations occur in the data. For example, Zhang's and Babu-Reddy's methods assume that the relationship between linear and nonlinear components is additive. If linear and nonlinear components are not additively associated and the relation is different (i.e., it can be multiplicative), the possible complex relationship between components might be overlooked and the forecasting performance might be degenerated. Another assumption that might not always hold is that residuals might not show nonlinear pattern property. Additionally, Zhang's and Khashei-Bijari's methods do not actually decompose data into linear and nonlinear components, but they assume that linear component can be extracted by ARIMA and error sequences shows nonlinear pattern. As a result; such assumptions may lead

to low forecasting performances when unexpected scenarios happen. Our ARIMA-ANN hybrid method proposed in Section 3.1 avoids making those assumptions and creates more general models and outperforms the other examined methods.

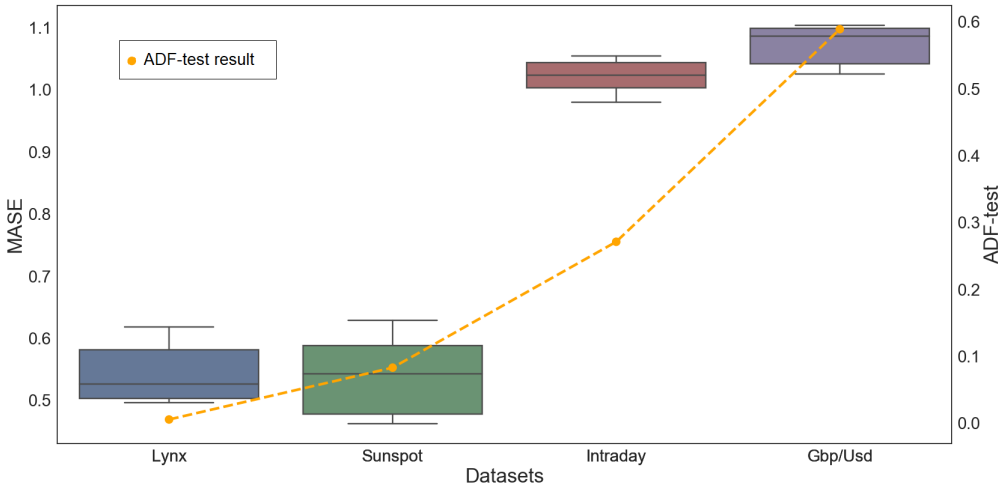

Figure 3.11: Comparison of MASE distributions of the ARIMA-ANN hybrid method with ADF test results for all datasets

The other proposed feature-based hybrid method in Section 3.2, which is built on our ARIMA-ANN hybrid method, uses extra statistical and structural features which summarize characteristics of the time series in addition to the historical values. Although the average results of the feature-based method improves the ARIMA-ANN hybrid method for all datasets, it cannot produce significantly different results for small datasets such as Sunspot and Lynx as seen in p-values results of the t-test analysis showed in Table 3.5. The reason of this might be due to the fact that the final ANN model of the proposed method with a high number of inputs cannot capture the patterns of small datasets well. Sontag et al. [85] states that the lower bound on the minimal number of training examples in a neural network (Vapnik–Chervonenkis (VC) dimension of neural networks) to obtain robust generalization performance results is p^2, given that p is the number of input features. The feature-based hybrid method which increases the number of inputs with additional statistical and structural features does not satisfy the lower bound rule of VC dimension for small datasets. Because of that, the feature-based method might not produce significantly different

results from the ARIMA-ANN hybrid method for these small datasets.

Figure 3.11 compares the MASE distribution of the ARIMA-ANN hybrid method proposed in Section 3.1 for all datasets with ADF test results of the corresponding dataset. In this figure, boxplots are drawn by using MASE values of all examined methods for each dataset, presented in Table 3.1. When we compare error results among time series data by using a scale-invariant error metric MASE, it is observed that the more non-stationarity in a dataset leads to a higher error value. For example, Lynx dataset, which turns out to be the most linear among all datasets according to ADF test results, has the lowest MASE results. On the other hand, GbpUsd dataset, which shows the most non-linearity according to ADF test results, has the highest MASE results. As a result of this, we can conclude that having more regular data distribution in a time series leads to more accurate results in forecasting.

CHAPTER 4

METHODS USING EMPIRICAL MODE DECOMPOSITION

> I have seen the future and it is very much like the present only longer.
>
> *Kahlil Gibran*

In this chapter, two methods using Empirical Mode Decomposition (EMD) are proposed [4, 86, 87], which improve the prediction performance results of the hybrid method proposed in the previous chapter. These methods take basically advantage of stationary EMD components which are easy to predict. While the first method uses the MA and EMD decompositions respectively to achieve linear sub-components, the second method directly applies the hybrid ARIMA-ANN method proposed in Section 3.1 to all EMD components. Moreover, the enhancement of these methods with the method selection algorithm is also explained in this chapter [88, 89].

4.1 Motivation

Irregular patterns in a time series is cause of volatility which shows the variation of a time series over time and reduces the stationarity. Therefore, high variance or equivalently, high volatility implies that the time series is more non-stationarity. As a result of the experiments carried out in Chapter 3, it is observed that more non-stationarity in a dataset leads to a higher error value. Namely, having more regular data patterns in a time series provides more accurate results in forecasting. Moreover, several studies in literature [8, 9] state that stationarity is required as a precondition to have high accuracy predictions.

This conclusion motivates us to propose an improvement on our already best performing hybrid methods introduced in the previous chapter. This improvement aims to produce more stationary sub-series from a given time series by using a multi-scale decomposition technique. In the literature there are several multi-scale decomposition methods such as Empirical Mode Decomposition (EMD), Wavelet Packet Decomposition (WPD) [90], Fourier Transform (FT) [91] and etc. [92]. EMD has two main advantages which makes it preferable for time series analysis. Firstly, EMD does not make a priori assumption about the stationarity of the time series unlike WPD and FT. Therefore, it can decompose any non-stationary or non-linear data into more regular and autocorrelated components. Secondly, the EMD procedure considers only local extrema (minimum and maximum) points. Therefore, it becomes highly efficient and self-adaptive as compared to WPD and FT. Because of these advantages of EMD, it is a more preferable technique than the others for decomposing time series [93].

4.2 Empirical Mode Decomposition (EMD)

Empirical Mode Decomposition (EMD) is an efficient and effective technique to analyze the nonlinear and non-stationary time series [93]. Unlike WPD and FT, the EMD is completely data-driven, is not a theoretical tool and does not need a prior defined time series [94]. The main principle of the EMD is to extract the intrinsic oscillatory modes by their local characteristic time scales in the data via iterative steps. These intrinsic modes are represented as an Intrinsic Mode Function (IMF), which satisfies following two properties:

1. IMFs have the same numbers of extrema and zero-crossings or differ at most by one over the whole time series.

2. IMFs are symmetric with respect to local zero mean, so that an IMF has a mean value of zero.

Those two conditions ensure that an IMF is stationary and this simplifies its prediction and analysis. However, IMFs might have various sizes of amplitude and changing frequencies. Moreover, EMD can perfectly reconstruct the original series without

information loss or distortion with the sum of all extracted IMFs and the residual.

EMD decomposes given $x(t)$ dataset into the components as shown in Equation 4.1.

$$x(t) = \sum_{n} x_n(t) + r(t) \tag{4.1}$$

where $x_n(t)$ represents the n^{th} IMF component and $r(t)$ is the residual component. The detailed EMD procedure (see Figure 4.1), called *sifting process*, is as follows:

1. Identify all local extrema maxima and minima points of time series $x(t)$.
2. Construct upper and lower envelopes, $e_{min}(t)$ and $e_{max}(t)$ for extrema points, using cubic spline interpolation.
3. Compute mean value $m(t)$ of the envelopes:

$$m(t) = \frac{e_{min}(t) + e_{max}(t)}{2} \tag{4.2}$$

4. Take difference of mean and given time series and estimate as first component:

$$h(t) = x(t) - m(t) \tag{4.3}$$

5. Check the properties of $h(t)$
   (a) If $h(t)$ satisfies two above mentioned conditions, $h(t)$ can be considered as an IMF. The $h(t)$ is denoted as $x_n(t)$ in Equation 4.1. Then, let residual $r(t) = x(t) - h(t)$ and replace $x(t)$ with the residual $r(t)$.
   (b) If $h(t)$ does not satisfy the conditions to be IMF, let $x(t) = h(t)$ and return step 1.
6. Repeat steps 1-5 until the residual satisfies termination criterion which check whether residual $r(t)$ is a monotonic function or a local extrema.

When the EMD procedure is completed, no more IMFs can be extracted from residual $r(t)$. In the end, EMD procedure limits the total number of IMFs to $logN$ where N is the length of the time series. Although this procedure ensures that the extracted components are stationary, they might have different changing frequencies and amplitudes. While the first component $x_1(t)$ has the highest frequency change, the residual

$r(t)$ has the slowest change in frequency. As a result, the *sifting process* extracts intrinsic modes from high frequency to low frequency. Furthermore, the amplitudes of the IMFs grow with the frequency change from high to low. For example, the amplitude of the residual is larger than the amplitudes of first IMF.

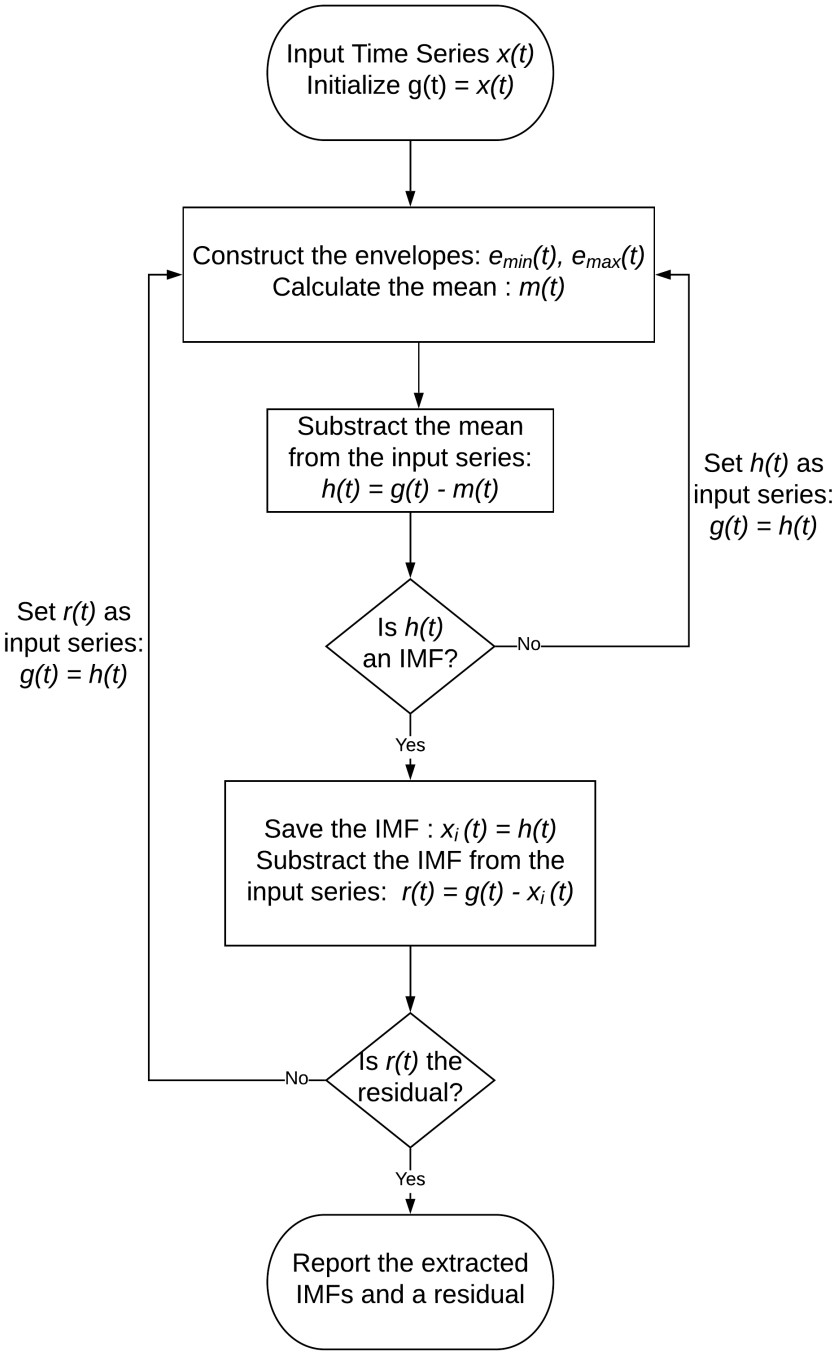

Figure 4.1: Flow Chart of the EMD Algorithm

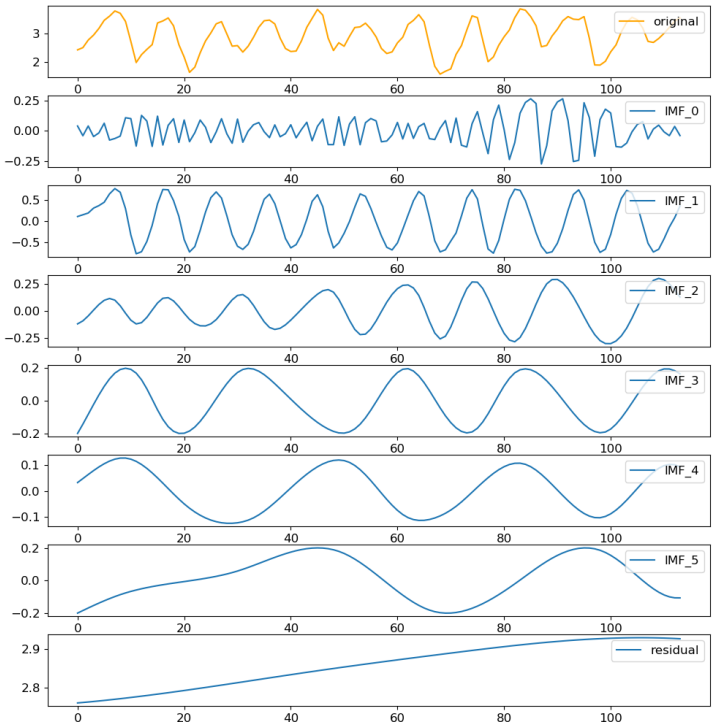

Figure 4.2: A sample EMD decomposition: The Lynx dataset and its components

Figure 4.2 shows the components obtained as a result of an example EMD decomposition. In this example, the Lynx dataset shown in Figure 2.2 is used. As can be seen from the figure, 6 components and a residual are extracted from the Lynx dataset. All IMFs have the same number of zero crossing , which satisfies the first property of being an IMF. Similarly, all of them have a mean value of zero , which satisfies the second property of being an IMF. The residual as seen in Figure 4.2 is a monotonic function from which no more IMF can be extracted and is the end of the *shifting process*. Although all extracted IMFs are stationary, each of them has its own frequency change rate and amplitude. For example, while the first IMF called *IMF_0* has the highest frequency change rate, the last IMF called *IMF_5* has the lowest one. In addition, the second IMF component called *IMF_1* has the largest amplitude with the value of 0.5, whereas the *IMF_4* has the smallest one with the value of 0.1. The amplitude of the residual is the largest among all other IMFs.

Since EMD was introduced in 1998 [93], it has been widely utilized in various areas such as engineering [95, 96], medicine and biology [97, 98], image processing [99, 100] and process control [101, 102]. Moreover, EMD has been extensively studied in

time series forecasting problems [103, 104, 105, 106, 107, 108, 109].

Xu et al. [103] aim to predict the demand of commodities after natural disasters. The authors propose to use the integration of EMD and ARIMA for these nonlinear and irregular datasets, instead of using only traditional statistical methods such as ARIMA.

Wang et al. [104] aim to forecast streamflow for water resource application and management. The study decomposes monthly streamflow data, which show irregular and non-stationarity patterns, into IMFs by using EMD. After that, ARIMA method is employed on each component separately. Finally, aggregated forecast results show the effectiveness of the proposed EMD-ARIMA method to improve the prediction results.

Wei et al. [105] propose to combine neural networks and EMD for short-term passenger flow forecasting. The flow data is firstly decomposed by using EMD into subcomponent, and they are then modelled by neural networks, separately. The experimental results of the study indicated the proposed method performs well and stably in forecasting results.

Yu et a. [106] propose to develop an early-warning system for financial crisis which is a rare kind of an event. The proposed method firstly decomposes a specified currency exchange rate series into sub-components. Subsequently, their internal correlations structures are explored by neural networks. By the neural network weights, the important sub-components are selected to be used in the final neural network. The experimental results reveal that the proposed method can significantly improve the generalization performance as compared to conventional neural networks.

