

ELECTRICITY LOAD AND PRICE FORECASTING OF TURKISH
ELECTRICITY MARKETS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

OĞUZ KALAY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
FINANCIAL MATHEMATICS

OCTOBER 2018

Approval of the thesis:

**ELECTRICITY LOAD AND PRICE FORECASTING OF TURKISH
ELECTRICITY MARKETS**

submitted by **OĞUZ KALAY** in partial fulfillment of the requirements for the degree of **Master of Science in Department of Financial Mathematics, Middle East Technical University** by,

Prof. Dr. Ömür Uğur
Director, Graduate School of **Applied Mathematics**

Prof. Dr. A. Sevtap Selçuk Kestel
Head of Department, **Financial Mathematics**

Prof. Dr. Bülent Karasözen
Supervisor, **Scientific Computing, METU**

Examining Committee Members:

Prof. Dr. Ömür Uğur
Scientific Computing, METU

Prof. Dr. Bülent Karasözen
Scientific Computing, METU

Assoc. Prof. Dr. Kasırğa Yıldırak
Actuarial Sciences, HACETTEPE UNIVERSITY

Date: _____

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

Name, Last Name: OĞUZ KALAY

Signature :

ABSTRACT

ELECTRICITY LOAD AND PRICE FORECASTING OF TURKISH ELECTRICITY MARKETS

Kalay, Oğuz

M.S., Department of Financial Mathematics

Supervisor : Prof. Dr. Bülent Karasözen

October 2018, 66 pages

Progress on the forecasting techniques of market clearing price (MCP) and system marginal price (SMP) conveys millions of dollars profit to the electricity generator corporations. Therefore, more accurate forecasting system on total electricity load (TEL) and market prices enhances profits of the market participants. System of the Turkish Electricity Market has three interdependent platforms; Day Ahead Market (DAM), Intra-Day Market (IDM) and Balancing Power Market (BPM). MCP and SMP are the most important decision variables at power distribution strategies of companies, which are determined by DAM and BPM respectively. The trading activity should be estimated a day in advance for the delivery of loads on the market operator side, optimum output and best trading strategies on the power generator company side.

Wind as a renewable energy source affects MCP. Our models contain the wind energy as a predictor, because wind energy is the most important renewable sources for energy production. The aim of this thesis is to estimate wind energy production using the numerical weather predictions (NWP) and achieving more accurate prices compared to the traditional methods used in the market. Estimation of renewable energy sources like the wind energy is more complicated than the other energy sources. They require accurate weather forecasts for accurate estimation. Because the prediction of wind is complicated, the estimation of wind energy production affects MCP forecasting values and real time MCP values so that they become more volatile.

Some regression models are developed and compared on MATLAB programme to predict MCP, wind energy production and TEL. Hourly estimations of the TEL, and predictions for average of six-hour periods MCP and wind energy production for the next day, can be used in investment strategies of the market participants, like energy consumers and producers. Moreover, the Turkish Electricity Transmission Company(TETC), which manages the balancing power market as the system operator, can improve its planning schedule and take necessary precautions about imbalances a day in advance the forecasting results of TEL.

Keywords : electricity load, market clearing price, forecasting, numerical weather prediction, regression

ÖZ

TÜRKİYE ELEKTRİK PIYASALARINDA ELEKTRİK ÜRETİMİ VE FİYAT TAHMINİ

Kalay, Oğuz

Yüksek Lisans, Finansal Matematik Bölümü

Tez Yöneticisi : Prof. Dr. Bülent Karasözen

Ekim 2018, 66 sayfa

Piyasa takas fiyatı ve sistem marjinal fiyatı tahminleme yöntemlerindeki gelişmeler elektrik üreticisi firmalara milyonlarca dolar kar sağlamaktadır. Bu nedenle elektrik yükü ve elektrik fiyatlarında yapılacak iyileştirmeler piyasa katılımcılarının karlarını yükseltir. Türkiye Elektrik Piyasası birbiriyle bağlantılı olan üç piyasadan oluşur: gün öncesi piyasası, gün içi piyasası ve dengeleme güç piyasası. Sırasıyla gün öncesi piyasada ve dengeleme güç piyasasında oluşan piyasa takas fiyatı ve sistem marjinal fiyatı, firmaların güç dağıtım stratejilerinde karar verirken kullanmış oldukları en önemli değerlerdir. Market operatörünün güç transferlerini planlanması ve elektrik üreticisi firmaların en fazla güç satışını, en iyi yatırım stratejisi ile gerçekleştirebilmeleri için yatırım hareketleri bir gün öncesinden tahmin edilmelidir.

Rüzgar, yenilenebilir enerji kaynağı olarak piyasa takas fiyatını etkilemektedir. Modelimiz rüzgar enerjisi değerlerini kestirici olarak kullanmakta, çünkü rüzgar günümüzde yenilenebilir enerji kaynakları arasında en önemli olanıdır. Bu tezin amacı, sayısal hava tahminlerini kullanarak rüzgar enerjisinden toplam elektrik üretimini tahmin ederek, geleneksel yöntemlere göre daha iyi fiyat tahminleri elde etmektir. Rüzgar gibi yenilenebilir enerji kaynakları hava koşullarına bağlı olmaları nedeniyle üretim tahmini yapabilmek diğer enerji kaynaklarından üretilen elektrikle karşılaştırıldığında daha zordur. Daha iyi elektrik üretimi tahmini yapabilmek için daha iyi hava durumu tahmini gerekir. Rüzgar enerjisi üretiminde yapılan tahminlerin yanlış payları, piyasa takas fiyatlarının tahmininden elde edilen değerleri ve piyasa takas fiyatının

kendisini daha dengesi hale getirmektedir.

Bu tezde, piyasa takas fiyatının, toplam rüzgar enerjisinden elektrik üretiminin ve toplam elektrik yükünün tahmin edilmesinde, MATLAB programı kullanılarak çeşitli regresyon modelleri geliştirilmiş ve karşılaştırılmıştır. Rüzgar enerjisinden elektrik üretiminin saatlik, piyasa takas fiyatının ve toplam elektrik yükünün ise altışar saatlik üretim ortalamaları tahminleri iyi sonuçlar vermiş olup, bu tahminler market katılımcılarının bir gün sonrası için yapacak oldukları yatırım stratejilerinde kullanılabilir. Ayrıca, toplam elektrik yükünün tahmini sayesinde, elektrik sistem operatörü ve dengeleme güç piyasasının işletim lisansına sahip olan Türkiye Elektrik İletim A.Ş., planlama programını geliştirebilir ve bir gün öncesinden olası yük dengesizliklerine karşı gerekli önlemleri alabilir.

Anahtar Kelimeler: elektrik yükü, piyasa takas fiyatı, tahmin etme, sayısal hava tahmini, regresyon

To My Family

ACKNOWLEDGMENTS

I would like to express my gratitude to my thesis advisor Prof. Dr. Bülent Karasözen for his endless help and patience. I could not finish this thesis without his encouragement and guidance.

I would like to thank Assoc. Prof. Dr. Kasırga Yıldırak for helping me to find my thesis subject and provide some resources and information about models.

I would like to thank my friends and my family for their support, patience and motivation.

I would like to thank the examining committee members, Prof. Dr. Bülent Karasözen, Prof. Dr. Ömür Uğur and Assoc. Prof. Dr. Kasırga Yıldırak for their suggestions, corrections and comments.

TABLE OF CONTENTS

ABSTRACT	vii
ÖZ	ix
ACKNOWLEDGMENTS	xiii
TABLE OF CONTENTS	xv
LIST OF FIGURES	xix
LIST OF TABLES	xxi
LIST OF ABBREVIATIONS	xxiii
CHAPTERS	
1 INTRODUCTION	1
2 TURKISH ELECTRICITY MARKET	3
2.1 Overview of the Turkish Electricity Market Structure	3
2.2 Turkish Electricity Market	4
2.2.1 Day-Ahead Market (DAM)	4
2.2.2 Intra-Day Market (IDM)	5
2.2.3 Balancing Power Market (BPM)	5
2.3 Formation	9
2.3.1 Formation of Electricity Loads	9
2.3.2 Formation of Market Clearing Price (MCP)	12

3	WIND ENERGY PRODUCTION	15
3.1	Introduction of Wind Energy	15
3.2	Wind Energy in Turkey and in the World	16
3.3	Dependant Factors of Wind Energy	20
4	FORECASTING	23
4.1	Forecasting Models	23
4.1.1	Regression Tree Model	23
4.1.2	Ensemble of Regression Trees Model	25
4.1.3	Robust Regression Model	26
4.1.4	Neural Network Model	28
4.1.5	Mathematical Model of A Single Neuron	28
4.1.6	Mathematical Model of Multiple Neuron	29
4.2	Data	31
4.2.1	Historical Market Data	31
4.2.2	Weather Data	31
4.2.3	Wind Farms Data	31
4.3	Wind Energy Production Forecasting	32
4.4	Electricity Load Forecasting	35
4.5	Market Clearing Price(MCP) Forecasting	39
5	CONCLUSION	45
	REFERENCES	47
	APPENDICES	

A	Weather Data Acquisition and Processing	49
A.1	NWP data codes	51
A.2	Expanding and adapting dataset of NWP	56
A.3	Grid appointment of wind plants	60
A.4	Grid appointment of cities	62
A.5	Weighted NWP averages of cities	62

