

IMPROVED WIND POWER FORECASTING USING COMBINATION
METHODS

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

BY

CEYDA ER KÖKSOY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

SEPTEMBER 2015

Approval of the thesis:

**IMPROVED WIND POWER FORECASTING USING
COMBINATION METHODS**

submitted by **CEYDA ER KÖKSOY** in partial fulfillment of the requirements
for the degree of **Master of Science in Computer Engineering Department,**
Middle East Technical University by,

Prof. Dr. M. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering**

Dr. Ayşe Nur Birtürk
Supervisor, **Computer Engineering Department, METU**

Assoc. Prof. Dr. Pınar Karagöz
Co-supervisor, **Computer Engineering Dept., METU**

Examining Committee Members:

Prof. Dr. Nihan Çiçekli
Computer Engineering Department, METU

Dr. Ayşe Nur Birtürk
Computer Engineering Department, METU

Assoc. Prof. Dr. Pınar Karagöz
Computer Engineering Department, METU

Prof. Dr. İsmail Hakkı Toroslu
Computer Engineering Department, METU

Assoc. Prof. Dr. Osman Abul
Computer Engineering Department, TOBB ETU

Date: 01/09/2015

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, Last Name: CEYDA ER KÖKSOY

Signature :

ABSTRACT

IMPROVED WIND POWER FORECASTING USING COMBINATION METHODS

Köksoy, Ceyda Er

M.S., Department of Computer Engineering

Supervisor : Dr. Ayşe Nur Birtürk

Co-Supervisor : Assoc. Prof. Dr. Pınar Karagöz

September 2015, 86 pages

Wind is an important renewable energy source to produce electricity thanks to its reliable, omnipresent and economically feasible characteristics and it has a growing proportion in overall energy production worldwide. However, integration of the generated wind power into the existing transmission grid is an issue due to inherently volatile and intermittent behavior of wind. Moreover, the power plant owners need reliable information about day-ahead power production for market operations. Therefore, wind power forecasting approaches have been gaining importance in renewable energy research area. There are many applicable wind power forecasting models including physical model, several statistical models such as ANN and SVM, and hybrid models. However, all of them have different advantages and disadvantages according to the wind characteristic of wind power plant region. At this point, forecast combination approaches stand out not to rely on a single forecast model, and also utilize forecast diversification. A combined forecast should be better than the individual forecasts,

or at least be equal to the best performed one in order to be regarded as an ideal combination. Within the scope of this thesis, various forecast combination methods are proposed to provide ideally combined forecasts. These combination methods have been verified on forecasts data of The Wind Power Monitoring and Forecast System for Turkey (RİTM). The experimental results show that all of the applied combination methods give better forecast error rates for most of the wind power plants compared to individual forecasts.

Keywords: Wind Power Forecasting, Forecast Combination, L_p -Norm Estimators, Fuzzy Soft Sets, Regression Trees, Random Forest

ÖZ

KOMBİNASYON YÖNTEMLERİ KULLANILARAK RÜZGAR GÜCÜ TAHMINLERİNİN GELİŞTİRİLMESİ

Köksoy, Ceyda Er

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi : Dr. Ayşe Nur Birtürk

Ortak Tez Yöneticisi : Doç. Dr. Pınar Karagöz

Eylül 2015 , 86 sayfa

Emniyetli, her yerde ve her zaman hazır olarak bulunabilen ve ekonomik açıdan elverişli olan rüzgar, dünya çapındaki toplam elektrik üretimindeki payı giderek artan önemli bir yenilenebilir enerji kaynağıdır. Ancak, rüzgar doğası gereği değişken ve devamlılığı olmayan bir yapıdadır. Bu nedenle rüzgardan üretilen gücün, kararlı yapıda olan elektrik şebekesine entegrasyonu sorun oluşturabilmektedir. Ayrıca, santral sahipleri enerji piyasasına bildirmek üzere gün öncesinden rüzgar gücü üretim bilgilerine ihtiyaç duymaktadır. Bu sebeplerden dolayı, rüzgar gücü tahmin çalışmaları yenilenebilir enerji alanında önemli bir yer kazanmıştır. Günümüzde kullanılan ve uygulanabilir olan birçok rüzgar gücü tahmin modeli bulunmaktadır. Fiziksel model, ANN ve SVM gibi istatistiksel modeller ile hibrit modeller bunlara örnek olarak verilebilir. Ancak, rüzgar gücü santralinin bulunduğu bölgedeki rüzgar karakteristiği göz önüne alındığında, bu yöntemlerin birbirlerine göre bazı avantaj ve dezavantajları vardır. Bu noktada, rüzgar

gücü tahmin sistemini tek bir tahmin modeline dayandırmak yerine, tahmin modellerinin çeşitliliğinden yararlanan bir yaklaşım benimsenmektedir. Bu yaklaşım tahminlerin birleştirilmesi ya da tahminlerin kombinasyonu olarak isimlendirilebilir. Başarılı bir tahmin kombinasyonu, kendini oluşturan özgün tahminlerinin her birinden daha iyi sonuç vermeli ya da, en azından en iyi sonucu veren özgün tahmin ile eşit sonuç vermelidir. Bu tez kapsamında, ideal tahmin kombinasyonunu elde edebilmek için çeşitli tahmin kombinasyon yöntemleri geliştirilmiştir. Geliştirilen bu yöntemler, Türkiye Rüzgar Gücü İzleme ve Tahmin Merkezi'nin (RİTM) tahmin verileri üzerinde test edilmiştir. Deneysel sonuçlar neticesinde, test edilen rüzgar gücü santrallerinin birçoğu için, önerilen tüm kombinasyon yöntemleri özgün tahminlere kıyasla daha iyi tahmin hata oranları ile sonuçlanmıştır.

Anahtar Kelimeler: Rüzgar Gücü Tahmini, Tahminlerin Kombinasyonu, L_p -Norm Tahmincileri, Bulanık Yazılımsal Kümeler, Regresyon Ağaçları, Rastgele Orman

To my little brother Çağın Alp

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my supervisor Dr. Ayşenur Birtürk and my co-supervisor Assoc. Prof. Dr. Pınar Karagöz for their guidance, criticism, and engagement on this study.

I wish to thank Prof. Dr. Nihan Çiçekli, Prof. Dr. İsmail Hakkı Toroslu and Assoc. Prof. Dr. Osman Abul for their valuable comments during my thesis defense presentation.

I wish to express my thanks to TÜBİTAK MAM Energy Institute, for providing me with all the necessary facilities for this thesis.

I am grateful to Dr. Turan Demirci, Dr. Dilek Küçük, Serkan Buhan and Mehmet Barış Özkan, who provided great motivation during this study and also contributed with their knowledge and wisdom.

My sincere thanks also goes to my family and my friends for their support, patience, and encouragement throughout this thesis.

I also place on record, my sense of gratitude to one and all, who directly or indirectly have contributed this study.

TABLE OF CONTENTS

ABSTRACT	v
ÖZ	vii
ACKNOWLEDGMENTS	x
TABLE OF CONTENTS	xi
LIST OF TABLES	xiv
LIST OF FIGURES	xvi
LIST OF ABBREVIATIONS	xviii
CHAPTERS	
1 INTRODUCTION	1
1.1 Overview	1
1.2 Motivation	3
1.3 Combination of Forecasts	4
1.3.1 Concepts in Forecast Combination	4
1.3.2 Proposed Combination Methods and Contributions	5
1.4 Organization of the Thesis	6
2 BACKGROUND	7

2.1	Wind Power Forecasting Domain	7
2.1.1	Numerical Weather Predictions	9
2.1.2	Wind Power Forecasting Methods	10
2.1.2.1	Physical Models	11
2.1.2.2	Statistical Models	11
2.1.2.3	Hybrid Models	12
2.1.3	Performance Evaluation of Power Forecast Systems	12
2.2	Predictive Modeling	15
2.2.1	Multiple Linear Regression	16
2.2.1.1	L_p -Norm Estimators	18
2.2.2	Tree-Based Models	19
2.2.2.1	Classification and Regression Trees	20
2.2.2.2	Random Forests	21
2.2.2.3	Variable Importance	22
2.2.3	Fuzzy Soft Sets	23
3	ARCHITECTURE OF RITM SYSTEM	27
3.1	An Overview	27
3.2	Wind Power Monitoring and Forecast Center	28
3.3	Data Sources	30
3.3.1	Wind Power Analyzers	30
3.3.2	Wind Masts	31

3.3.3	Medium Scale NWP	31
3.4	Forecast Models	33
3.4.1	Physical Model	33
3.4.2	ANN Model	33
3.4.3	SVM Model	35
3.4.4	Cluster Based Model	35
3.5	User Software Applications	36
4	PROPOSED COMBINATION METHODS	41
4.1	L_p -Norm Based Forecast Combination	42
4.2	Tree-Based Forecast Combination	44
4.3	FSS Based Forecast Combination	46
5	EXPERIMENTAL RESULTS	49
5.1	Combination Results and Evaluations	53
5.1.1	L_p -Norm Based Combination Results	53
5.1.2	Tree-Based Combination Results	65
5.1.3	FSS Based Combination Results	72
5.2	Comparison of the Combination Methods	76
6	CONCLUSION AND FUTURE WORK	79
	REFERENCES	81

LIST OF TABLES

TABLES

Table 2.1	Multiple Linear Regression Dataset	18
Table 2.2	Tabular representation of the soft set (F, E)	24
Table 2.3	Tabular representation of the fuzzy soft set (f, A)	25
Table 4.1	Tabular Representation of FSS	48
Table 5.1	WPPs' Installed Capacities	50
Table 5.2	Individual Forecasts of RITM	52
Table 5.3	NMAEs of L_p -Norm Estimators Based Forecast Combination	60
Table 5.4	NRMSEs of L_p -Norm Estimators Based Forecast Combination	61
Table 5.5	NMBs of L_p -Norm Estimators Based Forecast Combination .	62
Table 5.6	Variable Importance List of Regression Tree of WPP-9	65
Table 5.7	NMAEs of Tree-Based Forecast Combination	69
Table 5.8	NRMSEs of Tree-Based Forecast Combination	70
Table 5.9	NMBs of Tree-Based Forecast Combination	71
Table 5.10	NMAEs of FSS Based Forecast Combination	73
Table 5.11	NRMSEs of FSS Based Forecast Combination	74
Table 5.12	NMBs of FSS Based Forecast Combination	75

Table 5.13 NMAE Comparison among Forecast Combination Methods . .	77
---	----

LIST OF FIGURES

FIGURES

Figure 3.1	The General Architecture of RITM	28
Figure 3.2	A panoramic view from the Center	29
Figure 3.3	A View from Wind Power Analyzer	31
Figure 3.4	A View from Wind Mast	32
Figure 3.5	An example to 3D modeling for physical forecasts of a particular WPP	34
Figure 3.6	Flow Diagram of the Training Algorithm of Cluster Based Forecasts Model	37
Figure 3.7	The General Structure of User Subsystem of RITM	38
Figure 3.8	Real-time Monitoring Software	39
Figure 4.1	The General Structure of Combination Process	42
Figure 4.2	Timeframe of Each Daily Run For Combination	43
Figure 5.1	A Comparative View to Q-Q Plots of L_p Norm Estimators	55
Figure 5.2	Histogram Plots of L_p Norm Estimators	56
Figure 5.3	Kernel Density Plots of L_p Norm Estimators	57
Figure 5.4	Comparable Kernel Density Plots of L_p Norm Estimators	58

Figure 5.5	Residuals vs. Fitted Values	59
Figure 5.6	Series of Combination Model Forecasts and the Actual Power	64
Figure 5.7	Pruned Regression Tree of WPP-9	68

LIST OF ABBREVIATIONS

RİTM	Wind Power Monitoring and Forecast Center for Turkey
TÜBİTAK	The Scientific and Technological Research Council of Turkey
YEGM	General Directorate of Renewable Energy of Turkey
TSO	Transmission System Operator
WPP	Wind Power Plant
ANN	Artificial Neural Networks
SVM	Support Vector Machines
NWP	Numerical Weather Prediction
FW	Fixed Weights
FSS	Fuzzy Soft Sets
MM5	Mesoscale Model Version 5
CFD	Computational Fluid Dynamics
MOS	Model Output Statistics
AR	Autoregressive
MA	Moving Average
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
KF	Kalman Filters
NNs	Neural Networks
MB	Mean Bias
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
MAPE	Mean Absolute Percentage Error
NMB	Normalized Mean Bias
NMAE	Normalized Mean Absolute Error
NRMSE	Normalized Root Mean Squared Error
LAD	Least Absolute Deviations
LAE	Least Absolute Errors

LAV	Least Absolute Value
LAR	Least Absolute Residuals
CART	Classification and Regression Trees
TEİAŞ	Turkish Electricity Transmission Company
WPA	Wind Power Analyzer
PQ	Power Quality
ADSL	Asymmetric Digital Subscriber Line
GPRS	General Packet Radio Service
GFS	Global Forecast System
ECMWF	European Center for Medium range Weather Forecasting
WRF	Weather Research and Forecasting Model
DMİ	General Directorate of Meteorology of Turkey
ALADIN	Air Limit Adaptation Dynamic Development International
WAsP	Wind Atlas Analysis and Application Program
PCA	Principal Component Analysis

CHAPTER 1

INTRODUCTION

1.1 Overview

Environmental aspects and quality of life indicate that environmental pollution (of air, water, etc.) and climate changes are largely related to the excessive usage of conventional energy sources based on oil, coal, and natural gas. These traditional energy sources not only threaten the human health, but also damage economic progress, since obtaining energy from these sources is a costly process. However, the main challenge is that energy demand of the world is increasing, while it is running short of fossil-fuel based energy sources day by day. All the mentioned negative aspects of these non-renewable sources redound to importance of renewable resources in overall energy production [1, 2].

Renewable energy resources can be generally defined as the sources which are continually replenished by nature and derived directly from the sun (such as thermal, photo-chemical, and photo-electric), indirectly from the sun (such as wind, hydropower, and photosynthetic energy stored in biomass), or from other natural movements and mechanisms of the environment (such as geothermal and tidal energy). In comparison to the fossil fuels, these sources do not have any negative effects on the environment and human health, while turning into usable forms of energy such as electricity, heat and fuels [3].

Among all renewable energy sources, wind has been ranked as the first renewable energy resource in the world as regards its potential to produce electricity [2]. It is a type of clean energy source with its environment-friendly, renewable and

sustainable nature. In addition, it is a cost effective energy source with its feasible rich potential. However, a large-scale introduction of wind power causes a number of challenges for electricity market and power system operators.

Unlike most of the other renewable energy sources, wind is inherently variable and uncertain. Therefore, it is a fluctuating source of electrical energy which causes difficulties for Transmission System Operators (TSOs) and wind power plant (WPP) owners who need reliable information for unit commitment, dispatching and trading in electricity markets [4]. At this point, forecasting the behavior of wind has emerged as an important issue in renewable energy research. Moreover, accurate wind power prediction systems are required for WPP owners to reduce the financial and technical risk of the uncertainty of wind power production, since variations in the estimated wind power will influence the clearing prices for both energy and operating reserves [5].

A great number of studies have been conducted in order to benefit from wind energy in all over the world. Generally, countries are building wind power monitoring and forecast centers in order to handle the variability of wind. Experience has shown that accurate and reliable wind power forecasting systems are widely recognized as a major contribution for increasing wind penetration [6]. Accordingly in our country, a large-scale wind power management system for Turkey, called Rüzgar Gücü İzleme ve Tahmin Merkezi (RİTM), is developed by The Scientific and Technological Research Council of Turkey (TÜBİTAK) for General Directorate of Renewable Energy of Turkey (YEĞM) with the features of on-line monitoring of wind power plants and short-term forecasts up to 48 hours.

In recent years, various short-term wind power forecasting models have been developed using physical or statistical approaches. The physical approach attempts to predict local wind speed by using atmospheric behavior, and then the corresponding power generated at the wind farm, while the statistical approach attempts to determine the relationship between a set of explanatory variables and the power generated at the wind farm by using historical data [7].

Forecast models of RİTM includes a physical model, two of the commonly used statistical models, namely Artificial Neural Network (ANN) and Support Vec-

tor Machine (SVM), and a hybrid model composed of statistical methods and data mining approach, which will be referred to as cluster based forecasts in the following sections. All these models are applied to multiple numerical weather prediction (NWP) sources independently. Thus, a certain number of power forecasts can be obtained that are corresponding to each hourly power generation.

1.2 Motivation

Faced with multiple forecasts of the same variable, which one would be the best to choice? In particular, should a single dominant forecast be identified as the best one, or should a combination of the outstanding forecasts be used to produce a pooled summary measure? Theoretically, unless one can identify a specific forecast model that frequently outputs lower forecast errors when compared to its competitors, combination of forecast models tends to give better results rather than relying on a single model, because of the advantage of diversification [8].

The concept of combining forecasts started with the study in [9], which claimed that a suitable linear combination of forecasts could achieve higher forecasting accuracy than a policy of trying to identify and use the single best forecast, in the sense of a smaller error variance. Over the years, applications of combining forecasts have been found in many fields such as meteorology, economics, insurance and forecasting sales and price [10].

In the wind energy industry, the availability of alternative forecasts is rather usual. There are various NWP sources from several prediction agencies, and also several applicable methods for wind power forecasting. So, it is difficult to decide which one is the most appropriate to use among a set of alternative forecasts [11].

Combination of wind power forecasts becomes an issue as discussed in the literature to obtain a final single forecast output, if multiple forecast outputs exist. Moreover, an individual forecast may not always output satisfactorily, while a combined forecast that takes the advantages of the individual ones may give

better results. As a consequence, the researchers who study on wind power prediction not only use multiple NWP sources on their proposed forecast methods, but also try different procedures on a particular NWP source. Then, they try to combine all applied forecasts instead of choosing the best individual one.

Combining several wind power predictions is valuable in exploiting the fact that each forecast model has strengths and weaknesses in different situations [12, 13]. Ideally, a combined forecast has to be better than the individual predictions, or at least be equal to the best performed prediction to be regarded as a successful combination of forecasts [11].

There are a number of studied forecast models within the scope of RITM, therefore, this thesis' motivation is applying different forecast combination methods on all these forecast models' outputs to find the optimal combination solution, and also to obtain better wind power predictions for RITM.

1.3 Combination of Forecasts

In this section, the idea behind the forecast combination, and forecast combination approaches in the literature are presented in a general view. Moreover, the proposed combination methods with their contributions to RITM are briefly mentioned.

1.3.1 Concepts in Forecast Combination

A final single forecast value, obtained from a combination of the individual forecasts, is the output of a function that computes the sum of the weighted forecasts produced by different forecast models. The weights (presented between 0 and 1) show the contribution of corresponding forecast models used in the combination method. With n individual forecasts, f_{it} , for $i = 1, \dots, n$, the combined forecast

is

$$f_{Ct} = \sum_{i=1}^n w_i f_{it} \quad (1.1)$$

with $\sum w_i = 1$, if the individual forecasts are unbiased and this is also desired for the combined forecast [14].

There are two ways to construct the weight of each individual forecast. The first one is finding the optimal fixed weights (FWs), and the second is changing weights from time to time. Both fixed and time-varying weights can be determined by using various optimization methods, or taking some assumptions into account [15, 16].

A growing number of studies have appeared in the literature over the last decades to find the optimal forecast combinations. In order to obtain accurate forecasts, the key point in the combination procedure is determination of the weights and the type of combination function. The combination function can be chosen as a linear or a non-linear function. Conventional approaches presented in forecast combination studies are generally focused on linear combination methods, such as various types of unconstrained or constrained regression techniques, successive averaging, and multiple objective linear programming models. [15, 17].

1.3.2 Proposed Combination Methods and Contributions

RITM has a naive combination module to publish a final single forecast only from Cluster-based Forecasts. However, it needs to be improved and expanded to cover all the studied forecast models including also physical model, ANN model, and SVM model. Therefore, different combination methods are proposed in this study to combine N -individual forecast models and evaluated in terms of normalized mean bias, normalized mean absolute errors, normalized root mean square errors of generated forecasts.

Within the scope of this thesis, there are three main proposed approaches for wind power forecast combination that are generally based on time-varying weight

determination. The first one uses L_p -norm estimators for multiple linear regression estimation, the second one uses tree-based methods including regression trees and random forests, and the last one uses fuzzy soft sets (FSS) to determine weights of short-term forecasts. However, L_p -norm estimators based combination method is stated as more feasible solution for RITM combination problem, compared to tree-based methods and FSS based method.

1.4 Organization of the Thesis

The rest of this thesis is organized as follows.

In Chapter 2, background information about wind power forecasting domain is given that includes NWPs and commonly used wind power forecasting methods. In addition, some of the predictive modeling techniques used in statistics, data mining and machine learning are presented in this chapter in order to introduce the proposed forecast combination methods.