Ren et al. [107] aim to forecast wind speed in their studies. Since wind is intermittent and highly volatile, they propose to integrate EMD and Support Vector Regression (SVR) methods. In their experimental results, they report that EMD-SVR approach outperforms other examined methods with respect to accuracy or computational complexity.

Fan et al. [108] aim to forecast electric load by incorporating SVR and auto regression (AR) in EMD decomposition. The authors propose to employ a nonlinear modelling

method such as SVR for all IMFs, whereas they forecast the monotonic residual with a simple and linear method AR.

Monjoly et al. [109] performs forecasting of short-term solar radiation series. The authors propose to employ different time series forecasting methods for different EMD components. While the components which have small time scales are modelled by ANN, the remaining components in which their time scale is long are modelled by ARIMA. The study shows that the proposed method significantly improves the accuracy results for solar prediction.

4.3 Two-level Decomposition Method

The one of the most important outputs of the studies in Chapter 3 is to make given time series data as stationary as possible in order to increase forecasting performance results. For this purpose, we aim to develop a method in this section that obtains more stationary and regular components from a given time series with the two-stage decomposition technique [87]. The proposed method extracts stationary and linear sub-components from a given time series by using Moving Average (MA) filter and Empirical Mode Decomposition (EMD) techniques and then applies a linear method such as ARIMA to obtain components, separately. In the final, prediction results of each individual ARIMA model are summed up to achieve the final prediction of the original time series.

4.3.1 Proposed Method

The proposed method includes three main steps as shown in Figure 1. In the first step, the method assumes that time series data y_t is sum of linear L_t and nonlinear N_t components, given in the form of:

$$y_t = L_t + N_t \tag{4.4}$$

These components are discovered as we use in the ARIMA-ANN hybrid method (see Section 3.1) by using MA filter. Similarly, the length of the MA filter is adjusted by utilizing ADF stationarity test. In the second step, the nonlinear component which

is the output of the MA filter is more sophisticatedly decomposed into linear sub-components called Intrinsic Mode Functions (IMFs) by using EMD technique. The nonlinear component N_t is m sum of IMFs $IMF(t)$ and a residue $Res(t)$ are obtained as seen below equation:

$$N_t = \sum_{k=1}^{m} IMF(t)_k + Res(t)_m \tag{4.5}$$

In the final, the achieved linear components including IMFs are separately modeled by ARIMA, and prediction results are then summed up to present the final result.

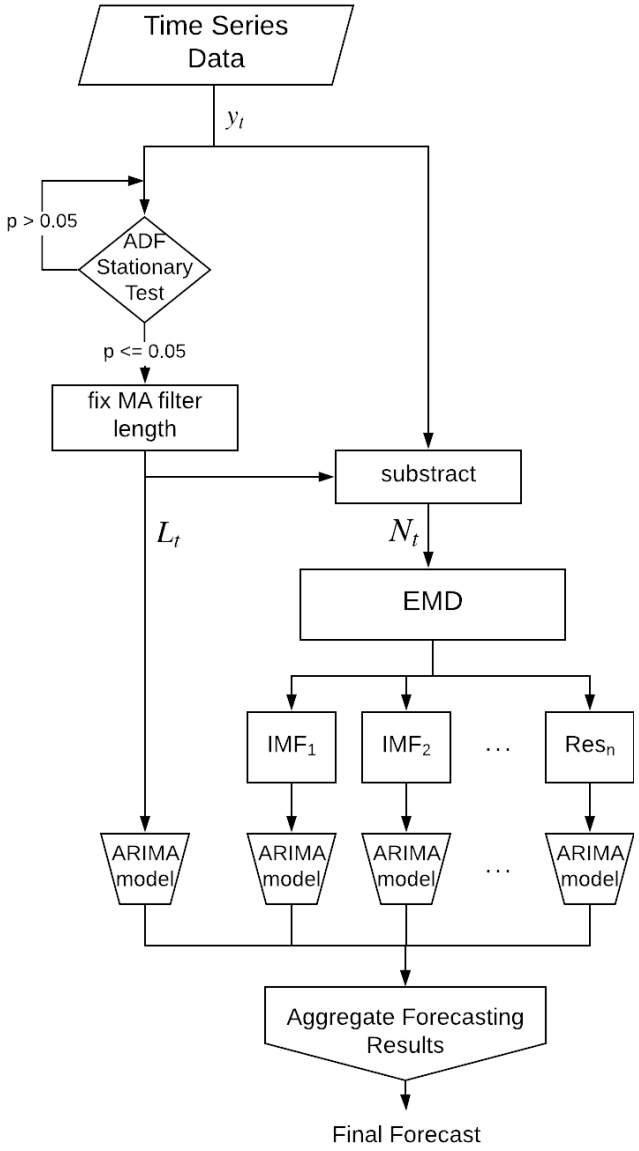

Figure 4.3: The architecture of the proposed Two-level decomposition method

4.3.2 Experiments and Results

In order to evaluate the performance of the developed two-level decomposition method, the same experimental environment is used as described in Section 3.1. The experimental results are summarized in the Table 4.1. The performance results of the proposed method for all datasets are compared with the results of the individual methods such as ANN, ARIMA and the hybrid methods such as Zhang, Khashei-Bijari, Babu-Reddy, our proposed hybrid method (see Section 3.1). As seen from results, the proposed two-level decomposition method obtains even better results than our ARIMA-ANN hybrid method, which is already best performing (see Table 3.1), by taking advantage of the sophisticated data decomposition.

Table 4.1: Performance comparisons of the proposed Two-level decomposition method for all datasets

Datasets	**Methods** / **Metrics**	ANN	ARIMA	Zhang's Method	K-B's Method	B-R's Method	Proposed Hybrid Method	Proposed Two-level Decomp.
Sunspot	*MAE*	14.23	13.37	13.14	10.62	11.39	10.48	**7.60**
	MSE	353.12	306.97	289.31	205.08	239.90	194.29	**94.62**
	MASE	0.629	0.591	0.581	0.470	0.504	0.463	**0.335**
Lynx	*MAE*	0.1249	0.1198	0.1003	0.1025	0.1102	0.1013	**0.0763**
	MSE	0.0241	0.0231	0.0173	0.0175	0.0189	0.0162	**0.00996**
	MASE	0.6185	0.5932	0.4966	0.50757	0.5457	0.5016	**0.3778**
GbpUsd	*MAE*	428.55	435.72	429.52	406.22	436.34	404.90	**145.87**
	MSE	3.4681	3.5272	3.4496	3.1053	3.5053	2.9538	**0.3517**
	MASE	1.085	1.103	1.087	1.028	1.104	1.025	**0.3693**
Intraday	*MAE*	20.10	20.22	19.16	19.79	19.50	18.81	**10.73**
	MSE	617.46	652.72	594.09	600.93	619.67	581.38	**156.66**
	MASE	1.048	1.054	0.999	1.031	1.016	0.980	**0.559**

* Khashei-Bijari's method is shown as K-B's method.

** Babu-Reddy's method is shown as B-R's method.

*** MAE and MSE results are multiplied with 10^5 in GbpUsd dataset

The examined hybrid methods alleviate unpredictable volatility of the time series data, so that they are able to improve the accuracy results of the individual methods.

However, they use insufficient levels of decomposition. For this purpose, we propose Two-level decomposition method using MA filter and EMD to achieve linear sub-components to feed into a linear modelling technique ARIMA. High accuracy results and the simplicity of the method suggest that the new proposed method has a good potential for time series prediction.

4.4 EMD-based Hybrid Method

In order to have more regular data distribution in a time series, the Two-level decomposition method (see Section 4.3) uses EMD technique for nonlinear time series component. Since the sub-components obtained by EMD are relatively stationary and have simpler frequency range, the prediction results are highly improved as compared to the ARIMA-ANN hybrid method (see Section 3.1). In this proposed method, each EMD sub-components are considered as separate time series data which are separately modelled by the already best performing our hybrid method [4]. The studies in the literature generally use EMD in time series prediction problems using singular methods such as ARIMA [103, 104], ANN [105, 106] and SVR [107, 108]. Unlike, in this proposed method it is aimed to combine the hybrid and empirical mode decomposition technique [4, 86].

4.4.1 Proposed Method

The success of the proposed hybrid method which is examined in Section 3.1 and the effectiveness of using empirical mode decomposition in time series prediction which is shown in Section 4.3 motivates us to combine them. The proposed EMD-based hybrid method includes three main steps, as seen in Figure 4.4: In the first step, the original time series data is decomposed into IMFs and a residue. In the second step, each sub-component called IMF is treated as a single time series and forecasting is performed by using our ARIMA-ANN hybrid method on each IMF. In the last step, prediction results of each individual hybrid model are summed up to achieve the final prediction of the original time series. We use an additive function in the end to capture the additive relation of IMFs with the original data (see Equation 4.1).

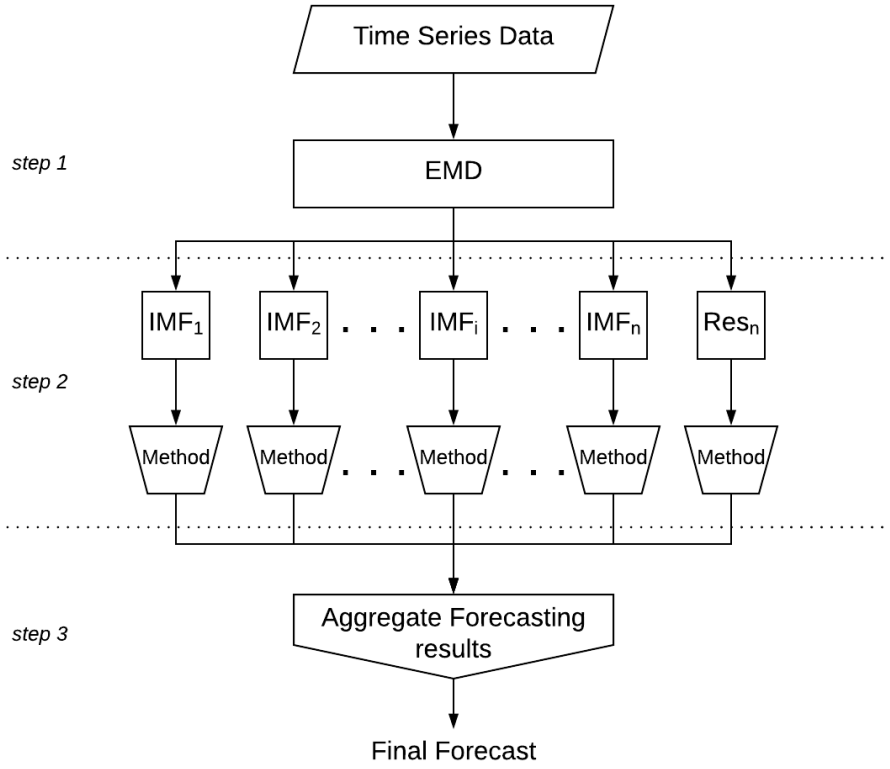

Figure 4.4: The architecture of the proposed EMD-based hybrid method

4.4.2 Experiments and Results

The EMD-based method is evaluated on the same datasets by using the experimental setup used in Section 3.1. In order to evaluate the effect of EMD, all examined single and hybrid methods such as ANN, ARIMA, Zhang's method, Khashei-Bijari's method, and Babu-Reddy's method are executed in the second step of the algorithm (see Figure 4.4). Table 4.2 gives the forecasting results of all examined methods with EMD on four all datasets. The table additionally includes the results of the Two-level decomposition method which is proposed in Section 4.3.

When the performance results of the experiments shown in Table 4.2 are compared with the previous ones shown in Table 3.1, the methods with EMD give significantly higher accuracies. The improvements at each dataset varies between 23% and 89% for all error metrics. For example, MASE error results in Sunspot, Lynx, GbpUsd, Intraday datasets are improved between 26-42%, 25-35%, 55-66%, 45-52% , respectively, for all methods using EMD. In addition, the EMD-based hybrid method is also

Table 4.2: Performance comparisons of methods using EMD for all datasets

Datasets	**Methods** / ***Metrics***	ANN	ARIMA	Zhang's Method	K-B's Method	B-R's Method	Two-level Method	Proposed Method
Sunspot	*MAE*	8.33 ±0.58	7.72	7.46 ±0.57	7.76 ±0.48	7.92 ±0.52	7.60	**7.28** ±0.47
	MSE	120.14 ±9.80	99.07	90.04 ±8.47	99.17 ±11.5	100.64 ±9.20	94.62	**87.86** ±9.11
	MASE	0.368 ±0.025	0.341	0.330 ±0.025	0.343 ±0.014	0.350 ±0.023	0.335	**0.322** ±0.020
Lynx	*MAE*	0.0912 ±0.0046	0.0772	0.0751 ±0.0034	0.0782 ±0.0043	0.0788 ±0.0045	0.0763	**0.0760** ±0.0035
	MSE	0.0131 ±0.0010	0.0099	0.0099 ±0.0009	0.0099 ±0.0010	0.0101 ±0.0008	0.0099	**0.0092** ±0.0008
	MASE	0.4516 ±0.0230	0.3822	0.3718 ±0.0168	0.3872 ±0.0213	0.3902 ±0.0226	0.3778	**0.3763** ±0.0176
GbpUsd	*MAE*	190.08 ±8.25	146.03	142.92 ±6.13	146.11 ±7.66	147.18 ±9.87	145.87	**141.38** ±6.87
	MSE	0.5610 ±0.0408	0.3578	0.3479 ±0.0271	0.3581 ±0.0319	0.3633 ±0.0292	0.3517	**0.3285** ±0.0280
	MASE	0.4812 ±0.0209	0.3697	0.3618 ±0.0155	0.3699 ±0.0194	0.3726 ±0.0250	0.3693	**0.3579** ±0.0174
Intraday	*MAE*	10.96 ±1.39	10.79	10.37 ±1.36	9.88 ±1.03	10.05 ±1.20	10.73	**8.93** ±1.04
	MSE	182.86 ±19.60	158.64	158.88 ±17.60	162.55 ±15.24	165.41 ±16.33	156.66	**130.79** ±14.91
	MASE	0.57 ±0.07	0.56	0.54 ±0.07	0.51 ±0.05	0.52 ±0.06	0.559	**0.45** ±0.05

* Khashei-Bijari's method is shown as K-B's method.

** Babu-Reddy's method is shown as B-R's method.

*** MAE and MSE results are multiplied with 10^5 in GbpUsd dataset

**** The values shown with $\pm$ give the standard deviation values.

able to achieve better accuracy results as compared to the Two-level decomposition method which also uses EMD.

The t-test analysis of the experimental results reported in Table 4.2 are presented

Table 4.3: p-value test result comparisons of the EMD-based hybrid method for all datasets

Datasets	**Methods** / ***Metrics***	ANN	Zhang's Method	Khashei-Bijari's Method	Babu-Reddy's Method
Sunspot	*MAE*	0.001	0.016	0.001	0.001
	MSE	0.001	**0.233**	0.001	0.001
Lynx	*MAE*	0.001	**0.136**	0.001	0.002
	MSE	0.001	0.001	0.001	0.001
GbpUsd	*MAE*	0.001	**0.328**	0.011	0.001
	MSE	0.001	0.004	0.003	0.001
Intraday	*MAE*	0.001	0.001	0.001	0.001
	MSE	0.001	0.001	0.001	0.001

in Table 4.3 for all datasets and error metrics. As Table 4.3 suggests, the proposed method using EMD is statistically significant than at least three of the examined methods for all experimented datasets.

Finally, we would like to emphasize that our proposed hybrid method with EMD gives the best results compared to other methods (see Table 4.2). Sub-components obtained from EMD are relatively stationary as compared to original data however, they still show fluctuations in their frequency range. Performing one more decomposition on each of these sub-components using MA filter and constituting functional relation between the stationary part, residuals, and the original data values outperforms all other remaining methods.

4.5 Method Selection Approach

Empirical Mode Decomposition technique guarantees the linearity of the components obtained by the decomposition process on a given time series. These linear components have simpler frequency range which make them strongly correlated in themselves. Thanks to these properties, high performance results are achieved in the methods using EMD proposed in Section 4.3 and Section 4.4. In these hybrid methods,

all components are modeled by the same methods. While the Two-level decomposition method (see Section 4.3) uses only ARIMA for all obtained components, the EMD-based hybrid method (see Section 4.4) employs only the ARIMA-ANN hybrid method. However, each EMD component is generated at its own frequency range, at different frequency change rate. These different characteristics of the components can be observed in the example shown in Figure 4.5 which shows the components of the Sunspot dataset decomposed by EMD. The first subplot represents the original Sunspot time series, whereas the remaining subplots show the IMFs and the residual. First of all, it can be clearly seen that the original dataset changes in a very wide frequency range, while the components change in a narrower range. Moreover, it is clear that the components have a mean value of zero by meeting the condition of being an IMF. These properties simplify the analysis and prediction of an IMF. However, it should be noted that the fact that each component actually has different characteristics with different frequency change rate [94]. For example, the frequency change rate of the first component, called *IMF_0*, is the fastest among all other IMFs of the Sunspot dataset, whereas the 9^{th} component, called *IMF_8*, has the slowest frequency change rate. By considering these differences of the components, modeling with a suitable method for each of them can improve prediction performance [110] [109].