LIST OF FIGURES

Figure 2.1	Area Control Error (ACE)	6
Figure 2.2	Frequency level of the electrical system below 50 Hz	7
Figure 2.3	Frequency level of the electrical system above 50 Hz	7
Figure 2.4	Installed capacity and total production from 1971 to 2016 [3] . . .	9
Figure 2.5	Hourly electricity consumption level in between 15 May 2017 and 22 May 2017	11
Figure 3.1	Wind velocity at 100 m. height on Turkey [17]	16
Figure 3.2	Density of wind power at 100 m. Height on Turkey [17]	17
Figure 3.3	Wind Energy Production on world [14]	18
Figure 3.4	Development of Wind Energy Production on Turkey [5]	19
Figure 3.5	Wind power output of a wind turbine with steady wind speed . . .	21
Figure 4.1	The structure of regression tree	24
Figure 4.2	Non linear autoregressive neural network model with external input	30
Figure 4.3	Wind Production Data and Model Prediction	34
Figure 4.4	Hourly electricity load averages in December 2017	36
Figure 4.5	Real Time Total electricity production and Model Prediction with neural network	38
Figure 4.6	Boxplot of the Electricity Load Predictions	39
Figure 4.7	Hourly Average MCPs of December 2017	40
Figure 4.8	Real Time MCP and Forecasted Prices	42
Figure 4.9	Plot of Real and forecasted MCP with robust fit model in first 10 days of December 2017	43

Figure A.1 Republic of Turkey surface temperature forecast map with grids resolution on 14.04.2016 at 00:00	50
---	----

LIST OF TABLES

Table 2.1	Example of an offer on DAM	5
Table 2.2	Example of a participants hourly offer on DAM	12
Table 2.3	Example of a block offer	13
Table 2.4	Example of a flexible offer	13
Table 3.1	Wind potentials of some regions in Turkey [7]	17
Table 4.1	Forecasted amounts of wind energy production for 1 December 2017	34
Table 4.2	<i>Mape</i> and <i>Mae</i> values of regression models for 1 December 2017 .	35
Table 4.3	<i>Mape</i> and <i>Mae</i> values for electricity load forecasting in December 2017	38
Table 4.4	<i>Mape</i> and <i>Mae</i> values of regression models for MCP in 1 December 2017	42

LIST OF ABBREVIATIONS

MCP	Market Clearing Price
SMP	System Marginal Price
EXIST	Energy Exchange Istanbul
TETC	Turkish Electricity Transmission Company
TWEA	Turkish Wind Energy Association
RSM	Renewables Supporting Mechanism
ECMWF	The European Centre for Medium-Range Weather Forecasts
DAM	Day Ahead Market
IDM	Intra-Day Market
BPM	Balancing Power Market
TEL	Total Electricity Load
Hz	Hertz
ACE	Area Control Error
NUP	Negative Unbalance Price
PUP	Positive Unbalance Price
NUA	Negative Unbalance Amount
PUA	Positive Unbalance Amount
TL	Turkish Lira
b.c	before Christ
<i>MAPE</i>	Mean Absolute Percent Error
<i>MAE</i>	Mean Absolute Error
<i>V</i>	Speed of Wind
<i>f</i>	Frequency
<i>MW</i>	Megawatt
<i>MWh</i>	Megawatthour
<i>kW</i>	kiloWatt
<i>m/s</i>	meter per second

CHAPTER 1

INTRODUCTION

Renewable energy sources have begun to replace of old and environmentally unfriendly energy sources such as fossil fuels, which affects the electricity trading. Since, renewable energy sources are based on production from an indirect and uncertain energy source, planning and distribution needs to proceed in a more systematic way, in order to avoid damages caused by system imbalances. In the early 2000's similar to the developed countries, preparations in Turkey are started to establish the electricity market, which took its present form.

Turkish electricity market has three main branches, which are Day Ahead Market (DAM), Intra-day Market (IDM) and Balancing Power Market (BPM). These markets are described in Chapter 2 in terms of their functioning, formation of total electricity load and market clearing price. Electricity transactions are made a day in advance, in order to schedule production plans and the electrical flow in lines. Total electricity load is an important variable that determines the price flow of the market. It shows how much electrical power should be installed to the system. This value is important for two sides, system operator and market participants. System operator needs to know total electrical power of tomorrow to be able to see possible imbalances, to calculate the total transaction volume of market and to plan the necessary electricity distribution. On the other hand, forecasting the electricity load give a chance to the market participants to maximize their profits and minimize the cost. However, the market participants need to know a reference price of the market, which is called market clearing price. It is the price of electricity for 1 MWh, occurring about intersection of supply and demand surplus in the day ahead market. Market participants can choose to sell or purchase their power in different ways. They can sell or buy their power in one of the day ahead, intra-day or balancing power markets. There are different prices at the same time period, but it is not possible to know prices of the next day, since these prices are affected from some conditions and supplies and demands of other firms. However, forecasting the market clearing price give a chance to market participant about making strategy. These strategies are also explained in Chapter 2.

The increasing share of the renewable energy sources has an important effect on the electricity market development. The high presence of renewable energy sources create one of the biggest problem for the market participants and system operators. Especially, wind, as a renewable energy source, depends on several factors. This means that estimation of the wind energy generation has lot of uncertainties and errors. Moreover,

amount of wind energy is not stable because of the volatility of wind energy. Our main aim is to find a relationship between the wind energy and the market clearing price. As a result of these inferences, history of wind energy, development of wind energy production on Turkey and World and working principles and dependent variables of the wind energy are explained in Chapter 3.

It is well known that the wind energy has an important impact on the prices and therefore forecasting of the prices. Using the numerical weather predictions we calculate total energy produced by the wind plants. It usually has a lot of errors and uncertainties. Our aim is to capture the price changes and fluctuations of wind energy and find a relationship between the wind energy production and Market Clearing Price (MCP). Moreover, we use numerical weather predictions and statistics of the previous data to predict total electricity load (TEL) and search for a non-linear relationship between our predictor values and TEL. For numerical forecasting we use some regression models of MATLAB; these are regression tree model, assembling of regression trees model, robust fit model and neural network model, and we compare the result of these models. Moreover, we evaluate the adequacy of the data we obtained from the estimates in Chapter 4. Furthermore, in the Chapter 5, a general evaluation and suggestions for future studies are given.

CHAPTER 2

TURKISH ELECTRICITY MARKET

2.1 Overview of the Turkish Electricity Market Structure

In the late 1990's energy systems became marketable in the industrialized countries due to enhancing technologies, growing demand for the energy. All developed countries established their own energy markets. Since then all countries make huge investments to improve their energy production systems. As a developing country, Turkey had a big leap in energy systems in the early 2000's. In 2 February 2001, the government introduced the electricity market law with number 4628 and led the foundations of current markets. By way of this statute new power plants were started to set, electrical generation and electrical conduction institutes' mechanisms were regularized. And then they became controllable [11]. In November 2003, temporary balancing and reconciliation regulation was issued. At that time the unbalanced energy production was controlled only by Turkish Electricity Transmission Company (TETC). In December 2003 market financial settlement center was set up. A pilot scheme were practiced between December 2003 and July 2006. After that, the ultimate scheme was set gradually. In December 2009, Day-Ahead planning mechanism was available in the market and after that time electricity market has two divisions "Day-Ahead Planning " and "Balancing Power Market". "Day-Ahead planning" announces the price of the total electricity power production and total electricity demand prediction, the day before. Balancing power market indicates the price of the forecast deviations in real time. In January 2011 Day-Ahead Market Mechanism was established instead of Day-Ahead Planning and it became the second electricity market of Turkey.

As in developed countries the aim is to establish an energy exchange market, so that the electricity buying and selling contracts can be easily exchanged. For this purpose, by the electricity market law numbered 4628 was replaced by electricity market law numbered 6446. The Energy Exchange Istanbul (EXIST) was established in 12 March 2015 with the aim to plan, establish, schedule, conduct and improve the energy market with its market operation licence[12]. EXIST aims to increase the trading volume of the market and the liquidity level. To fulfill this aim IDM was opened in the 1 July 2015 as improved and liberal Turkish electricity market. Thanks to IDM, producers can easily obtain the unbalanced power from the market in real time. Before establishment of IDM, TETC took or transfer unbalanced power from companies instead of producers. Today they have an option to cancel out their unbalanced power

themselves. This market gives a chance to producers about evaluating their energy in real time, which improve liquidity. In this way TETC is supported by balancing the electricity grid with individual balancing option for an each producing company. In 1 September 2015, EXIST took operation licenses of DAM and IDM, and TETC got operation license of BPM. Today both of the foundations are operating in the market.

2.2 Turkish Electricity Market

There are many agreements of electricity transaction in Turkey such as a reciprocal one between a customer and a producer. There will be a price which agreed for one Megawatt on an hourly basis. For the specified price of a Megawatt on an hourly basis the producer may sign bilateral agreements. Other options of the market are: selling the produced power the day before, in real time or offer to sell the power on a BPM. A customer can buy the power, if needed, by using the same procedure of the producer. The three main markets, "Day-ahead Market (DAM)", "Intra-Day Market (IDM)" and "Balancing Power Market (BPM)" will be explained in this chapter.

2.2.1 Day-Ahead Market (DAM)

DAM offers an option to the producers and customers to dispose of their energy excesses or deficiencies for a day later. Market transactions are made a day in advance under the name of the bilateral agreements, which are the long-term selling and buying energy agreements of price on a period between two companies. Generator company has an obligation to transmit that energy to customer. Similarly, customer company has a commitment about absorbing that energy from the producer. Agreements can be made for at most 60 days later.

Operations are made on an hourly basis. Every day has an equal 24 time zone, that starts from 00:00 and end on 00:00 the day after. Market offers can be made for a day later for next five days. Prices, which are used on the reconciliation, are determined for every hour on a daily basis. After reconciliation part, payment information is collected by the system operator (EXIST). Participants have to make warrant payments to give offers for the next day. Security payment notices are finished before 10:30.