The general architecture of RITM is given in Chapter 3. The details of the data used in the forecast models of RITM, including data acquisition and data storage procedures are explained in this section. Moreover, forecast models of RITM, which participate in combination methods, are presented.

In Chapter 4, the proposed combination methods, namely L_p -norm estimators based combination, tree-based combination, and FSS based combination are described in detail, together with their implementations to RITM. In addition, related studies on forecast combination inspiring to proposed combination methods, that are not only specific to wind power forecasting, but also used in different areas taking place in the literature are summarized.

The evaluation results of all the proposed forecast combination methods are given in Chapter 5. Also, a comparison among these methods are presented in that chapter.

The thesis is concluded with further remarks and possible future work in Chapter 6.

CHAPTER 2

BACKGROUND

2.1 Wind Power Forecasting Domain

Variability of wind is the biggest challenge in implementing wind energy in a reliable electricity system that necessitates a perfect balance between supply and demand in real time. Wind power is produced according to available wind, i.e., the power output fluctuates with wind speed variations and it needs to be balanced in terms of supply and demand. At this point, an accurate wind power forecasting system has a significant role to deal with this challenge providing economic and technical advantages. It reduces the risk of uncertainty and provides better grid planning and integration of wind into power systems [4, 18].

Wind power forecasting is a complex task, since a wind speed time series can be a highly nonlinear random process, which changes its mean and standard deviation at any time. So, no typical patterns can be directly found from the series. However, in the literature, many studies have introduced different wind forecasting technologies and experience [6].

A forecasting system is characterized by its time horizon, which is the future time period for which the wind power generation will be predicted. In other power system forecasting problems, like load forecasting, the forecasting system is characterized according to its time horizon as very short-term, short-term, medium-term, or long-term. Wind forecasting has mainly focused on the very short-term of a few hours, the short-term ranging from hours up to 2 or 3 days and the medium-term from 3 up to 7 days. These time frames represent the

time intervals that power system operations are carried out within, including regulation, load flow, balancing, unit commitment and scheduling. Presently, due to the economic value of forecasting, most of the commercial and research forecast systems are used for time horizons ranging from 36 to 72 hours ahead [4, 5].

Wind power prediction is always site dependent. This is because power output of a wind turbine directly depends on the wind speed, which is affected by large-scale atmospheric conditions and topography of the surface landscape. Therefore, the characteristics of the local wind profile, terrain type and climate conditions should be considered while wind power predictions are being generated. If there are small-scale changes in atmosphere for a particular region, wind speed would be more stable, which means that wind power forecasting would be more accurate in that region. Also, wind forecasting is more complex in a terrain of onshore wind farm than an offshore one [6].

Relationship between wind speed and wind power is nonlinear, basically cubic. It can be expressed with the following equation:

$$P = \frac{1}{2} \rho A v^3 \quad (2.1)$$

where P (W) is wind power (wind energy per unit time), ρ (kg/m^3) is density of air, which depends on temperature and pressure of air, A (m^2) is swept area of wind turbine, and v (m/s) is wind speed. Therefore, a small error in wind speed prediction will actually cause a larger (cubic) error in wind power prediction. Moreover, considering the entire wind farm, this relation is more sophisticated, since different turbines use multiple wind directions and speed to perform optimal operation [18].

In wind power forecasting concept, wind speed is predicted firstly, and then estimated wind speed is converted into wind power prediction. Although using the manufacturers' power curve is the easiest approach to map wind speed into wind power, the optimal way is to use a power curve created by using measured wind at the terrain. Power curve is a data matrix specific to each wind farm,

which comprises all possible combinations of wind direction and wind speed corresponding to different wind power outputs. Power curve is constructed through simulation software tools by using historical data [6, 18].

Another point worth mentioning is that because of the differences in each local characteristics of wind, different results can be obtained from different wind power prediction systems, and also it is difficult to compare prediction systems based on available results [6].

2.1.1 Numerical Weather Predictions

NWPs are general purpose models that, in principle, can be used for whatever problem concerning atmospheric physics. The output of an NWP model is a detailed forecast of the state of the atmosphere at a given time. Thus, NWP forecasts are not produced for only the electricity industry, but also used by a variety of industries, sectors and government agencies [4, 19].

NWPs are rendered on supercomputers as they need high number of computations [18]. The development of powerful computers has enabled the implementation of reliable and high-resolution NWP models, which can reach approximately two kilometres resolution at maximum in their non-hydrostatic version. NWP models can be classified according to their space-temporal scale. Each NWP model tries to monitor the evolution of the atmosphere for its specific scale, although high spatial resolution cannot be combined with high temporal resolution. In general, a NWP model with high spatial resolution (small spatial scale) will have a low temporal validity for its predictions (small temporal scale), while a NWP model with with a low spatial resolution will have a much greater temporal validity. NWP models with great spatial and temporal scales (also known as global models) usually make predictions for the whole world, which is valid over one week. Short-term wind power forecasting (hours up to 2 or 3 days) needs predictions from a NWP model with high spatial resolution [7, 19].

It is well-known that the accuracy of the NWP model has a major impact on the error of the wind power predictions. An error on wind speed prediction

by the NWP model becomes cubically larger when it is converted into wind power as can be seen in Equation 2.1. Therefore, using NWP models for wind power forecasting is even more critical than for other applications. Customarily, NWPs are run more than once a day due to the difficulty of gaining information in short-time. Moreover, accurate predictions can be obtained when weather conditions are stable [13, 18].

NWP models mostly tend to exhibit systematic errors in the prediction of certain meteorological parameters such as wind speed and direction, especially near the surface and onshore because of a weak representation of the local effects on the airflow. Increasing the model resolution may provide improvement in the representation of smaller-scale flow characteristics, nevertheless, it is still uncertain that the use of higher-resolution models would improve the forecast skill considerably, or compensate excessive usage of computational resources required for these applications. In other words, predictions with high resolution would improve accuracy slightly, but would cost more. Therefore, a variety of approaches that are generally based on statistical methods can be used to reduce the influence of systematic errors in the output of NWP model [19].

2.1.2 Wind Power Forecasting Methods

A number of different approaches have been applied to forecast wind power produced by wind farms. These approaches can be divided into two main categories: One is physical modeling approach and the other is statistical modeling approach. Physical models focus on the description of the wind flow around and inside the wind farm by considering terrains, obstacles, pressures, and temperatures to estimate wind power. On the other hand, statistical models construct mathematical models by using historical data to predict near future generation of wind power. The physical models have advantages in predictions for long (larger than 6 hours ahead) time horizons, while statistical models perform well in predictions for very short (several minutes to one hour) time horizons [4, 5, 6, 20].

In addition to physical and statistical approaches, there is a third group named

hybrid model, which is the combination of physical and statistical models, and it associates the advantages of both physical and statistical approaches [4, 5]. Details of these three approaches are presented in the following sub-sections.

2.1.2.1 Physical Models

Physical models mainly use NWP forecasts, and they downscale NWP forecasts at a certain grid point to the particular site and turbine hub height, by taking into consideration geographical conditions of WPP area including surface roughness, obstacles, effects of orography and wind farm layouts. However, collecting the information of WPP area conditions is one of the main difficulties in the implementation of physical models [4, 5, 6].

Several sophisticated flow modeling methodologies, such as mesoscale meteorological model (MM5) and computational fluid dynamics (CFD) have been used for wind speed prediction for physical models. These advanced models have the potential to improve the modeling of the wind flow, especially in complex terrains. Additionally, model output statistics (MOS) are often used to reduce systematic forecasting errors for post-processing power predictions [5, 6].

There is a famous physical wind power forecast model called Prediktor [21], which is developed by Landberg in Denmark and it is one of the oldest forecast tools in the literature. It is integrated to wind power forecasting systems in Spain, Denmark, Republic of Ireland, Northern Ireland, France, Germany, USA, Scotland and Japan [4].

2.1.2.2 Statistical Models

Statistical models construct a statistical model by analyzing a vast amount of historical data, and the relation between historical power production and NWP forecast is determined to predict the future power output. Constructed model is also updated dynamically with new incoming data.

Unlike the physical models, statistical models include only one step to translate

input variables into wind generation. This one-step statistical block can apply one or more statistical linear and nonlinear models of different types. A sub-classification consisting of time series based models and soft computing based approaches, exists for statistical models .

Traditional time series based models are developed using historical values. They are easy to model, inexpensive, and capable to provide timely prediction. Commonly used time series models include autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) models, and Kalman Filters (KF). Few other time series based models include grey predictors, linear predictions, exponential smoothing, etc. [18].

Soft computing approaches learn from the relationship between the predicted wind and observed power output using a long-time frame historical data. Therefore, they are also known as learning approaches. Some examples of learning approaches called “black-box” models, which include most of the artificial intelligence based models, such as Neural Networks (NNs) and Support Vector Machines. Other types of models are the “grey-box” models, which learn from experience and for which prior knowledge can be injected [4, 5].

2.1.2.3 Hybrid Models

The aim of hybrid models is to benefit from the advantages of physical and statistical models, and obtain a globally optimal forecasting performance. These models use NWP sources, in addition to physical characteristics of the terrain, and take into consideration the historical power generation data [5, 6].

2.1.3 Performance Evaluation of Power Forecast Systems

Performance of a power forecast model is evaluated by its forecast error, which can be defined as any difference between the actual and predicted power values. Obtained value describes how much the predictions deviated from the actual power productions.

In present, a number of performance metrics have been proposed and employed to evaluate the wind power forecast accuracy. However, a single universal standard is still not recognized, and this deficiency actually complicates the performance comparison of different forecasting models. Therefore, several statistical metrics are used in order to assess the performance of forecast models such as mean bias (MB), mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) [22, 23].

$$\text{MB} = \frac{1}{T} \sum_{t=1}^T (f_t - y_t) \quad (2.2)$$

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |f_t - y_t| \quad (2.3)$$

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (f_t - y_t)^2} \quad (2.4)$$

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{f_t - y_t}{y_t} \right| \quad (2.5)$$

where f_t represents the predicted value at time t , y_t represents the actual observation value at time t , and T is the number of observations that corresponds to total time.

In statistics, bias refers to the tendency to overestimate or underestimate the value of population parameter, and it is used to measure instantaneous performance. For wind power forecasting assessment, MB gives the average bias value by which the predicted power differs from the actual power over an entire time series of n -values as shown in Equation 2.2. Therefore, this metric gives insight to whether the wind tends to be over or under-predicted [24].

A commonly used metric, MAE, is obtained by taking the average of all absolute errors in a set of predictions. More clearly, it is the mean over the verification sample of the absolute values of the difference between prediction and

corresponding observation. MAE is a linear score that means, all the individual differences are weighted equally in the average. MAE also gives a more comprehensive idea about the average magnitude of the forecast errors over an entire dataset without the effect of canceling positive and negative errors unlike mean bias metric. However, this advantage is gained with the sacrifice of error directionality, which can be significant when large amounts of wind power are integrated into the transmission grid [22, 24]. MAE can be calculated by Equation 2.3.

RMSE is a quadratic metric given in Equation 2.4 that gauges the average magnitude of the forecast errors, such as MAE, but it gives relatively high weight to large errors because of the squared errors. That means, RMSE is the most useful metric when large errors are particularly undesirable for system planners. On the other hand, this disables the error directionality specification contrary to MAE [22, 24].

MAPE is a measure that expresses accuracy as a percentage, given in Equation 2.5. The absolute difference between the predicted value and the observed value is divided by the observed value, and this calculation presents individual absolute forecast errors. Then, the final score is calculated by taking the average of sum of the all absolute forecast errors. MAPE has a drawback that, if there are any observed values being equal to zero, there will be a division by zero, which is undefined [22].

Wind power forecast accuracy is represented with a number between 0% and 100%, and errors are normalized to the installed capacity of the power plant in order to obtain relative differences, since it enables to compare performance of different type of wind turbines with variant installed capacities [25]. Therefore, MB, MAE, and RMSE metrics are used to validate forecast models by normalizing to installed capacity in scope of RITM. Normalized MB (NMB), normalized MAE (NMAE), and normalized RMSE (NRMSE) equations are respectively given by Equations 2.6, 2.7, and 2.8, where C denotes installed capacity of

WPP.

$$\text{NMB} = \frac{\frac{1}{T} \sum_{t=1}^T (f_t - y_t)}{C} * 100 \quad (2.6)$$

$$\text{NMAE} = \frac{\frac{1}{T} \sum_{t=1}^T |f_t - y_t|}{C} * 100 \quad (2.7)$$

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{T} \sum_{t=1}^T (f_t - y_t)^2}}{C} * 100 \quad (2.8)$$

2.2 Predictive Modeling

Predictive modeling allows us to interpret historical data and make predictions about the future. More specifically, predictive modeling involves a set of mathematical techniques, which aim to find a mathematical relationship between a target (response), or namely dependent variable and various predictors, or namely independent variables in order to predict future values of the dependent variable by using the identified relationship. These models can be utilized in several disciplines and industries including marketing, engineering, finance, psychology, medicine and healthcare [26, 27].

Predictive modeling solutions are commonly involved in statistics, machine learning and data mining literature, in that they work by analyzing historical and current data. In a predictive model, data is collected, a statistical model is formulated, predictions are made, and then the model is validated or revised by new incoming data.

Classification and regression are commonly used predictive modeling techniques that classification predicts a categorical response, while regression predicts a response with meaningful magnitude. Almost any regression model can be used

as a predictive model. Moreover, various soft computing approaches including fuzzy logic, neural networks, genetic algorithms, and rough sets, have been applied as predictive modeling in recent years. Therefore, a list of useful predictors can be fitted to the available data, then they are evaluated by means of simplicity and how well they fit to the data [28].

A number of predictive models, which are the basis of each proposed wind power forecast combination method, are explained in the following subsections in detail.

2.2.1 Multiple Linear Regression

Linear regression is a concept of finding the best-fitted straight line through n data points (observations). Simple linear regression model can be defined with the following equation:

$$y_i = \alpha + \beta X_i + \epsilon_i \quad (i = 1, 2, \dots, n). \quad (2.9)$$

where y_i is the response variable and X_i is the explanatory variable of i^{th} data point, α is the intercept and β is the slope of best-fitted line, and ϵ_i is the error, which accounts for the fact that in most practical situations, data points, (X_i, y_i) , are not fitted in a straight line.

In order to identify the estimates of intercept and slope, ordinary least squares solution is commonly proposed, that is, for any potential intercept α and slope β , deviations (residuals) are squared and added through n data points. The α and β are chosen as to minimize the sum of squares of these residuals. It should be indicated that least squares estimator assumes that the errors, ϵ_i , show a normal distribution (or Gaussian distribution) with zero mean and some variance σ^2 [26].

Simple linear regression analysis finds the coefficient β , for the one explanatory variable, X . On the other hand, when there are more than one explanatory variable, multiple linear regression can be used for examining the relationship of a collection of explanatory variables (or regressors/predictors) with a single

response variable. This type of problem is also known as multicollinearity. An example to dataset structure of multicollinearity, or multiple linear regression problem is given in Table 2.1.

The model for multiple linear regression for the given k independent variables and n observations can be expressed as follows:

$$y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_k X_{ik} + \epsilon_i \quad (i = 1, 2, \dots, n). \quad (2.10)$$

Least squares regression model can be used to estimate the regression coefficients, $\beta_1, \beta_2, \dots, \beta_k$ when $k < n$. It generates coefficients by minimizing the sum of the residual's squares. Multiple least squares problem is preferably solved using matrix calculus [29]. The model expressed in Equation 2.10 can be rewritten in a matrix form as follows:

$$y = X\beta + \epsilon, \quad (2.11)$$

where

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{k1} & X_{k2} & \dots & X_{kn} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Then, the least squares estimator $\hat{\beta}$, which fit the linear equations "best", is given as follows:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2.12)$$

where X^T is the transpose of X .

Ordinary least squares regression is very popular in statistics literature for a long time because of its simple, well-developed and documented characteristics. There is a wide range of available computer packages, which enables to apply least squares regression model on predefined dataset [30]. In a nutshell, least squares solution is the commonly used regression model to fit observed data and predict the response value of new observations.

Table 2.1: Multiple Linear Regression Dataset

Observations, i	Response, Y	Predictors (Regressors)			
		x_1	x_2	...	x_k
1	y_1	x_{11}	x_{12}	...	x_{1k}
2	y_2	x_{21}	x_{22}		x_{2k}
.
.
.
n	y_n	x_{n1}	x_{n2}		x_{nk}

The least squares regression is the optimum solution for estimating unknown parameters of a model, when the data is normally distributed and errors are independent. However, it is far from the optimal in many non-normal distributions and/or if outliers are stood out in data, since least squares estimator is strongly sensitive to the outliers.

Various regression approaches are proposed as alternatives to ordinary least squares method, in order to handle the mentioned drawback of least squares. On the other hand, it is claimed in [30] that these alternative approaches may also be required even if the errors follow a normal distribution. Using L_p -norm estimators is one type of such alternatives, which offers a robust regression.

2.2.1.1 L_p -Norm Estimators

L_p -norm estimators are used for estimating coefficients in multiple linear regression models and they minimize the sum of the p^{th} power of the absolute deviations of the observations from the regression function. Mathematical representation of L_p -norm is given in the following equation:

$$\|X\|_p = \left(\sum_{i=1}^n |X_i|^p \right)^{\frac{1}{p}} \quad (p \in \mathbb{R}) \quad (2.13)$$

In practice, L_p -norm is the generalization of the ordinary least squares (L_2 norm, the case of $p=2$), which is explained in detail in Section 2.2.1. However, for L_p

norm estimators, $\hat{\beta}$ cannot be expressed as an explicit function of X and y as given in Equation 2.12, when p is not equal to 2.

In addition to L_2 norm corresponding to least squares, a quite common norm, L_1 norm is worth to mention, since it represents a robust regression estimation. L_1 norm has different names, such as: least absolute deviations (LAD), least absolute errors (LAE), least absolute value (LAV), and least absolute residuals (LAR). LAD regression minimizes the sum of the absolute values of the residuals. Whereas L_2 regression estimate the mean of the response variable, y_i given the explanatory variable, X_i , and L_1 regression estimates the median of y_i given the value of X_i [31]. L_1 regression, namely LAD problem can be solved using any linear programming method in order to find the optimal values of regression parameters.

Selecting p value different from 2 is more suitable for data with non-normally distributed errors [32]. In general, values of p between 1 and 2 are suggested in L_p -norm estimation studies such as [31], due to its robustness and efficiency. In addition, a number of studies have been made to investigate robustness and efficiency characteristics of L_p -norm estimators for $p < 1$.

2.2.2 Tree-Based Models

Tree-based models are commonly used for classification and regression problems in which a number of predictor variables with a corresponding response variable exist. Regression is used to estimate a continuous response variable (typically real number), while classification deals with categorical response variables from newly coming predictor variables [33].

A tree-based classification or regression problem consists of four main components. The first one is the response variable as a categorical output or a dependent variable which will be predicted based on the second component of the problem, which are predictors, or independent variables. In general, there are many possible predictor variables. The third component is the training dataset, which includes both dependent and independent variables of n observations.

Finally, the fourth component of the problem is the test dataset required to validate the constructed prediction model [34].

In the following subsections, two tree-based methods, which have been directly used in the proposed combination approaches of this study, are presented in detail.

2.2.2.1 Classification and Regression Trees

Classification and Regression Trees (CART) is a tree-based method, and it is generally used in statistics, data mining, and machine learning. This method presents a tree structure, which partitions the data into homogeneous and separated groups (nodes) regarding the response variables. Tree starts with a root node containing all the observations, and then it is divided into nodes by binary recursive partitioning. The term "binary" states that each node can be split into two children nodes, in which case the original node is called a parent node, while "recursive" implies that partitioning is applied over and over again [34, 35].

Classification trees are used in applications dealing with categorical data that the response variable can take a finite set of values. On the other hand, regression trees are used when the response variable takes continuous values. Classification trees differ from regression trees with respect to the data splitting and aggregating techniques [36]. Regression trees are focused henceforward, since proposed tree-based combination method is concerned only with the numerical variables in the scope of this thesis.

Regression tree constructs a set of decision rules by recursively partitioning data based on a single explanatory variable. Each split point is the one that maximizes the homogeneity of the two children nodes with respect to the response variable. This suggestion is provided by impurity reduction defined as the total sum of squares of the response values around the mean of each node. The procedure initially grows maximal tree which is overfitted, and it can be pruned in order to reduce the tree size without impairing prediction accuracy [35, 36].

Cross-validation is a commonly used model validation method, in which the

original training data is randomly partitioned in to subsets to be used as test data. Each data subset is reserved for testing, while its complemented data is used as training data in order to construct the tree model. Then, the model is tested by using reserved data subset, and the error is recorded. Finally, the average error over all subsets refers to the cross-validated error.

2.2.2.2 Random Forests

In recent years, several methods have been proposed to generate an ensemble of classifiers including bagging and boosting, which are the most widely used ones. Bagging (or bootstrap aggregating) is based on training many classifiers on bootstrapped samples from a training dataset, which has been shown to reduce the variance of classification. On the other hand, boosting is based on sample re-weighting in which the incorrectly classified samples are given increased weighting iteratively [37].