4.5.1 Method Selection Algorithm

In this section, we propose to improve prediction results of EMD based methods proposed in Section 4.3 and Section 4.4 with method selection approach [88, 89]. In this approach, the obtained EMD components are modeled separately by appropriate methods according to the frequency change rates which is calculated by Fourier analysis [111] as shown in Equation 4.6.

$$w_m = \frac{\int_0^\infty E_m(f)\,df}{\int_0^\infty f E_m(f)\,df} \tag{4.6}$$

where $E_m(f)$ shows the Fourier spectrum of m^{th} component. The w_m represents mean frequency value which is greater for faster fluctuating components and smaller for slower fluctuating components. A high mean frequency value implies that the given component includes fast fluctuation and shows more irregularity. In order to effectively handle these irregularities, we propose to use the ARIMA-ANN hybrid

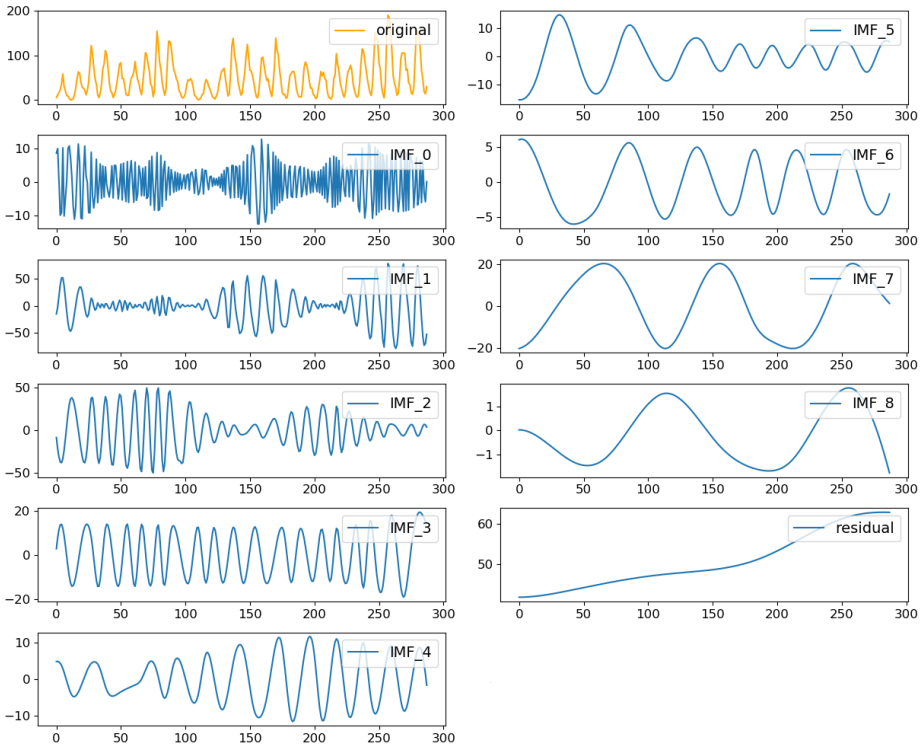

Figure 4.5: EMD components of the Sunspot dataset with different characteristics

method introduced in Section 3.1. The hybrid method overcomes the limitations of traditional hybrid methods and individual methods such as ANN, ARIMA by eliminating strong assumptions that they made. Thus, it is able to give high accuracy performance results in the prediction of datasets containing irregular patterns. On the other side, the slow fluctuating components are stationary and show linear patterns. Thanks to these features, we propose to model them by a linear method such as ARIMA.

The method selection algorithm is applied on two EMD based methods as shown in Figure 4.6: The Two-level decomposition method and the EMD-based hybrid method which are introduced in Section 4.3 and Section 4.4 , respectively.

4.5.2 Experiments and Results

The experiments are performed on the same experimental setup used in Section 3.1 in order to make the results comparable. As previously stated, the method selection

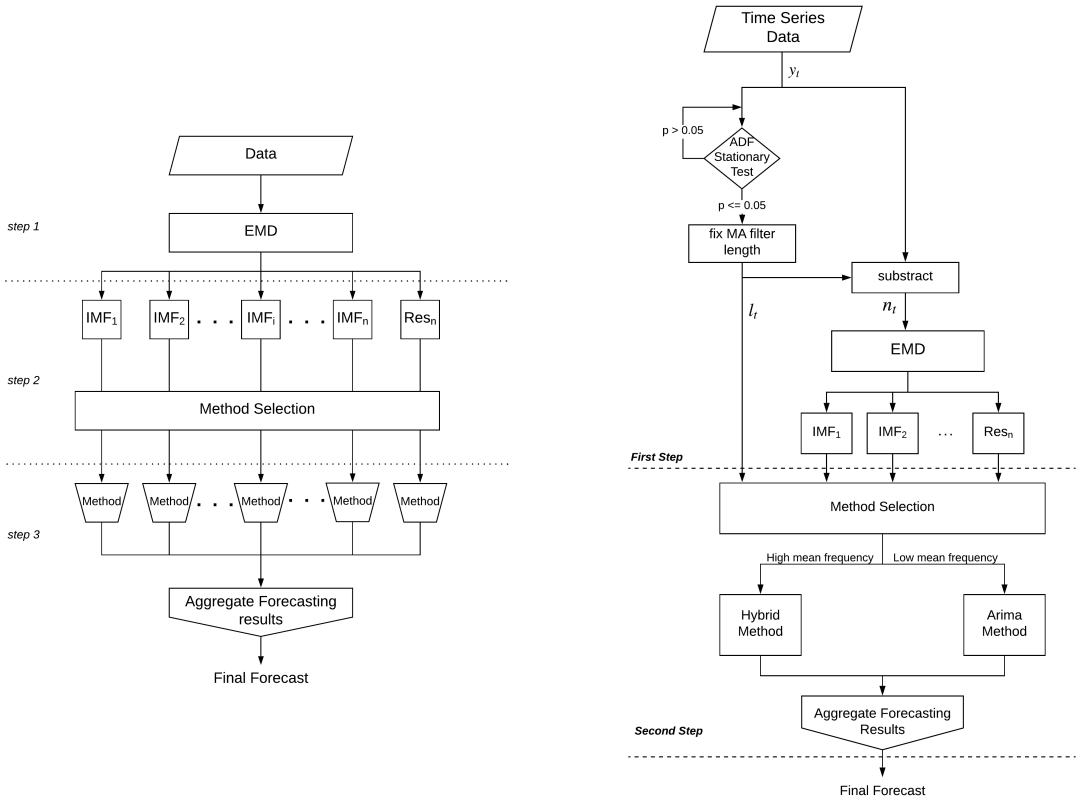

Figure 4.6: EMD-based methods with method selection approach

algorithm is adopted into two EMD based methods. The performance results for all examined datasets are shown in Table 4.4.

The table includes performance results of three different prediction approaches for both EMD based methods: In the first approach, all obtained components are modelled by ARIMA, whereas the second approach uses the ARIMA-ANN hybrid method introduced in Section 3.1 only for all components. The last approach is the method selection approach proposed in this section.

When the methods are compared for the same approaches (see Table 4.4), it is observed that the EMD-based method gives slightly better accuracy results than the Two-level decomposition method overall. Moreover, the ARIMA-ANN hybrid method is able to produce better performance results as compared to ARIMA in both EMD based methods. Although the results for Lynx, which is a linear dataset, are close to each other, in the nonlinear datasets, especially in the Intraday data, the hybrid

method clearly achieves better accuracy results than ARIMA.

Furthermore, the results of the method selection approach generally improves the results of the approaches that apply the same method (ARIMA or the ARIMA-ANN hybrid method) to each component. For example, the method selection algorithm improves the absolute accuracies of the EMD-based hybrid method for three datasets such as Sunspot, Lynx, and Intraday. Similarly, the method selection algorithm reduces the squared errors of the Two-level decomposition method for these three datasets.

Table 4.4: Performance comparisons of EMD based methods with different approaches for all datasets

Datasets	**Methods** / **Metrics**	Two-level Decomposition & ARIMA	Two-level Decomposition & Hybrid	Two-level Decomposition & Method Selection	EMD-based & ARIMA	EMD-based & Hybrid	EMD-based & Method Selection
Sunspot	*MAE*	7.83	7.60	7.21	7.72	7.28	7.20
	MSE	101.13	94.62	89.08	99.07	87.86	88.92
	MASE	0.345	0.335	0.317	0.341	0.322	0.318
Lynx	*MAE*	0.0774	0.0763	0.0738	0.0772	0.0760	0.0740
	MSE	0.01022	0.00996	0.00967	0.00996	0.00923	0.00972
	MASE	0.383	0.3778	0.365	0.3822	0.3763	0.366
GbpUsd	*MAE*	145.99	145.87	145.91	146.03	141.38	142.56
	MSE	0.3598	0.3517	0.3588	0.3578	0.3285	0.3374
	MASE	0.3696	0.3693	0.3694	0.3697	0.3579	0.360
Intraday	*MAE*	10.84	10.73	8.88	10.79	8.93	8.85
	MSE	131.67	156.66	127.95	158.64	130.79	125.11
	MASE	0.564	0.559	0.462	0.562	0.456	0.451

* MAE and MSE results are multiplied with 10^5 in GbpUsd dataset

In the light of these results, it can be said that ARIMA is not as successful as the hybrid method in components with high frequency variation. In the same way, ARIMA can be superior to the components which have slow frequency change rate. Therefore, selecting the most proper method for each component can improve overall forecasting results.

4.6 Discussion

Empirical Mode Decomposition is a multi-scale decomposition technique, incorporating the decomposition process with the main aim of mitigating modeling difficulty by decomposing any non-stationary and nonlinear time series data into relatively simple components [110]. In addition, the extracted components have their own frequency range which shows simple complexity [112]. With such helpful properties of multi-scale decomposition technique, the methods using EMD is able to alleviate the modeling complexity of a given time series data, thereby significantly enhancing the performance results [113].

Table 4.5: Percentage improvement in the methods when EMD is used

Datasets	**Methods** / **Metrics**	ANN (%)	ARIMA (%)	Zhang's Method (%)	Khashei-Bijari's Method (%)	Babu-Reddy's Method (%)	Proposed Method (%)
Sunspot	*MAE*	41.1	42.2	43.2	26.9	30.4	30.5
	MSE	65.9	67.7	68.8	51.6	58.0	54.7
	MASE	41.1	42.2	43.2	26.9	30.4	30.5
Lynx	*MAE*	26.9	35.5	25.1	23.7	28.4	28.4
	MSE	45.6	57.1	42.7	43.4	46.5	43.2
	MASE	26.9	35.5	25.1	23.7	28.4	28.4
GbpUsd	*MAE*	55.6	66.4	62.3	64.0	66.2	65.0
	MSE	83.8	90.0	90.1	88.7	89.7	89.1
	MASE	55.6	66.4	62.3	64.0	66.2	65.0
Intraday	*MAE*	45.4	46.6	44.4	50.0	48.4	52.5
	MSE	70.3	75.6	73.2	72.9	73.3	77.5
	MASE	45.4	46.6	44.4	50.0	48.4	52.5

The percentage improvement in prediction performance of the examined single and hybrid methods combined with EMD as shown in Section 4.4 for all datasets are presented in Table 4.5. The improvements at each dataset varies between 23% and 89% for all error metrics. We also provide a bar chart (see Figure 4.7) that shows the MASE results for each method which are averaged over the results of all datasets. The chart indicates that methods with EMD achieve remarkably less error in their forecasts. Further analysis shows that EMD-based methods give even greater improvements in

the accuracies for non-stationary datasets. This is due to the fact that EMD is able to resolve non-stationarity and alleviate high volatility problems in time series data. For example, while our hybrid method with EMD improves MASE results of relatively stationary datasets Sunspot and Lynx 30% and 28% respectively, this improvement jumps over 50% in GbpUsd and Intraday datasets which are highly non-stationary (see Table 4.5). Another indication of EMD's capability of solving volatility problem is that ARIMA is able to achieve better results than ANN for all datasets (see Table 4.5), although ANN without using EMD was better in non-stationary datasets (see Table 3.1).

In experiments (see Table 4.2), it seems that the best performance results of time series forecasting are achieved by incorporating EMD with our hybrid method. Even if the components obtained with EMD are meaningful and simple, each of them has distinctive patterns and periodicity. Modeling each of these components as a time series independent from each other by utilizing the advantages of the ARIMA-ANN hybrid method presented in Section 3.1 outperforms all other remaining methods.

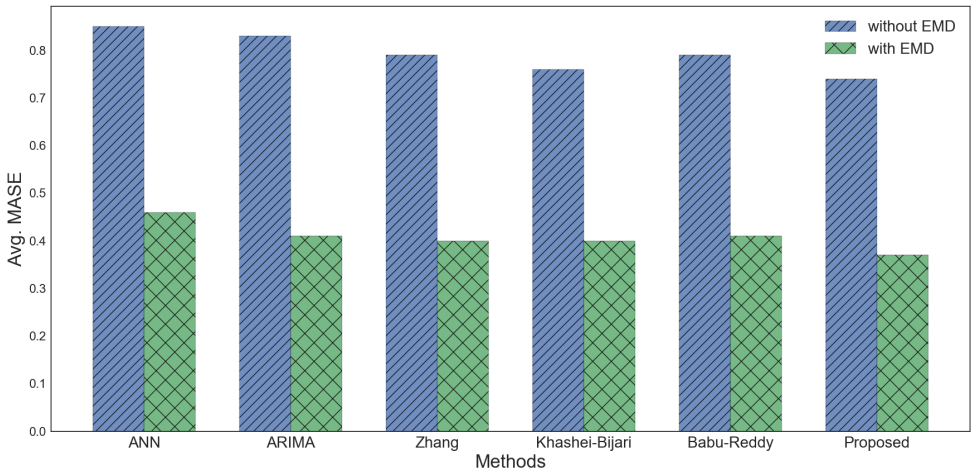

Figure 4.7: Average MASE results of the methods for all datasets with/without using EMD

In this chapter, a method selection approach, which takes into account the different frequency change rate of EMD components, has additionally been developed. In the experiments, it has been shown that considering the different characteristics of the components can yield better generalization performance.

CHAPTER 5

RECURSIVELY DECOMPOSITION METHOD

> If you can look into the seeds of time,
> and say which grain will grow and
> which will not, speak then unto me.
>
> *William Shakespeare*

Empirical Mode Decomposition guarantees the generation of stationary and linear components from any given time series. On the other hand, each component has a different frequency change rate, that is fluctuation rate, in itself. The method selection approach proposed in the previous chapter suggests that the high variance in fast fluctuating components should be differently considered; thereby providing a certain improvement in performance results. In this chapter, a new method is proposed [114], which performs the decomposition process to these fast fluctuating components until slower fluctuating components with more regular patterns, which are easy to predict, are obtained. As a result of this process, a number of components fluctuating as slowly as possible are obtained , which can be modeled with high performance by hybrid method. The experimental results given at the end of the chapter shows the effectiveness of this recursive method

5.1 Motivation

EMD can reduce any data from nonlinear and non-stationary process into simple Intrinsic Mode Functions (IMFs) as described in Section 4.2, which have to met following two conditions:

1. An IMF has a mean value of zero.

2. An IMF has only one extremum between two subsequent zero crossing. In other words, the number of local minima and maxima differs at most by one.

The first condition implies that IMFs are stationary time series which simplify their analysis and forecasting operations. However, the second condition says that each IMF might have different frequency change rate, namely fluctuation rate. The fast fluctuating components show more irregularity which tends to degenerate forecasting accuracies. In this section, we show that those fast fluctuating components can further be decomposed into more regular and easy-to-predict sub-components. For this purpose, the developed algorithm repeatedly invokes EMD decomposition for fast fluctuation components until it achieves slow fluctuating sub-components. These obtained more regular sub-components can be modeled with higher performance for time series forecasting.

5.2 Proposed Method

In this study, we propose a recursive EMD-based method for time series forecasting which aims to obtain more regular sub-components from the fast fluctuating components in order to achieve better prediction accuracies [114]. The architecture of the proposed method is given in Figure 5.1.