Market attenders give their offers till 12:30, for every hour of the next day. Participants give their bidding amount on maximum 32 price levels between 0 TL (zero Turkish Lira) and 2000 TL (two thousand Turkish Lira). Tenders can be made as lots. 1 Lot is equal to 0,1Mwh. The negative amounts show loading energy to the system and positive amounts indicates receiving energy from that price level. Price level can not include both of the buying and the selling offer. Participants want to gain maximum so that when price is low they want to get more energy and when price is high they want to sell more energy. Therefore, they will give their offers on a ascending offer. Example of an offer illustrated in Table 2.1 . Bidding firm gives different lots for different price levels. For instance let the price for 0-1 timezone is issued as 150TL. Thus, the firm accepts to give 20MWh power to system. Total money that will be earned from that

transaction is $20 \times 150 = 3000$ TL.

Table 2.1: Example of an offer on DAM

HOUR	PRICE(TL/MWh)									
	0	49,99	50	79,99	80	109,99	110	199,99	200	2000
0-1(Lot)	600	600	400	400	0	0	-200	-200	-1000	-1000
1-2(Lot)	300	300	300	300	300	200	0	0	-200	-2000

Participants may give their bids as blocks at least 4 and at most 24 hours. When the offer is accepted, it can not be divided to hours. Load should be given without interruption. Market clearing price (MCP) is the reference price for energy-money transactions, which is determined by intersection of supply and demand curves in DAM. EXIST collects supply and demand levels for all prices using an optimization program, which determines a MCP at 14:00. Payments of the transmits, that will be taken tomorrow are made immediately today. In addition, the objections about the offers and the payments are made between 13:50 and 14:00 and EXIST changes them if the objection is accepted.

2.2.2 Intra-Day Market (IDM)

IDM was established in 1 July 2015 with the aim of becoming a liberal market, and of the fact that participants take a well-balanced and active positions on marketplace. In addition to the DAM and BPM, IDM will give an opportunity to its participants about trading almost real-time and balancing their portfolios in the short-term. IDM has a critical role between DAM and BPM. Market participants, who do not take a position in DAM or do not sell all the power in it's hands, have an option to sell its power on the IDM in real time. On the opposite way the customer who need power and do not take its position on DAM can absorb the power that needed in IDM. In the DAM, there will be maximum 36 and minimum 12 hours of difference between the time that entering offers and actual electrical conduction. On the other hand in IDM participants can give an offer 90 minutes before transmission. Therefore, IDM plays an important role of diminishing the grid unbalances and losses of the market attenders if there exists some unexpected situations such as power plant breakdowns, unpredictable weather conditions on the renewable energy sources and imbalances on the electrical consumption.

2.2.3 Balancing Power Market (BPM)

The basis of the BPM was founded on the early 2000 with the purpose of real time electricity unbalances. Its main aim is to provide the equality between supply and demand. Manager of the market is TETC. This company has responsibility to all generators and users to secure the power transmissions without loss. If there is a vulnerability in the system, TETC must take an action to compensate the difference. Real time balancing

is formed with ancillary services and BPM. BPM provides the energy that the system need for the real time balancing. Power resources of the market must be ready to supply electricity to grid in maximum 15 minutes. Ancillary services provides frequencies of the system and demand checks. If there is a problem in the frequencies or the company using the power, the BPM operator takes an action and avoid the system breakdowns. BPM transactions are made on hourly basis. All of the electricity generator participants must report their disposable capacity, which is the maximum energy that can be produced by a power plant on an hourly basis. Companies are trying to produce as same as their disposable capacity but generally generate less than that level due to the fact that energy production is a volatile system and during production sessions there may be a system failures. Furthermore bids of the companies are not offset all of their production and they will work under their capacities.

Although DAM ensures a balanced market, there can be improbable events in the market in reality. For example, as shown in Figure 2.1 the error in the area control is desired when it is kept close to zero and it gives the value of instant electricity unbalance. If it is not close to zero, the value is determined as wrong. For instance, the possible power cuts caused by the frequency deviations and the power plant shutting down are some of the main problems in the market. In these situations, TETC uses the offers given in BPM to take control of the market.

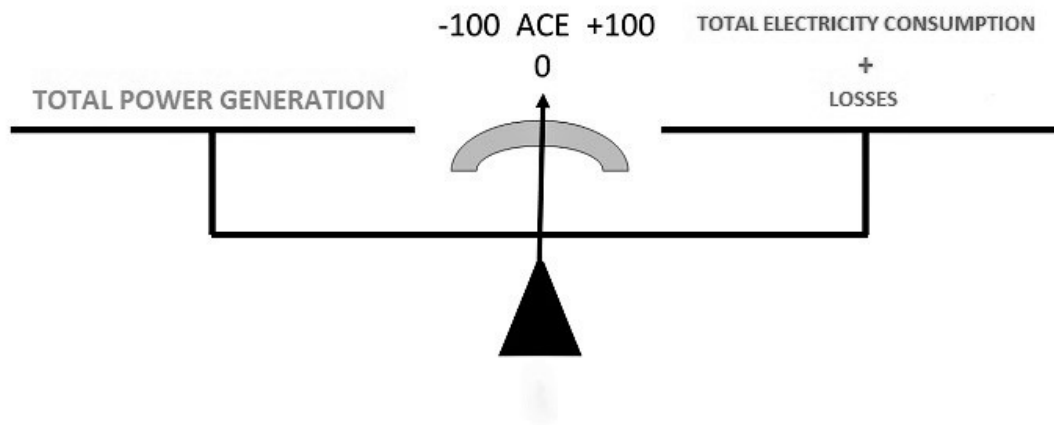


Figure 2.1: Area Control Error (ACE)

The system frequency should be held in 50 Hz in the Turkish electricity system as illustrated in Figures 2.2 and 2.3. While one side of the balance has the total power generated by power plants, the other side includes total electricity consumption and losses in transmission. The case of the systems frequency below 50 Hz means that the production is less than the consumption in Figure 2.2. To fix this problem the power generation should be increased or the consumption should be decreased. The case of the systems frequency above 50 Hz means in Figure 2.3 production is more than consumption. The only option to resolve this problem is to reduce generation.

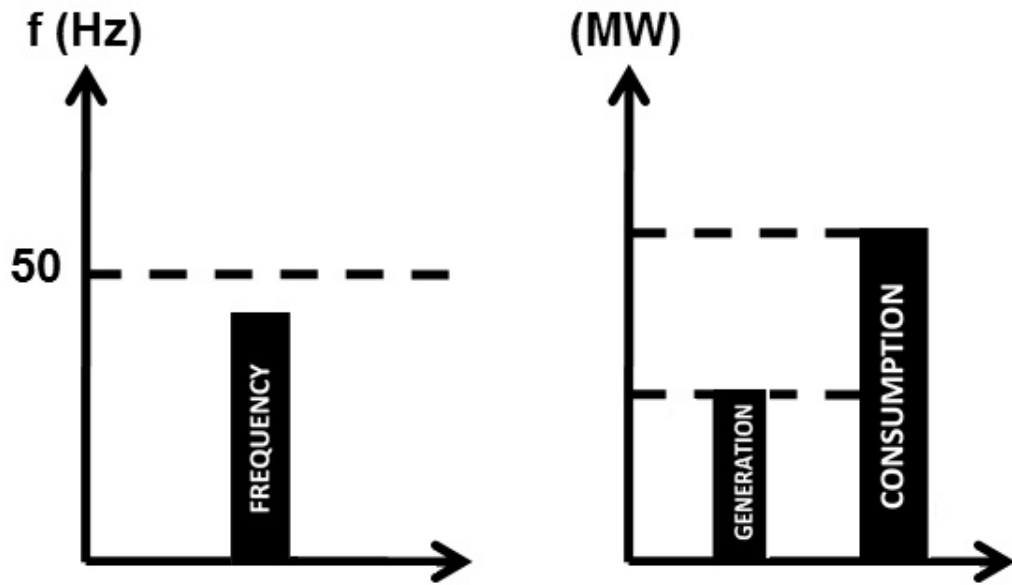


Figure 2.2: Frequency level of the electrical system below 50 Hz

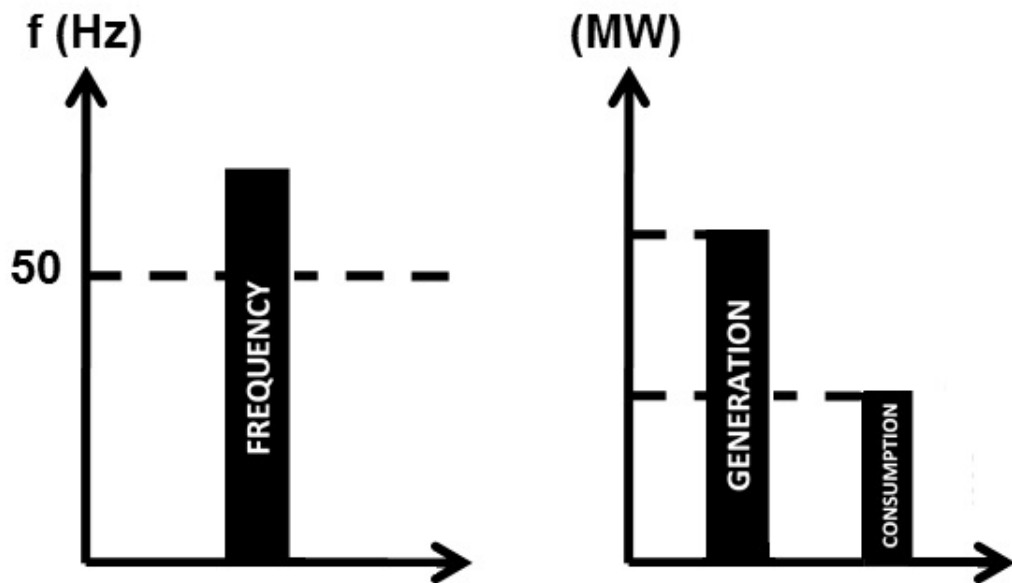


Figure 2.3: Frequency level of the electrical system above 50 Hz

BPM starts after the end of DAM. Market participants report their conclusive daily production schedules for the next day and give load pick-up or load throw-off offers till 16:00. All of the offers are gathered on the system and ordered from lower price to higher price for 24 hours independently. After 17:00, the system manager (TETC) starts to foreseen the energy imbalances and gives proper load pick-up and load throw-off instructions. Load pick-ups and load throw offs are valued as negative integers and

positive integers, respectively. All of the instructions are gathered together with codes 0 (zero), 1 (one) or 2 (two). Code 1 (one) and Code 2 (two) instructions are given for the system constraints and creating secondary reserve for the system (for ancillary services), respectively. 0 (Zero) code is given for the total of load pick-ups and load throw-offs that are given as code 1 and code 2. If system direction shows energy gap, operator puts into use of the lowest priced load pick-up instruction and if system direction shows hyperenergy condition, operator takes the opposite control mechanism. It neutralises supply and demand surpluses with one another and pick-up or throw-off excessive amount. Because system operator wants to make optimal loose from that transactions in order to represent the "loose" to market participants.