Boosting is considerably more accurate than bagging in most cases, since it generally reduces both the variance and the bias of the classification. However, it has several drawbacks. First of all, it becomes computationally demanding and consumes more time than bagging. Moreover, it is sensitive to noise and it can also cause overtraining.

More recently, random forests have been proposed to ensemble tree-based classifiers [38]. It is a supervised learning method like CART, however it is better to handle lots of observations and lots of predictor variables.

Random forest approach uses a similar but improved method of bagging to form an ensemble of CART classifiers. In Breiman's study [38], it is claimed that random forests are comparable to boosting in terms of accuracy, but without the drawbacks of boosting. In addition, random forests are computationally less demanding than boosting.

A random forest consists of a large number, of *n-trees*, for example 1000. They are random because each tree is built with a random subset of observations, also each split within each tree is created based on random subset of candidate

variables. This randomness creates different individual trees' predictions and the final prediction is the average of them [39].

Random forests approach can handle thousands of explanatory variables M , since it specifies a random number $m \ll M$ to grow each tree in the forest. The specified value of m is held constant during the forest growth. In addition, each tree is grown to the largest extent as much as possible without the need of pruning, because random forests do not overfit [38].

2.2.2.3 Variable Importance

In some situations, tree-based approaches can be used to find out the so-called "variable importance", which is a ranking process to represent each predictors' contribution to the response variable. Variable importance gives an idea about which predictor/predictors are associated with the response, how they relate, and perhaps even which ones are interact with the others in terms of predicting the response [33].

CART and random forests are generally used to map observations without class label to their predicted class or target value. However, they can also be used to select from a large number of explanatory variables, which are most important to determine the response variable to be explained. This can be done by ranking variables using some measure of importance [35].

In classification trees, a well-known metric for variable importance, the Gini index proposed by Corrado Gini in [40] is used to rank explanatory variables. On the other hand, importance of an explanatory variable is simply the total reduction in sum of squares achieved by all splits on that variable in regression trees [36, 39].

It is also worth mentioning that an explanatory variable in CART can be considered highly important even if that variable never appears as a node splitter in the tree structure. Because, CART keeps track of surrogate variables, which take the place of the primary splitter, in case the primary variable is missing in an observation. If an explanatory variable appears as neither a primary nor a

surrogate splitter, then the importance score is assigned to zero. Since such a variable does not contribute to the tree structure to predict the response variable, it can be eliminating from the dataset.

2.2.3 Fuzzy Soft Sets

Most of the conventional mathematical tools for modeling, reasoning, and computing are deterministic and precise, so-called "crisp". However, lots of complicated problems in economics, engineering, social science, medical science, environment, etc., are inherently uncertain, imprecise and vague. In order to deal with this type of problems, a number of theories have been proposed including theory of probability, theory of fuzzy sets, and theory of interval mathematics. However, all these theories have their own difficulties due to the inadequacy of the parameterization as pointed out in [41], where the concept of soft set theory is introduced as a new approach to handle uncertainties, since the theory is free from the mentioned difficulties [16, 42].

A soft set can be defined as the collection of approximate descriptions of an object. Each approximate description is a composition of a predicate and an approximate value set [43].

Definition 2.1. A pair (F, E) is called a soft set (over a initial universal U) if and only if F is a mapping of E into the set of all subsets of U [16, 42].

In the view of above discussions, an example is given below.

Example 2.1. Suppose that $U = \{h_1, h_2, h_3, h_4\}$ is a set of house, and $E = \{e_1, e_2, e_3, e_4, e_5\}$ is a set of parameters such that each parameter corresponds to a fuzzy expression which represents the status of houses and stands for "expensive", "beautiful", "wooden", "cheap", "in the green surroundings", and "modern", respectively. Consider F be a mapping of E into the set of all subsets of set U . Now suppose a soft set (F, E) , which describes the "attractiveness of houses for purchase". According to the data collected, the soft set (F, E) is given by

$$(F, E) = \{(e_1, \{h_1, h_2, h_4\}), (e_2, \{h_1, h_2\}), (e_3, \{h_1, h_3\}), (e_4, \{h_2\}), (e_5, \{h_3, h_4\})\}$$

Table 2.2: Tabular representation of the soft set (F, E)

U	e_1	e_2	e_3	e_4	e_5
h_1	1	1	1	0	0
h_2	1	1	0	1	0
h_3	0	0	1	0	1
h_4	1	0	0	0	1

where

- $F(e_1) = \{h_1, h_2, h_4\}$,
- $F(e_2) = \{h_1, h_2\}$,
- $F(e_3) = \{h_1, h_3\}$,
- $F(e_4) = \{h_2\}$,
- $F(e_5) = \{h_3, h_4\}$.

A two-dimensional table is used in order to store the represented soft set (F, E) as shown in Table 2.2.

Fuzzy sets have been specifically proposed in [44], in order to mathematically represent uncertainty and impreciseness which are frequently encountered in real world applications that the classes of objects do not have precisely defined criteria of membership. For example, "the class of beautiful women", or "the class of tall men" do not constitute classes or sets in traditional mathematical tools. However, such imprecisely defined classes play a significant role in the pattern recognition, communication of information, and abstraction domains. Therefore, the concept of fuzzy set is proposed to handle this circumstance, in which a fuzzy set can be defined as a class with a continuum of grades of membership [44].

Definition 2.2. A fuzzy set A in initial universe U , is defined by a membership function $\mu : U \rightarrow [0, 1]$ whose membership value $\mu_A(x)$ specifies the degree to which $x \in U$ belongs to fuzzy set A , for $x \in U$ [16, 42].

Table 2.3: Tabular representation of the fuzzy soft set (f, A)

U	Blackish	Reddish	Green
h_1	0.4	1.0	0.5
h_2	0.6	0.5	0.6
h_3	0.5	0.3	0.8
h_4	0.8	1.0	0.8
h_5	1.0	0.7	0.4

The concept of fuzzy soft sets theory, which is a generalized combination of soft set and fuzzy set, has been firstly introduced in [45].

Definition 2.3. Let $A \subset E$ and $\mathcal{F}(U)$ be the set of all fuzzy sets in U . Then, a pair (f, A) is called a fuzzy soft set over U , where $f : A \rightarrow \mathcal{F}(U)$ is a function [16, 42].

From the definition, it is obvious that $f(a)$ is a fuzzy set in U , for each $a \in A$, and the membership function if $f(a)$ is presented by $f_a : U \rightarrow [0, 1]$. This idea can be illustrated with the following example.

Example 2.2. Let universe $U = \{h_1, h_2, h_3, h_4, h_5\}$ be a set of houses, and $A = \{blackish, reddish, green\}$ be a set of parameters. Then, fuzzy soft set (f, A) over U is given by following, and it can also be expressed in a tabular form as in Table 2.3.

- $f(blackish) = \{(blackish), \{h_1/.4, h_2/.6, h_3/.5, h_4/.8, h_5/1\}\},$
- $f(reddish) = \{(reddish), \{h_1/1, h_2/.5, h_3/.3, h_4/1, h_5/.7\}\},$
- $f(green) = \{(green), \{h_1/.5, h_2/.6, h_3/.8, h_4/.8, h_5/.4\}\}.$

In this thesis, mentioned predictive modeling techniques have been used to combine RİTM forecasts. L_p -norm estimators have been utilized to find combination weights of individual models derived from regressors' coefficients. On the other hand, regression trees and random forests have been used to rank individual forecast models by the help of variable importance, so that the weights could be assigned proportional to variable importance ranking. Finally, FSS has been

used to designate weights of individual forecasts by means of fuzzy membership function.

CHAPTER 3

ARCHITECTURE OF RİTM SYSTEM

In this chapter, the general architecture of RİTM is described with its hardware infrastructure, internal and external data sources, wind power forecast models and user software applications. Note that the explanations on forecast models of RİTM need more attention in this chapter, because their predictions are directly participated in forecast combination methods proposed in the scope of this thesis.

3.1 An Overview

RİTM is a project on wind power monitoring and forecast system for Turkey [46], which is designed and implemented by TÜBİTAK in cooperation with YEGM. The main aim of the RİTM project is to provide large-scale integration of wind power energy to Turkey Electricity System properly.

RİTM project has been started in 2011 with 14 pilot WPPs, and currently, it is extended to cover approximately 70 WPPs. Turkish Electricity Transmission Company (TEİAŞ) has become one of the stakeholders of the RİTM project in 2013, which is a government agency that manages the flow of the energy sources in Turkey.

RİTM is a valuable wind power monitoring and forecast system to provide accurate forecasts, since TSOs have to plan regular power ahead of time, and the more accurate forecasts, the less regulation power is needed. Moreover, the forecast outputs of RİTM are currently supplied to the WPP owners, who are

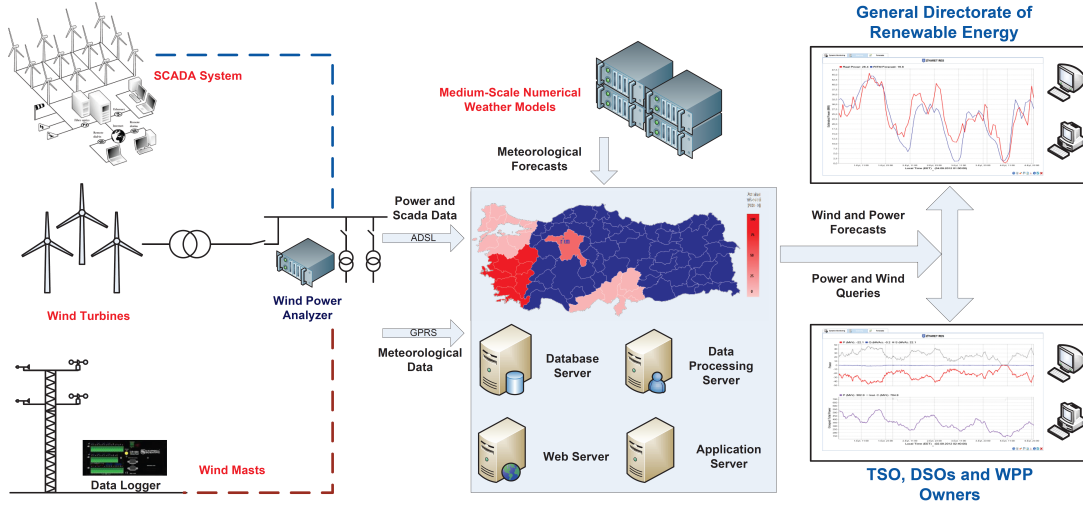


Figure 3.1: The General Architecture of RİTM

responsible to declare their estimated production to the energy market for the following 48-hours period. They also need reliable forecasts not only to not pay penalty, but also get huge profit.

The system is equipped with related hardware and software, acquires power/power quality and meteorological measurements in addition to several NWP, and processes the acquired data through physical and statistical methods to produce high-quality short-term wind power forecasts for the included WPPs. General architecture of RİTM is presented in Figure 3.1. Although the design of the system is similar to related works on wind power monitoring and forecast systems in [47] and [48], there are several outstanding features including the ease of extensibility, the range of real-time data sources, and the range of forecasts produced [2].

3.2 Wind Power Monitoring and Forecast Center

Wind power monitoring and forecast center is established at YEGM, and the all data acquisition, storage, processing and publishing operations are done in this center physically. Also, a DLP video wall system is integrated to this center in order to monitor power production of involved WPPs in real time with



Figure 3.2: A panoramic view from the Center

corresponding forecast outputs. A panoramic view from the center is given in Figure 3.2.

A server room has been placed in the center to satisfy the data storage and processing requirements of the overall system. It includes two database servers, a data processing server, an application server, and a Web server.

Database servers have a huge storage capacity and one of them is duplicated to serve as a backup. PostgreSQL open-source database management system [49] is run on these servers to store acquired data. Also, there is a server designated as the application server, which basically hosts a set of applications in order to acquire data from various sites and communicate with the user software applications. Another server is utilized as the data processing server on which the forecast applications are periodically run. The application server and the data processing server have ability to handle the operations for 200 WPPs at the same time. Finally, a web server with comparatively less processing and storage capacity, hosts an e-mail server and the RITM project Web page [2].

3.3 Data Sources

The main data sources used in the RITM system consist of wind power analyzers, wind masts and medium-scale NWP. Wind power analyzers and wind masts are genuine components, which basically provide data in binary files subsequently parsed and stored in the system database while medium-scale NWP are directly acquired from several weather forecasts centers. These data sources are presented with their detailed features in the following subsections, but also there is a third-party data source, SCADA systems of WPPs, which are used to obtain wind speed and wind direction measurements with the status information on the basis of particular wind turbine or the whole wind farm.

3.3.1 Wind Power Analyzers

A wind power analyzer (WPA) is a genuine device, which is based on the multi-purpose PQ⁺ analyzer presented in [50]. Thanks to its ability to collect 3-phase voltage and current data from two feeders at the same time, the analyzer simultaneously collects raw data from 12 channels and performs the calculations for the power quality (PQ) parameters and electrical quantities including frequency, magnitudes of the supply voltage and current, flicker, unbalance, voltage/current harmonics/interharmonics, PQ events consisting of voltage sags, swells, and interruptions, and power parameters (active/reactive/apparent power and power factor). A view of an WPA is given in Figure 3.3 both as unconnected and in panel.

WPAs installed in involved WPPs, calculate the PQ parameters and electrical quantities, and then send to the RITM center via an ADSL connection in every 3 seconds in order to be stored database servers. The analyzers are also capable of storing these measurements on itself to prevent data loss in case of a network problem. When the communication problem is solved, stored measurements in tar files are sent to the center.

Another point worth mentioning is that among the all analyzed data, the most significant parameter is the active power with regard to wind power forecasting



Figure 3.3: A View from Wind Power Analyzer

that corresponds to the wind power generation of WPPs.

3.3.2 Wind Masts

Wind masts are the systems with hardware and software modules, and they are installed to WPPs in order to measure instant meteorological parameters. In the scope of RİTM, the sensors on the wind masts mainly carry out measurements for wind speed/direction, humidity, temperature, and pressure in 1-second resolution. The measured parameters are stored by the data logger, and then they are sent to the RİTM center over a GPRS modem with a bandwidth of up to 256 Kbps. Views of a typical wind mast used in the RİTM system is given in Figure 3.4.

3.3.3 Medium Scale NWP

Medium-scale NWP are the basic data sources in order to generate 48-hours wind forecasts within the scope of the RİTM project. There are three main NWP sources, which are currently used in the all forecast models of RİTM. Two of them are GFS [51] and ECMWF [52] global weather forecasts used as the inputs to WRF [53] meteorological model. These weather forecasts are



Figure 3.4: A View from Wind Mast

periodically downloaded to the RİTM center, and the WRF model is run on data processing server to generate the WRF forecasts. The remaining data source for NWP is Turkish State Meteorological Service (DMİ), which presents forecasts of ALADIN [54] meteorological model directly.

The parameters of NWP consist of u , v components of wind speed in different altitudes, in addition to temperature (K) and pressure (Pa). There are six different predictions for wind speed components corresponding to different altitudes above ground level that the difference between each level is 10 km. Among them, predictions at the highest level that is closest to the turbine level are used in forecast models.

All the obtained data involving three different NWP of all grid points in Turkey, are stored in the central database. However, the distance between two grid points is 6 km for GFS and ECMWF while it is 4 km for DMİ, which represents the resolution of NWP. Therefore, all these three data sources are utilized

separately in all forecast models of RİTM.

3.4 Forecast Models

Within the scope of RİTM project, different wind power forecasting models are studied in order to find optimal solution including physical model, ANN model, SVM model, and the cluster based forecasts model.

3.4.1 Physical Model

Physical model of RİTM is developed on the basis of downscaling NWP's to the turbine's hub height. Firstly, the best representative four grid points of the WPP area are determined as a result of CFD studies. Then, 48-hours NWP's of these grid points are translated to the reference point of the WPP by using WindSIM [55] and WAsP [56] software tools. These tools serve to model WPP area with turbines in 3D space, and take into account topography of WPP area, terrain characteristics, turbine locations, terrain roughness, and obstacles while transforming NWP's. Model maps are generated for a $20 \times 20 \text{ km}^2$ area bordered by moving 10km away from WPP's center point. Example outputs of the model is given in Figure 3.5. Finally, transformed NWP's are mapped to wind powers using power curves, thus, 48-hours power predictions can be obtained.

This model is also run for once in order to generate power curve of the WPP, which pretends a look-up table. Power curve is a 360×251 matrix that is specific to each WPP and includes corresponding estimated wind power values for each possible wind speed and direction combination.

3.4.2 ANN Model

ANNs are simple, but powerful and flexible processing units, which are interconnected and designed in a way to model how the human brain performs a particular task. They are statistical based tools for forecasting, which are relatively easy to implement and reveal good performance being less time consuming than

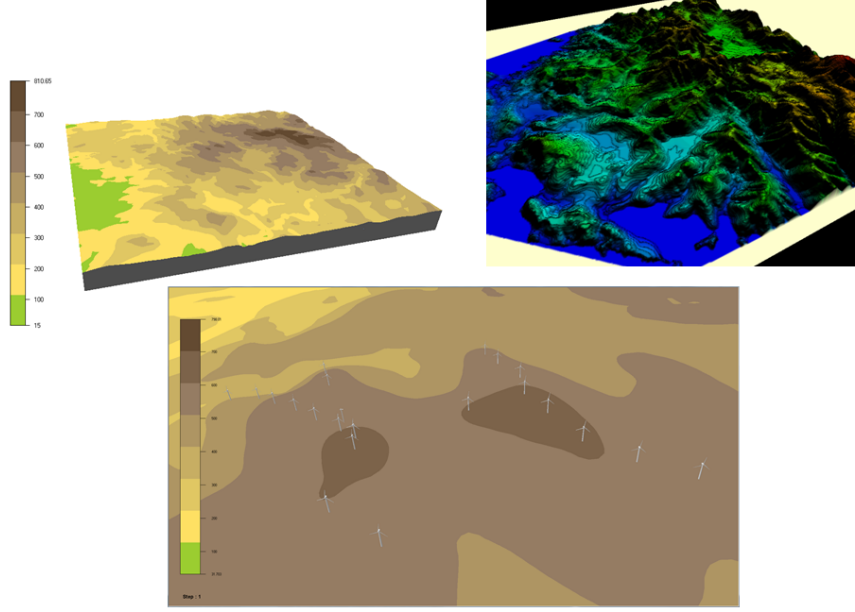


Figure 3.5: An example to 3D modeling for physical forecasts of a particular WPP

traditional time series techniques, however they need enough data for training [57, 58].

ANN model of RİTM consists of ensemble NWP inputs and a multistage structure to reach a final wind power forecast. In this model, a multilayer feed forward ANN with three layers including one hidden layer is found to be sufficient to handle the input/output relationship. For each NWP input, ANN model applies the WPP power data to build a weight/bias structure. According to the available WPP power data the training stage size is extended up to 2 years. At the testing stage 48-hours or 72-hours NWP data is used as input to the formed ANN structure to get a final wind power forecast for each NWP model. Finally, three ANN wind power forecasts are obtained at the output of ANN model. Performance of the model depends critically on parameter selection of ANN models. In this study, lots of combination of layer numbers and neuron numbers are tried, and best results are obtained by three layers ANN structure with a neuron number varying between 15 and 25 with respect to WPP. Because of wind being either zero or a positive value, the activation function is selected to be logarithmic.

3.4.3 SVM Model

SVM has been successfully applied to the problems of pattern classification in literature. The idea behind the SVM classification is to translate data into a higher-dimensional space using non-linear mapping in order to classify data more explicitly that class of patterns as far as possible from each other. Various kernel functions are used to split data categories via a nonlinear line (support vectors) in hyperplane [59, 60].

SVM model of RITM is operated similar to ANN model as a statistical regression method. Three different NWP dataset are preprocessed to refine training data. Then, for each NWP model data and WPP power data, an SVM structure is established and support vector parameters are saved at training stage. Finally, at testing stage, again similar to ANN model, three different 48-hours or 72-hours NWP dataset are processed by SVM model and three final SVM forecasts are obtained. The most important parameter of this structure is the selection of the best kernel function. In order to find the best kernel function, linear, polynomial, sigmoid and radial basis function kernels are tried out. Although other kernel functions are also applied to the model, because of nonlinear relation between wind speed and power, the kernel function that gives best performance is found as radial basis function.

3.4.4 Cluster Based Model

This model is a statistical hybrid wind power forecasting approach which is composed of two phases as training and test that are applied to three NWP sources separately.

At the start of the training phase, best NWP data are determined among NWPs of NxN grid points surrounding plant's center using historical data. During this process, linear regression is respectively applied on historical NWP data of NxN grid points by considering historical power production at the same time period, and the best linear correlation coefficients with minimum NMAE are stored for each grid point. Then, the best grid point with minimum NMAE among NxN

grid points is determined.