The proposed method recursively calls the EMD decomposition technique for each component until it satisfies one of the following two conditions:

1. The fluctuation of the component becomes less than of the given series.

2. Correlation between the component and the given series is higher than a threshold.

The first condition measures frequency change rate values of each component. This is made by using Fourier Analysis [111] for each component obtained by EMD as shown in Equation 4.6. As described in Section 4.5, the equation gives the mean frequency value w_m , which is greater for faster fluctuating components and smaller for

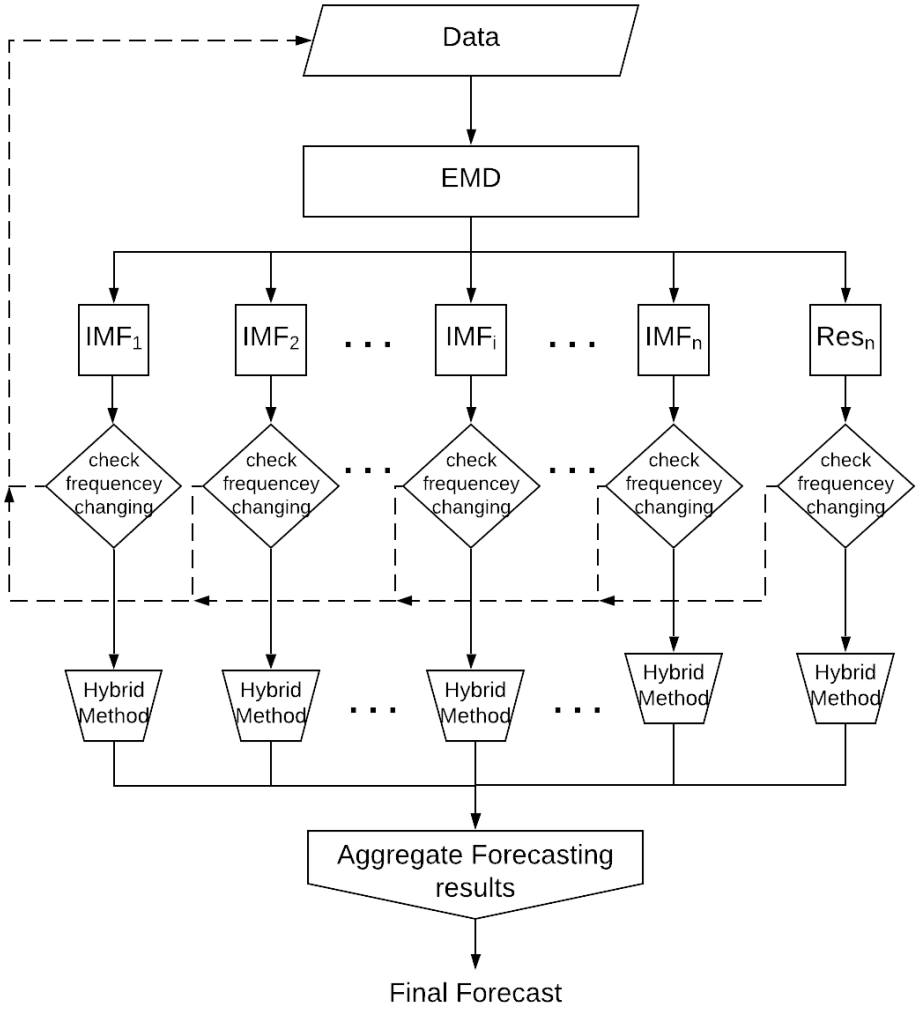

Figure 5.1: The architecture of the proposed Recursive EMD-based hybrid method

slower fluctuating components. A high mean frequency value implies that the given component includes fast fluctuation that might degrade forecasting performance. Therefore, in the algorithm these components are once again decomposed into their own sub-components via EMD. Same procedure is applied for sub-components until at least one of the two above conditions are satisfied.

The second condition uses Pearson correlation [115] to measure the similarity between given component and series. High correlation is a stop criteria for this proposed recursive algorithm. If a sub-component is highly correlated with its upper component, regardless of its fluctuation rate, recursive decomposition is terminated.

The pseudo-code of the proposed method is indicated in Algorithm 1. After the given data is decomposed into components via EMD at line 2, all components are called in the loop at line 4 where each component is analyzed. At line 5, frequency change rate of each component, which is the first condition of the proposed algorithm, is calculated by using Equation 4.6. At line 6, the second condition of the proposed algorithm which measures the correlation between upper component and the sub-component is calculated. If the frequency change rate of the sub-component is higher than its upper component and the correlation between them is low, the given component must be re-decomposed by EMD. In this case, another call for the *RecursiveMethod* function is done for this sub-component at line 8. Otherwise, it can be said that no more decomposition is required for the sub-component. In this case, the forecast result is generated for this component by the ARIMA-ANN hybrid method which is examined in Section 3.1. At line 12, all forecasts results generated from all components are aggregated and final forecast result is constituted.

Algorithm 1 Recursive EMD-based method

```
1:  procedure RECURSIVEMETHOD(data, mean_freq)
2:      IMFs ← EMD(data)
3:      forecasts ← Ø
4:      for each component imf in IMFs do
5:          imf_meanfreq ← MeanFrequency(imf)
6:          imf_corr ← Correlation(imf, data)
7:          if imf_meanfreq < mean_freq and imf_corr < 0.9 then
8:              imf_forecast ← RECURSIVEMETHOD(imf, imf_meanfreq)
9:          else
10:             imf_forecast ← HYBRIDMETHOD(imf)
11:         forecasts ← forecasts + imf_forecast
12:     final_forecast ← sum(forecasts)
13:     return final_forecast
```

The proposed recursive algorithm produces components with low volatility from fast fluctuating components which would have otherwise been difficult to predict.

5.3 Experimental Results

The datasets used to evaluate the performance of the proposed recursive method are composed of four different time series. Three of them are the Wolf's sunspot data, the British pound/US dollar exchange rate, and electricity intraday market dataset, that we use in other chapters. Similarly, the experimental setup of these datasets are the same as in other chapters. The last dataset used in this method is the electricity consumption time series data which is publicly available [116] . The electricity consumption dataset contains daily data of the electricity demand of Turkey. The dataset consists of 1110 data points between 2016 and 2019. In order to show the effectiveness of the proposed recursive method for time series forecasting, different machine learning algorithms such as Gradient Boosting Machine (GBM), Random Forecast, and Linear Regression are also evaluated.

Table 5.1: Parameters of the methods used in proposed recursive algorithm for all datasets

Model	**Parameters**	**Datasets**			
		Sunspot	Consumption	GbpUsd	Intraday
ARIMA	*(p,d,q)*	(9,0,0)	(12,0,0)	(0,1,0)	(9,0,0)
ANN	*architecture*	7x14x1	9x9x1	9x18x1	17x17x1
	activation	ReLU	ReLU	ReLU	ReLU
	optimizer	Nadam	Nadam	Nadam	Nadam
Hybrid Method	*MA length*	15	5	40	6
	(a,b)	(4,2)	(5,3)	(5,3)	(8,8)
Random Forest	*n_estimators*	480	800	800	800
	min_samples_split	2	3	3	2
	learning_rate	0.04	0.01	0.01	0.01
GBM	*n_estimators*	600	720	800	800
	max_depth	2	3	3	3
	learning_rate	0.02	0.01	0.01	0.01

In order to asses the forecasting performance of the different methods, each dataset is divided into three sets: Training, validation, and test sets. Training set is used for

fitting the model. Validation set is used to tune the hyper-parameters of the methods. These parameters are shown in Table 5.1 for each dataset. In order to determine model selection of ARIMA(p,d,q), an automatic stepwise approach is used [117]. The parameters related to ANN, Random Forest, GBM, and hybrid method are determined through a grid search approach. Test set is only used to assess the performance of a fully-trained model. Table 5.2 shows the sample size and divisions for training, validation, and test sets.

Table 5.2: Datasets setup in the experimental evaluation of the proposed recursive method

Time	**Size**			
Series	Training Set	Validation Set	Test Set	Total Sample
Sunspot	177	44	67	288
Consumption	816	204	90	1110
GbpUsd	543	136	52	731
Intraday	424	157	60	581

The experiments are performed for only *one-step-ahead forecasting*. The performance of the proposed method is compared with the other well-known single and hybrid methods. Table 5.3 and Table 5.4 show the forecasting results of the method on four time series datasets in two parts. We provide results for 4 different types of approaches. The first type of approaches include the single methods such as ARIMA, ANN, GBM, Random Forecast, Linear Regression, and naïve method. The naïve method ($y_t = y_{t-1}$) uses the value at the previous time step $t-1$ to predict the expected outcome at the current time step t. It is a useful method to have a baseline comparison. The second type of approach is the hybrid method examined in Section 3.1, which combines ARIMA and ANN single methods. The third type of approaches include incorporating the EMD technique to the single methods and the hybrid method. Finally, the last type of approaches are the Recursive EMD-based methods where we apply each single method on recursively decomposed EMD components.

Table 5.3: Performance comparisons of the proposed recursive EMD-based method for all datasets (*part 1: individual and hybrid methods*)

Method	**Datasets** / **Metrics**	Sunspot	Consumption	GpbUsd	Intraday
Naive	*MAE*	22.96	1692	512.41	21.54
	MSE	920.73	5321	3.97	797.03
	MASE	1.01	1.09	1.29	1.12
ANN	*MAE*	14.23	723	428.55	20.10
	MSE	353.12	969.9	3.47	617.46
	MASE	0.63	0.46	1.08	1.05
ARIMA	*MAE*	13.37	734	435.72	20.22
	MSE	306.97	977.0	3.52	652.72
	MASE	0.59	0.47	1.10	1.05
LR	*MAE*	14.84	1257	423.46	24.19
	MSE	405.84	3609	3.269	851.13
	MASE	0.65	0.80	1.06	1.25
RF	*MAE*	15.49	743	427.84	20.09
	MSE	509.45	981.1	3.23	710.52
	MASE	0.68	0.48	1.10	1.04
GBM	*MAE*	14.95	994	411.75	21.5
	MSE	482.06	1956	3.23	780.75
	MASE	0.66	0.64	1.03	1.11
Hybrid	*MAE*	10.48	711	404.90	18.81
	MSE	194.29	943.1	2.95	581.38
	MASE	0.46	0.45	1.02	0.98

* MAE and MSE results are multiplied with 10^5 in the GbpUsd dataset.

** MSE results are multiplied with 10^{-3} in the Consumption dataset.

*** Linear Regression is showed as LR, Random Forecast is showed as RF, Gradient Boosting Machine is showed as GBM.

Table 5.4: Performance comparisons of the proposed recursive EMD-based method for all datasets (*part 2: methods using EMD*)

Method	**Datasets** / **Metrics**	Sunspot	Consumption	GpbUsd	Intraday
ANN with EMD	*MAE*	8.33	478	190.08	10.96
	MSE	120.14	539.1	0.56	182.86
	MASE	0.37	0.31	0.48	0.57
ARIMA with EMD	*MAE*	7.72	447	146.03	10.79
	MSE	99.07	520.8	0.36	158.64
	MASE	0.34	0.28	0.37	0.56
LR with EMD	*MAE*	11.83	971	178.16	9.61
	MSE	288.12	2120	0.48	125.6
	MASE	0.52	0.62	0.44	0.50
RF with EMD	*MAE*	15.19	624	223.52	15.23
	MSE	493.01	793.0	0.90	353.27
	MASE	0.67	0.40	0.56	0.79
GBM with EMD	*MAE*	14.66	648	261.45	15.61
	MSE	479.77	821.4	1.18	379.14
	MASE	0.64	0.41	0.65	0.81
Hybrid with EMD	*MAE*	7.28	441	141.38	8.93
	MSE	87.86	509.2	0.33	130.79
	MASE	0.32	0.28	0.36	0.45
ANN with Recursive EMD-based	*MAE*	8.04	470	140.98	9.42
	MSE	102.06	536.9	0.33	138.93
	MASE	0.35	0.30	0.35	0.47
ARIMA with Recursive EMD-based	*MAE*	7.53	440	143.14	9.89
	MSE	90.63	497.5	0.34	137.41
	MASE	0.33	0.29	0.35	0.49
LR with Recursive EMD-based	*MAE*	10.39	915	170.11	9.42
	MSE	197.88	3324	0.43	125.63
	MASE	0.45	0.58	0.42	0.49
RF with Recursive EMD-based	*MAE*	14.80	598	214.04	14.71
	MSE	482.09	735.8	0.81	339.37
	MASE	0.65	0.38	0.53	1.01
GBM with Recursive EMD-based	*MAE*	14.2	612	250.81	15.04
	MSE	470.19	7597	1.09	349.09
	MASE	0.63	0.39	0.62	1.03
Hybrid with Recursive EMD-based	*MAE*	**7.12**	**431**	**139.25**	**8.20**
	MSE	**81.32**	**474.3**	**0.31**	**114.55**
	MASE	**0.31**	**0.28**	**0.34**	**0.41**

* MAE and MSE results are multiplied with 10^5 in the GbpUsd dataset.

** MSE results are multiplied with 10^{-3} in the Consumption dataset.

*** Linear Regression is showed as LR, Random Forecast is showed as RF, Gradient Boosting Machine is showed as GBM.

Forecasts for Sunspot dataset

The Sunspot dataset has 0.083 p-value of the ADF stationarity test which is greater than the threshold 0.05. Therefore this dataset can be regarded as non-stationary time series. The original dataset is shown in the first subplot of Figure 5.2. The remaining subplots of the figure depict the IMFs and residual components of the dataset. As can be seen in the figure, EMD decomposes the dataset into 9 IMFs and a residual.

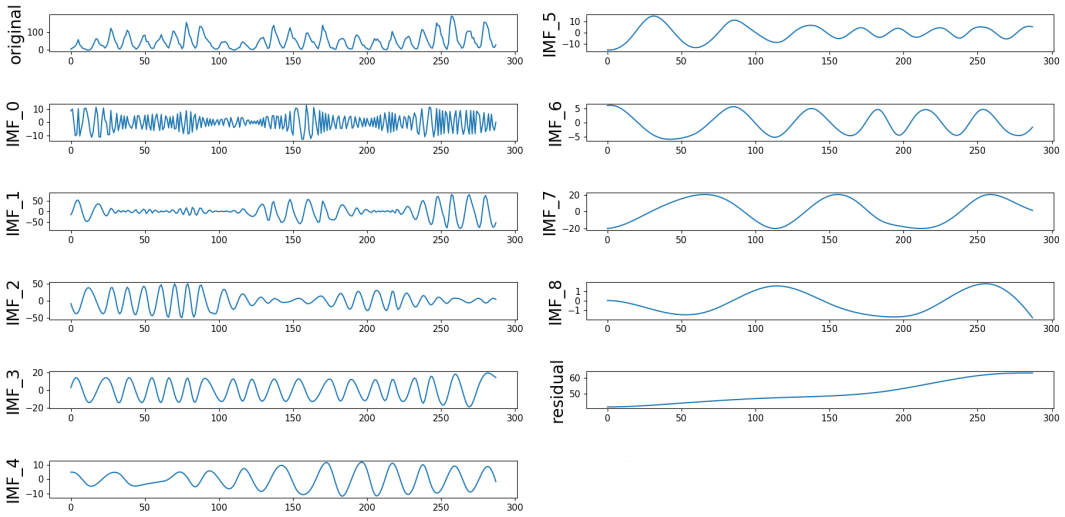

Figure 5.2: The IMFs and residual of the Sunspot series (1700-1987)

All components generated by EMD, shown in Figure 5.2, are stationary components of which their ADF-test results are less than 0.001. However the frequency change rate of these components can be faster than the original series as seen in the figure. For example, the first IMF component of the Sunspot dataset has greater mean frequency than the original series. A high mean frequency value means that given component includes fast fluctuations which might decrease the forecasting performance [1, 2]. Therefore, the proposed method aims to decompose the sub-component into its own sub-components using EMD recursively. This process is performed until (i) components with low mean frequencies are obtained, (ii) the correlation between the component and the sub-component exceeds a certain value.

When the numerical results of Sunspot dataset given in Table 5.3 and Table 5.4 are analyzed, it is clearly seen that the hybrid method produces better results as com-

pared to single methods such as ARIMA, ANN, Linear Regression, Random Forest, and GBM. Moreover, when those algorithms are combined with EMD, they achieve significantly higher accuracies than their single use cases. For example, the hybrid method with EMD reduces the mean absolute error of the Sunspot dataset via a significant value of 30% as compared to the single usage of the hybrid method. Furthermore, when the hybrid method with EMD is adopted to the proposed recursive algorithm, we again observe remarkable increases in the performance. For example, the proposed recursive approach improves absolute accuracy (MAE) of the EMD-based hybrid method by 4.6%, while mean squared errors (MSE) by 7.4%. These results indicate the positive effect of decomposing time series into their sub-components having slower frequency change rates.

Forecasts for Consumption dataset

The Consumption dataset has 0.11 p-value of the ADF stationarity test which is greater than the threshold 0.05. Thus, this dataset can be regarded as non-stationary time series. The original dataset is shown in the first subplot of Figure 5.3. EMD decomposes the dataset into 10 IMFs and a residual. The remaining subplots of Figure 5.3 show the IMFs and residual components of the dataset.