System operator penalizes market participants causing the energy unbalance. In Turkish electricity system, dual pricing method is used for penalizing. Dual pricing is an encouraging system for participants to minimise their energy unbalances. Price that will be paid by participant is determined dependently to system direction. An unbalance is called negative unbalance if it feeds the total grid unbalance and increases the total unbalance. If unbalance helps to reduce total grid unbalance, and to decrease the value of unbalance, then it is named as positive unbalance. Positive unbalances and negative unbalances are priced differently in dual pricing system.

Negative imbalances are valued with negative unbalance price (NUP), which are determined as multiplication of negative unbalance amount (NUA), maximum of MCP and SMP for that hour and factor of 1,03.

$$NUP = NUA \times \max(MCP, SMP) \times 1.03. \quad (2.1)$$

Positive unbalances are priced similar as the other one with two difference, the factor that is used for this pricing is 0,97 and the price is calculated with minimum of MCP and SMP. The following equation shows calculation of positive unbalance price (PUP) with positive unbalance amount (PUA)

$$PUP = PUA \times \min(MCP, SMP) \times 0.97. \quad (2.2)$$

Therefore, it can be said that participants do not want to make energy unbalance on purpose. Because participants' unbalance may cause to lose money from their maximum profits or to pay extra money than their actual power needs. In any case of MCP or SMP is bigger than the other, participant is a possible money loser. SMP is the energy transaction price that determined on the BPM and is calculated with total instruction amounts in market. Total instruction amount is adjusted by taking the difference between keep-on and throw off instructions and shows the system direction. This amount is supplied by grid operator with the ordered price offers. Price of the last evaluated offer is SMP. It will be particularly explained in the next section.

BPM is a risky market and it is not preferred as trading by market participants. However there is an obligation about contribution to that market. All of the power plants and consumers, who can keep-on or throw-off at least 10MW power in 15 minutes, have

an obligation to participate this market. Market participants schedule their production or consuming plans and notify these information with market system.

2.3 Formation

This section describes how values of TEL and MCP are formed.

2.3.1 Formation of Electricity Loads

Electricity is the instantaneous power that should be consumed or stored on time it is generated. Since cost of storing electrical power is very high, market participants wants to schedule their transactions and market operator wants to codify all of the electricity loads. In this way there will be a chance to plan loads on the distribution networks, and avoids excessive loads on these networks. In Turkish electricity market, TETC is responsible from all of the electricity transmissions larger than 1 MW. Each year, electricity production and usage level of Turkey has been enhancing each passing year. Figure 2.4 shows the development of Turkish electricity systems installed capacity and production levels from 1971 to 2016.

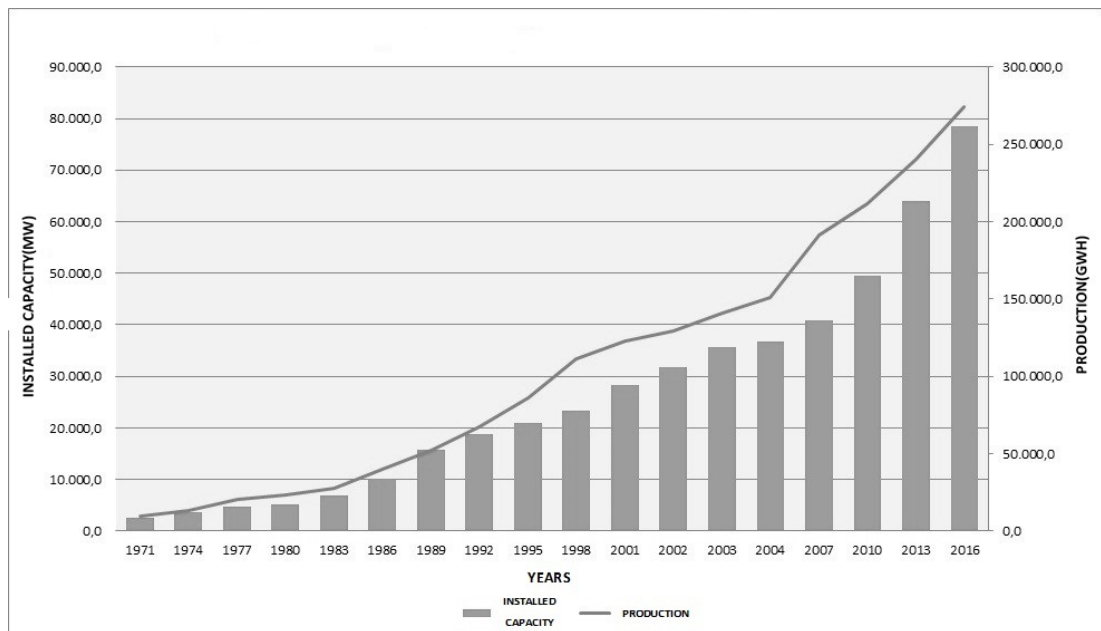


Figure 2.4: Installed capacity and total production from 1971 to 2016 [3]

Electricity can be produced from some different sources, which are separated as non-renewable and renewable sources. Non-renewable sources are fossil fuels such as fuel oil, lignite, bituminous coal, asphaltite coal and natural gas. On the other hand, renewable energy sources are waste heat, biomass, geothermic, hydro power with dam, stream, wind and solar power. All of power plants that use non-renewable energy

sources can produce energy dependent to its source limits. It can be scheduled, programmed and their production plans are given with exact predictions. Some of the renewable energy sources (waste heat, biomass) can be similarly programmed, since their levels can be obtained and they are certain. However, geothermic, hydro power with dam, stream, wind and solar power energy sources are dependent to natural events and power plants, regulating their electricity production programmes directly connected to their sources.

In early 2000's electrical energy production with non-renewable energy sources using started to decrease with their harmful effects on the environment and their increasing prices. They give their places to renewable energy sources because these sources are environment friendly, free of charge and government promotes the production from these plants. Therefore, new power plants are opened generally to generate energy from renewable energy sources. Nowadays, wind as a renewable energy source, is the most popular energy source in Turkey. Wind energy is explained in detail in the next chapter.

Electricity is consumed in various ways such as industry, agricultural irrigation, illumination, communication, service sector and personal use. Therefore, there are many factors affecting the level of electricity load. The main factor that effects this level is time. Electricity usage changes in different hours of the day and different days of the week. It diminishes when the time people sleep and not at work. On the contrary, when people are at work, level is on the top because most of the factories work flow continues parallel to people's work and electricity. Similarly, service sector and communication services workflows are peak when time is hour of work. Furthermore, the electricity usage levels on weekends and the days the country having public holidays such as their communities lifes, establishment anniversaries or religious beliefs are on rock bottom. For these reasons, when predicting the electricity load level it should be considered that there will be different electricity load levels on day time and night time. In addition, weekdays, weekends and public holidays should be also considered while calculating the electricity load level. Figure 2.5 shows total electricity consumption level of one week from 15 May 2017 to 22 May 2017. Decreasing on consumption level is starts from 19 May since it is public holiday on Turkey. Moreover the level is also at low levels in the following two days, since it is a weekend.

The second factor that effects electricity usage is weather conditions. When the temperature is high in summer, people wants to open air conditioners and when temperature shows cold degrees people want to open infrared heaters, or similarly open their air conditioners. Furthermore, when the weather cold, heating a water needs more energy than normal temperature degrees. In the same way, when the weather is not on its normal level (in between 15 – 25 centigrade degrees), more electricity is used by people to continue their works. Agricultural irrigation systems needs different energy and water levels in different conditions. So that, some of the weather analysis data should be used, while calculating electricity load.