Subsequently, a well-known technique, Principal Component Analysis (PCA) [61] is applied on the best grid's NWP's. This data transformation process is applied to reduce 3-dimensional training dataset to one dimension. Attributes of the 3-dimensional dataset are u (u component of wind speed), v (v component of wind speed) and p (pressure), which are selected among all NWP parameters that are assumed to be more effective on the wind speed. Among them, u and v components are directly used for calculation of wind speed and wind direction.

After the PCA implementation, a commonly used data mining approach, K-means clustering is applied on transformed dataset for $k=3$. The cluster number 3 is determined with regard to the results of dynamic clustering approach proposed in [62] that applies clustering dynamically for each cluster number between 2 to 15. This time, linear regression is applied on each cluster's data to find linear correlation coefficients specific to each cluster. As the outputs of training process, coefficients, cluster centroid points and calculated NMAEs for each cluster are saved to use in test phase. The flowchart of the detailed training algorithm is given in Figure 3.6.

Test process is run to generate 48-hours power predictions. Start time of the test period is always next 48-hours to end time of training period.

At the test phase, 48-hours NWP's of the best grid point determined during the training process are used as test dataset. First of all, PCA is applied on test data as in training phase, and then, cluster IDs are assigned by calculating Euclidian distance between the centroids of clusters to each data. Then, estimated power is calculated by using the correlation coefficient of assigned cluster.

3.5 User Software Applications

A number of user software applications have been developed and implemented within the scope of the RITM project in order to present the measured data and generated forecast data. These applications include a real-time monitoring

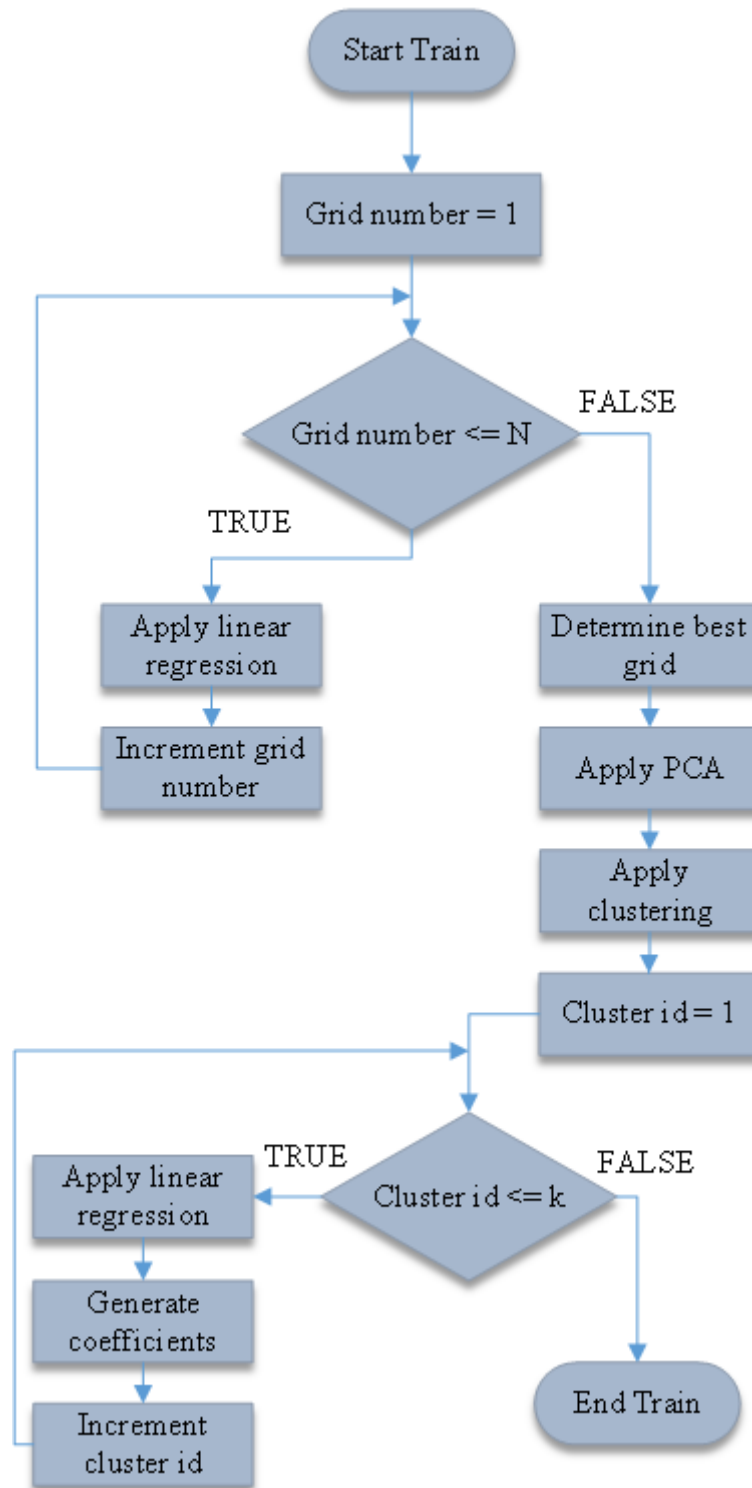


Figure 3.6: Flow Diagram of the Training Algorithm of Cluster Based Forecasts Model

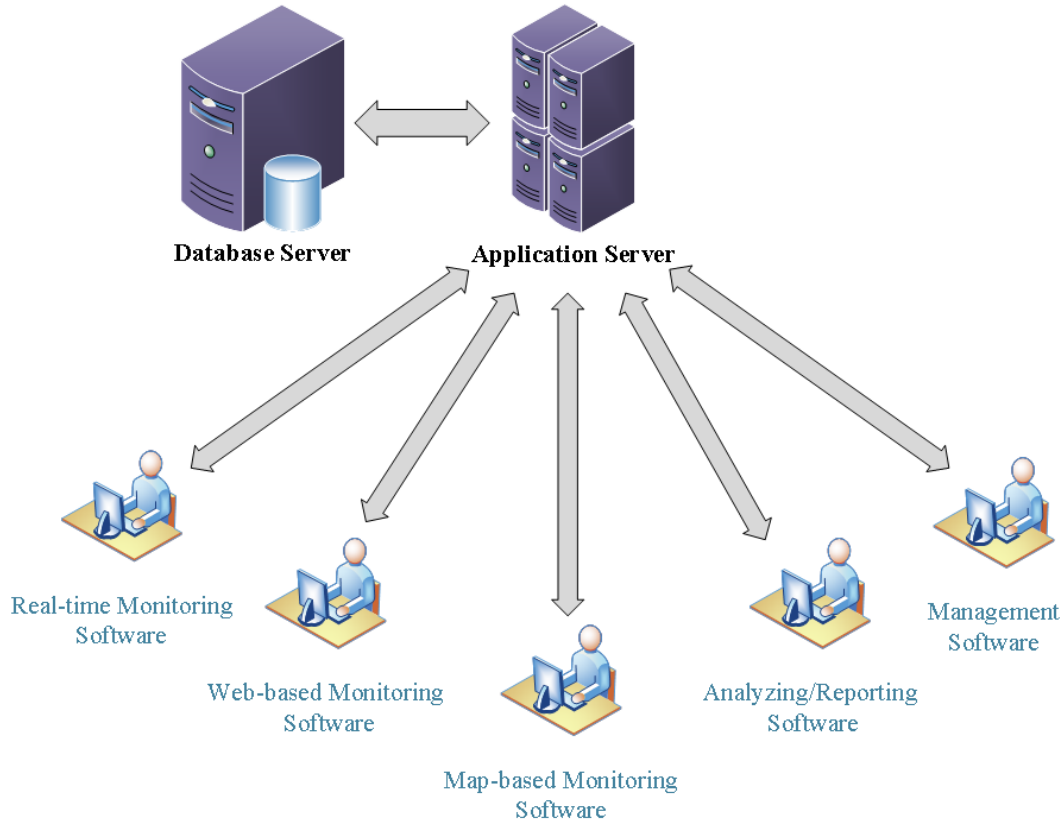


Figure 3.7: The General Structure of User Subsystem of RITM

software, a Web-based monitoring software, a map-based monitoring software, and an analyzing/reporting software. Additionally, a management software is developed for authorized users, which enables system operations. All the user applications communicate with application server located in RITM center that is illustrated in Figure 3.7.

Real-time monitoring software is a Java application which consists of dynamic monitoring, querying and forecast interfaces, therefore, it is also called as monitoring, querying and forecast software. This application enables to monitor power and forecast data of selected WPP dynamically or query wind power production/forecasts, in addition to PQ parameters retrospectively to the authorized users. Requested data is displayed on convenient graphs with an option for reporting plotted data. A snapshot of the forecast interface is shown in Figure 3.8 where production and forecast data are displayed as 1-hour averages

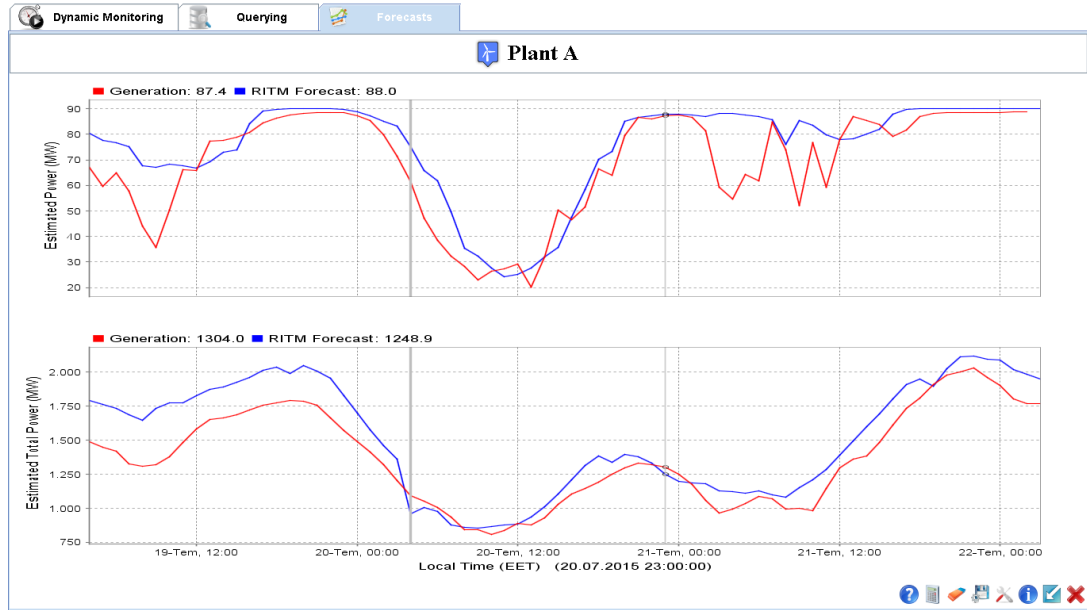


Figure 3.8: Real-time Monitoring Software

both for selected WPP and all WPPs in total.

Web-based software can be ubiquitously accessed through RITM Web page [46] with a valid authentication, and it provides similar but limited query options to such in real-time monitoring software in order to visualize data on graphics.

Map-based monitoring software is also implemented in Java and it was initially designed for the DLP video wall panel. It presents a general 3D view of involved WPPs thanks to the integration of Google Earth application. Instant productions of WPPs are also displayed via bar charts respectively. In present, this user software is adjusted to be used in personal computers.

Analyzing/reporting software of RITM facilitates the visualization of the power/PQ measurements including voltage/current, frequency, active/reactive/apparent power, power factor, unbalance, and harmonics/interharmonics carried out by the WPAs, and PQ events including sags, swells, and interruptions. This software also provides detailed reports for the mentioned parameters.

CHAPTER 4

PROPOSED COMBINATION METHODS

In RİTM, there are 12 different power predictions corresponding to the same power generation due to 4 forecast models and 3 NWP sources. Each forecast model produces 48-hour short term power forecasts. Choosing the best performed model and NWPs is not easy, since all of them can be outstanding for different time periods and different WPPs. In order to obtain an optimum power estimation from N -individual forecast models, a combination method is required as a part of RİTM. The general structure of combination process is presented in Figure 4.1.

In fact, forecast combination is based on weight determination, and the determined weights can be fixed or dynamic as mentioned in Section 1.3.1. Therefore, different combination approaches are developed within the scope of this thesis, which are fundamentally based on FW determination. However, the weights are reassigned at the end of each daily training, in other words, at the start of each daily test, which actually represents a dynamic weighting. Thus, it can be referred as time-varying weighting. The illustration of the time periods for reassigning weights can be seen in Figure 4.2.

The proposed combination approaches can be grouped in three classes. One of them is L_p -norm based forecast combination, which uses different p values for multiple linear regression estimation. The second one comprises of tree-based approaches to construct forecast combination models. Regression trees and random forests are in the scope of this group. The last one uses a soft computing method, namely FSS for combination of forecasts. All these combination meth-

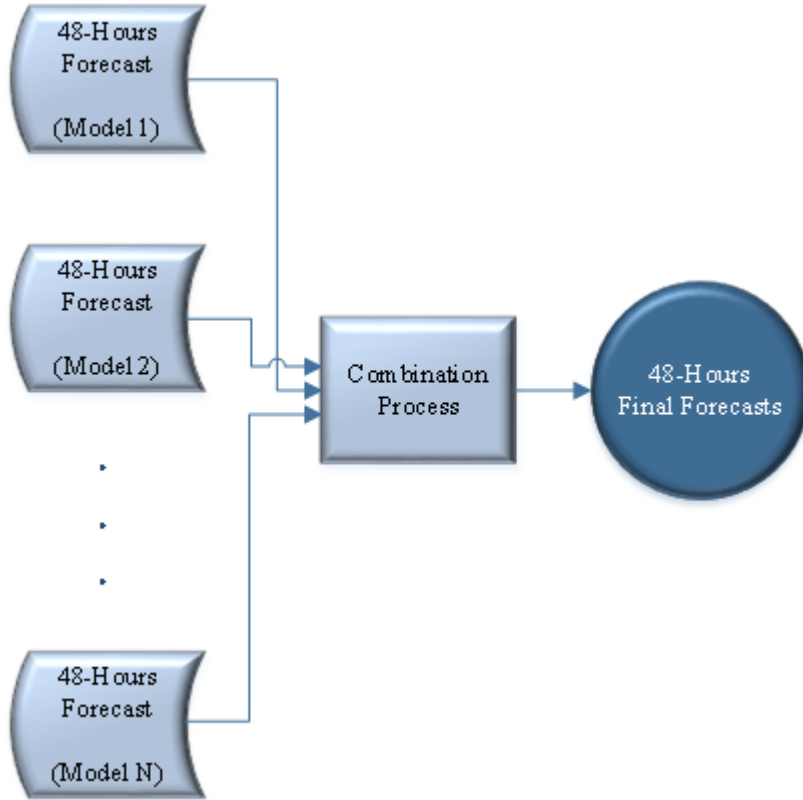


Figure 4.1: The General Structure of Combination Process

ods are presented in detail in the following subsections.

4.1 L_p -Norm Based Forecast Combination

In order to find the best combination of alternative forecasts, linear regression methods have been commonly proposed in literature to find the optimal combination weights. These methods are mainly based on a traditional method introduced by Bates and Granger [9]. In this pioneer study, the weights of the individual forecasts are set to be inversely proportional to the forecast error variances, so that the residual variance of the linear combination is minimized. After that, Granger and Ramanathan proposed a method in [63], which selects the optimal weights by minimizing the sum of squared forecast errors of individual forecasts.

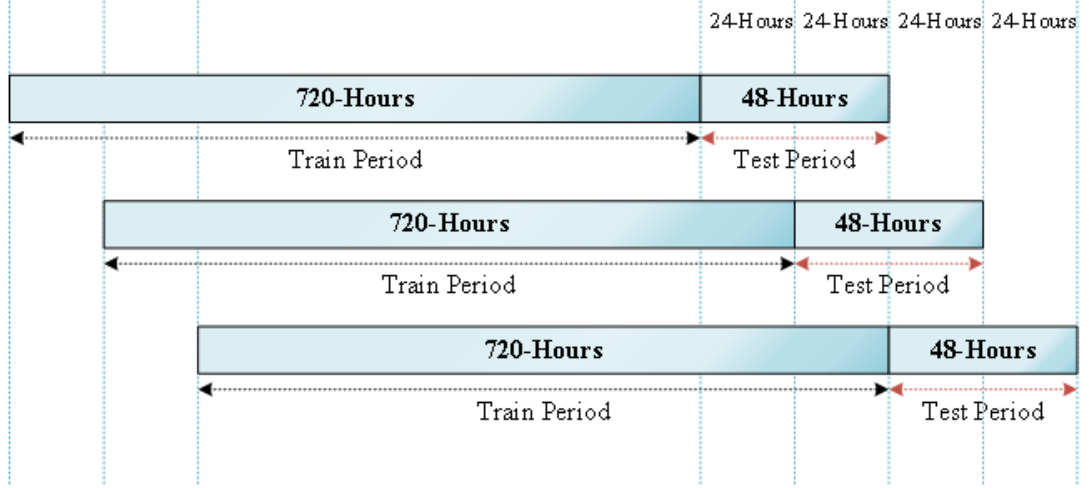


Figure 4.2: Timeframe of Each Daily Run For Combination

When considering the combination of wind power forecasts, linear regression of actual power and power predictions of individual forecast models can be applied to find the linear relationship among them. Therefore, a combination study with application to wind energy forecasts has been proposed by Sanchez [11]. His study is based on least squares regression estimation, which is the most typically used regression estimator.

In the light of the related works, a generalized linear regression model is proposed to combine alternative forecasts of RITM, and it is called as L_p -norm based forecast combination. In this approach, different norms are used in order to construct a robust regression model.

The objective of L_p -norm estimators is to minimize L_p -norm function, which is given in Equation 2.13. In terms of wind power forecasts combination, X_i in the equation represents the predictions of individual forecasts. L_p -norm estimators aim to minimize the sum of the prediction errors by assigning optimal coefficients to these individual forecasts.

As a result of L_p -norm based multiple regression estimation, N different coefficients can be obtained, each of which corresponds to an individual forecast. These coefficients refer to the weights of forecast combination. Some of the resulting coefficients can be lower than zero, which indicates a negative correlation

between these forecast models and the actual power generation. In this case, weights of those particular forecasts are assigned as zero. Then, the sum of the remaining coefficients are proportioned to 1, and assigned as weights of the corresponding forecast models.

The final combined forecast of time t , F_{ti} is calculated by using the following equation:

$$F_{ti} = \sum_{i=1}^n (M_{ti} * w_i) \quad (i = 1, 2 \dots n) \quad (4.1)$$

where n is the individual forecasts number, M_{ti} is the power prediction of i^{th} model at time t , and w_i is the weight of the i^{th} individual forecast model.

It is mentioned in Section 2.2.1 that if the errors are normally distributed, least squares estimators (L_2 norm), being maximum likelihood estimators, give satisfying results. Nevertheless, other norms different from L_2 may also be required even if the errors follow a normal distribution. Although the estimated wind power data of forecast models of RITM show basically a normal distribution, which means no prominent outliers can be observed, impacts of different p values on multiple linear regression based combination are considered in the scope of this thesis. Therefore, 0.5, 1, and 1.5 values are assigned to p apart from $p = 2$, and L_p -norm estimators model is rerun.

4.2 Tree-Based Forecast Combination

Tree-based forecast combination approach is implemented by taking inspiration from the fact that the primary variables of the constructed tree can be considered to have the best predictions. In this way, these variables can be used as the major components of the combined forecast.

As mentioned in Section 2.2.2.3, variable importance is an output of both CART and random forests applications, which gives information about the contribution of each predictor to the response variable. Therefore, resulting variable impor-

tance lists of both regression tree and random forests can be utilized with the intent of combining N independent wind power forecasts. Variable importance is usually available in software packages of tree-based approaches.

In order to build a regression tree, training dataset is prepared from explanatory and response variables. In the case of combination of wind power forecasts, N individual forecasts are identified as the explanatory variables. On the other hand, actual power observations are assigned as the response variables.

When regression tree is grown to its largest possible size, it is recommended to prune the tree in order to avoid overfitting. Typically, a tree size is selected for pruning, which minimizes the cross-validated error. Software tools for CART usually include cross-validated errors table, which becomes available when the tree is grown. This table involves the complexity parameters with corresponding cross-validated errors, and the complexity parameter associated with minimum error is given as an input to the pruning function.

After the tree is pruned, variable importance can be extracted, which gives an order of importance of predictor variables. In order to utilize this information in forecast combination, the importance score of each variable in the list is proportioned to the total variable importance score. Then, the obtained weights are assigned to each corresponding individual forecast model.