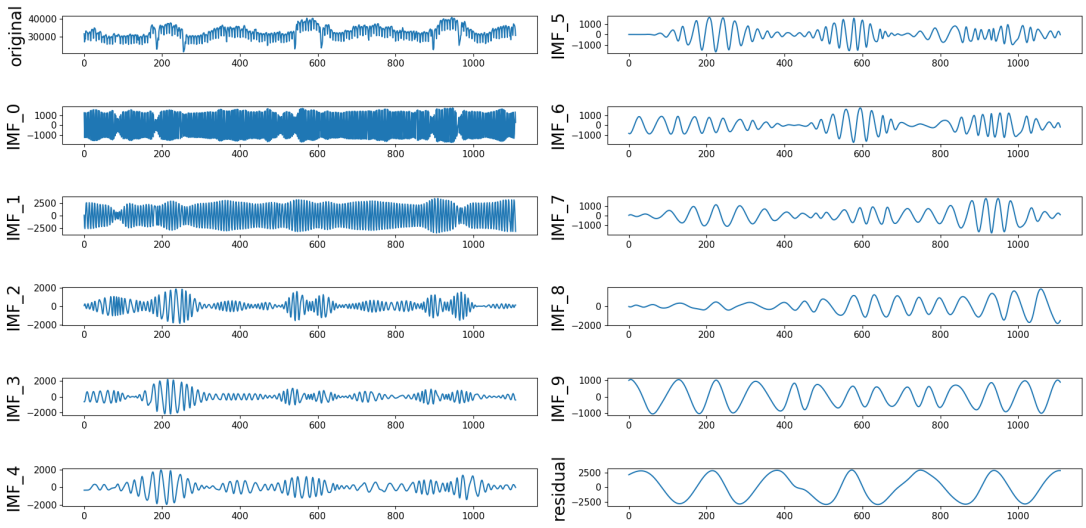

Figure 5.3: The IMFs and residual of the Consumption Series (2016-2019)

The components of the dataset generated by EMD are stationary components with having less than 0.001 ADF-test results. On the other side, their frequency change rate can be faster than the original series as seen in the figure. For example, first four components (*IMF_0, IMF_1, IMF_2, IMF_3*) have greater mean frequency value than the original series. Therefore, these four components are recursively decomposed into their own sub-components , until one of the stop conditions of the proposed method is met.

When the numerical results of the Consumption dataset given in Table 5.3 and Table 5.4 are analyzed, it is clearly seen that the proposed recursive method is able to achieve the highest accuracy results among all examined methods. For example, the hybrid method with recursive EMD-based algorithm reduces the mean absolute error of the Consumption dataset via a remarkable value of 39% as compared to the single usage of the hybrid method. Similarly, MAE results of other methods such as ANN, ARIMA, Linear Regression, Random Forest, GBM are reduced by 34%, 41%, 34%, 27%, 19%, 38%, respectively. In addition, the recursive algorithm substantially reduces the squared errors (MSE) of the hybrid method with EMD-based algorithm by 6.8%.

Forecasts for GbpUsd dataset

GbpUsd dataset is highly volatile and non-stationary time series with 0.58 p-value of ADF-test result. Due to high volatility, forecasting with high accuracy for this dataset is an important, yet difficult task.

The dataset and its 10 components obtained by EMD decomposition, are plotted on Figure 5.4. The non-stationarity of the original dataset can be seen from the first subplot of Figure 5.4 which shows numerous changing turning points in the series. Furthermore, when the IMFs in the figure are analyzed, it is clearly seen that first three components (*IMF_0, IMF_1, IMF_2*) have faster fluctuation than the original series. The mean frequency values of these first three components, which are 0.213, 0.914, 0.125, respectively, are greater than the mean frequency value of the original series which is 0.124. Therefore, these components are recursively decomposed into less volatile sub-components as proposed in the developed method.

Comparing the results in Table 5.3 and Table 5.4 for GbpUsd dataset, it is seen that the proposed recursive algorithm achieves very high performance compared to single and hybrid methods. For example, the absolute errors (MAE) of the single and hybrid methods are reduced between 39-65% by the recursive algorithm. In addition, the recursive algorithm is able to significantly reduce squared error results (MSE) from 3.47 to 0.33 for ANN, from 3.52 to 0.34 for ARIMA, from 3.26 to 0.43 for Linear Regression, from 3.23 to 0.81 for Random Forest, from 3.23 to 0.62 for GBM, from 2.95 to 0.31 for the hybrid method. These remarkable improvements are due to the fact that the proposed algorithm is able to produce easily-predictable sub-components with low volatility from fast fluctuating GbpUsd dataset. Moreover, the proposed algorithm makes the forecasting performance of highly non-stationary datasets and stationary datasets comparable. For example, when we compare the results in a scale-invariant error metric, such as MASE, the result of the proposed algorithm with the hybrid method in a highly non-stationary dataset GbpUsd (0.34) is very close to the result in a relatively stationary dataset Sunspot (0.31). On the other hand, in the single usage of the hybrid method there is a wider difference in MASE results of the Sunspot and GbpUsd datasets, which are 0.46 and 1.02, respectively.

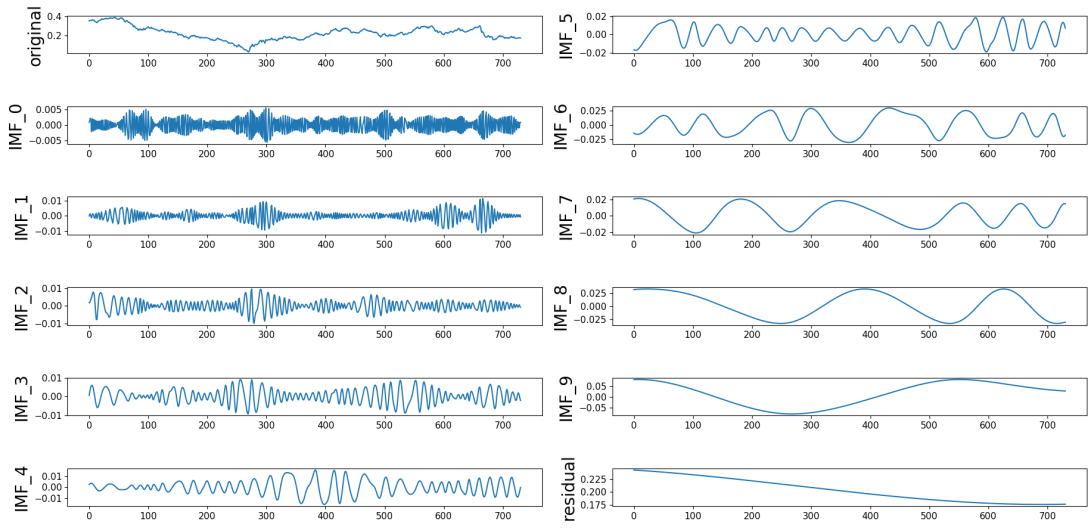

Figure 5.4: The IMFs and residual of the GbpUsd Series (1980-1993)

Forecasts for Intraday dataset

Intraday dataset is highly fluctuating time series like GbpUsd dataset (see the first subplot of Figure 5.5). The p-value of the ADF stationarity test result is 0.27 which is highly greater than the threshold of 0.05.

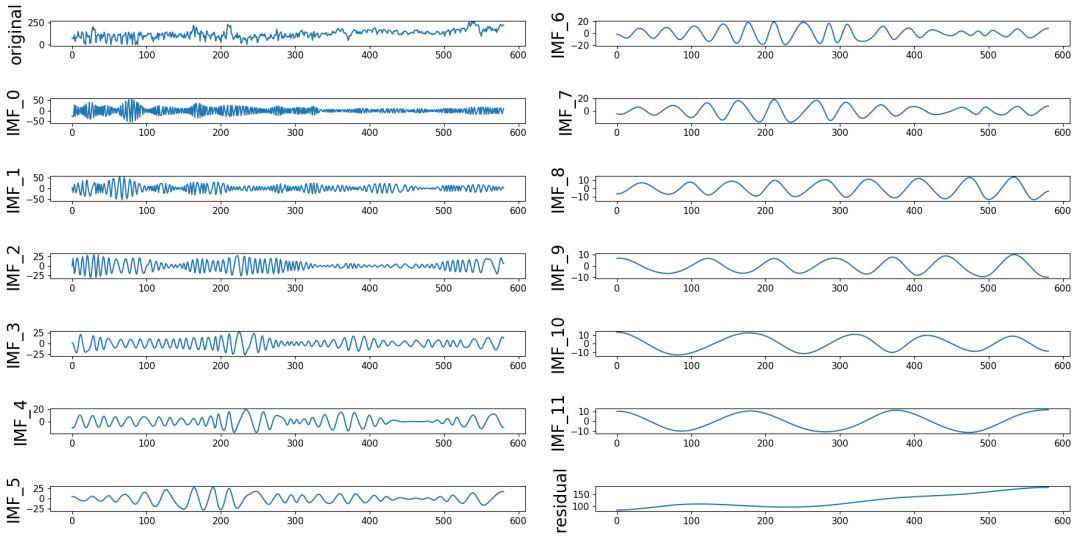

Figure 5.5: The IMFs and residual of the Intraday Series (Jul. 2015 - Dec. 2017)

Figure 5.5 additionally shows the 12 IMFs and a residual component of the Intraday dataset obtained by EMD decomposition. As can be seen, although each IMFs is stationary with a mean value of zero, each of them has different changing frequency rate. For example, the frequency change rate of the first two components (*IMF_0* and *IMF_1*) are clearly faster than the original time series. The mean frequency values of these components, 0.206 and 0.191, respectively, are higher than the original series, which is 0.188. Therefore, these two components are recursively decomposed as proposed in the developed method.

Intraday dataset is highly volatile and non-stationary which cannot be effectively modeled by using only a linear model. Although the hybrid method alleviates this volatility and gives better results, the proposed method, which uses the hybrid method in the recursively decomposed sub-components, provides remarkably superior accuracy in all error metrics compared to the single use of hybrid method. For example, MAE and MSE improves more than 55% and 75%, respectively. Similarly, the per-

formance results of the other single methods remarkably increase. For example, the proposed algorithm improves the MAE results of ANN, ARIMA, Linear Regression, Random Forecast, GBM by 53%, 53%, 60%, 26%, 30%, respectively. In addition, the recursive algorithm clearly improves the EMD-based algorithm as well. For example, the proposed recursive algorithm improves the MAE and MSE results of EMD-based hybrid method by 8.1% and 12.3%, respectively.

5.4 Discussion

EMD guarantees linearity for the decomposed time series components. Thus, it is able to resolve non-stationarity and alleviate high volatility problem in time series data. However, each component has a different frequency change rate and some of those even might have faster fluctuations than the original series which show more irregularity, tending to degenerate forecasting accuracies. In this chapter, the developed algorithm recursively decomposes those fast fluctuating components into more regular and easy-to-predict sub-components through EMD.

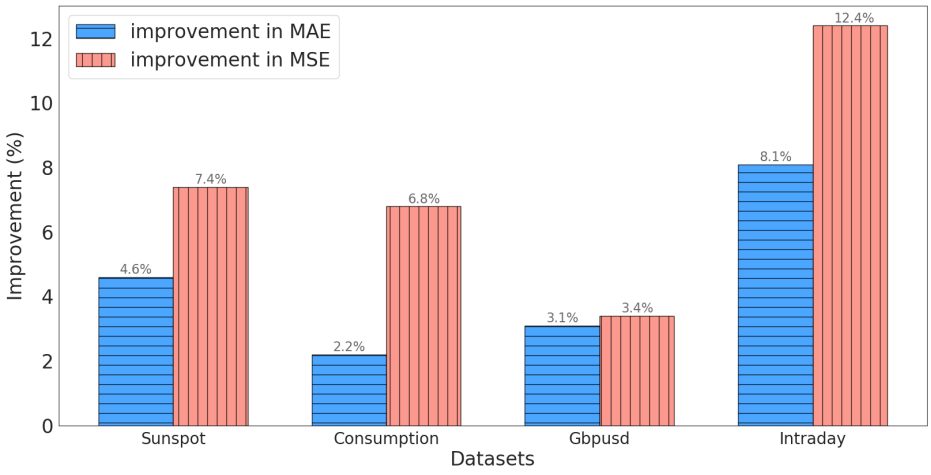

Figure 5.6: Improvement in MAE and MSE for all datasets with the Recursive EMD-method as compared to the EMD-based method

As a result of the experiments, it can be seen that the proposed recursive EMD-based hybrid method provides a significant improvement as compared to other examined

methods. In all examined datasets, the proposed recursive algorithm increases prediction accuracy results of the EMD-based hybrid algorithm between approximately 3% and 15%. However, this improvement varies in different error metrics. It is observed from the experimental results that the improvement in MSE is higher than that of MAE. Figure 5.6 compares the improvement of both error metrics for all examined datasets. For example, in the Consumption dataset, the improvement in MAE is 2.2%, while the improvement in MSE is 3 times more, i.e. 6.8%. Although this difference is not significant for the GbpUsd dataset, the improvement in MSE is clearly higher than MAE in the Sunspot and Intraday datasets, like in the Consumption dataset. MSE aims to punish large errors by squaring all errors. Therefore, the higher improvement in MSE implies that the proposed method reduces large errors. This achievement is actually an expected result, because the proposed recursive algorithm produces less volatile and stationary sub-components of which their sharp movements are reduced. In the light of these results, we can say that the proposed method reduces major errors more while improving the accuracy results.

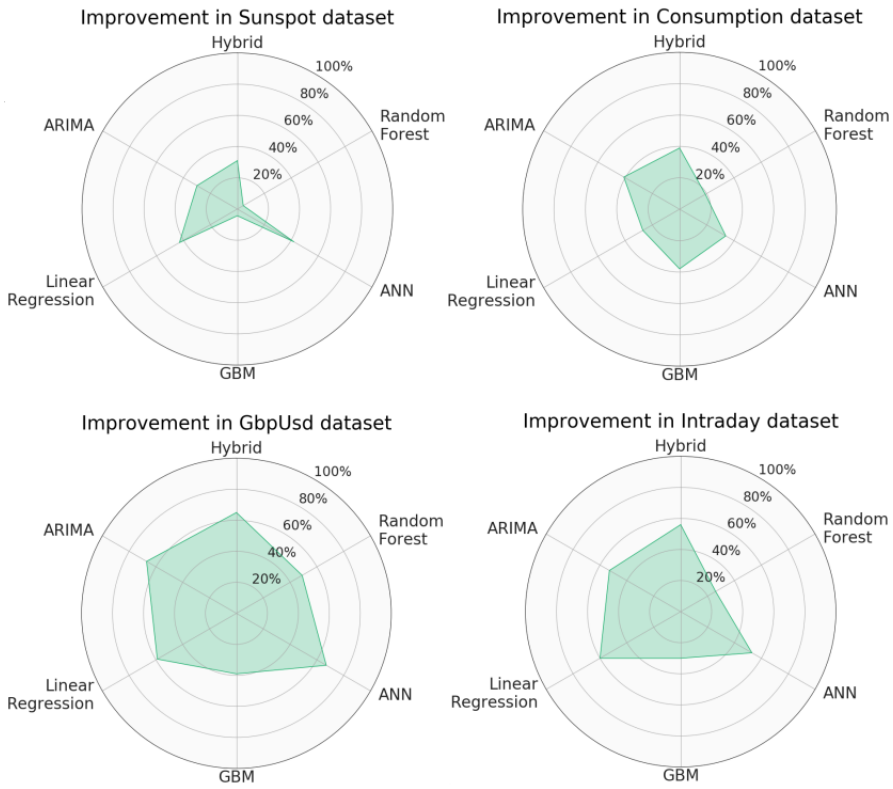

Figure 5.7: Improvement of MAE results of the Recursive EMD-based algorithm

Figure 5.7 shows the improvement of the proposed recursive EMD-based algorithm in the MAE error metric for each method, as compared to their single usages for all datasets. As seen in the figure, there is a significant increase in absolute accuracies when the recursive method is deployed for Consumption, GbpUsd, and Intraday dataset. This improvement varies between 25% and 65% for those datasets. In the Sunspot dataset, accuracy performance is also highly improved, with the exception of Random Forest and GBM. For example, the results of ANN, ARIMA, Linear Regression, the hybrid methods are improved by 42%, 30%, 43%, 31%, respectively.

Another remarkable point that emerges from Figure 5.7 is that the recursive method shows more performance improvement in highly non-stationary and volatile datasets. For example, the proposed method achieves 40-65% improvement in non-stationary dataset, such as GbpUsd; while the improvement in the relatively stationary dataset such as, Sunspot remains between 4-43%.