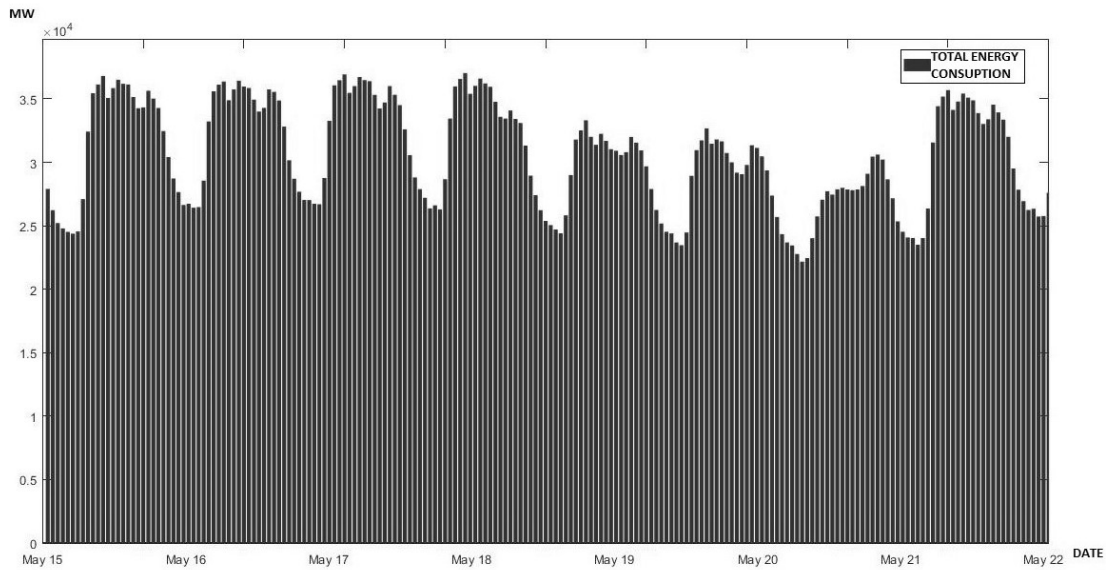


Figure 2.5: Hourly electricity consumption level in between 15 May 2017 and 22 May 2017

The third factor of the electricity level is price. Out of the personal use of electricity, the price level of the industrial usage of electricity has been effected from price of the electricity. When price is low, people want to use more so that the production amount is high. On the other hand when price is high people don't want to use electricity and decrease their production amounts. As it is explained before, in DAM and IDM participants give their offers for different price levels. On the consumers side, while low price levels, participants want to get more power. However on the high price levels, they decrease their needs and use less energy then their needs. On the contrary, producers want to generate more energy when price is high and generate less energy or do not want to produce energy (since cost of the electricity is low and then the price). When the price is low we can say that producers and consumers are attracted by the different price levels according to their perspectives. Thus, the total electricity level is affected from these conditions.

System operator should forecast total electricity load with minimum error in order to maximize its profit. For instance, let the total electricity need for an hour is forecasted as 40000MW and this amount of electricity is scheduled to produce with the aim of compensating the demand with MCP, 150 TL/MW. However, it turns out that the actual need is realized as 39000MW for that hour and there is a 1000MW open amount that should be consumed or stored by some participants. This open amount is purchased by MCP and if it is not used there is total 150000TL, that is overpayment for excessive amount of energy. This energy is sold in BPM, but that may not meet the total loss for an hour, if SMP in BPM is not equal to MCP.

Therefore, electricity load should be forecasted before closing market, which is DAM. As it is said before, DAM finishes its transactions a day in advance. In this way, hourly electricity load should be predicted 24 hours before all energy transfers are done.

2.3.2 Formation of Market Clearing Price (MCP)

DAM is managed by EXIST collecting hourly offers for the next day from market participants and adding power demand and supplying to their sides. In this way, supply and demand curves are formed and MCP is determined on the intersection of these supply and demand curves. Hourly buy and sell amount is named as Equilibrium Quantity (EQ). Trading volume is found by multiplication of EQ and MCP.

The market participants in DAM can give different offers on portfolio basis as hourly, block or flexible by considering their bilateral agreements. For instance, a participant schedules its generation plan for the next day in between 12:00 and 13:00 as 400 megawatt (MW) and let its bilateral agreement for that hour is 300 MW and its consuming level is 200 MW. Then, its portfolio has 100 MW open amount and this participant may obtain that amount from DAM.

Offers are given as price and amount duals. Amount notices are given as positive and negative integers per lots. 1 lot is equal to 0,1 megawatt per hour (MWh). Negative amounts shows power flow from participant to market, on the other hand positive amounts show the behaviour vice-versa.

As in table 2.2, a participant gives its offers for different price levels of the next day in between 12:00 and 13:00. While prices are increasing, lots should be ordered as equal or less than the previous one. In this example, lots are truly given. Lets all of the offer are given and the system is closed for the next day, and MCP is determined hourly. MCP is founded as 130 TL/MWh. Since that participant is not give an amount about this 130 TL/MWh price level, the gap amount is found by interpolation method between participant's offers. In this way participant's offer for price level is -300 lot thanks to interpolation method. -300 lot means a 30 MW electricity power is sold to market for the next day of an hour in between 12:00 and 13:00. For that transaction participant takes 130 TL/MWh for its 1 MW and since, it gives 30 MW to system, it will take $30 \times 130 = 3900TL$ from Market Operator (EXIST).

Table 2.2: Example of a participants hourly offer on DAM

HOUR	PRICE (TL/MWh)								
	0	10	50	80	100	110	150	200	2000
12-13(Lot)	2000	1000	400	200	0	-200	-400	-1000	-2000

As can be seen, hourly offer is a function that takes all price levels to a power amount that is bought or sold. In other words, let p and q be price and amount, respectively. The function f can be identified as $q = f(p)$ where $f : [0, 2000] \rightarrow R$

Block offers, which are accepted by the market operator, can be given as timeliness from 3 to 24 hours and contain price, amount and time zone. They can not be divided and their maximum size is 50 in a day.

Example of a block offer can be shown in Table 2.3. Market participant commits hourly 300 MW of electrical power with a fixed price 130TL/MWh for 5 hours; starting at 17:00 and finalizing at 22:00. If it is accepted, market participant's offer will be

processed as dual priced. For this time zone, if the offer price is lower than the average MCP of that time zone, then participant will take the price it is offered. On the other hand, if average MCP of that time zone is bigger than the offered price, then participant will take its allowance from the average price. Average price of that time zone is 140TL/MWh, then participant will take totally $300 \times 5 \times 140 = 210000$ TL. Thus, it can be said that participant has a chance to get extra money than his offer, if the offer is the block type. The power plants which can not change its production amount easily prefer giving their offers as block so that they are not affected from production fluctuations caused by hourly offers. The advantages of the block offers are that the plants have a fixed amount to produce and do not need to reduce their working programme.

Table 2.3: Example of a block offer

Timezone	Price (TL/MWh)	Amount (Lot)
17:00-22:00	130	-3000

The flexible offers can be given for any hour of the next day as the power selling to the system. They can not be divided as the block offers and their maximum size is 10 a day. They should be accepted or refused completely. If this offer's price is lower than the maximum MCP of the next day, it is accepted and this offer is fixed for an hour of the next day. Thus, it is priced according to that which one of MCP and the offer price is greater.

Example of a flexible offer is shown in Table 2.4. Participant offers to give 20 MW of power from price 170TL/MWh. This offer is accepted and arranged to be taken on the system between 17:00 and 18:00 hours, and MCP of that hour is 150 TL/MWh. Then participant will take $20 \times 170 = 3400$ TL from that transaction. When participants give flexible offers, they need to adjust themselves after the time DAM is closed. A producer, who has a production reserve in a day for a nonspecific hour, can get maximum profit with flexible offers.

Table 2.4: Example of a flexible offer

Time	Price (TL/MWh)	Amount (Lot)
17:00	170	-200

As it is explained before, participants can get or supply electricity power by bilateral agreements between themselves. These agreements consist of 24 hours power flow of the next day. While picking-up powers are written as positive lots, throw-off powers are entered in the system as negative lots. Both sides of the agreement enter amounts as opposite directions. Agreements of the sides are accepted by Market Operator (EXIST) only if they give the same values on the system. These offers can be given for maximum 60 days later.

In conclusion, each offer consists of quantity-price pairs (QPPs). Each pair represents either a supply or a demand of a electrical power. Demand and supply quantities are given in positive and negative values, respectively. Let p denote the price and q denote the quantity of QPP. In any offer the pair (p, q) shows that the participant wants to pay at most p per MWh with an amount of q for purchasing or want to take at least p per

MWh with an amount of q for selling.

EXIST collects all bid from market participants with different types for the next day and decides on the accepted quantities for each offer. It determines matching quantities for each offer with the purpose of daily market surplus is maximized. During the maximization part, operator needs to balance supply and demand for each period. Each hourly offer is a piecewise linear function formed as quantity-price duals.

CHAPTER 3

WIND ENERGY PRODUCTION

3.1 Introduction of Wind Energy

In between 1950 and 1973, oil prices were stable and most of the energy. However, after the oil crisis in 1973, the prices of the fossil fuels went sky-high. For instance, while the price was \$2,5, it increased over \$30 in 1980. Therefore, countries had to find some alternative energy sources. Since fossil fuels are expensive and cause to a environmental pollution, renewable energy sources such as wind energy became an alternative. The wind is free to use and renewable and natural energy source. Its operational costs are very low since its turbines do not need continual maintenance. Developments in the wind turbine technologies diminish electrical production costs and most of the countries on world use this power plants to generate electricity. For many years wind turbines worked with small installed capacities and countries determined to produce modern wind turbines that produce electricity as megawatts in an hour as a consequence of the crisis. Installed capacity of wind turbines is continually increasing. In Turkey, installed capacity of a wind turbine was 500kW for an hour at the first use of wind technology, but today it is 3MW for an hour. Hence, field of the power plants can be used effectively so that total generated electrical power has been increasing with the same number of turbines.

In ancient times, the Egyptians around 3000s b.c. and the Persians field around 200s b.c. used the wind energy for agricultural irrigation and grain milling. Wind wheel became widespread around Mesopotamia in 11th century. It was taken from Anatolian to Europe with crusades in 13th century. Around 17th century it was used for vessels, wind mills in Holland and it affected the development of the country. [15, 19] The first electricity production was made on Denmark in 1890 and the first modern wind turbine was founded by Palmer C. Putnam on America in 1941. In between 1950 and 2000 England, Germany, Denmark, U.S.A. and Canada made important studies in this issue and wind turbines took today's shape [15, 22].

Developments in the wind turbine technologies diminish electrical production costs and most of the countries on world use this power plants to generate electricity. After the oil crisis, as the other renewable energy sources, wind energy researches were expedited. For many years wind turbines worked with small installed capacities and countries determined to produce modern wind turbines that produce electricity as megawatts

in an hour as a consequence of the crisis. Installed capacity of wind turbines is continually increasing. In Turkey, installed capacity of a wind turbine was 500 kW for an hour at the first use of wind technology, but today it is 3 MW for an hour. Hence, field of the power plants can be used effectively so that total generated electrical power has been increasing with the same number of turbines.