It should not be forgotten that, variable importance list is not supposed to include all the explanatory variables. If one or more missing variables are noticed in the list, that means the importance scores of these variables have been assigned to zero. Therefore, the weights of corresponding individual models are also assigned to zero.

Similar to the regression tree-based combination, random forest is built by using the same training dataset. However, in contrast to regression trees, random forests do not need to be pruned as aforementioned in Section 2.2.2.2. Therefore, right after the forest is grown, variable importance output can be used to determine forecast weights.

Random forest construction requires an input n , which represents the tree num-

ber in the forest. This number is typically chosen as a few hundred to several thousand. However, it is clarified in [64] that setting the number of trees in a trial and error basis is considered as more favorable. Because, it fundamentally depends on the size and the nature of the training dataset. In general, if the number of observations or the number of explanatory variables are large, then the tree number is set to higher values in order to prevent underfitting. Otherwise, some observations may not be predicted, or some variables may be missed in all subsets. Both cases cause a decrease in predictive power of random forest. On the other hand, increasing tree number too much, may result in only spending more computational power for almost no performance gain [64].

In this study, neither the number of observations nor the number of predictors is high. However, tree number n is assigned as 50, 100, 500, and 1000 in order to investigate the effects of increasing tree number on the tree-based combination of wind power forecasts.

After the weights of the individual forecasts are determined, the final combined forecasts of regression tree-based combination and random forest based combinations are calculated as in L_p -norm based combination methods, using Equation 4.1.

4.3 FSS Based Forecast Combination

Fuzzy Soft Set Theory is a mathematical approach that provides an approximate framework to represent vague concepts by allowing partial membership as mentioned in Section 2.2.3 in detail.

In terms of the forecasts combination, Xiao et al. [16] claimed that the individual forecasts employed in combined forecasts are a description of the actual time series, which is fuzzy. Besides, determination of the optimal weights is a decision-making problem to find a vector, by which the combined forecasts can minimize forecast errors. Therefore, they proposed a forecast combination approach to combine individual forecasts of export and import volumes in international trade. In this study, the relative error of each individual forecast has been used to

construct the membership function of fuzzy sets. Then, it has been reflected in a tabular form of FSS.

Inspiring from the mentioned related work, a wind power forecasting combination method is proposed for the wind energy industry within the scope of this thesis. In FSS based wind power forecasts combination approach, power productions and N -individual forecast models' predictions of each training time interval are used to construct FSS model.

$$f_{ti} = \max(1 - |M_{ti} - P_t|/P_t, 0) \quad (4.2)$$

By using the given membership fuzzy function in Equation 4.2, the fuzzy variable is calculated for each observation. The membership function of each fuzzy variable, f_{ti} , is the forecast accuracy of an individual forecast at a time-point, in which M_{ti} denotes the predicted power value of the i^{th} model, and P_t denotes the actual power at time t .

In some rare situations, fuzzy variable may be negative. Since the accuracy is interpreted as a number between 0% and 100%, a maximum operation with 0 is used in Equation 4.2 in order not to assign a negative value to the fuzzy variable. Obviously, $f_{ti} \in [0, 1]$, and it states how close is the predicted value to the actual value [16].

All the calculated fuzzy variables are placed on a $d \times n$ matrix, namely FSS. Tabular representation of the constructed FSS is given in Table 4.1.

At the end of the each training phase, FWs of the individual forecasts are assigned by using the constructed FSS. First, the sum of the fuzzy variables of each individual forecast is generated by Equation 4.3. Then, the weight of each individual forecast is calculated by Equation 4.4.

$$\gamma_j = \sum_{t=1}^d f_{tj} \quad (4.3)$$

Table 4.1: Tabular Representation of FSS

	Individual Forecast Models				
Time	M_1	M_2	M_3	...	M_n
t_1	0.67	0.29	0.81	...	0.77
t_2	0.78	0.37	0.96	...	0.73
t_3	0.81	0.44	0.96	...	0.92
t_4	0.88	0.46	0.76	...	0.98
t_5	0.75	0.50	0.98	...	0.90
t_6	0.62	0.49	0.99	...	0.90
t_7	0.72	0.52	0.99	...	0.91
t_8	0.88	0.46	0.81	...	0.88
t_9	0.94	0.48	0.83	...	0.96
t_{10}	0.96	0.74	1.00	...	0.93
t_{11}	0.93	0.63	0.90	...	0.93
t_{12}	0.91	0.62	0.91	...	0.93
...

$$w_i = \frac{\gamma_i}{\sum_{j=1}^n \gamma_j} \quad (4.4)$$

The final combined forecast, F_{ti} , is calculated in the same way as L_p -norm based combination and tree-based combination, using Equation 4.1.

CHAPTER 5

EXPERIMENTAL RESULTS

This chapter presents the evaluation results of the proposed combination methods. However priorly, some of the implementation details of proposed combination methods are explained.

14 pilot WPPs of RİTM are selected to validate combination methods, because these WPPs have been monitoring since the RİTM project started. Therefore, adequate historical data of these WPPs are available in the central database of RİTM. Mentioned WPPs' installed capacities are given in Table5.1. However, they are referred anonymously because of the agreement restriction of WPP owners.

The time period from April 19th, 2014 until August 16th, 2014 is used to obtain experimental results. This time interval represents four month (precisely 2880-hours) test period. On the other hand, the training period for all proposed combination methods is set to one month (precisely 720 hours), which starts from March 19th, 2014 for the first run. Training and test time intervals are particularly selected between these dates, in spite of the data discontinuity in these intervals. Missing data in training dataset affects the accuracy of forecast models and also combination models negatively. Therefore, obtained results from the discontinuous test data are unreliable.

Because of the fact that short-term power forecasts of RİTM are generated for next 48-hours at the end of each daily run, retrospective forecast combination study is applied as if the combined forecasts are published at the end of the

Table 5.1: WPPs' Installed Capacities

Wind Power Plant	Installed Capacity (MW)
WPP-1	15
WPP-2	90
WPP-3	89.7
WPP-4	35
WPP-5	10.2
WPP-6	14.9
WPP-7	39.2
WPP-8	135
WPP-9	240
WPP-10	48
WPP-11	30
WPP-12	50
WPP-13	76.25
WPP-14	12

same run. Therefore, training and test periods are shifted day by day (24-hours) until four month combined forecast data is collected. Then, collected data is continuously merged by taking last 24-hours of each 48-hours combined forecasts. This process is illustrated in Figure 4.2. Note that, training time interval can be changed empirically in order to improve combination models' accuracies, while test time interval for each run should be fixed as 48-hours in scope of RITM.

The algorithm of aforementioned procedure, which is implemented to obtain four month combined wind power forecasts for evaluation is given in Algorithm 1 below.

There are 3 different NWP sources as DMI, GFS, ECMWF, and 4 different forecast models which are ANN, SVM, physical model and cluster based forecasts model as mentioned in Chapter 3 and Chapter 4. These ($N = 3 \times 4$) independent forecasts are listed in Table 5.2, and they have been respectively referred as M_1, M_2, \dots, M_{12} in given results.

Most of the proposed combination models are developed in Java platform in

Algorithm 1 Algorithm of Combination of Forecasts

```
testStart  $\leftarrow$  test start in unixtime
testEnd  $\leftarrow$  test end in unixtime
trainingStart  $\leftarrow$  testStart - (720 * 3600)     $\triangleright$  a month ago from testStart
trainingEnd  $\leftarrow$  testStart
for i = 1 to 14 do
    while (testStart  $\leq$  testEnd) do
        initialize combinedAllForecastsArr[]
        initialize combined48HoursForecastsArrList
        Query real power data between trainingStart and trainingEnd
        for i = 1 to N do
            Query forecast data of model N between trainingStart and
            trainingEnd
        end for
        for j = 1 to k do     $\triangleright$  There are k different combination methods
            Run kth combination model
            combined48HoursForecastsArr[]  $\leftarrow$  combined forecasts
            combinedAllForecastsArrList  $\leftarrow$  combined48HoursForecastsArr
            [24 : 48]
        end for
        testStart  $\leftarrow$  testStart + (24 * 3600)
        trainingStart  $\leftarrow$  trainingStart + (24 * 3600)
        trainingEnd  $\leftarrow$  trainingEnd + (24 * 3600)
    end while
    testStart  $\leftarrow$  test start in unixtime
    Query real power data between testStart and testEnd
    Calculate NMAEs%, NRMSEs%, and NMBs%, and store in file
end for
```

Table 5.2: Individual Forecasts of RITM

Model Id	Model Name
M_1	ANN.DMI
M_2	ANN.GFS
M_3	ANN.ECMWF
M_4	SVM.DMI
M_5	SVM.GFS
M_6	SVM.ECMWF
M_7	PHYSICAL.DMI
M_8	PHYSICAL.GFS
M_9	PHYSICAL.ECMWF
M_{10}	CLUSTERED.DMI
M_{11}	CLUSTERED.GFS
M_{12}	CLUSTERED.ECMWF

cooperation with R [65], which is a free software environment for statistical computing and graphics. The generated R codes for all combination models except FSS, are executed in Java code thanks to a TCP/IP server, namely Rserve which allows other programs including C/C++, PHP, and Java to use the facilities of R. On the other hand, FSS based combination method is only developed in Java.

Rserve stands by for R queries received from Java, and when the request is taken, it executes the queries and returns the obtained results. On the side of Java, data is queried from the central database, and all combination models are run in sequence over R. Returned results are parsed by the help of particular libraries, which also allow calling R functions from Java.

Historical power production and forecast data of all the individual models are queried from the central database of RITM in order to apply the proposed combination methods. After all methods are run for 14 WPPs, new generated combination predictions have been evaluated in terms of NMAE, NRMSE, and NMB metrics, which are explained in Section 2.1.3.

5.1 Combination Results and Evaluations

In order to validate proposed combination methods, each of them is separately evaluated in this section. On the other hand, the results are compared in Section 5.2 in order to make an inference about which one is more suitable for the final implementation to RITM.

5.1.1 L_p -Norm Based Combination Results

A plug in library including the fitted linear model with exponential power distribution errors, is used to run L_p -norm estimators over R. L_p -norm estimators result with regression coefficients, which corresponds to weights of predictors in the linear model.

After L_p -norm estimators based model is applied for different p values as $p = 0.5$, $p = 1$, $p = 1.5$ and $p = 2$, constructed models are evaluated by means of graphical tools. R environment enables the user to visualize linear regression model results in order to show how well the model fits the data via several plots including scatter plot, Q-Q plot, histogram plot, scale-location plot, kernel density plot, and influence plot. Therefore, some of them are utilized to analyze L_p -norm estimators models with regards to residual distributions and outliers.

Note that the following graphical representations of this chapter, between Figure 5.1 and Figure 5.7, cover the data of WPP-9 among 14 tested WPPs as an example. It is chosen due to its installed capacity, which is the maximum among 14 WPPs. In addition, training data (involves 720 hours) of only last 48 hours of four month test time period, which is the final training period of combination models is used in plotted R graphics, because of the fact that 120 different training results exist for the whole combination time period.

In statistics, residuals state the distance of the data points from the fitted regression line, and the ratio of the residuals to their standard error called the standardized residuals. They have zero mean and standard deviation 1. On the other hand, the studentized residuals are similar to standardized residuals, but

they take into account that the variance of the estimated value is not constant. So, they are divided by an estimate of the standard deviation. This adjustment makes the deviation of regression assumptions more visible, so it becomes useful to identify outliers. Another point worth mentioning is that, if 95% of the studentized residuals fall between the range of ± 2 , the distribution of the errors are considered approximately normal. Otherwise, they could be considered unusual. In addition, the studentized residuals out of this range should be more closely analyzed, due to the fact that these points could be outliers. Figure 5.1 represents the Q-Q plots of studentized residuals vs. theoretical quantiles for $p = 0.5$, $p = 1$, $p = 1.5$ and $p = 2$.

As can be seen in Figure 5.1, residual distributions are almost normal for all examined p values of L_p -norm estimators. Since, the data points approximately follow the assumed distribution, which is emphasized by the red line, and the most of the data is in the range of ± 2 . These plots also indicate that there are no apparent outliers.

Histograms and kernel density plots can also be used to check whether the data is normally distributed or not. A histogram with symmetric bell-shaped pattern, which is evenly distributed around zero implies that the normality assumption is likely to be true. Kernel density plots are also an effective way to view the distribution of data in the same manner.

Histogram plots and kernel density plots of residuals of L_p -norm estimators model are presented in Figure 5.2 and Figure 5.3, respectively. In addition, a single density plot, which involves all p valued regression models' results are given in Figure 5.4 for the sake of comparability. According to this plot, it can be seen that the density of residuals slightly increases around zero, while p value is decreased. Therefore, $L_{0.5}$ norm based regression model residuals show slightly more normal distribution compared to other norms.