CHAPTER 6

AN APPLICATION: TURKEY HYDROPOWER PRODUCTION PREDICTION

> Real generosity towards the future lies
> in giving all to the present
>
> *Albert Camus*

Hydropower is one of the most efficient and effective renewable energy resources for Turkey's economic development. Therefore, forecasting hydropower production plays an important role for making decisions in power system planning and operation. In this chapter, time series prediction methods, developed in previous chapters, are used and compared for Turkey hydropower production forecasting [89]. The experiments show that the successively developed methods improve the performance results step by step.

6.1 Introduction

Sustainability and stability of energy are vital for financial development and economic growth. Supplying the energy, which is inevitable input for all industries, through sustainable resources becomes an important part of national economic strategies. It is well known that renewable energy sources, including hydropower, are usable and sustainable in relatively long-term timelines as compared to the limited fossil resources [118]. In developing countries such Turkey, electricity generation from hydropower is substantially greater than any other renewable energy technology. Figure 6.1 shows the percentage distribution of resources in electricity generation in Turkey. As seen from the figure, 20% of electricity production of Turkey in 2018

was obtained from hydropower, whereas only 9% was obtained from other renewable resources (wind and solar) [119]. In addition to this, the largest of the power plant investments in Turkey were made to the hydropower plants. Turkey has more than 500 operational hydropower plants in which approximately 80 of them have more than equal 10 MW installed capacity [120]. Similarly, the largest hydropower plants in the country are Atatürk and Keban dams, with an installed power of 2400 and 1800 MW, respectively. As a result; forecasting hydropower production, which has such a large share and importance in total electricity generation, plays an important role in the Turkey electric industry, since it provides considerable knowledge for making decisions in power system planning and operation.

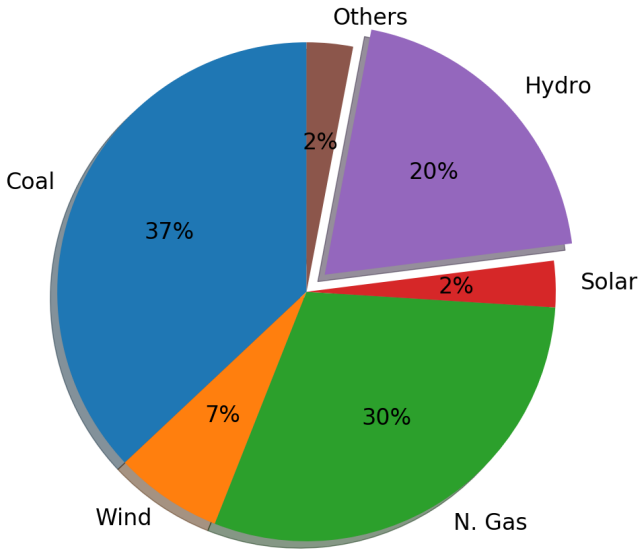

Figure 6.1: The percentage distribution of resources in electricity generation of Turkey in 2018

In literature, the well-accepted and widely used hydrological forecasting method is Hydrologiska Byråns Vattenbalansavdelning (HBV) which basically performs rainfall-runoff model [121]. The HBV model firstly developed for Scandinavians catchments, however then this method and its different versions has been applied in more than 30 countries all over the world such as Switzerland [122], Ethiopia [123], China [124] and the others. The HBV model basically calculates the energy value of water for a catchment area by considering all hydrological processes components such as pre-

cipitation, evapotranspiration. The water balance equation of the method including numerical descriptions of these hydrological processes is shown as follows:

$$P - E - Q = \frac{\mathrm{d}}{\mathrm{d}t}[SP + SM + UZ + LZ + lakes] \tag{6.1}$$

where P is precipitation, E is evapotranspiration, Q is runoff, SP is snow pack, SM is soil moisture, UZ is upper ground zone, LZ is lower ground zone, $lakes$ is lake volume. The method requires precipitation, temperature and evaporation as inputs. On the other hand, it creates various sub-models to obtain other hydrological processes.

There are three main challenges of the HBV method to forecast hydropower production of the whole country. The first challenge is that the method is developed at catchment scale. In order to model the whole country, all catchment should be separately calculated. This process is not easy for a country such as Turkey which has many catchments. The second challenge is that the method requires various exogenous data for each basin such as temperature, precipitation, evaporation. Finally, the performance of the method depends heavily on performances of sub-models e.g. snowpack model, soil moisture model, etc. As a result, the HBV method, taking into various water cycle processes into account at the catchment scale, is not an effective and efficient method for forecasting the total hydropower production of Turkey. Instead, we considered this problem as a univariate time series prediction in this study. In the literature, various methods have been deployed such as ARIMA, ANN by considering hydro production data as a univariate time series [125, 126, 127]. We propose to use our univariate time series forecasting methods (the ARIMA-ANN hybrid method in Section 3.1, the EMD-based hybrid method in Section 4.4, the EMD-based method with method selection in Section 4.5, the Recursive EMD-based method in Section 5.2) that we have developed successively in the previous sections .

6.2 Experimental Results

For Turkey hydropower forecasting, hourly production data of the last two years is used (from Jan. 2017 to Dec. 2018). This data is publicly available [128]. As seen in

Figure 6.2, the production data is highly volatile and includes irregular movements.

In the experiments, the univariate time series prediction methods that we proposed by improving successively the performance results are used. They are the ARIMA-ANN hybrid method, the EMD-based hybrid method , the EMD-based method with method selection approach, and the Recursive EMD-based method, proposed in Section 3.1, Section 4.4, Section 4.5, Section 5.2, respectively. In addition, naïve, ARIMA and ANN singular methods are also evaluated for the sake of comparison. The parameters of the used method are indicated in Table 6.1. While modelling with the examined methods, the dataset is split into three sets such as train, validation and test. The validation set used to adjust hyperparameters of the methods is the 20% of the train set. The test set used to present the performance of the methods is the last three months of the dataset.

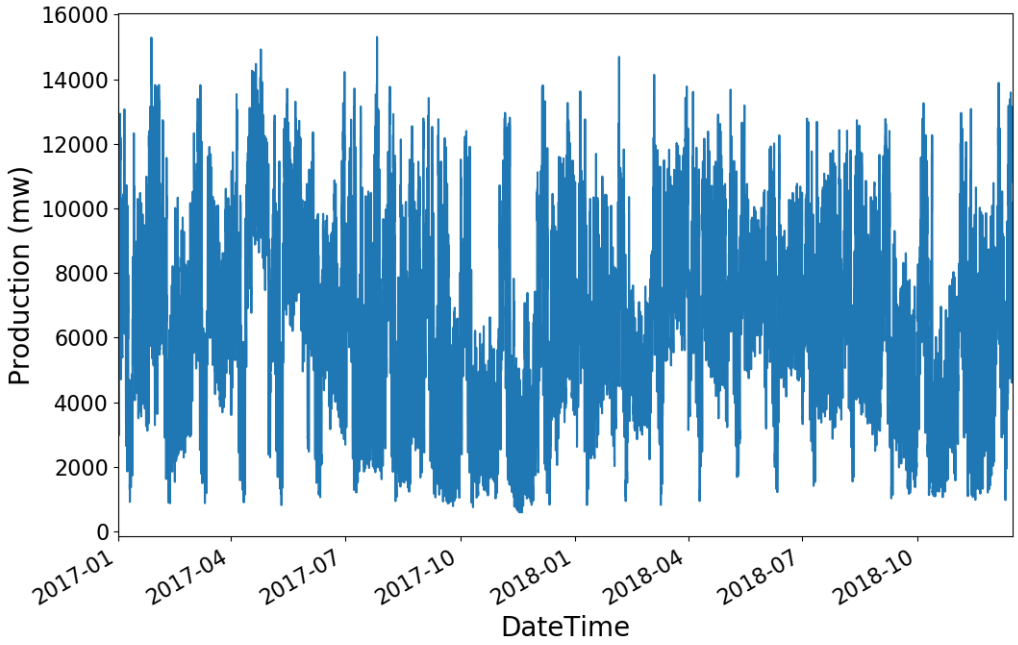

Figure 6.2: The hydropower production of Turkey between Jan. 2017 and Dec. 2018

The methods perform day-ahead forecasting and the evaluation is performed on hourly prediction results. In the evaluation, Normalized Mean Absolute Error (NMAE) and Normalized Mean Squared Error (NMSE) are used. These error metrics present the

normalization of MAE and MSE error metrics by Turkey's installed capacity of hydropower plants which is 28358 MW [120]).

Table 6.1: The parameters of the examined methods for forecasting Turkey Hydropower production

Model	**Parameters**	**Values**
ARIMA	*(p,d,q)*	(11,1,1)
ANN	*architecture*	9x19x1
	activation	ReLU
	optimizer	Nadam
Hybrid	*MA length*	18
Method	*(a,b)*	(6,2)

Table 6.2 shows the forecasting performance results of Turkey hydropower production for all examined methods in terms of NMAE and NMSE.When the results in the table are analyzed, firstly it is seen that all examined methods remarkably improve the naïve method results. Thus, we can say that past movement of the hydro production can be used to predict its future movement by more sophisticated methods. Moreover, it is clearly seen that the ARIMA-ANN hybrid method achieves better accuracy results as compared to individual methods such as ARIMA and ANN by taking unique strength of individual methods. For example , the NMAE results of ANN and ARIMA are 2.81 and 2.80 respectively, whereas the hybrid method is able to produce more accurate forecasts with 2.65 NMAE result.

It is apparent that the methods using EMD technique obtain impressive results as compared to individual methods and combination of both. For example, the mean absolute errors are approximately decreased by 40% when the hybrid method is used with EMD. Similarly, NMAE results of ANN and ARIMA methods are improved by 38% and 40%, respectively, when they are employed with EMD. This improvement is due to the fact that EMD is able to reduce high volatility problems by generating stationary components in time series data. Another indication of EMD's capability of alleviating volatility problems is that ARIMA is able to achieve better results than

Table 6.2: The experimental results of the all examined methods for forecasting Turkey Hydropower production

Methods	**Metrics**	
	NMAE	*NMSE*
Naïve	3.27	6.20
ANN	2.81	3.74
ARIMA	2.80	3.71
The Hybrid Method *proposed in Section 3.1*	2.65	3.53
EMD-based ANN	1.74	1.24
EMD-based ARIMA	1.61	1.14
EMD-based Hybrid *proposed in Section 4.4*	1.59	1.13
EMD-based Method Selection *proposed in Section 4.5*	1.55	1.09
Recursive EMD-based *proposed in Section 5.2*	**1.52**	**1.06**

* NMSE results are multiplied with 10^3 in the table

ANN, although their individual performance results without EMD is almost the same.

When the EMD-based method is enhanced with method selection algorithm as proposed in Section 4.5, there is a clear improvement in forecasting performance results. The method selection algorithm selects appropriate methods according to the character of each component obtained with EMD. The ones with high frequency change rates are modelled by a hybrid method whereas the ones with slow change rate is modelled by ARIMA. Thus, the proposed method which selects the appropriate method for the components is able to improve the accuracy of the models that apply the same method to all components such as EMD-based ARIMA and EMD-based Hybrid methods.

Table 6.3 shows the frequency change rate analysis of the hydropower data and its IMF components. As seen in Figure 6.3, EMD generates 16 IMF components and a

Table 6.3: Frequency change rate analysis of the EMD components of Turkey Hydropower production data

Data	**Frequency Change Rate**	**Selected Method**
Original data	0.090	-
IMF_0	0.223	Hybrid
IMF_1	0.159	Hybrid
IMF_2	0.091	Hybrid
IMF_3	0.052	ARIMA
IMF_4	0.036	ARIMA
IMF_5	0.025	ARIMA
IMF_6	0.016	ARIMA
IMF_7	0.011	ARIMA
IMF_8	0.006	ARIMA
IMF_9	0.004	ARIMA
IMF_10	0.003	ARIMA
IMF_11	0.002	ARIMA
IMF_12	0.001	ARIMA
IMF_13	0.001	ARIMA
IMF_14	0.001	ARIMA
IMF_15	0.001	ARIMA
Residual	0.001	ARIMA

residual from the data. Although EMD guarantees that the obtained components are stationary, they might have different frequency change rate. For example, the first component, called *IMF_0*, has the highest frequency change rate with 0.223 which is even greater than the frequency change rate of the original data. Since fast fluctuation could be the cause of more irregularity, the forecasting performance can decrease for the component which has faster frequency change rate. Thus, for those components whose frequency change rate are faster than the original data are modelled by our

hybrid method proposed in Section 3.1 which has a high performance for nonlinear datasets. As a result of this analysis, the first three components which have fast fluctuations are modelled by the hybrid method, whereas the remaining components with more regular patterns are modelled by ARIMA in the method selection approach.

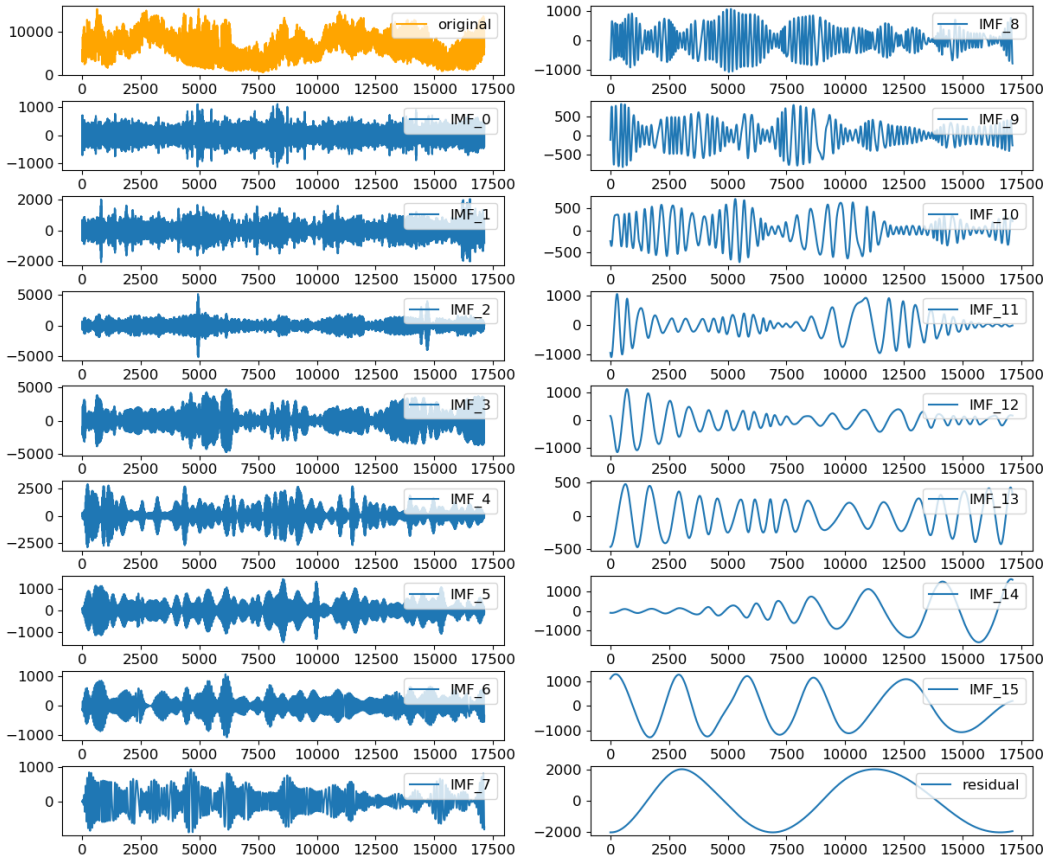

Figure 6.3: The IMFs and residual of the Turkey Hydropower productin data

It is proposed in Section 5.2 to re-decompose these three fast fluctuating components in the Recursive EMD-based method, until it obtains slow fluctuating components. When the iteration terminates, we are able to achieve numerous stationary and slow fluctuating components which are easy to predict. In this way, we can improve the performance results of our other univariate time series forecasting methods. For example, the recursive algorithm improves the NAME results of ARIMA-ANN hybrid method presented in Chapter 3.1 by 42.64%, the NAME results of the EMD-based hybrid method presented in Chapter 4.4 by 4.40%, and the NAME results of the EMD-based method with the method selection approach presented in Section 4.5 by 2.03%.

CHAPTER 7

CONCLUSION AND FUTURE WORK

> Science illuminates the paths of truth
> and any road that doesn't follow
> science, ends in darkness.
>
> *Hacı Bektaş-ı Veli*

In this thesis, we proposed novel methods, that improve performance results successively, for forecasting univariate time series where only one variable changes over time. By evaluating the proposed methods on benchmark and publicly-available datasets, we were able to compare them to the existing well-known methods and make them suitable to be used in future studies. The developed methods can be categorized into three main subjects.