3.2 Wind Energy in Turkey and in the World

Turkey depends on external energy resources. Turkey produces huge amount of the electricity with natural gas which is bought from its neighboring countries. Sometimes, there is an arrival problem because of the foreign relations such as political, technical or economical. So that a supply deficit problem happens and the price of the natural gas is so high. Therefore, it slows down the production amount of electricity and price of the electricity will be so high. As a result of this there is possible fluctuations in the electricity production because of the external dependance to a raw material.

Wind power, as a renewable and free energy, is a great opportunity for power generation and Turkey does not benefit from its advantages well enough. In the last five years, Turkey has a great development in wind power systems and rich windy places since it surrounded on three sides by the sea.

As seen in Figure 3.1 north of Aegean region, Marmara region of the Turkey shows red and orange type of wind speeds, i.e., average wind speed in this region is higher than 8m/s at 100m. height. It is known that the wind turbines start to produce electricity between 3 – 4 m/s and 10 m/s which is the speed the wind turbines reach its optimal yield. Hence, Aegean and Marmara regions of Turkey are so convenient to build wind turbines to produce more electricity.

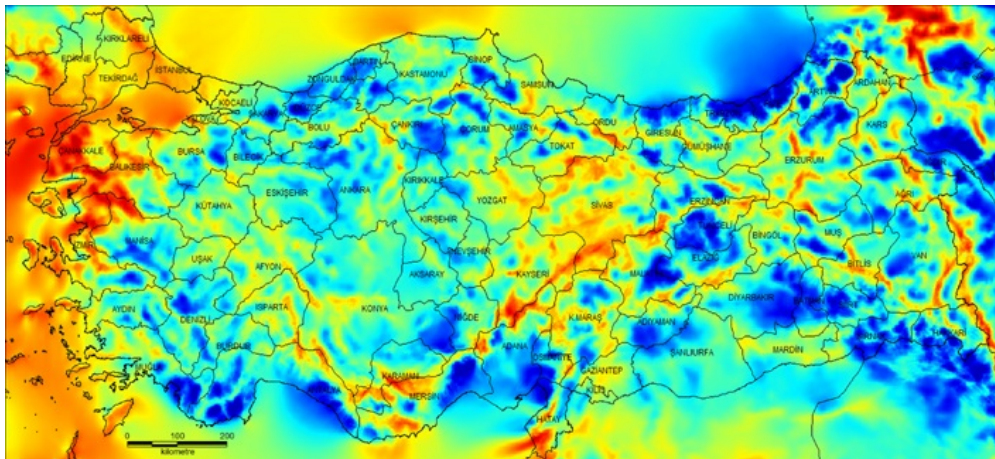


Figure 3.1: Wind velocity at 100 m. height on Turkey [17]

Density of a wind is the air particles counted per square meter. The wind speed can be used effectively in the places having higher density to produce more electricity. In other words, it can be said basically that the higher density helps the wind to push

blades of the wind turbines with more power. The wind density varies in proportion to the change in air temperature and humidity. As seen in the Figure 3.2 average density of the northern section of Aegean region is more than 700W/m^2 . This result supports us in the establishment of wind farms in the Aegean region. Moreover, other regions of Turkey can be examined to setup wind farms to meet the regional electrical power needs.

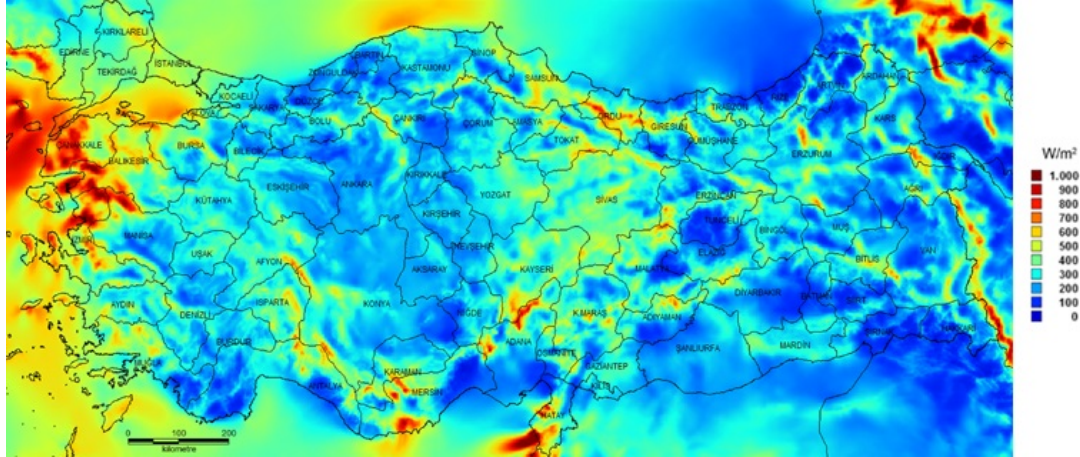


Figure 3.2: Density of wind power at 100 m. Height on Turkey [17]

Table 3.1 in below shows some places and their wind speeds and wind densities and these fields have big potentials about wind power. Turkey has a lot of wind farms established and also, it is planned to set new wind turbines in these fields [7].

Table 3.1: Wind potentials of some regions in Turkey [7]

	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec	avg
Akhisar	5.2	6.2	2.7	2.7	4.3	5.4	8.7	9.0	5.0	7.1	5.6	8.3	5.85
Bababurnu	6.0	6.0	5.9	6.0	3.0	5.3	6.2	6.4	4.5	5.7	5.9	8.2	5.75
Bandırma	-	-	8.1	6.6	6.5	6.8	8.2	9.6	6.1	8.4	8.1	12.0	8.04
Belen	4.8	4.9	5.4	6.0	5.0	9.3	11.5	10.3	8.2	5.3	5.6	5.8	7.41
Datca	4.6	5.1	4.7	4.5	6.5	5.6	8.0	9.9	6.2	5.5	5.3	5.4	5.90
Foca	5.6	5.5	5.6	5.3	4.8	5.5	5.8	6.3	4.7	5.8	6.0	7.4	5.70
Gelibolu	7.4	6.7	8.3	6.0	5.2	5.8	5.7	7.6	5.7	7.4	7.0	9.1	6.80
Gokceada	7.7	6.9	8.1	6.7	5.3	5.9	6.2	7.8	5.7	7.5	7.6	10.8	7.20
Kocadag	8.0	7.7	8.0	7.0	7.0	8.6	8.6	6.5	7.2	5.3	7.1	-	7.36
Sinop	3.9	4.2	4.9	4.7	4.5	4.8	4.4	4.1	-	-	5.3	5.2	4.60

As seen in Figure 3.3, the cumulative installed capacity of wind energy between 2011 and 2017 has been increasing from 23.900MW to 539.581MW and cumulative capacity between these years shows that China is the first country at all countries on world and has 35% of total capacity. The success of China also can be seen by looking the new installed capacity in 2017 by establishing new wind turbines. Their total new capacity is 37% of the total new establishments on world. Turkey is also in the list with 1% of total created new wind energy installed capacity on the world with totally new 766, 05

MW of power capacity was setup. The Figure 3.3 shows the need in wind power energy of the world, which is one of the most important renewable energy source [14].