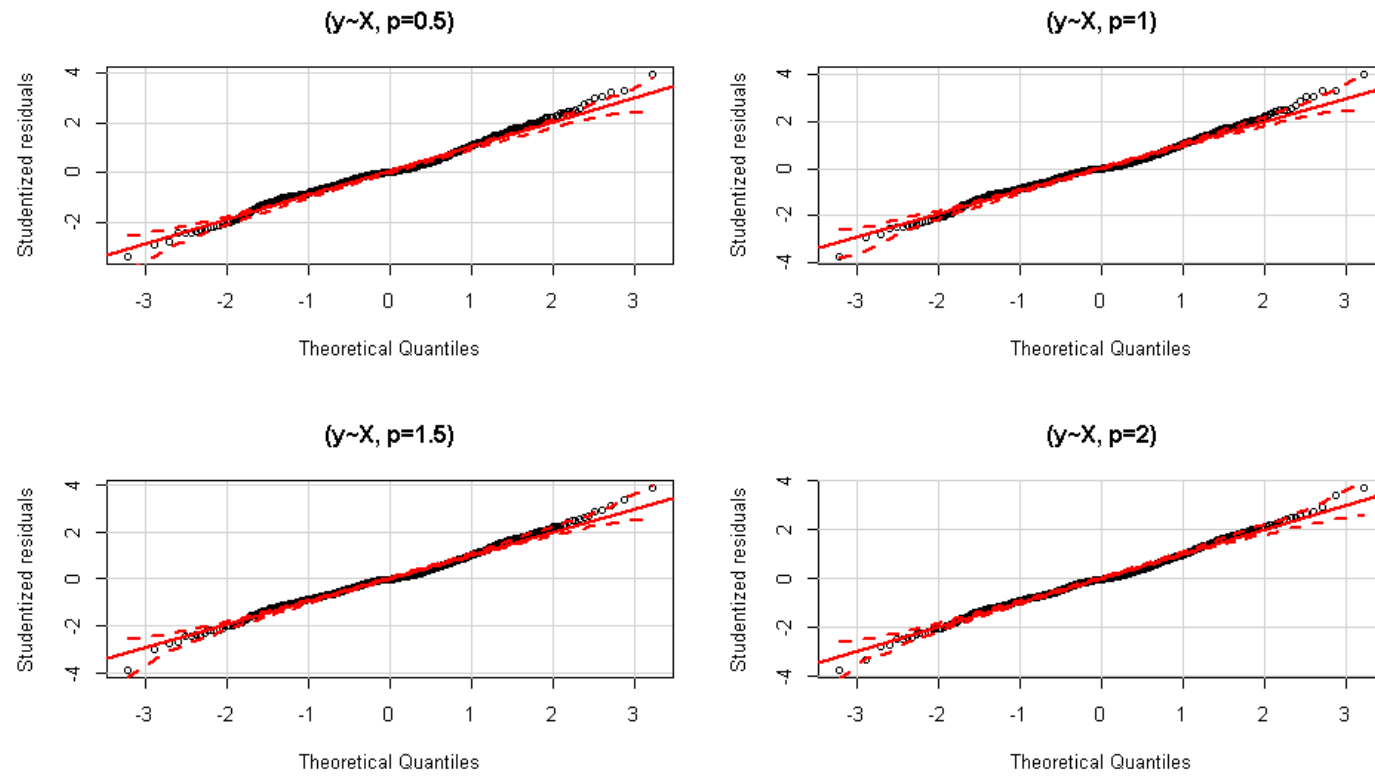


Figure 5.1: A Comparative View to Q-Q Plots of L_p Norm Estimators

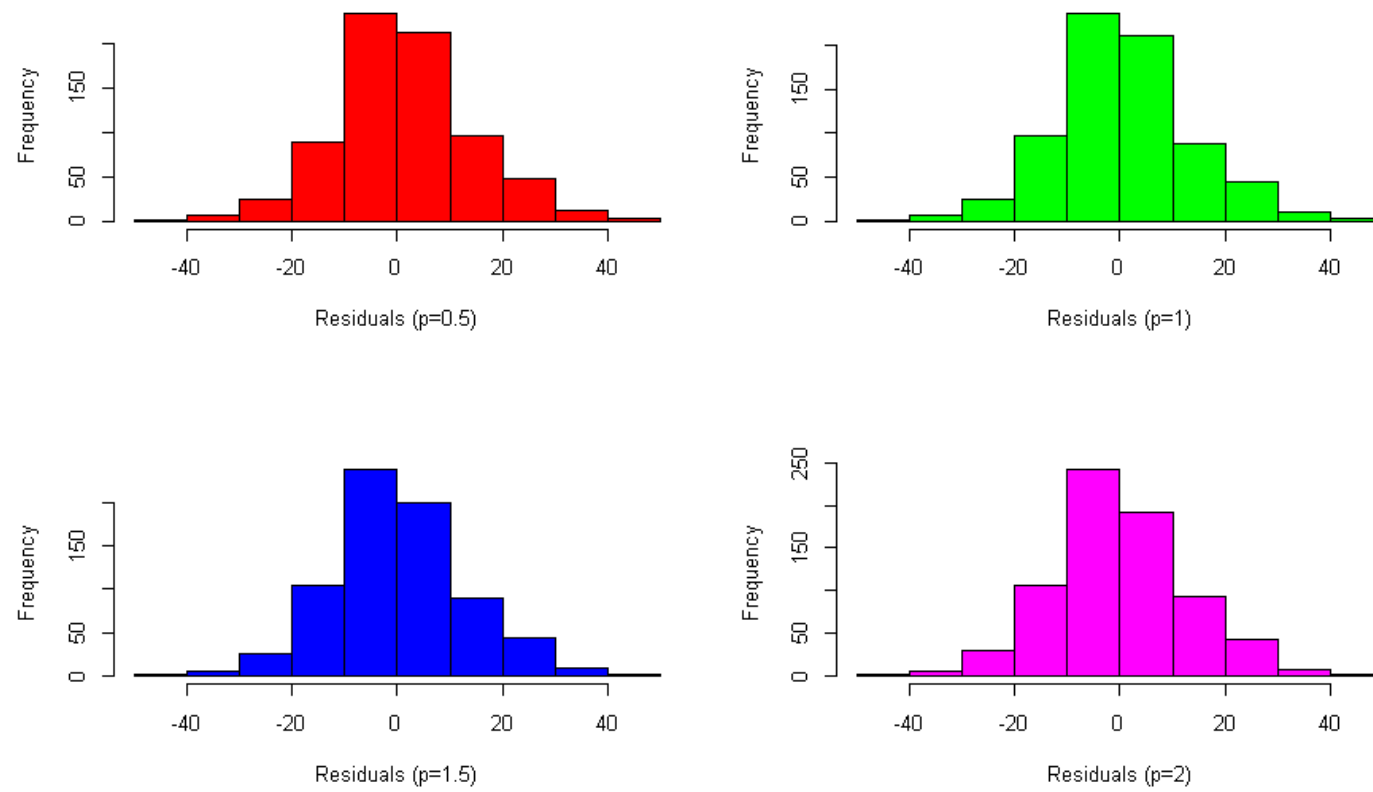


Figure 5.2: Histogram Plots of L_p Norm Estimators

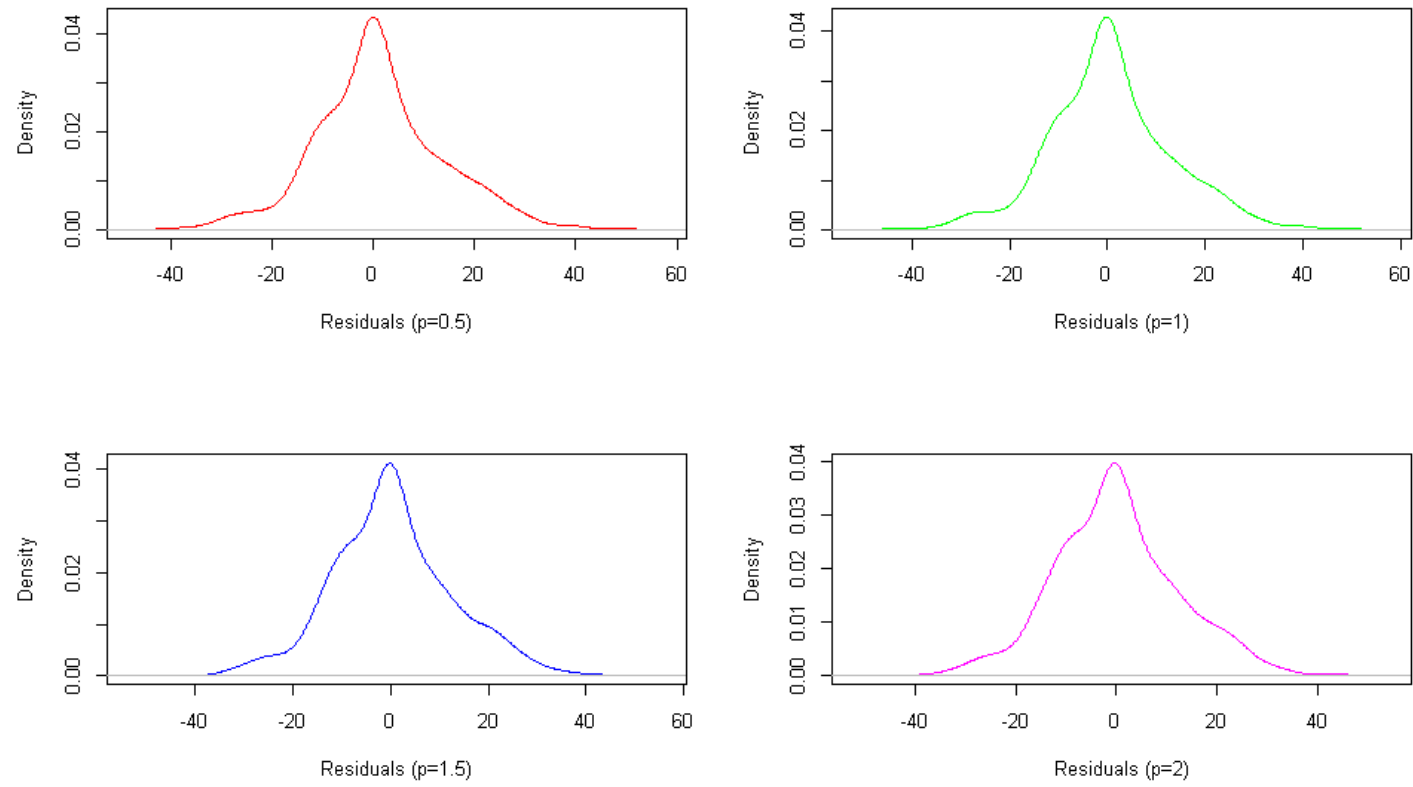


Figure 5.3: Kernel Density Plots of L_p Norm Estimators

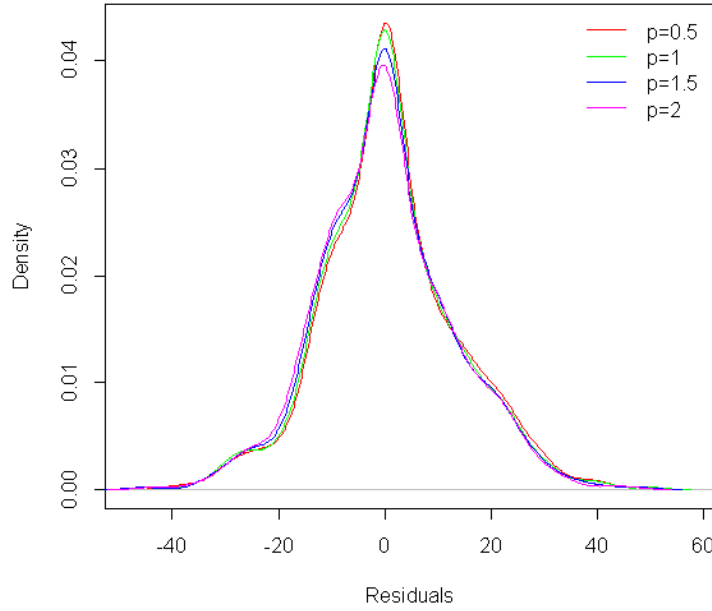


Figure 5.4: Comparable Kernel Density Plots of L_p Norm Estimators

An alternative way to observe distribution of constructed regression models' residuals is to display scatter plot of residuals versus fitted values. This plot should produce a distribution of points scattered randomly about zero, regardless of the size of the fitted values. However, increasing residual values with increasing fitted values pattern is most often observed, since large values in the response variable results in greater variation. Residuals against fitted values plot of L_p -norm estimators model is given in Figure 5.5. In this plot, it can be shown that the residuals corresponding to lower fitted values are more closer to zero. But still, residuals can be accepted as normally distributed according to these plots.

The resulting final forecasts of 14 WPPs, that obtained at the end of the 4 month test period, are evaluated in terms of NMAE, NRMSE, and NMB metrics. Obtained combination error rates comparable to individual forecast models' rates are respectively presented in Table 5.3, Table 5.4, and Table 5.5.

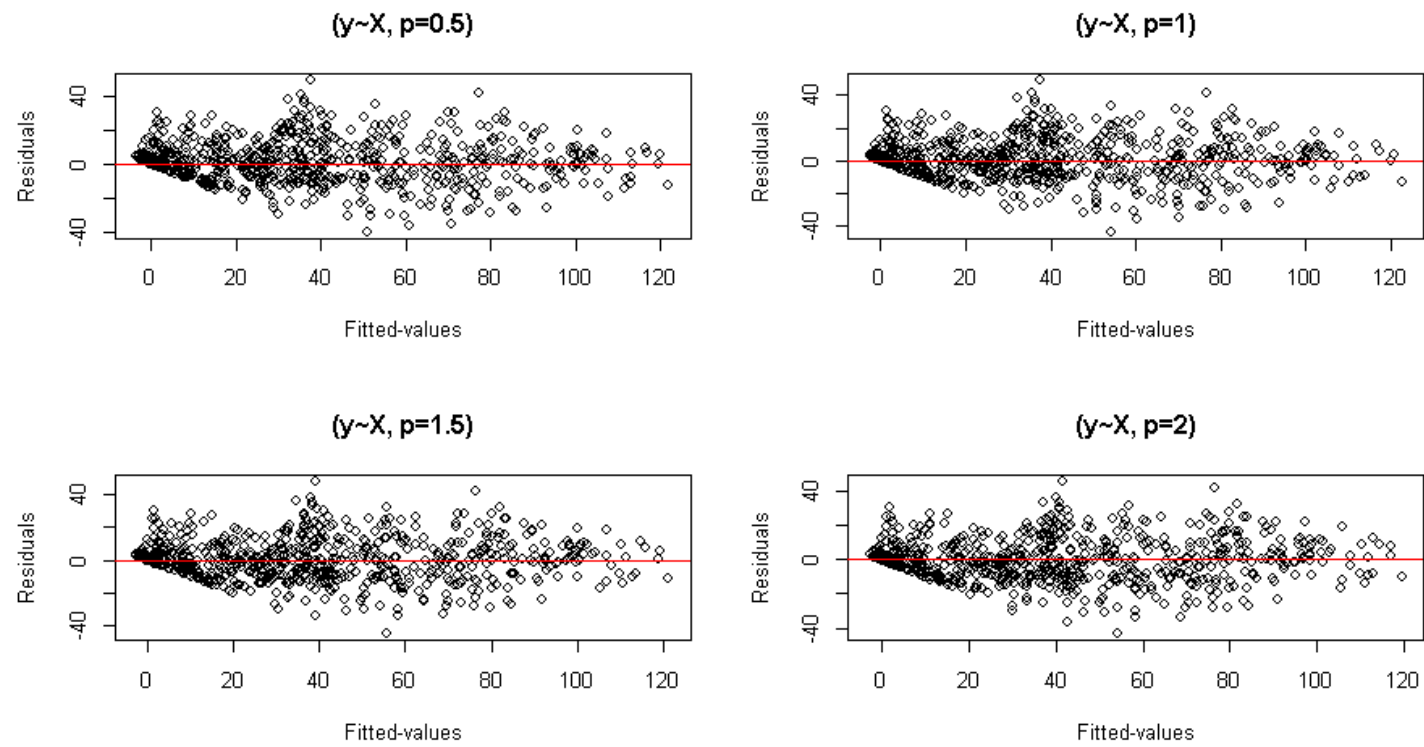


Figure 5.5: Residuals vs. Fitted Values

Table 5.3: NMAEs of L_p -Norm Estimators Based Forecast Combination

Plant ID	NMAEs (%)															
	Individual Forecasts												L_p -Norm Estimators			
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	$L_{0.5}$	L_1	$L_{1.5}$	L_2
1	12.41	11.55	11.57	12.77	13.15	13.09	12.49	13.67	13.21	12.29	11.36	11.19	9.68	9.75	9.81	9.85
2	11.63	11.01	10.59	12.34	15.25	13.30	11.45	13.71	10.07	12.11	12.78	10.78	9.61	9.54	9.53	9.56
3	8.70	8.91	9.22	9.19	10.14	10.43	9.68	10.74	10.83	8.48	9.27	9.51	7.74	7.77	7.82	7.86
4	14.95	15.13	14.76	14.90	17.35	16.96	13.98	14.84	14.59	15.08	15.80	15.51	12.50	12.49	12.73	12.88
5	12.42	12.10	10.94	12.81	12.53	11.23	12.33	11.97	10.52	11.45	11.81	10.12	9.84	9.77	9.87	9.91
6	10.82	10.60	11.20	11.79	12.18	12.70	11.55	16.24	14.95	10.12	10.50	10.41	8.67	8.70	8.75	8.82
7	12.04	11.62	11.00	12.74	12.48	12.27	14.05	14.20	12.79	12.36	11.81	11.95	9.69	9.73	9.74	9.84
8	9.86	9.42	8.56	14.62	9.68	10.90	18.83	23.71	23.51	12.89	11.61	11.78	8.13	8.11	8.13	8.16
9	5.55	5.60	5.38	5.91	6.21	6.33	12.62	10.89	12.07	5.63	5.42	5.73	4.73	4.86	4.80	4.83
10	13.78	11.85	11.28	14.61	13.66	12.44	19.27	19.27	16.78	14.92	13.56	12.36	10.40	10.43	10.48	10.46
11	16.76	15.91	15.77	19.46	18.85	18.49	20.66	21.49	21.23	18.55	18.00	16.35	15.70	15.77	15.86	16.01
12	10.78	10.07	10.89	11.71	10.96	11.16	11.12	14.44	15.61	13.63	13.45	12.81	10.19	10.11	10.07	10.12
13	14.20	15.88	14.03	15.34	17.54	16.37	24.00	21.93	21.86	15.54	14.72	15.16	14.37	14.44	14.77	15.24
14	13.90	13.77	13.21	14.92	14.37	13.77	14.28	14.43	14.74	13.36	14.45	14.12	11.59	11.54	11.56	11.57

Table 5.4: NRMSEs of L_p -Norm Estimators Based Forecast Combination

Plant ID	NRMSEs (%)															
	Individual Forecasts												L_p -Norm Estimators			
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	$L_{0.5}$	L_1	$L_{1.5}$	L_2
1	17.21	16.99	17.10	17.76	18.81	18.90	19.38	20.31	19.86	19.04	16.81	16.88	13.94	13.91	13.84	13.87
2	16.93	16.92	16.20	18.19	22.32	20.27	17.83	21.27	15.98	19.71	19.77	16.81	14.67	14.57	14.53	14.56
3	12.23	13.11	13.49	12.54	13.70	14.04	14.62	15.99	16.04	13.20	13.79	13.99	11.10	11.16	11.07	11.03
4	21.53	22.10	21.68	21.41	23.74	23.69	22.56	22.80	22.04	23.93	23.92	23.32	18.46	18.37	18.46	18.51
5	18.88	18.28	16.94	19.31	18.39	17.09	19.63	18.94	16.40	18.58	18.50	15.99	15.06	14.99	15.09	15.16
6	15.73	15.45	16.27	16.71	17.05	17.52	17.46	23.59	21.82	15.94	15.87	15.88	12.66	12.69	12.69	12.74
7	17.50	17.12	16.14	18.19	18.01	17.65	21.27	21.64	19.10	19.07	17.32	18.48	14.50	14.59	14.57	14.69
8	12.99	12.85	11.65	17.59	12.98	14.04	23.62	30.76	30.74	17.04	15.40	15.87	10.72	10.72	10.76	10.79
9	7.91	7.93	7.60	8.19	8.44	8.66	17.32	14.96	16.09	8.41	7.89	8.38	6.69	6.83	6.76	6.80
10	17.72	15.96	15.41	18.65	17.58	16.34	24.49	26.00	22.24	20.18	18.28	17.38	13.93	13.97	14.00	13.97
11	22.83	23.07	23.02	25.42	25.38	25.09	29.77	31.47	30.90	27.29	26.34	24.44	21.82	21.83	21.75	21.77
12	15.70	14.67	15.65	16.49	15.62	15.42	17.03	22.01	23.17	20.53	20.26	19.06	14.99	14.87	14.78	14.85
13	19.39	21.23	19.58	19.85	21.40	20.77	30.02	28.44	27.97	21.72	20.35	20.96	19.83	19.95	20.18	20.66
14	19.94	20.29	19.33	21.14	20.96	20.17	21.99	21.39	21.69	20.64	21.41	21.06	16.95	16.82	16.81	16.85

Table 5.5: NMBs of L_p -Norm Estimators Based Forecast Combination

Plant ID	NMBs (%)															
	Individual Forecasts												L_p -Norm Estimators			
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	$L_{0.5}$	L_1	$L_{1.5}$	L_2
1	4.57	1.48	1.83	5.01	1.62	3.08	0.08	2.29	2.13	-2.04	-2.33	-0.44	1.44	1.59	1.88	2.04
2	0.43	-1.05	-0.46	2.68	-1.62	-1.24	0.28	-9.68	-1.95	-0.24	-4.54	-1.83	-1.13	-1.01	-1.02	-0.94
3	1.45	-0.19	1.64	1.82	1.99	3.09	-2.30	1.47	3.69	-0.49	-1.29	-0.35	0.32	0.44	0.71	0.94
4	4.58	-0.96	2.61	4.77	2.39	5.13	-2.71	-2.71	1.06	0.70	0.07	0.09	0.60	0.64	1.25	1.61
5	2.49	0.31	2.02	2.58	0.12	1.99	-1.30	-5.12	0.26	-2.75	-3.86	-2.16	-0.68	-0.75	-0.50	-0.31
6	1.74	-1.08	1.06	1.89	-0.14	1.51	2.70	9.71	8.92	-3.89	-3.02	-3.20	-0.51	-0.33	0.02	0.29
7	-0.28	-3.35	-0.50	-0.79	-3.62	-2.15	1.10	-1.50	0.26	-1.34	-1.10	-1.53	-0.88	-0.92	-0.93	-0.88
8	0.47	1.38	1.10	-4.17	0.55	2.04	-11.49	4.59	7.30	-2.32	-1.14	-2.13	0.17	0.44	0.41	0.36
9	-0.67	-0.84	-0.09	0.36	-1.40	-0.72	8.06	7.49	10.20	-0.32	-0.73	0.54	0.93	1.11	1.12	1.21
10	-6.63	-2.32	-1.50	-6.26	-5.01	-1.54	10.40	1.59	6.55	-2.23	-0.44	-0.34	-2.36	-2.15	-2.25	-2.28
11	-0.07	-3.57	-2.46	-3.98	-6.53	-5.65	-10.03	-13.80	-9.36	-7.80	-7.30	-7.81	-6.03	-5.93	-5.67	-5.52
12	-3.76	-3.75	-4.48	-2.08	-3.17	-3.09	-0.34	-12.47	-15.09	-12.02	-11.57	-10.75	-6.40	-6.24	-5.99	-6.01
13	0.71	-0.74	1.27	-7.13	-10.06	-8.27	16.77	9.22	6.01	0.40	1.81	1.48	1.44	1.98	2.19	2.61
14	0.21	-1.67	0.78	1.81	-2.90	-1.00	1.37	1.28	3.91	-0.83	-2.61	-2.27	0.12	0.43	0.44	0.22

As can be seen in Table 5.3, for most of the WPPs (11 of 14), NMAE rates of L_p -norm based combination forecasts are lower than all other individual forecast models. For the remaining three plants namely WPP-11, WPP-12, and WPP-13, the best performed individual forecast has equal or slightly lower NMAE compared to L_p -norm estimators' rates. This is an expected result, since some of the individual forecasts which have over 20% NMAE, have performed the worst predictions. On the other hand, it can be seen in Table 5.4 that the squared errors are lower than all individual forecast models for 12 of 14 WPPs, since the large errors are majorly considered in terms of NRMSE.

NMB scores are calculated in order to find out either individual and combined forecast models tend to overestimate or underestimate. As can be seen in Table 5.5, WPP-11 and WPP-12 has higher NMBs with negative direction, which refers that some of the individual forecast models tend to underestimate the particular WPPs compared to other WPPs. Thus, L_p -norm estimators based combinations implicitly result with slightly underestimation.

Among the results of different p valued estimators, $L_{0.5}$ norm estimators model a little bit more outstanding with regard to performing minimum forecast combination error in terms of NMAE with the rate of 8/14 WPPs. This implication has been also comprehended from density plot of residuals in Figure 5.4 as aforementioned. Nevertheless, there is no explicit difference among different p valued estimators, since the data is almost normally distributed. Therefore, robust regression for combination is not essential in this situation.

Time series of the actual power and L_p norm based combination model forecasts of WPP-9 are presented in Figure 5.6, which covers the last one month of test time period. It can be seen that the series of all the p valued estimators pretty much overlap.

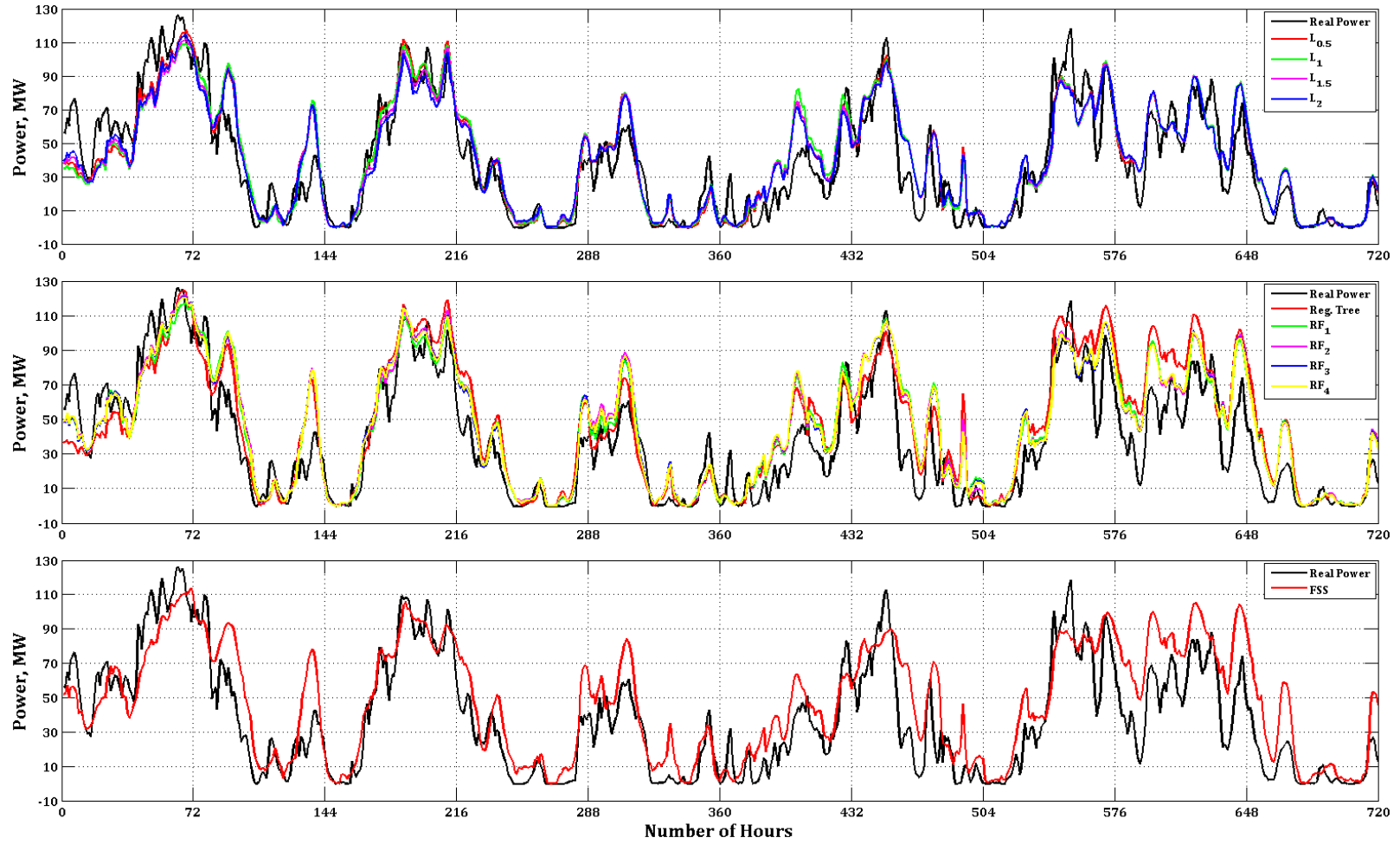


Figure 5.6: Series of Combination Model Forecasts and the Actual Power

Table 5.6: Variable Importance List of Regression Tree of WPP-9

Individual Forecast Models (Combination Predictors)	Variable Importance
PHYSICAL.DMI	587914.83
CLUSTERED.DMI	557372.41
ANN.DMI	433205.32
SVM.DMI	414265.43
PHYSICAL.GFS	303469.92
PHYSICAL.ECMWF	296385.80
CLUSTERED.ECMWF	119271.40
ANN.ECMWF	79967.33
ANN.GFS	64259.77
CLUSTERED.GFS	30497.49

5.1.2 Tree-Based Combination Results

A regression tree is constructed from each training data with 720 observations. To grow a tree, "rpart" plugin library is used, and then the tree is pruned in order to avoid overfitting. An example of the resulting tree structure is given in Figure 5.7, which belongs to the WPP-9 training data. After the pruning, variable importance list is obtained as a result. This ranking list of the combination predictors (actually, individual forecast models) is used to generate weights of the individual forecasts models as mentioned in Section 4.2.