Firstly, we proposed a novel ARIMA-ANN hybrid method, which uses the linear model ARIMA and nonlinear model ANN methods separately on the respective components and combines forecasting results. Although hybrid methods in literature give better results in general as compared to the cases where single methods are individually used, they generally suffer from the assumptions they make while constructing their model. These assumptions lead to produce inconsistent results and give low accuracies in overall if unexpected situations occur. The hybrid method that we introduced in this thesis aims to improve forecasting results by overcoming three main assumptions made by traditional hybrid models. Firstly, the proposed method removes the assumption that the linear component is the ARIMA model output of the given data. Rather, it extracts the linear component by using MA filter. It is known that data showing linear characteristics can be more accurately modeled by linear methods. Therefore, properly decomposed data yields more accurate linear forecast-

ing and consequently more accurate final results in the hybrid methods. Secondly, the proposed method does not directly model residuals via a nonlinear method ANN, since the assumption of that residuals might comprise valid nonlinear patterns, does not always hold. Thirdly, the proposed method does not restrict linear and nonlinear component modeling and also combining the results of them. Rather, it can capture structures of the linear and nonlinear components in a better way, and produces more general models than those existing hybrid models. The evaluations results showed that eliminating these assumptions remarkably improves the results of both the traditional hybrid methods and singular results. This method was then extended with a feature-based hybrid approach where the objective is to present a meaningful summaries of a given time series to a hybrid method. To find the most descriptive features of a given series, the importance scores of each feature is calculated using GBT. The feature-based hybrid method takes the most informative features, which have high importance score, as input and succeeded in reducing the average error values.

In the light of our experimental results of hybrid methods, we concluded that forecasting performance becomes better if more stationary time series data is provided. This result motivated us to make original time series data more stationary in order to improve accuracy results. For this purpose, secondly, we suggested two different methods using EMD multi-scale data decomposition. The first one, called Two-level decomposition method, uses MA filter and EMD to obtain linear components to feed into a linear modelling technique, ARIMA. The second method employs the developed ARIMA-ANN hybrid method on EMD components. We showed that when EMD multi-scale data decomposition is combined with all examined individual and hybrid methods, accuracy results can be remarkably improved. Moreover, our experimental results indicated that our hybrid method with EMD gives remarkably superior accuracy as compared to all other examined hybrid methods. The components obtained by EMD multiscale decomposition ensure stationarity, so that the methods using EMD provide high performance in time series forecasting. However, each of these components has different variations and fluctuations in themselves. We hypothesized that the fast fluctuating components show more irregularity and volatility which prevents them from having high accuracy forecasting, even though they are stationary. For that purpose we suggested a method selection algorithm for EMD-based

methods. The algorithm computes the frequency change rate of the components, i.e. fluctuation rates, by fourier analysis. After that, our high-performance ARIMA-ANN hybrid method for nonlinear datasets is applied to fast-fluctuating components and linear prediction method ARIMA is applied to slow-fluctuating components. The obtained promising results from empirical experiments showed that particular consideration of the components with irregular patterns improves forecasting results.

Thirdly, we proposed another novel method which recursively employ EMD decomposition for those fast fluctuating components, which include irregular patterns, until it achieves more regular and easy-to-predict sub-components. The obtained stationary and slowly fluctuating components are then modelled by our ARIMA-ANN hybrid method which provides high accuracy performance results in time series forecasting. The experimental results indicated that the proposed recursive method achieves remarkably superior accuracy values as compared to single uses of all other examined methods. In addition, the recursive method even outperforms our other developed methods.

When a detailed analysis of the experiment results was made, it was seen that the methods using EMD does not only improve the forecast results, but also reduces the large errors. Based upon the comparison of different error metrics of the methods using EMD, it was seen that the improvements in squared error (MSE), which penalizes the higher error values more, are greater than absolute error (MAE). Furthermore, since the recursive algorithm reduces fast fluctuation rates in the highly non-stationary datasets, the performance improvements in those kind of datasets are more significant than those of the stationary datasets. Consequently, we showed that appropriate decomposition in time series in various levels help improving the forecasting performance significantly.

In addition, step-by-step developed methods in this thesis for prediction in univariate time series, were employed for forecasting Turkey hydropower production. The results demonstrated that making the predicted univariate time series as stationary as possible, based only on historical data, significantly improves forecast performance.

There are several future research directions that we identified for univariate time series forecasting.

In univariate time series forecasting, since there is not any available exogenous series that is explanatory for the original data, only the observed values are used. The feature-based hybrid method we propose in Section 3.2 does not only use historical data, but also extracts some statistical and structural features from those data and selects the most descriptive ones. The method, which gives promising performance results, has three main future research directions that could be applied step-by-step. Firstly, the feature set can be enriched with different time series features and their variations by using different window sizes. Secondly, it will be important to constitute an automated feature selection method to reveal the most descriptive ones from the feature set which grows even more in the first step. Finally, after automating feature selection, this method can be combined with EMD multi-scale decomposition, as proposed in Section 4.4, for better performance results.

The Recursive EMD-based method, introduced in Section 5.2, applies the hybrid method, that we proposed in Section 3.1, to all obtained components. However, according to the results we obtained in Section 4.5 and Section 6.2, modeling by selecting the most appropriate method for each EMD component provides accurate predictions. Another research direction will be to combine these two algorithms, Recursive EMD-based method and method selection approach, respectively. This combined method aims to better understand the design patterns of the components obtained with the Recursive EMD-based method and to determine more suitable methods for each. For this purpose, in addition to the frequency change rate used by our method selection algorithm, the other time series features such as autocorrelation, entropy etc. can be used to determine the most appropriate method.

Finally, deep learning methods achieving successful results in almost all areas today can be another future research direction. The types of Recurrent Neural Networks are specifically designed to capture the temporal relationship in a given dataset. Thus, it may be more appropriate to infer the temporal relationship of a given series through such deep learning methods. These methods, which require huge amount of data, may better model temporal relationships on larger time series.

REFERENCES

[1] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003.

[2] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting," *Applied Soft Computing*, vol. 11, pp. 2664–2675, Mar. 2011.

[3] C. N. Babu and B. E. Reddy, "A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data," *Applied Soft Computing*, vol. 23, pp. 27–38, Oct. 2014.

[4] U. C. Buyuksahin and S. Ertekin, "Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition," *Neurocomputing*, vol. 361, pp. 151–163, Oct. 2019.

[5] D. S. de O. Santos Júnior, J. F. L. de Oliveira, and P. S. G. de Mattos Neto, "An intelligent hybridization of ARIMA with machine learning models for time series forecasting," *Knowledge-Based Systems*, vol. 175, pp. 72–86, July 2019.

[6] "Intraday EPIAS." `https://www.epias.com.tr/en/intra-day-market/introduction`. [Online; accessed 2018-05-14].

[7] T. C. Mills, *Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting*. Academic Press, Feb. 2019.

[8] B. H. Baltagi, *Econometrics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[9] C. Chatfield, *Time-Series Forecasting*. Department of Mathematical Sciences, University of Bath, UK, 2000.

[10] J. H. Cochrane, *Time Series for Macroeconomics and Finance*. Graduate School of Business, University of Chicago, spring, 1997.

[11] R. S. Tsay, *Financial Time Series and Their Characteristics*. John Wiley & Sons, Ltd, Graduate School of Business, University of Chicago, 2010.

[12] D. Dickey and W. Fuller, "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *JASA. Journal of the American Statistical Association*, vol. 74, June 1979.

[13] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?," *Journal of Econometrics*, vol. 54, pp. 159–178, Oct. 1992.

[14] P. C. B. Phillips and P. Perron, "Testing for a Unit Root in Time Series Regression," *Biometrika*, vol. 75, no. 2, pp. 335–346, 1988.

[15] C. R. Rao, H. Toutenburg, Shalabh, and C. Heumann, "Linear Models and Generalizations: Least Squares and Alternatives," 1997.

[16] Denton James W., "How good are neural networks for causal forecasting," *The Journal of Business Forecasting Methods and Systems*, vol. 14, no. 2, pp. 17–20, 1995.

[17] T. H. Hann and E. Steurer, "Much ado about nothing? Exchange rate forecasting: Neural networks vs. linear models using monthly and weekly data," *Neurocomputing*, vol. 10, pp. 323–339, Apr. 1996.

[18] J. L. Callen, C. C. Y. Kwan, P. C. Y. Yip, and Y. Yuan, "Neural network forecasting of quarterly accounting earnings," *International Journal of Forecasting*, vol. 12, pp. 475–482, Dec. 1996.

[19] H. Kantz and T. Schreiber, "Nonlinear Time Series Analysis: Contents," Cambridge University Press, 1997.

[20] W. R. Foster, F. Collopy, and L. H. Ungar, "Neural network forecasting of short, noisy time series," *Computers & Chemical Engineering*, vol. 16, pp. 293–297, Apr. 1992.

[21] M. C. Brace, J. Schmidt, and M. Hadlin, "Comparison of the forecasting accuracy of neural networks with other established techniques," in *Proceedings*

of the First International Forum on Applications of Neural Networks to Power Systems, pp. 31–35, July 1991.

[22] S. Aras and I. D. Kocakoc, "A new model selection strategy in time series forecasting with artificial neural networks: IHTS," *Neurocomputing*, vol. 174, pp. 974–987, Jan. 2016.

[23] J. Contreras, R. Espinola, F. Nogales, and A. J. Conejo, "ARIMA Models to Predict Next-Day Electricity Prices," *Power Engineering Review, IEEE*, vol. 22, pp. 57–57, Oct. 2002.

[24] E. Gonzalez-Romera, M. A. Jaramillo-Moran, and D. Carmona-Fernandez, "Monthly Electric Energy Demand Forecasting Based on Trend Extraction," *IEEE Transactions on Power Systems*, vol. 21, pp. 1946–1953, Nov. 2006.

[25] V. K P, P. Sahu, B. Dhekale, and P. Mishra, "Modelling and Forecasting Sugarcane and Sugar Production in India," *Indian Journal of Economics and Development*, vol. 12, p. 71, Jan. 2016.

[26] W.-c. Wang, K.-w. Chau, D.-M. Xu, and X.-Y. Chen, "Improving Forecasting Accuracy of Annual Runoff Time Series Using ARIMA Based on EEMD Decomposition," *Water Resources Management*, vol. 29, pp. 2655–2675, June 2015.

[27] V. Ediger and S. Akar, "ARIMA forecasting of primary energy demand by fuel in Turkey," *Energy Policy*, vol. 35, pp. 1701–1708, Feb. 2007.

[28] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, "Forecasting," in *Time Series Analysis*, pp. 137–191, John Wiley & Sons, Inc., 2008.

[29] Q. Yao and P. J. Brockwell, "Gaussian Maximum Likelihood Estimation For ARMA Models. I. Time Series," *Journal of Time Series Analysis*, vol. 27, pp. 857–875, Nov. 2006.

[30] R. SHIBATA, "Selection of the Order of an Autoregressive Model by Akaike Information Criterion," *Biometrika*, vol. 63, pp. 117–126, Apr. 1976.

[31] R. H. Jones, "Fitting Autoregressions," *Journal of the American Statistical Association*, vol. 70, pp. 590–592, Sept. 1975.

[32] P. Zhang, E. Patuwo, and M. Hu, "Forecasting With Artificial Neural Networks: The State of the Art," *International Journal of Forecasting*, vol. 14, pp. 35–62, Mar. 1998.

[33] M. J. Campbell and A. M. Walker, "A Survey of Statistical Work on the Mackenzie River Series of Annual Canadian Lynx Trappings for the Years 1821-1934 and a New Analysis," *Journal of the Royal Statistical Society. Series A (General)*, vol. 140, no. 4, pp. 411–431, 1977.

[34] A. Lapedes and R. Farber, "Nonlinear Signal Processing Using Neural Networks: Prediction and System Modelling," *1. IEEE international conference on neural networks, 1987, San Diego*, June 1987.

[35] W.-S. Chen and Y.-K. Du, "Using Neural Networks and Data Mining Techniques for the Financial Distress Prediction Model," *Expert Syst. Appl.*, vol. 36, pp. 4075–4086, Mar. 2009.

[36] D. Singhal and K. S. Swarup, "Electricity price forecasting using artificial neural networks," *International Journal of Electrical Power & Energy Systems*, vol. 33, pp. 550–555, Mar. 2011.

[37] S. F. Ardabili, B. Najafi, S. Shamshirband, B. M. Bidgoli, R. C. Deo, and K. wing Chau, "Computational intelligence approach for modeling hydrogen production: a review," *Engineering Applications of Computational Fluid Mechanics*, vol. 12, no. 1, pp. 438–458, 2018.

[38] B. R. Chang and H. F. Tsai, "Novel hybrid approach to data-packet-flow prediction for improving network traffic analysis," *Applied Soft Computing*, vol. 9, pp. 1177–1183, June 2009.

[39] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals and Systems*, vol. 5, pp. 455–455, Dec. 1992.

[40] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, Jan. 1989.

[41] W. Zhang and A. T. C. Goh, "Multivariate adaptive regression splines and neural network models for prediction of pile drivability," *Geoscience Frontiers*, vol. 7, pp. 45–52, Jan. 2016.

[42] T. Taskaya-Temizel and M. C. Casey, "A comparative study of autoregressive neural network hybrids," *Neural Networks: The Official Journal of the International Neural Network Society*, vol. 18, pp. 781–789, July 2005.

[43] N. Terui and H. Van Dijk, "Combined Forecasts from Linear and Nonlinear Time Series Models," *International Journal of Forecasting*, vol. 18, pp. 421–438, Mar. 2000.

[44] M. Hibon and T. Evgeniou, "To combine or not to combine: selecting among forecasts and their combinations," *International Journal of Forecasting*, vol. 21, pp. 15–24, Jan. 2005.

[45] D. Ömer Faruk, "A hybrid neural network and ARIMA model for water quality time series prediction," *Engineering Applications of Artificial Intelligence*, vol. 23, pp. 586–594, June 2010.

[46] E. Cadenas and W. Rivera, "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model," *Renewable Energy*, vol. 35, pp. 2732–2738, Dec. 2010.

[47] T. Koutroumanidis, K. Ioannou, and G. Arabatzis, "Predicting fuelwood prices in Greece with the use of ARIMA models, artificial neural networks and a hybrid ARIMA–ANN model," *Energy Policy*, vol. 37, pp. 3627–3634, Sept. 2009.

[48] R. Adhikari and R. Agrawal, "A Combination of Artificial Neural Network and Random Walk Models for Financial Time Series Forecasting," *Neural Computing and Applications*, vol. Accepted, Mar. 2013.

[49] M. Khashei and M. Bijari, "An artificial neural network (p,d,q) model for time-series forecasting," *Expert Systems with Applications*, vol. 37, pp. 479–489, Jan. 2010.

[50] K.-Y. Chen and C.-H. Wang, "A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan," *Expert Systems with Applications*, vol. 32, pp. 254–264, Jan. 2007.

[51] J. J. Ruiz-Aguilar, I. J. Turias, and M. J. Jiménez-Come, "Hybrid approaches based on SARIMA and artificial neural networks for inspection time series forecasting," *Transportation Research Part E: Logistics and Transportation Review*, vol. 67, pp. 1–13, July 2014.

[52] B. Zhu and Y. Wei, "Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology," *Omega*, vol. 41, pp. 517–524, June 2013.

[53] R. Shanmugam and R. Chattamvelli, "Skewness and Kurtosis," in *Statistics for Scientists and Engineers*, pp. 89–110, John Wiley & Sons, Ltd, 2016.

[54] P. R. A. Firmino, P. S. G. de Mattos Neto, and T. A. E. Ferreira, "Error modeling approach to improve time series forecasters," *Neurocomputing*, vol. 153, pp. 242–254, Apr. 2015.

[55] S. Panigrahi and H. S. Behera, "A hybrid ETS–ANN model for time series forecasting," *Engineering Applications of Artificial Intelligence*, vol. 66, pp. 49–59, Nov. 2017.

[56] L. Zhou, L. Yu, Y. Wang, Z. Lu, L. Tian, L. Tan, Y. Shi, S. Nie, and L. Liu, "A Hybrid Model for Predicting the Prevalence of Schistosomiasis in Humans of Qianjiang City, China," *PLOS ONE*, vol. 9, no. 8, p. e104875, 2014.

[57] K. Jeong, C. Koo, and T. Hong, "An estimation model for determining the annual energy cost budget in educational facilities using SARIMA (seasonal autoregressive integrated moving average) and ANN (artificial neural network)," *Energy*, vol. 71, pp. 71–79, July 2014.

[58] R. Wongsathan and I. Seedadan, "A Hybrid ARIMA and Neural Networks Model for PM-10 Pollution Estimation: The Case of Chiang Mai City Moat Area," *Procedia Computer Science*, vol. 86, pp. 273–276, Jan. 2016.

[59] H. Nie, G. Liu, X. Liu, and Y. Wang, "Hybrid of ARIMA and SVMs for Short-Term Load Forecasting," *Energy Procedia*, vol. 16, pp. 1455–1460, Jan. 2012.

[60] S. Liu, X. Wang, M. Liu, and J. Zhu, "Towards better analysis of machine learning models: A visual analytics perspective," *Visual Informatics*, vol. 1, pp. 48–56, Mar. 2017.