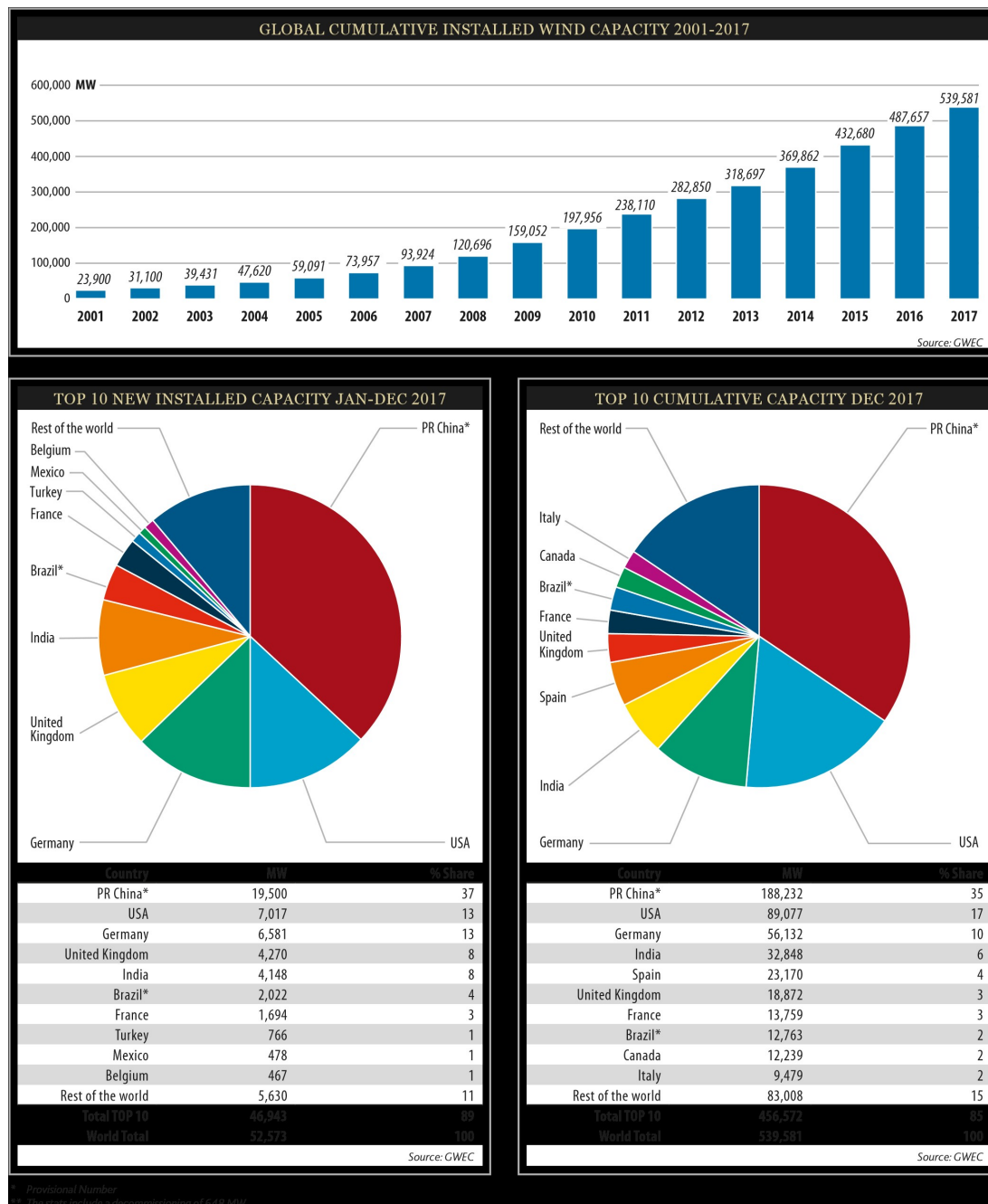


Figure 3.3: Wind Energy Production on world [14]

Turkey as a developing country determined to enhance its renewable energy capacity too late. It's total installed capacity in 2007 was only 146,30 MW. Installed capacity of total electricity production in Turkey was over 40000MW as shown in Figure 2.4. Hence, wind energy installed capacity was only about 0,4% of total capacity in 2007.

This situation changed after 2007 such that Turkey had total 6872,10MW installed capacity of all 81.553,02 MW total capacity at the end of 2017, i.e., the wind energy is about 8.4% of total capacity. It is seen that Turkey develop its point of view after 2007 and started to interest in renewable energy sources [5].

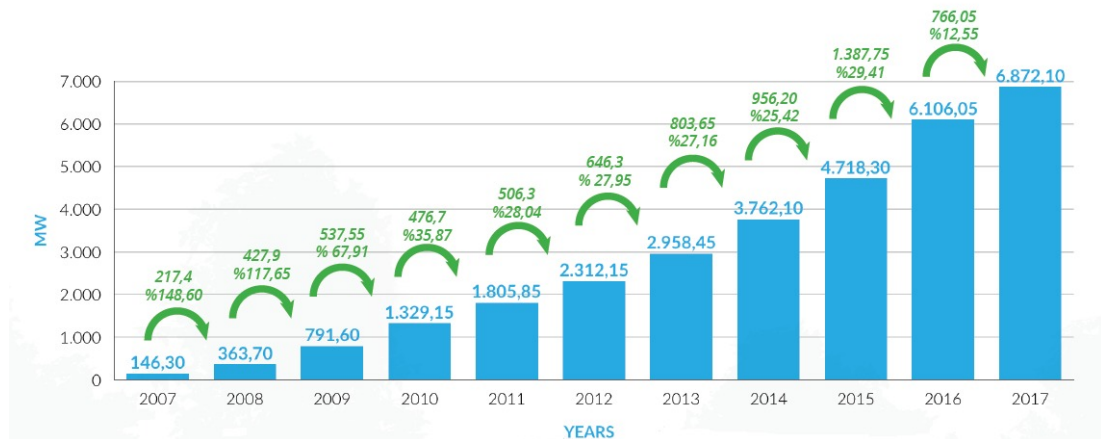


Figure 3.4: Development of Wind Energy Production on Turkey [5]

Furthermore, Turkey has some encouragements about producing electrical energy from renewable energy sources. All of this encouragements are named as Renewables Supporting Mechanism (RSM). These supports were guaranteed by the law 5346 in 2005. Their purposes are disseminating the use of renewable energy resources for electric energy production, making the economics of these resources in a reliable, economic and quality way, increasing resource diversity, reducing greenhouse gas emissions, utilizing waste, protecting the environment and improving the manufacturing sector. This mechanism ensures that the state receives the electricity generated from renewable energy sources. Regardless of the amount in proportion to the production potential, the produced electricity is taken by the government from determined price. It is optional for the power plants to participate in this mechanism. They can make use of the advantages of this mechanism for 10 years after they are incorporated [1, 2]. The main advantage for power plants is that they are exempted from the imbalance cost when they benefits from this mechanism. Thus, they do not need to forecast their exact production plans because of purchase guarantee.

Since forecasting electrical generation from the wind energy is more complex than the other production types, wind power plants choose to participate in the scope of this mechanism. Thus, they offer less production than they will produce tomorrow and they sell their production surplus of tomorrow under the price of mechanism.

On the other hand, there is an extra imbalance in the system due to the fact that renewable energy power plants do not care about imbalances. This is reflected as an additional cost to electricity prices because of purchase guarantees. This additional cost is named as RSM cost. It is reflected in consumers but electricity companies must estimate this cost. If their estimates are low, they lose money. If their estimates are high, they can not give competitive offers during sales.

3.3 Dependant Factors of Wind Energy

After the solar energy has fallen into earth, pressure temperature differences, with the formation of different warming zones are main source of wind. From this point of view, wind energy that can be considered as a form of solar energy contain 2% energy coming from the sun as a kinetic energy form. The kinetic energy of this moving mass is calculated as follows:

$$P = \frac{1}{2}mV^2, \quad (3.1)$$

where m is moving mass of air flow, V is speed of wind. Moreover, m can be calculated by

$$m = \rho AV, \quad (3.2)$$

where A is the perpendicular cross-sectional area to the wind direction, ρ is density of air. Therefore, substituting Equation (3.2) into (3.1)

$$P = \frac{1}{2}\rho AV^3. \quad (3.3)$$

As in Equation (3.3), the wind energy has proportional to the third degree of wind speed. Thus, the small changes in the speed can affect dramatically the wind energy.

Let the initial entrance speed and the exit speed of the wind turbine be V_i and V_o , respectively. The energy produced in the wind turbine is also proportional to the difference between V_i and V_o .

Let V_{avg} be $\frac{V_i+V_o}{2}$. According to this, the moving mass of air flow over turbine blades is calculated as

$$m_w = \rho A_T V_{avg}, \quad (3.4)$$

where A_T is the perpendicular cross-sectional areas of the wind turbine blades to the wind direction.

Therefore, the energy produced in the wind turbine is calculated using Equation (3.3) as

$$P_o = \frac{1}{2}m_w(V_i^2 - V_o^2) = \left(\frac{1}{2}\right)(\rho A_T V_{avg})(V_i^2 - V_o^2) = \frac{1}{2}(\rho A_T V_i^3) \frac{(1 + \frac{V_o}{V_i})(1 - \frac{V_o}{V_i})^2}{2}. \quad (3.5)$$

Let $C_P = \frac{(1 + \frac{V_o}{V_i})(1 - \frac{V_o}{V_i})^2}{2}$ then Equation (3.5) can be rewritten as

$$P_o = \frac{1}{2}(\rho A_T V_i^3) C_P. \quad (3.6)$$

According to the law of Betz, the maximum power coefficient C_p is 0.593 at the ratio $\frac{V_o}{V_i} = \frac{1}{3}$.

In other words, the law of betz says that the maximum production yield of a wind turbine is 59,3% [8, 16].

Therefore, Equation (3.6) can be rewritten as

$$P_o = \frac{1}{2}(\rho A_T V_i^3) 0.593. \quad (3.7)$$

The C_P value is around 40% for an ideal wind turbine that is well designed using today's technology. Therefore, we can say that all of the wind turbines have different power coefficient and so that different productivity levels with the variety of turbines, distinguishing with sizes and types.

Figure 3.5 shows the working principles of the wind turbines with respect to wind speed. Wind turbines can start to produce electrical energy only at a certain wind speed called "cut-in". Below "cut-in" speed, the system completely stops. At a nominal wind speed for each wind turbine model, the power from the system reaches the greatest value called "nominal power". Even if the wind speed is higher than nominal wind speed of the wind turbine, the system can produce nominal power. Automatic shutdown of the wind turbines after a certain wind speed called as "cut-out" is ensured so that the system is not damaged. Modern wind turbines have "cut-in" speeds of 3 – 4 m/s, "nominal" speeds of 11 – 15 m/s and "cut-out" speeds of 25 – 30 m/s.

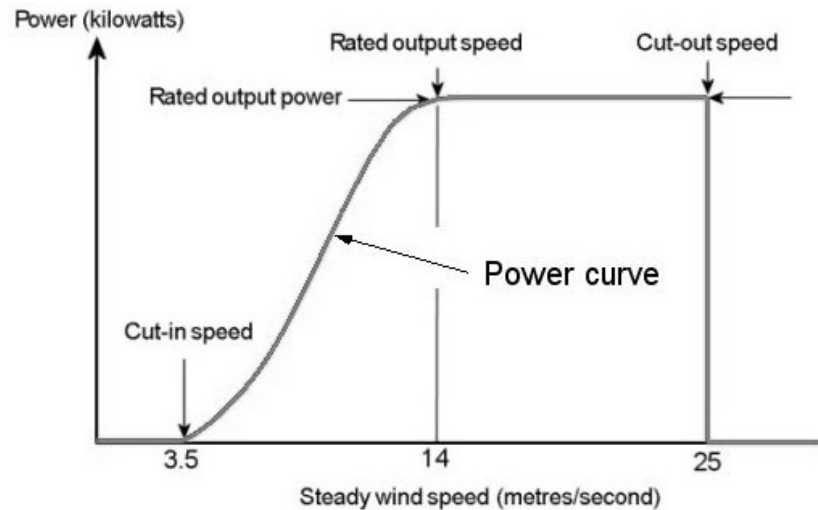


Figure 3.5: Wind power output of a wind turbine with steady wind speed

CHAPTER 4

FORECASTING

In this Chapter we introduce regression tree model, assembling of regression trees model, robust fit model and the artificial neural network are described for forecasting of wind energy production, load, and MCP. Customized tools and functions of MATLAB are used for the data, which has complex non-linear behaviour like the wind data. To control market imbalances and profits/losses better forecasts are needed from the perspectives of market operator and participants. After that, the models are evaluated and compared.