The corresponding variable importance list to the exemplified regression tree in Figure 5.7 is given in Table 5.6. It can be seen from the table that the individual model, namely *PHYSICAL.DMI*, has the maximum variable importance value among the all other individual forecast, and it has been used as the root node of the constructed regression tree. The second important predictor of the tree has been determined as *CLUSTERED.DMI* model, which has placed in the first level of the tree. Although most of the remaining variables placed in the list, they do not appear in the tree structure. That means these variables are also important, but they are specified as surrogate splitters. On the other hand, there are two individual models, namely *SVM.GFS* and *SVM.ECMWF*, which are not placed in variable importance list, since they never appear as either a

primary or a surrogate splitter while the tree is growing. Due to this absence, they have been weighted with zero while the final combined forecast has been generated for its corresponding test period.

Random forests are constructed in the similar way as regression tree construction that, the training data with 720 observation is used to grow each random forest. To generate a random forest, which consists of n regression trees, "party" plugin library is used by passing tree number parameter as an input. Number of trees in the forest, n is respectively assigned as 50, 100, 500, and 1000, then random forests are constructed. As a result of the random forest learning, variable importance result is used to generate weights of the individual forecasts models.

After the tree-based combination methods are applied on 14 WPPs data involved 4 month test period, the resulting final forecasts are evaluated in terms of NMAE, NRMSE, and NMB metrics. Obtained combination error rates comparable to individual forecast models' rates are presented in Table 5.7, Table 5.8, and Table 5.9 respectively.

According to the given NMAEs in Table 5.7, regression tree based combination method results with lower error rates than all other individual forecast models for 12 out of 14 WPPs. And the remaining two WPPs, namely WPP-9 and WPP-12, has slightly lower NMAEs on the part of individual forecast models. On the other hand, this rate is 13/14 in terms of NRMSE, as can be seen in Table 5.8.

Random forests based combination methods with different tree numbers give similar NMAEs and NRMSEs in comparison with each other. For 11 out of 14 WPPs, they result with lower NMAEs than all other individual forecast models. On the other hand, this rate is 13/14 in terms of NRMSE. In addition, it is clear that increasing the tree number, n , has not provided an explicit improvement on combination and also has consumed more time, even the random forests with lower tree numbers has performed better results for a number of WPPs.

When a comparison is considered between regression tree based method and random forests based method, regression tree based method has performed the

combination with minimum NMAEs and also NRMSEs for 4/14 WPPs, while this rate is 10/14 for random forests based one. Therefore, it can be inferred that the random forests based combination method is more preferable than regression tree based combination method with regard to RITM forecast models.

Both regression tree based and random forest based combined forecasts tend to perform a little underestimation for WPP-11 and WPP-12 according to the given NMB scores in Table 5.9. Indeed, this result derives from some of the individual forecast models, which have under -7% NMB.

Time series of the actual power and the both regression tree and random forests based combination model forecasts of WPP-9 are presented in Figure 5.6, which covers the last one month of test time period. Series of combined forecasts based on random forests with 50, 100, 500, and 1000 trees are respectively referred as RF_1 , RF_2 , RF_3 , and RF_4 in the figure.

It can be seen from the Figure 5.6, that the series of combination models based on random forests with different tree numbers pretty much overlap. In addition, the series of regression tree based model shows a slightly different trend than random forests based ones.

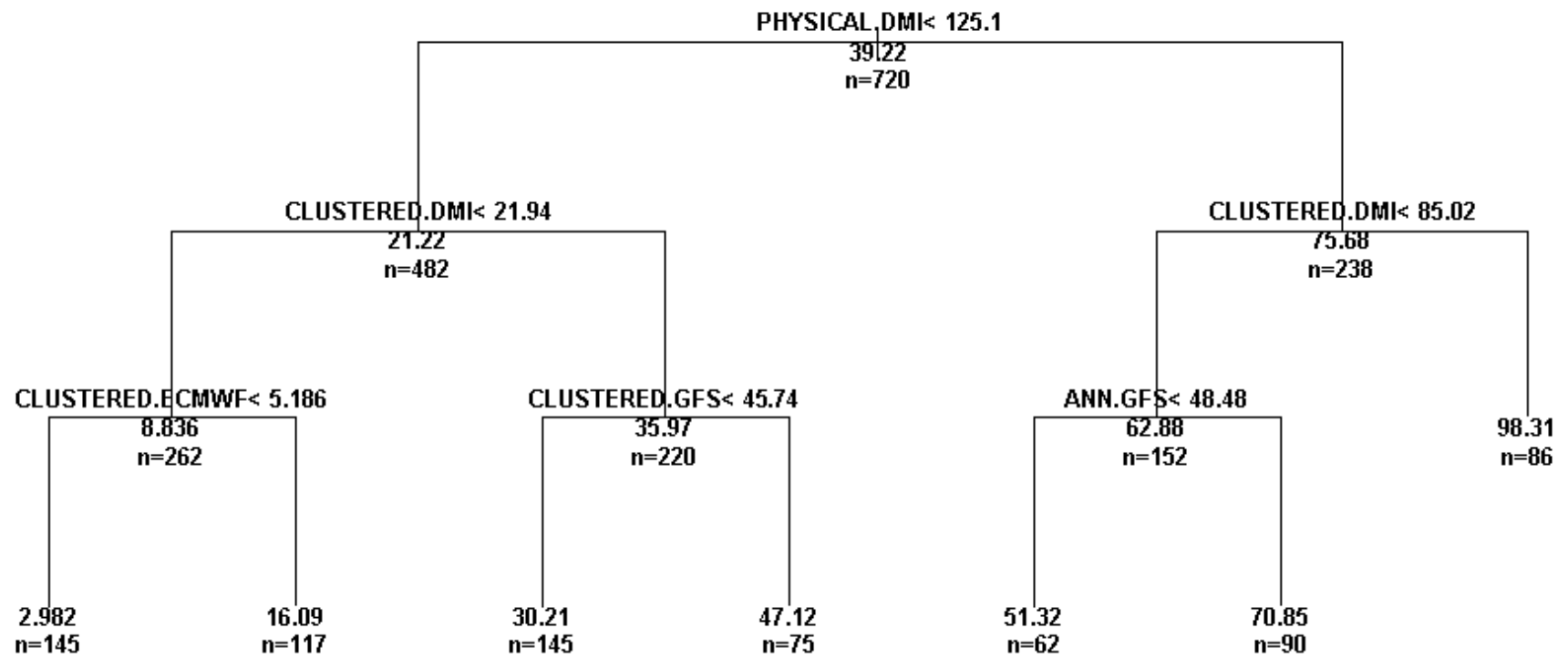


Figure 5.7: Pruned Regression Tree of WPP-9

Table 5.7: NMAEs of Tree-Based Forecast Combination

Plant ID	NMAEs (%)																
	Individual Forecasts												Tree-Based Combination				
	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈	M ₉	M ₁₀	M ₁₁	M ₁₂	Reg.Tree	n=50	n=100	n=500	n=1000
1	12.41	11.55	11.57	12.77	13.15	13.09	12.49	13.67	13.21	12.29	11.36	11.19	9.93	9.74	9.72	9.72	9.71
2	11.63	11.01	10.59	12.34	15.25	13.30	11.45	13.71	10.07	12.11	12.78	10.78	9.64	9.49	9.46	9.45	9.45
3	8.70	8.91	9.22	9.19	10.14	10.43	9.68	10.74	10.83	8.48	9.27	9.51	8.00	7.92	7.88	7.88	7.90
4	14.95	15.13	14.76	14.90	17.35	16.96	13.98	14.84	14.59	15.08	15.80	15.51	13.22	12.73	12.72	12.74	12.74
5	12.42	12.10	10.94	12.81	12.53	11.23	12.33	11.97	10.52	11.45	11.81	10.12	9.72	9.69	9.75	9.74	9.75
6	10.82	10.60	11.20	11.79	12.18	12.70	11.55	16.24	14.95	10.12	10.50	10.41	9.41	9.05	9.05	8.99	9.01
7	12.04	11.62	11.00	12.74	12.48	12.27	14.05	14.20	12.79	12.36	11.81	11.95	10.08	9.64	9.61	9.61	9.62
8	9.86	9.42	8.56	14.62	9.68	10.90	18.83	23.71	23.51	12.89	11.61	11.78	8.44	8.86	8.82	8.86	8.88
9	5.55	5.60	5.38	5.91	6.21	6.33	12.62	10.89	12.07	5.63	5.42	5.73	5.45	5.11	5.11	5.08	5.07
10	13.78	11.85	11.28	14.61	13.66	12.44	19.27	19.27	16.78	14.92	13.56	12.36	10.32	10.15	10.19	10.18	10.17
11	16.76	15.91	15.77	19.46	18.85	18.49	20.66	21.49	21.23	18.55	18.00	16.35	15.76	16.06	15.96	16.03	16.04
12	10.78	10.07	10.89	11.71	10.96	11.16	11.12	14.44	15.61	13.63	13.45	12.81	10.45	10.56	10.66	10.64	10.66
13	14.20	15.88	14.03	15.34	17.54	16.37	24.00	21.93	21.86	15.54	14.72	15.16	12.42	12.71	12.64	12.61	12.62
14	13.90	13.77	13.21	14.92	14.37	13.77	14.28	14.43	14.74	13.36	14.45	14.12	11.72	11.60	11.58	11.55	11.55

Table 5.8: NRMSEs of Tree-Based Forecast Combination

Plant ID	NRMSEs (%)																
	Individual Forecasts												Tree-Based Combination				
	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈	M ₉	M ₁₀	M ₁₁	M ₁₂	Reg.Tree	n=50	n=100	n=500	n=1000
1	17.21	16.99	17.10	17.76	18.81	18.90	19.38	20.31	19.86	19.04	16.81	16.88	14.29	13.70	13.72	13.74	13.74
2	16.93	16.92	16.20	18.19	22.32	20.27	17.83	21.27	15.98	19.71	19.77	16.81	14.70	14.35	14.33	14.30	14.34
3	12.23	13.11	13.49	12.54	13.70	14.04	14.62	15.99	16.04	13.20	13.79	13.99	11.19	11.07	11.00	11.06	11.07
4	21.53	22.10	21.68	21.41	23.74	23.69	22.56	22.80	22.04	23.93	23.92	23.32	18.98	18.36	18.33	18.37	18.34
5	18.88	18.28	16.94	19.31	18.39	17.09	19.63	18.94	16.40	18.58	18.50	15.99	14.95	14.86	14.89	14.81	14.81
6	15.73	15.45	16.27	16.71	17.05	17.52	17.46	23.59	21.82	15.94	15.87	15.88	13.26	12.98	12.89	12.85	12.86
7	17.50	17.12	16.14	18.19	18.01	17.65	21.27	21.64	19.10	19.07	17.32	18.48	14.96	14.25	14.34	14.26	14.29
8	12.99	12.85	11.65	17.59	12.98	14.04	23.62	30.76	30.74	17.04	15.40	15.87	10.94	11.32	11.32	11.27	11.28
9	7.91	7.93	7.60	8.19	8.44	8.66	17.32	14.96	16.09	8.41	7.89	8.38	7.57	7.15	7.11	7.07	7.07
10	17.72	15.96	15.41	18.65	17.58	16.34	24.49	26.00	22.24	20.18	18.28	17.38	13.97	13.79	13.80	13.76	13.75
11	22.83	23.07	23.02	25.42	25.38	25.09	29.77	31.47	30.90	27.29	26.34	24.44	21.61	21.75	21.94	21.84	21.85
12	15.70	14.67	15.65	16.49	15.62	15.42	17.03	22.01	23.17	20.53	20.26	19.06	15.13	15.58	15.64	15.63	15.62
13	19.39	21.23	19.58	19.85	21.40	20.77	30.02	28.44	27.97	21.72	20.35	20.96	17.77	17.98	17.96	17.90	17.97
14	19.94	20.29	19.33	21.14	20.96	20.17	21.99	21.39	21.69	20.64	21.41	21.06	17.11	16.81	16.77	16.77	16.75

Table 5.9: NMBs of Tree-Based Forecast Combination

Plant ID	NMBs (%)																
	Individual Forecasts												Tree-Based Combination				
	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈	M ₉	M ₁₀	M ₁₁	M ₁₂	Reg.Tree	n=50	n=100	n=500	n=1000
1	4.57	1.48	1.83	5.01	1.62	3.08	0.08	2.29	2.13	-2.04	-2.33	-0.44	1.41	1.60	1.66	1.65	1.67
2	0.43	-1.05	-0.46	2.68	-1.62	-1.24	0.28	-9.68	-1.95	-0.24	-4.54	-1.83	-1.46	-1.25	-1.32	-1.32	-1.31
3	1.45	-0.19	1.64	1.82	1.99	3.09	-2.30	1.47	3.69	-0.49	-1.29	-0.35	0.49	0.75	0.80	0.80	0.81
4	4.58	-0.96	2.61	4.77	2.39	5.13	-2.71	-2.71	1.06	0.70	0.07	0.09	1.15	1.39	1.43	1.36	1.42
5	2.49	0.31	2.02	2.58	0.12	1.99	-1.30	-5.12	0.26	-2.75	-3.86	-2.16	-0.25	-0.38	-0.43	-0.41	-0.41
6	1.74	-1.08	1.06	1.89	-0.14	1.51	2.70	9.71	8.92	-3.89	-3.02	-3.20	1.59	1.35	1.33	1.34	1.35
7	-0.28	-3.35	-0.50	-0.79	-3.62	-2.15	1.10	-1.50	0.26	-1.34	-1.10	-1.53	-0.97	-1.03	-1.11	-1.04	-1.04
8	0.47	1.38	1.10	-4.17	0.55	2.04	-11.49	4.59	7.30	-2.32	-1.14	-2.13	0.59	0.10	0.07	0.03	0.07
9	-0.67	-0.84	-0.09	0.36	-1.40	-0.72	8.06	7.49	10.20	-0.32	-0.73	0.54	2.05	2.03	2.04	2.03	2.04
10	-6.63	-2.32	-1.50	-6.26	-5.01	-1.54	10.40	1.59	6.55	-2.23	-0.44	-0.34	-1.99	-1.14	-1.22	-1.26	-1.25
11	-0.07	-3.57	-2.46	-3.98	-6.53	-5.65	-10.03	-13.80	-9.36	-7.80	-7.30	-7.81	-5.79	-6.02	-6.10	-6.08	-6.09
12	-3.76	-3.75	-4.48	-2.08	-3.17	-3.09	-0.34	-12.47	-15.09	-12.02	-11.57	-10.75	-5.68	-6.67	-6.74	-6.74	-6.75
13	0.71	-0.74	1.27	-7.13	-10.06	-8.27	16.77	9.22	6.01	0.40	1.81	1.48	0.78	1.10	1.07	1.10	1.10
14	0.21	-1.67	0.78	1.81	-2.90	-1.00	1.37	1.28	3.91	-0.83	-2.61	-2.27	0.14	0.13	0.08	0.14	0.12

5.1.3 FSS Based Combination Results

FSS based combination method only requires simple mathematical calculations as given in Section {FSS Based Forecast Combination. Therefore, this method is developed in Java without need of any library extension.

Each constructed FSS, which is in a matrix form, represents the contributions of individual models learning from 720 observations of one month training data. At the end of the each FSS construction, final combined forecasts of next 48 hours are generated through particular FSS. After the four month period is continued by shifting both training and test period day by day, obtained forecast combination outputs are evaluated in terms of NMAE, NRMSE, and NMB.

Table 5.10 presents the NMAE results of FSS based combination method compared to individual forecast models. As can be seen in this table, FSS based method outputs the minimum NMAEs for 7 WPPs among the total 14 WPPs. Moreover, it results with an equal NMAE for WPP-13, with regard to the individual model referred as M_3 . For the remaining 6 WPPs, NMAEs of FSS based combination forecasts are acceptable except WPP-8. Since the NMAE of this plant is too high that it is twice as much than the individual forecast model with minimum NMAE. In addition, this experimental results are similar in terms of NRMSE as can be seen in Table 5.11.

In order to validate FSS based combination method in terms of overestimation/underestimation, NMB scores are calculated such in the other proposed methods. It can be seen in Table 5.12, the forecasts of this proposed method tend to slightly underestimate for only WPP-11 with -5.53% NMB at maximum, among 14 WPPs.

Time series of the actual power and the FSS based combination model forecasts of WPP-9 are presented in Figure 5.6 that displayed time series belong to the last one-month of test time period.