[61] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: an analysis and review," *International Journal of Forecasting*, vol. 16, no. 4, pp. 437–450, 2000.

[62] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, p. 91, Feb. 2006.

[63] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," *Information Sciences*, vol. 191, pp. 192–213, May 2012.

[64] R. Adhikari and R. Agrawal, *An Introductory Study on Time series Modeling and Forecasting*. LAP Lambert Academic Publishing, Jan. 2013.

[65] R. D. Brooks and R. W. Faff, "A note on beta forecasting," *Applied Economics Letters*, vol. 4, pp. 77–78, Feb. 1997.

[66] C. Willmott and K. Matsuura, "Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance," *Climate Research*, vol. 30, p. 79, Dec. 2005.

[67] R. Hyndman, "Another Look at Forecast Accuracy Metrics for Intermittent Demand," *Foresight: The International Journal of Applied Forecasting*, vol. 4, pp. 43–46, Jan. 2006.

[68] J. Flaherty and R. Lombardo, "Modelling Private New Housing Starts in Australia," *6th Annual Pacific-Rim Real Estate Conference University of Technology Sydney*, Jan. 2000.

[69] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean Absolute Percentage Error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, June 2016.

[70] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures," *IEEE Signal Processing Magazine*, vol. 26, pp. 98–117, Jan. 2009.

[71] U. C. Buyuksahin and S. Ertekin, "A feature-based hybrid arima-ann model for univariate time series forecasting," *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, vol. 35, pp. 467–478, Oct. 2019.

[72] P. Baldi and A. D. Long, "A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes," *Bioinformatics (Oxford, England)*, vol. 17, pp. 509–519, June 2001.

[73] I. Moosa and K. Burns, "The unbeatable random walk in exchange rate forecasting: Reality or myth?," *Journal of Macroeconomics*, vol. 40, pp. 69–81, June 2014.

[74] X. Wang, K. Smith, and R. Hyndman, "Characteristic-Based Clustering for Time Series Data," *Data Mining and Knowledge Discovery*, vol. 13, pp. 335–364, Nov. 2006.

[75] X. Wang, K. A. Smith, and R. J. Hyndman, "Dimension Reduction for Clustering Time Series Using Global Characteristics," in *Computational Science – ICCS 2005* (V. S. Sunderam, G. D. van Albada, P. M. A. Sloot, and J. Dongarra, eds.), Lecture Notes in Computer Science, (Berlin, Heidelberg), pp. 792–795, Springer, 2005.

[76] B. D. Fulcher and N. S. Jones, "Highly Comparative Feature-Based Time-Series Classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, pp. 3026–3037, Dec. 2014.

[77] R. J. Hyndman, E. Wang, and N. Laptev, "Large-Scale Unusual Time Series Detection," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pp. 1616–1619, Nov. 2015. ISSN: 2375-9259.

[78] N. Laptev, J. Yosinski, L. E. Li, and S. Smyl, "Time-series Extreme Event Forecasting with Neural Networks at Uber," *International Conference on Machine Learning*, 2017.

[79] F. Aydin and Z. Aslan, "Yapay Öğrenme Yöntemleri ve Dalgacık Dönüşümü Kullanılarak Nöre Dejeneratif Hastalıkların Teşhisi," *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, vol. 32, pp. 0–0, Sept. 2017.

[80] H. Rao, X. Shi, A. Rodrigue, J. Feng, Y. Xia, M. Elhoseny, X. Yuan, and L. Gu, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Applied Soft Computing*, vol. 74, Nov. 2018.

[81] R. Leardi, "Application of a genetic algorithm to feature selection under full validation conditions and to outlier detection," *Journal of Chemometrics*, 1994.

[82] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[83] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics, New York: Springer-Verlag, 2 ed., 2009.

[84] F. J. O. R. Breiman, L. and Stone, "Classification and Regression Trees," *CRC Press*, Jan. 1984.

[85] E. Sontag, "VC Dimension of Neural Networks," *NATO ASI Series F Computer and Systems Sciences*, Aug. 1998.

[86] U. C. Buyuksahin and S. Ertekin, "Time series forecasting using empirical mode decomposition and hybrid method," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, May 2018. ISSN: null.

[87] U. C. Buyuksahin and S. Ertekin, "Time series prediction with two-level decomposition using moving-average filter and empirical mode decomposition," in *Digital Transformation & Smart Systems, DTSS 2018, Ankara Turkey*, pp. 1–4, Sept. 2018.

[88] U. C. Buyuksahin and S. Ertekin, "Prediction Model Selection with Frequency Check on Decomposed Time Series," in *2019 27th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, Apr. 2019. ISSN: 2165-0608.

[89] U. C. Buyuksahin and S. Ertekin, "Turkey hydropower production prediction based on empirical mode decomposition with method selection," in *Digital Transformation & Smart Systems, DTSS 2019, Ankara Turkey*, pp. 1–4, Sept. 2019.

[90] H. Liu, H.-q. Tian, D.-f. Pan, and Y.-f. Li, "Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Networks," *Applied Energy*, vol. 107, pp. 191–208, July 2013.

[91] P. Bloomfield, *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons, 2nd Edition, isbn: 978-0-471-65399-8, Apr. 2004.

[92] N. E. Huang, M.-L. Wu, W. Qu, S. R. Long, and S. S. P. Shen, "Applications of Hilbert–Huang transform to non-stationary financial time series analysis," *Applied Stochastic Models in Business and Industry*, vol. 19, pp. 245–268, July 2003.

[93] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903–995, Mar. 1998.

[94] A. Zeiler, R. Faltermeier, I. R. Keck, A. M. Tomé, C. G. Puntonet, and E. W. Lang, "Empirical Mode Decomposition - an introduction," in *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, July 2010. ISSN: 1098-7576.

[95] Z. Zhengkai, Y. Zhang, and Y. Zhu, "A new approach to analysis of surface topography," *Precision Engineering-journal of The International Societies for Precision Engineering and Nanotechnology - PRECIS ENG*, vol. 34, pp. 807–810, Oct. 2010.

[96] H. S. Tan, Y. X. Huang, and J. Meng, "Hilbert Statistics of Vorticity Scaling in Two-Dimensional Turbulence," *Physics of Fluids*, vol. 26, Jan. 2014. arXiv: 1401.4200.

[97] S. Charleston-Villalobos, R. González-Camarena, G. Chi-Lem, and T. Aljama-Corrales, "Crackle sounds analysis by empirical mode decomposition. Nonlinear and nonstationary signal analysis for distinction of crackles in lung sounds," *IEEE engineering in medicine and biology magazine: the quarterly*

magazine of the Engineering in Medicine & Biology Society, vol. 26, pp. 40–47, Feb. 2007.

[98] R. B. Pachori, "Discrimination between ictal and seizure-free EEG signals using empirical mode decomposition," *Research Letters in Signal Processing*, vol. 2008, pp. 14:1–14:5, Jan. 2008.

[99] H. Hariharan, A. Gribok, M. A. Abidi, and A. Koschan, "Image Fusion and Enhancement via Empirical Mode Decomposition," *Journal of Pattern Recognition Research*, vol. 1, no. 1, pp. 16–31, 2006.

[100] A. Linderhed, "Image empirical mode decomposition: a new tool for image processing," *Advances in Adaptive Data Analysis*, vol. 01, pp. 265–294, Apr. 2009.

[101] R. Srinivasan, R. Rengaswamy, and R. Miller, "A modified empirical mode decomposition (EMD) process for oscillation characterization in control loops," *Control Engineering Practice*, vol. 15, pp. 1135–1148, Sept. 2007.

[102] L. Luo, Y. Z. Yan, P. Xie, J. Sun, Y. Xu, and J. Yuan, "Hilbert–Huang transform, Hurst and chaotic analysis based flow regime identification methods for an airlift reactor," 2012.

[103] X. Xu, Y. Qi, and Z. Hua, "Forecasting demand of commodities after natural disasters," *Expert Systems with Applications*, vol. 37, pp. 4313–4317, June 2010.

[104] Z.-Y. Wang, J. Qiu, and F.-F. Li, "Hybrid Models Combining EMD/EEMD and ARIMA for Long-Term Streamflow Forecasting," *Water*, vol. 10, p. 853, July 2018.

[105] Y. Wei and M.-C. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 21, pp. 148–162, Apr. 2012.

[106] L. Yu, S. Wang, K. K. Lai, and F. Wen, "A multiscale neural network learning paradigm for financial crisis forecasting," *Neurocomputing*, vol. 73, pp. 716–725, Jan. 2010.

[107] Y. Ren, P. N. Suganthan, and N. Srikanth, "A Novel Empirical Mode Decomposition With Support Vector Regression for Wind Speed Forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, pp. 1793–1798, Aug. 2016.

[108] G.-F. Fan, L.-L. Peng, W.-C. Hong, and F. Sun, "Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression," *Neurocomputing*, vol. 173, pp. 958–970, Jan. 2016.

[109] S. Monjoly, M. André, R. Calif, and T. Soubdhan, "Hourly forecasting of global solar radiation based on multiscale decomposition methods: A hybrid approach," *Energy*, vol. 119, pp. 288–298, Jan. 2017.

[110] X. Zhang, K. K. Lai, and S.-Y. Wang, "A new approach for crude oil price analysis based on Empirical Mode Decomposition," *Energy Economics*, vol. 30, pp. 905–918, May 2008.

[111] P. Singh, S. D. Joshi, R. K. Patney, and K. Saha, "The Fourier Decomposition Method for nonlinear and nonstationary time series analysis," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, vol. 473, p. 20160871, Mar. 2017.

[112] L. Tang, S. Wang, K. He, and S. Wang, "A novel mode-characteristic-based decomposition ensemble model for nuclear energy consumption forecasting," *Annals of Operations Research*, vol. 234, Apr. 2014.

[113] J. Dong, W. Dai, L. Tang, and L. Yu, "Why do EMD-based methods improve prediction? A multiscale complexity perspective," *Journal of Forecasting*, Mar. 2019.

[114] U. C. Buyuksahin and S. Ertekin, "Recursive decomposition in univariate time series forecasting," in *IEEE Transactions on Neural Networks and Learning Systems*, Feb. 2020. [on review].

[115] J. Hauke and T. Kossowski, "Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data," *Quaestiones Geographicae*, vol. 30, pp. 87–93, June 2011.

[116] "Real time consumption EPIAS." https://seffaflik.epias.com.tr/transparency/tuketim/gerceklesen-tuketim/gercek-zamanli-tuketim.xhtml. [Online; accessed 2018-05-14].

[117] R. J. Hyndman and Y. Khandakar, "Automatic Time Series Forecasting: The forecast Package for R," *Journal of Statistical Software*, vol. 27, pp. 1–22, July 2008.

[118] O. Yuksek, M. Komurcu, I. Yuksel, and K. Kaygusuz, "The role of hydropower in meeting Turkey's electric energy demand," *Energy Policy*, vol. 34, pp. 3093–3103, Nov. 2006.

[119] "T.C. Enerji ve Tabii Kaynaklar Bakanlığı - Elektrik." http://www.enerji.gov.tr/tr-TR/Sayfalar/Elektrik. [Online; accessed 2019-01-01].

[120] "Hidroelektrik Santralleri (HES)." https://www.enerjiatlasi.com/hidroelektrik. [Online; accessed 2019-01-01].

[121] S. Bergström, *Development and Application of a Conceptual Runoff Model for Scandinavian Catchments*, vol. 134 pp. SMHI Norrköping, Jan. 1976.

[122] W. M. Organization (WMO) and W. M. O. (WMO), *Operational hydrology report (OHR), 23. Intercomparison of models of snowmelt runoff*. WMO, Geneva: WMO, 1986.

[123] Y. Faouzi, "Hydropower forecast in ethiopia using the hbv model - a case study in omo-gibe river basin," 2019. Master thesis in Department of Building and Environmental Technology Lund University.

[124] L. Liu, C. Xiao, L. Du, P. Zhang, and G. Wang, "Extended-Range Runoff Forecasting Using a One-Way Coupled Climate–Hydrological Model: Case Studies of the Yiluo and Beijiang Rivers in China," *Water*, vol. 11, p. 1150, June 2019.

[125] N. Ghadimi and M. H. Firouz, "Short-term management of hydro-power systems based on uncertainty model in electricity markets," *Journal of Power Technologies*, vol. 95, pp. 265–272, Dec. 2015.

[126] S. Wang, L. Yu, L. Tang, and S. Wang, "A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China," *Energy*, vol. 36, pp. 6542–6554, Nov. 2011.

[127] G. Li, Y. Sun, Y. He, X. Li, and Q. Tu, "Short-Term Power Generation Energy Forecasting Model for Small Hydropower Stations Using GA-SVM," *Mathematical Problems in Engineering*, vol. 2014, pp. 1–9, June 2014.

[128] "Real-Time Generation Generation | EXIST Transparency Platform." `https://seffaflik.epias.com.tr/transparency/uretim/gerceklesen-uretim/gercek-zamanli-uretim.xhtm`. [Online; accessed 2019-01-01].

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Büyükşahin, Ümit Çavuş
Nationality: Turkish (TC)
Date and Place of Birth: 03.06.1988, Kırşehir
Marital Status: Married

EDUCATION

Degree	Institution	Year of Graduation
M.S.	KTH Royal Institute of Technology, Sweden, and Universitat Politècnica de Catalunya (UPC), Spain	2013
B.S.	Dept. of Computer Engineering, METU, Turkey	2011
High School	Kırşehir Science High School	2006

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2017 - current	Vitus Commodities, Ankara, Turkey	Sr. Data Scientist
2015 - 2017	Arverus, Ankara, Turkey	Co-Founder,
2014 - 2015	Aurvis, Ankara, Turkey	Sr. R&D Engineer
2014 - 2015	Comodo, Ankara, Turkey	R&D Engineer
2013 - 2014	Intellica, İstanbul, Turkey	Consultant
2013 - 2013	Distributed System Group at UPC, Spain	Researcher

PUBLICATIONS

1. U. C. Buyuksahin and S. Ertekin, “Recursive decomposition in univariate time series forecasting,” in IEEE Transactions on Neural Networks and Learning Systems, Feb. 2020. (Under Revision).

2. U. C. Buyuksahin and S. Ertekin, “Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition,” Neurocomputing, vol. 361, pp. 151–163, Oct. 2019.

3. U. C. Buyuksahin and S. Ertekin, “A feature-based hybrid arima-ann model for univariate time series forecasting,” Journal of the Faculty of Engineering and Architecture of Gazi University, vol. 35, pp. 467–478, Oct. 2019.

4. U. C. Buyuksahin and S. Ertekin, “Turkey hydropower production prediction based on empirical mode decomposition with method selection,” in Digital Transformation & Smart Systems, DTSS 2019, Ankara Turkey, pp. 1–4, Sept. 2019.

5. U. C. Buyuksahin and S. Ertekin, “Prediction Model Selection with Frequency Check on Decomposed Time Series,” in 2019 27th Signal Processing and Communications Applications Conference (SIU), pp. 1–4, Apr. 2019.

6. U. C. Buyuksahin and S. Ertekin, “Time series prediction with two-level decomposition using moving-average filter and empirical mode decomposition,” in Digital Transformation & Smart Systems, DTSS 2018, Ankara Turkey, pp. 1-4, Sept. 2018.

7. U. C. Buyuksahin and S. Ertekin, “Time series forecasting using empirical mode decomposition and hybrid method,” in 2018 26th Signal Processing and Communications Applications Conference (SIU), pp. 1–4, May 2018.

8. A. M. Khan, U. C. Buyuksahin, F. Freitag, Incentive-based Resource Assignment and Regulation for Collaborative Cloud Services in Community Networks, In Journal of Computer and System Sciences (JCSS), 81(8), pp. 1479-1495. Dec 2015.

9. A. M. Khan, U. C. Buyuksahin, F. Freitag, Prototyping Incentive-based Resource Assignment for Clouds in Community Networks , The 28th IEEE International Conference on Advanced Information Networking and Applications. (AINA-2014). Victoria, Canada, May 13-16, 2014.

10. A. M. Khan, U. C. Buyuksahin, F. Freitag, Towards Incentive-based Resource Assignment and Regulation in Clouds for Community Networks, 10th International Conference on Economics of Grids, Clouds, Systems and Services, GECON-2013, Zaragoza, Spain, September 18-20, 2013.

11. A. M. Khan, U. C. Buyuksahin, F. Freitag, Support service for reciprocal computational resource sharing in wireless community networks, 2013 IEEE 14th International Symposium on A World of Wireless, Mobile and Multimedia Networks, Madrid, Spain, 2013.

12. U. C. Buyuksahin, Y. Gokirmak, IPv6 Saldırı Trafiginden İmza Üretme, IPv6 Konferansı, Ankara, Turkey, 12-13 Jan 2011.