Table 5.10: NMAEs of FSS Based Forecast Combination

Plant ID	NMAEs (%)												FSS
	Individual Forecasts												
	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈	M ₉	M ₁₀	M ₁₁	M ₁₂	
1	12.41	11.55	11.57	12.77	13.15	13.09	12.49	13.67	13.21	12.29	11.36	11.19	10.19
2	11.63	11.01	10.59	12.34	15.25	13.30	11.45	13.71	10.07	12.11	12.78	10.78	11.08
3	8.70	8.91	9.22	9.19	10.14	10.43	9.68	10.74	10.83	8.48	9.27	9.51	8.44
4	14.95	15.13	14.76	14.90	17.35	16.96	13.98	14.84	14.59	15.08	15.80	15.51	13.1
5	12.42	12.10	10.94	12.81	12.53	11.23	12.33	11.97	10.52	11.45	11.81	10.12	10.29
6	10.82	10.60	11.20	11.79	12.18	12.70	11.55	16.24	14.95	10.12	10.50	10.41	9.75
7	12.04	11.62	11.00	12.74	12.48	12.27	14.05	14.20	12.79	12.36	11.81	11.95	10.42
8	9.86	9.42	8.56	14.62	9.68	10.90	18.83	23.71	23.51	12.89	11.61	11.78	17.44
9	5.55	5.60	5.38	5.91	6.21	6.33	12.62	10.89	12.07	5.63	5.42	5.73	5.74
10	13.78	11.85	11.28	14.61	13.66	12.44	19.27	19.27	16.78	14.92	13.56	12.36	13.31
11	16.76	15.91	15.77	19.46	18.85	18.49	20.66	21.49	21.23	18.55	18.00	16.35	16.49
12	10.78	10.07	10.89	11.71	10.96	11.16	11.12	14.44	15.61	13.63	13.45	12.81	10.03
13	14.20	15.88	14.03	15.34	17.54	16.37	24.00	21.93	21.86	15.54	14.72	15.16	14.03
14	13.90	13.77	13.21	14.92	14.37	13.77	14.28	14.43	14.74	13.36	14.45	14.12	11.89

Table 5.11: NRMSEs of FSS Based Forecast Combination

Plant ID	NRMSEs (%)												
	Individual Forecasts												FSS
	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈	M ₉	M ₁₀	M ₁₁	M ₁₂	
1	17.21	16.99	17.10	17.76	18.81	18.90	19.38	20.31	19.86	19.04	16.81	16.88	14.26
2	16.93	16.92	16.20	18.19	22.32	20.27	17.83	21.27	15.98	19.71	19.77	16.81	16.14
3	12.23	13.11	13.49	12.54	13.70	14.04	14.62	15.99	16.04	13.20	13.79	13.99	11.59
4	21.53	22.10	21.68	21.41	23.74	23.69	22.56	22.80	22.04	23.93	23.92	23.32	18.64
5	18.88	18.28	16.94	19.31	18.39	17.09	19.63	18.94	16.40	18.58	18.50	15.99	15.56
6	15.73	15.45	16.27	16.71	17.05	17.52	17.46	23.59	21.82	15.94	15.87	15.88	13.80
7	17.50	17.12	16.14	18.19	18.01	17.65	21.27	21.64	19.10	19.07	17.32	18.48	15.16
8	12.99	12.85	11.65	17.59	12.98	14.04	23.62	30.76	30.74	17.04	15.40	15.87	27.90
9	7.91	7.93	7.60	8.19	8.44	8.66	17.32	14.96	16.09	8.41	7.89	8.38	7.81
10	17.72	15.96	15.41	18.65	17.58	16.34	24.49	26.00	22.24	20.18	18.28	17.38	21.67
11	22.83	23.07	23.02	25.42	25.38	25.09	29.77	31.47	30.90	27.29	26.34	24.44	22.08
12	15.70	14.67	15.65	16.49	15.62	15.42	17.03	22.01	23.17	20.53	20.26	19.06	14.24
13	19.39	21.23	19.58	19.85	21.40	20.77	30.02	28.44	27.97	21.72	20.35	20.96	19.53
14	19.94	20.29	19.33	21.14	20.96	20.17	21.99	21.39	21.69	20.64	21.41	21.06	17.15

Table 5.12: NMBs of FSS Based Forecast Combination

Plant ID	NMBs (%)												
	Individual Forecasts												FSS
	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈	M ₉	M ₁₀	M ₁₁	M ₁₂	
1	4.57	1.48	1.83	5.01	1.62	3.08	0.08	2.29	2.13	-2.04	-2.33	-0.44	2.72
2	0.43	-1.05	-0.46	2.68	-1.62	-1.24	0.28	-9.68	-1.95	-0.24	-4.54	-1.83	-0.55
3	1.45	-0.19	1.64	1.82	1.99	3.09	-2.30	1.47	3.69	-0.49	-1.29	-0.35	1.85
4	4.58	-0.96	2.61	4.77	2.39	5.13	-2.71	-2.71	1.06	0.70	0.07	0.09	2.34
5	2.49	0.31	2.02	2.58	0.12	1.99	-1.30	-5.12	0.26	-2.75	-3.86	-2.16	0.49
6	1.74	-1.08	1.06	1.89	-0.14	1.51	2.70	9.71	8.92	-3.89	-3.02	-3.20	2.66
7	-0.28	-3.35	-0.50	-0.79	-3.62	-2.15	1.10	-1.50	0.26	-1.34	-1.10	-1.53	-1.28
8	0.47	1.38	1.10	-4.17	0.55	2.04	-11.49	4.59	7.30	-2.32	-1.14	-2.13	0.16
9	-0.67	-0.84	-0.09	0.36	-1.40	-0.72	8.06	7.49	10.20	-0.32	-0.73	0.54	3.13
10	-6.63	-2.32	-1.50	-6.26	-5.01	-1.54	10.40	1.59	6.55	-2.23	-0.44	-0.34	-2.47
11	-0.07	-3.57	-2.46	-3.98	-6.53	-5.65	-10.03	-13.80	-9.36	-7.80	-7.30	-7.81	-5.53
12	-3.76	-3.75	-4.48	-2.08	-3.17	-3.09	-0.34	-12.47	-15.09	-12.02	-11.57	-10.75	-3.09
13	0.71	-0.74	1.27	-7.13	-10.06	-8.27	16.77	9.22	6.01	0.40	1.81	1.48	0.95
14	0.21	-1.67	0.78	1.81	-2.90	-1.00	1.37	1.28	3.91	-0.83	-2.61	-2.27	0.10

5.2 Comparison of the Combination Methods

After the results of all proposed combination methods are obtained, a comparison among them is considered with regard to combined forecast errors and time complexity.

In addition to the separately given results compared to the individual forecast models for each combination approach in the previous section, comparative NMAEs of all the combined forecasts are presented in Table 5.13. According to the given table, L_p -norm based combination methods have given the minimum NMAEs for 8 WPPs, while tree-based combination methods including regression tree based combination and random forests based combinations have given the minimum NMAEs for 5 WPPs. However, NMAE scores are very close to each other for these two proposed methods, namely L_p -norm based combination method and tree-based combination method. In other words, the difference between the NMAEs of best combined forecast and any combined forecast of these two proposed methods, is marginal. On the other hand, FSS based combination method has given the minimum NMAE for only one WPP. Also, its resulting NMAEs are over 1% more than the NMAEs of best combined forecast for 6 WPPs.

From the point of time complexity, tree-based combination method consumes more time than the L_p -norm based combination and FSS based combination methods. Especially, random forests construction increases time complexity considerably, so that the higher value of tree number is used to construct forest, the more time is consumed. On the other hand, it has been already declared in the previous section that increasing the tree number does not improve the performance of the combination anyway.

Table 5.13: NMAE Comparison among Forecast Combination Methods

Plant ID	NMAEs (%)									
	<i>L_p</i> -Norm Estimators				Tree-Based Combination					FSS
	L_{0.5}	L₁	L_{1.5}	L₂	Reg.Tree	n=50	n=100	n=500	n=1000	
1	9.68	9.75	9.81	9.85	9.93	9.74	9.72	9.72	9.71	10.19
2	9.61	9.54	9.53	9.56	9.64	9.49	9.46	9.45	9.45	11.08
3	7.74	7.77	7.82	7.86	8.00	7.92	7.88	7.88	7.90	8.44
4	12.50	12.49	12.73	12.88	13.22	12.73	12.72	12.74	12.74	13.10
5	9.84	9.77	9.87	9.91	9.72	9.69	9.75	9.74	9.75	10.29
6	8.67	8.70	8.75	8.82	9.41	9.05	9.05	8.99	9.01	9.75
7	9.69	9.73	9.74	9.84	10.08	9.64	9.61	9.61	9.62	10.42
8	8.13	8.11	8.13	8.16	8.44	8.86	8.82	8.86	8.88	17.44
9	4.73	4.86	4.80	4.83	5.45	5.11	5.11	5.08	5.07	5.74
10	10.40	10.43	10.48	10.46	10.32	10.15	10.19	10.18	10.17	13.31
11	15.70	15.77	15.86	16.01	15.76	16.06	15.96	16.03	16.04	16.49
12	10.19	10.11	10.07	10.12	10.45	10.56	10.66	10.64	10.66	10.03
13	14.37	14.44	14.77	15.24	12.42	12.71	12.64	12.61	12.62	14.03
14	11.59	11.54	11.56	11.57	11.72	11.60	11.58	11.55	11.55	11.89

CHAPTER 6

CONCLUSION AND FUTURE WORK

Wind has become one of the most significant energy sources in recent years, because of the increasing awareness of renewables to produce electricity. At this point, developing accurate wind power forecasting systems is one of the main concerns, which arises from the natural variability of wind. Therefore, many wind power forecasting methods are proposed in literature, most of which are based on numerical weather predictions (NWP). In addition to many alternative methods to generate wind power forecasts, there are several published NWPs from different sources, which lead to increase forecasts diversity. In order to utilize these alternative forecasts simultaneously, combining them is a frequently preferred approach in literature.

In this thesis, various methods for forecast combination, namely L_p -norm based combination, tree-based combination, and fuzzy soft sets (FSS) based combination, are presented to improve the performance of individual forecasts. In order to be regarded as a successful combination in wind power forecasting, the combined forecast should be better than the individual ones, or at least be as good as the best performed one. The evaluation results of this thesis study show that all proposed combination approaches considerably comply with this suggestion. They perform almost 6.5% improvement in terms of normalized mean absolute error (NMAE) compared to the worst individual forecast, and approximately 1% improvement compared to the best individual forecast for the average of 14 wind power plants (WPPs) under investigation. On the other hand, the rate is almost 1.8% on average in terms of normalized root mean squared er-

ror (NRMSE) compared to the best individual forecast, while it is about 9% in comparison with the worst individual forecast. Here, it should be kept in mind that 1% improvement of the forecasts is considered as a valuable contribution in wind energy industry.

According to the achieved results, L_p -norm based combination approach with lower p values such as 0.5, seems to be more suitable for RİTM forecasts combination. Tree-based combination is also valuable, but it consumes more computational time than L_p -norm based combination, especially for random forests with higher tree numbers. Besides, FSS based forecast combination has given acceptable results. Nevertheless, it is not as proper as the other combination methods in terms of consistency. Although it has resulted in lower forecast errors for many WPPs, it has twice as much errors, both in terms of NMAE and NRMSE, than other methods for a particular WPP, which is undesirable.

Within the scope of this thesis, one month training data is used during the implementation of the proposed methods to RİTM. It is because of the available data of individual forecasts in addition to available corresponding actual power data. However, RİTM is an on-going project, and new WPPs are integrated day by day. After the sufficient and continuous data is collected, training time interval can be empirically extended to over one month, such as three months or six months. Therefore, the effects of using different time horizons for training phase can be investigated for all proposed combination methods as a future study. Furthermore, it is also possible to make improvements on the individual forecasts employed in combined forecasts, because of the fact that the performance of the combined forecast depends on the accuracy of the individual forecasts eventually.

REFERENCES

- [1] A.K. Akella, R.P. Saini, and M.P. Sharma. Social, economical and environmental impacts of renewable energy systems. *Renewable Energy*, 34(2):390–396, 2009.
- [2] Erman Terciyanli, Turan Demirci, Dilek Kucuk, Serkan Buhan, Mehmet Baris Ozkan, **Ceyda Er Koksoy**, Erkan Koc, Ceren Kahraman, Ali Burhan Haliloglu, Tugba Demir, Mustafa Gunindi, Meltem Gokmen, and Zafer Karayilanoglu. The architecture of a large-scale wind power monitoring and forecast system. In *Power Engineering, Energy and Electrical Drives (POWERENG), 2013 Fourth International Conference on*, pages 1162–1167. IEEE, 2013.
- [3] Omar Ellabban, Haitham Abu-Rub, and Frede Blaabjerg. Renewable energy resources: Current status, future prospects and their enabling technology. *Renewable and Sustainable Energy Reviews*, 39:748–764, 2014.
- [4] Aoife M. Foley, Paul G. Leahy, Antonino Marvuglia, and Eamon J. McKeogh. Current methods and advances in forecasting of wind power generation. *Renewable Energy*, 37(1):1–8, 2012.
- [5] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelmann. Wind power forecasting : state-of-the-art 2009. Technical report, Argonne National Laboratory (ANL), 11 2009.
- [6] Yuan-Kang Wu and Jing-Shan Hong. A literature review of wind forecasting technology in the world. In *Power Tech, 2007 IEEE Lausanne*, pages 504–509. IEEE, 2007.
- [7] J. Ignacio Ramirez-Rosado, L. Alfredo Fernandez-Jimenez, Cláudio Monteiro, João Sousa, and Ricardo Bessa. Comparison of two new short-term wind-power forecasting systems. *Renewable Energy*, 34(7):1848–1854, 2009.
- [8] Allan Timmermann. Forecast combinations. *Handbook of Economic Forecasting*, 1:135–196, 2006.
- [9] J. M. Bates and C. W. J. Granger. Combination of forecasts. *OR*, 20(4):451–468, 1969.
- [10] Kuan-Yu Chen. Combining linear and nonlinear model in forecasting tourism demand. *Expert Systems with Applications*, 38(8):10368–10376, 2011.

- [11] Ismael Sánchez. Adaptive combination of forecast with application to wind energy forecast. In *Probabilistic Methods Applied to Power Systems 2006 International Conference on*, pages 1–6. IEEE, 2006.
- [12] **Ceyda Er Koksoy**, Mehmet Baris Ozkan, Dilek Kucuk, Abdullah Bestil, Sena Sonmez, Serkan Buhan, Turan Demirci, Pınar Karagoz, and Ayse Nur Birturk. Performance analysis of data mining techniques for improving the accuracy of wind power forecast combination. In *Proceedings of the International Workshop on Data Analytics for Renewable Energy Integration (DARE) of ECML/PKDD 2015*. Springer, 2015.
- [13] M. Lange, U. Focken, R. Meyer, M. Denhardt, B. Ernst, and F. Berster. Optimal combination of different numerical weather models for improved wind power predictions. In *6th International Workshop on Large-Scale Integration of Wind Power and Transmission Networks for Offshore Wind Farm*, 2006.
- [14] Jeremy Smith and Kenneth F. Wallis. A simple explanation of the forecast combination puzzle*. *Oxford Bulletin of Economics and Statistics*, 71(3):331–355, 2009.
- [15] Cagdas Hakan Aladag, Erol Egrioglu, and Ufuk Yolcu. Forecast combination by using artificial neural networks. *Neural Processing Letters*, 32(3):269–276, 2010.
- [16] Zhi Xiao, Ke Gong, and Yan Zou. A combined forecasting approach based on fuzzy soft sets. *Journal of Computational and Applied Mathematics*, 228(1):326–333, 2009.
- [17] A. Stathopoulos, L. Dimitriou, and T. Tsekeris. Fuzzy modeling approach for combined forecasting of urban traffic flow. *Computer-Aided Civil and Infrastructure Engineering*, 23(7):521–535, 2008.
- [18] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal. A review of wind power and wind speed forecasting methods with different time horizons. In *North American Power Symposium (NAPS), 2010*, pages 1–8. IEEE, 2010.
- [19] Federico Cassola and Massimiliano Burlando. Wind speed and wind energy forecast through kalman filtering of numerical weather prediction model output. *Applied Energy*, 99:154–166, 2012.
- [20] Hui Liu, Hong-Qi Tian, Chao Chen, and Yan-fei Li. A hybrid statistical method to predict wind speed and wind power. *Renewable Energy*, 35(8):1857–1861, 2010.
- [21] Lars Landberg and Simon J Watson. Short-term prediction of local wind conditions. *Boundary-Layer Meteorology*, 70(1-2):171–195, 1994.

- [22] Gong Li and Jing Sh. On comparing three artificial neural networks for wind speed forecasting. *Applied Energy*, 87(7):2313–2320, 2010.
- [23] Alexandre Costa, Antonio Crespo, Jorge Navarro, Gil Lizcano, Henrik Madsen, and Everaldo Feitosa. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, 12(6):1725–1744, 2008.
- [24] Mark F. Bielecki. Statistical characterization of errors in wind power forecasting. Master’s thesis, Northern Arizona University, Flagstaff, Arizona, USA, 2010.
- [25] Niya Chen, Zheng Qian, Xiaofeng Meng, and Ian T. Nabney. Short-term wind power forecasting using gaussian processes. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 2790–2796. AAAI Press, 2013.
- [26] David A. Dickey. Introduction to predictive modeling with examples. In *Statistics and Data Analysis, SAS Global Forum 2012*, 2012.
- [27] Michael S. Cousins, Lisa M. Shickle, and John A. Bander. An introduction to predictive modeling for disease management risk stratification. *Disease Management*, 5(3):157–167, 2002.
- [28] Thomas W. Miller. *Modeling Techniques in Predictive Analytics: Business Problems and Solutions with R*. Pearson FT Press, 2013.
- [29] Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to Linear Regression Analysis*, volume 5. Wiley, 2012.
- [30] Subhash C. Narula and John F. Wellington. The minimum sum of absolute errors regression: A state of the art survey. *International Statistical Review*, 50(3):317–326, 1982.
- [31] Francesca Pennecchi and Luca Callegaro. Between the mean and the median: the lp estimator. *Metrologia*, 43:213–219, 2006.
- [32] Gianna Agro. Maximum likelihood and lp-norm estimators. *Statistica Applicata*, 4(2), 1992.
- [33] Adele Cutler, D. Richard Cutler, and John R. Stevens. Tree-based methods. *High-Dimensional Data Analysis in Cancer Research*, pages 1–19, 2009.
- [34] Roger J. Lewis. An introduction to classification and regression tree (cart) analysis. In *Annual Meeting of the Society for Academic Emergency Medicine*, 2000.

- [35] F. Questier, R. Put, D. Coomans, B. Walczak, and Y. Vander Heyden. The use of cart and multivariate regression trees for supervised and unsupervised feature selection. *Chemometrics and Intelligent Laboratory Systems*, pages 45–54, 2005.
- [36] Anantha M. Prasad, Louis R. Iverson, and Andy Liaw. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, 9(2):181–199, 2006.
- [37] Pall Oskar Gislason, Jon Atli Benediktsson, and Johannes R. Sveinsson. Random forests for land cover classification. *Pattern Recognition Letters*, 27(4):294–300, 2006.
- [38] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [39] Ulrike Grömping. Variable importance assessment in regression: Linear regression versus random forest. *The American Statistician*, 63(9):308–319, 2009.
- [40] Gini Corrado. *Variabilità e Mutabilità (Variability and Mutability)*. 1912.
- [41] D. Molodtsov. Soft sets theory - first results. *Computers & Mathematics with Applications*, 37(4-5):19–31, 1999.
- [42] Bekir Tanay and M. Burç Kandemir. Topological structure of fuzzy soft sets. *Computers & Mathematics with Applications*, 61(10):2952–2957, 2011.
- [43] B. Ahmad and Athar Kharal. On fuzzy soft sets. *Advances in Fuzzy Systems*, 2009:6 pages.
- [44] L.A. Zadeh. Fuzzy sets. *Information And Control*, 8:338–353, 1965.
- [45] P.K. Maji, R. Biswas, and A.R. Roy. Fuzzy soft-sets. *J. of Fuzzy Math.*, 9(3):589–602, 2001.
- [46] Ritm web page. http://www.ritm.gov.tr/root/index_eng.php. Last accessed date: 2015-07-24.
- [47] C. Ensslin, B. Ernst, K. Rohrig, and F. Schlog. Online-monitoring and prediction of wind power in german transmission system operation centres. In *European Wind Energy Conference and Exhibition*, 2003.
- [48] Niya Chen, Zheng Qian, Xiaofeng Meng, and Ian T. Nabney. The integration of renewable energy and the system operation: The special regime control centre (cecre) in spain. In *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–6. IEEE, 2008.

- [49] Postgresql web page. <http://www.postgresql.org/>. Last accessed date: 2015-07-27.
- [50] T. Demirci, A. Kalaycioglu, D. Kucuk, O. Salor, M. Guder, S. Pakhuylu, T. Atalik, T. Inan, I. Cadirci, Y. Akkaya, S. Bilgen, and M. Ermis. Nationwide real-time monitoring system for electrical quantities and power quality of the electricity transmission system. *Generation, Transmission & Distribution, IET*, 5(5):540–550, 2011.
- [51] Global forecast system (gfs) web page. <http://www.emc.ncep.noaa.gov/index.php?branch=GFS>. Last accessed date: 2015-07-27.
- [52] European centre for medium-range weather forecasts (ecmwf) web page. <http://www.ecmwf.int>. Last accessed date: 2015-07-27.
- [53] Weather research and forecasting (wrf) model web page. <http://www.wrf-model.org/index.php>. Last accessed date: 2015-07-27.
- [54] Aladin model web page. <http://www.cnrm.meteo.fr/aladin>. Last accessed date: 2015-07-27.
- [55] Windsim web page. <http://www.windsim.com/>. Last accessed date: 2015-07-27.
- [56] Wasp web page. <http://www.wasp.dk/>. Last accessed date: 2015-07-27.
- [57] J.P.S. Catalao, H.M.I. Pousinho, and V.M.F. Mendes. An artificial neural network approach for short-term wind power forecasting in portugal. In *Intelligent System Applications to Power Systems, ISAP '09, 15th International Conference on*, pages 1–5. IEEE, 2009.
- [58] J.P.S. Catalao, S.J.P.S. Mariano, V.M.F. Mendes, and L.A.F.M. Ferreira. An artificial neural network approach for short-term electricity prices forecasting. In *Intelligent Systems Applications to Power Systems, ISAP 2007. International Conference on*, pages 1–6. IEEE, 2007.
- [59] Jianwu Zeng and Wei Qiao. Support vector machine-based short-term wind power forecasting. In *Power Systems Conference and Exposition (PSCE), 2011 IEEE/PES*, pages 1–8. IEEE, 2011.
- [60] Steve R. Gunn. Support vector machines for classification and regression. Technical report, University of Southampton, 05 1998.
- [61] Jolliffe I.T. *Principal Component Analysis*, volume 2. Springer, 2002.
- [62] Mehmet Baris Ozkan, Dilek Kucuk, Erman Terciyanli, Serkan Buhan, Turan Demirci, and Pinar Karagoz. A data mining-based wind power forecasting method: Results for wind power plants in turkey. In *Data Warehousing and Knowledge Discovery*, pages 268–276. Springer, 2013.

- [63] Clive W.J. Granger and Ramu Ramanathan. Improved methods of combining forecasts. *Journal of Forecasting*, 3(2):197–204, 1984.
- [64] Thais Mayumi Oshiro, Pedro Santoro Perez, and José Augusto Baranauskas. How many trees in a random forest? In *Proceedings of the 8th international conference on Machine Learning and Data Mining in Pattern Recognition*, pages 154–168. Springer, 2012.
- [65] The r project web page. <https://www.r-project.org/>. Last accessed date: 2015-08-05.