4.1 Forecasting Models

In this section basic structures of the models that we use are explained. The models that we use are regression tree model, ensembling of regression trees model, robustfit model and neural network model as mentioned.

4.1.1 Regression Tree Model

A decision tree is a machine learning algorithm that distribute the data into subsets. The distributing process starts with binary split and it ends when no further splits can be made. The statistical processes behind the classification and regression in tree analysis is so similar, but distinguishing these two thing is so important. For a response variable which has classes, often binary, we want to organize the dataset into groups by that variable(classification). When response variable is numeric or continuous we want to use that data to predict the target, and we will use regression trees in this situation. In other words, in the case where response variable does not have classes, a regression model is fit to each of the independent variables, determined these variables as nodes where their addition diminishes error.

An illustration of this process was shown in Figure 4.1. We want our explanation is so simple, and our given factors, $x_1, x_2, x_3, \dots, x_n$ in the domain X so that we want to predict the outcome of interest, Y . In figure 4.1 the tree is the domain of all factors related to Y in descending order of importance. In traditional regression models, it can be linear or polynomial, we develop a single model to represent the whole data

set. In this way, regression tree is an alternative method to this. Our data is partitioned into smaller subsets where variable interactions are more determined. Regression tree analysis uses this repeated partitioning to create a tree where each node T represents a cell of the distribution. Each cell is attached to node with a simple model, which is applied only for that cell. Where as we move down the nodes, of the tree we are conditioning on a certain variable. The final node is named as a leaf. Moreover, each of the terminal nodes A, B and C, suggest that after that split, more divisions of the data does not develop the regression about predicting Y .

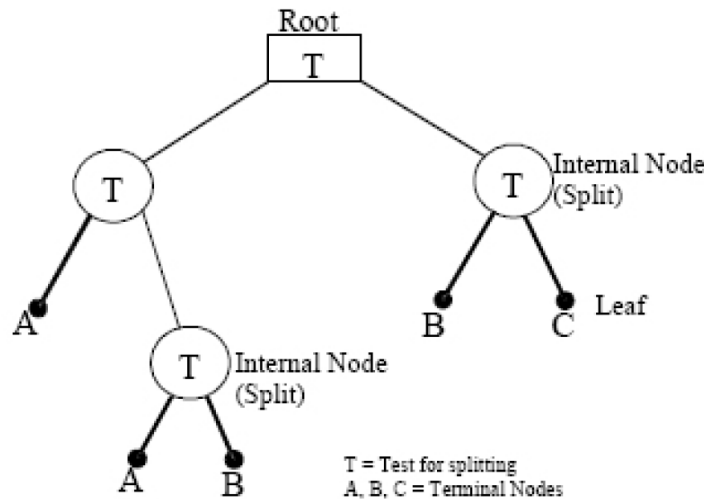


Figure 4.1: The structure of regression tree

Mathematical explanation of the process can be written as that:

We wish to find a function $t(x)$ maps our domain X to our target Y . Let L is the sample of n observations such that $L = (x_1, y_1), \dots, (x_n, y_n)$. Similar to the standard regression equations, while we choosing $t(x)$ we use the mean square prediction error $E(t(x) - E(y|x))^2$ as decision variable. For each leaf l and training sample c in the regression tree, the model is just $\hat{y} = \frac{1}{c} \sum_{c=1}^c y_1$, which is the sample mean of the response variable Y in that cell, so that it creates a piecewise constant model.

Regression trees are popular because they produce a model whose interpretation is so easy. They can handle the data of different types such as continuous, categorical, binary and ordinal. Data does not need transformation. Moreover, it is also useful while detecting importance of inputs on the response variable. We can easily made sensitivity analysis of the variables.

On the other hand, decision trees can be unstable with the small variations in the input data. This situation might give a completely wrong decision tree which cause to get an ineffective results on the test set.

Regression tree is one of the model that we use while predicting our target variable. Because, it is so fast about compilation time and easy to use. We use **fitrtree** function in MATLAB to train the sample and construct the decision trees. After that, the

predictions are obtained by predict function.

4.1.2 Ensemble of Regression Trees Model

In the regression context, ensembling models are used to build a comitee of regressors that may be the best regression method in our models. We use regression trees in ensembling model with least square ensembling method with least square boosting. We show this method below with its mathematical explanation.

Boosting is one of the ensembling methods in MATLAB. In this method, performance of the regression trees directly affects the accuracy of the predictors.

Suppose we have set of observations, (y_i, x_i) for $i = 1, 2, \dots, N_0$ where N_0 is the number of training set observations and x is an M-variate vector. We know that the probability density function of (y, x) is determined but unknown. Therefore we define a predictor $y^{(p)}(x)$ such that our prediction error PE and modeling error ME are founded as below:

$$PE = \frac{1}{N_1} \sum_{i=1}^{N_1} [y_i - y_i^{(p)}(x_i)]^2, \quad (4.1)$$

$$ME = \frac{1}{N_1} \sum_{i=1}^{N_1} [y_i^{(t)} - y_i^{(p)}(x_i)]^2, \quad (4.2)$$

where $y_i^{(p)}(x_i)$ is the prediction for the i 'th test example, y_i are the i 'th observation and $y_i^{(t)}$ is the real time value of y . We know that the parameters of $y^{(p)}(x)$ is acquired from N_0 training set observations. However the y_i and x_i in the above equations are obtained from a set of N_1 observations which means our test set is out of training sample.

Initially, let's we appoint a weight $w_i = 1$ for $i = 1, 2, 3, \dots, N_0$ to each training pattern such that the probability of the training sample i is in the training set is $p_i = w_i / \sum w_i$. Then we choose N_0 samples to form the training set. In the next step we have a regression model such that for each t , we have a regression machine $m_t : x \rightarrow y$ that takes predictor variables x to response variables y . Then every member of the training set is trained on that machines to obtain a prediction $y_i^{(p)}(x_i)$ for $i = 1, 2, \dots, N_0$.

In the next step, loss for each training sample is calculated as $L_i = L[|y_i^{(p)}(x_i) - y_i|]$. The loss L may be any functional form for $L \in [0, 1]$. The average loss can be written as $L_{avg} = \sum_{i=1}^{N_0} L_i p_i$ such that confidence level β in the predictor is calculated as $\beta = \frac{L_{avg}}{1 - L_{avg}}$. When β is low then the prediction has high confidence. Such that ensemble fits a that train when this β gives the lowest error for that prediction.

For a spesific x_i each of the machines produce predictions for $m_t, t = 1, 2, \dots, T$ such that we have the prediction for every T . Each machine m_t has the prediction $y_i^{(t)}$ and

the confidence β_t . We use that results while picking the ensemble regression tree. The picking method that we use in ensembling is least squares boosting and that method fits new weights for each step to minimize mean-squared error.

Detailed model that we use in MATLAB code is explained in the following sections with comparing the results of it with other regression models.

4.1.3 Robust Regression Model

We use **robustfit** function on MATLAB programme to predict the values of hourly electricity load, six-hour average wind energy production and six-hour average MCP of the Turkish Electricity Market. We compare results of it with the other regression model that we use. General principals of the model is explained below.

Robustfit function transfers predictor variables X to response variable Y by a $(p + 1)$ -by-1 vector \mathbf{b} , which is coefficient estimates for a robust multilinear regression. X is an n -by- p predictors variables at each of n observations and Y is n -by-1 vector of observed response variables.

The most accustomed method for robust regression is M-estimation. Estimators of that method can be considered as a generalization of maximum-likelihood estimation, i.e. "M"-estimation.

We have a linear model,

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i x'_i \beta + \epsilon_i, \quad (4.3)$$

for the i th of n observations. Given an estimator \mathbf{b} for β , the fitted model is

$$\hat{y}_i = a + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} + e_i = x'_i \mathbf{b}, \quad (4.4)$$

and the residuals are calculated as

$$e_i = y_i - \hat{y}_i. \quad (4.5)$$

The estimates for \mathbf{b} are determined by minimizing a particular objective function over all \mathbf{b} , so that

$$\sum_{i=1}^n \rho(e_i) x'_i \mathbf{b} = \sum_{i=1}^n \rho(y_i - x'_i \mathbf{b}). \quad (4.6)$$

The function ρ gives the addition of each residual to the objective function. The influence curve ψ is equal to the derivative of ρ . The system of $k + 1$ estimating equations

for the coefficients \mathbf{b} is generated as differentiating the objective function with respect to the \mathbf{b} and setting the partial derivatives to zero such that

$$\sum_{i=1}^n \psi(y_i - x_i' \mathbf{b}) x_i' = 0. \quad (4.7)$$

The weight function w is calculated as $w(e) = \psi(e)/e$, where $w_i = w(e_i)$. Then the equation can be written as

$$\sum_{i=1}^n \psi(y_i - x_i' \mathbf{b}) x_i' = 0. \quad (4.8)$$

To solve these estimating equations, we minimize $\sum w_i^2 e_i^2$, which is the same thing with the least squares problem. The weights are conditioned by the residuals, the residuals are conditioned by estimated coefficients \mathbf{b} and the estimated coefficients are conditioned by the weights. So we need an iterative solution which has the following steps,

- Initial estimates \mathbf{b}^0 should be chosen such as estimates of least squares.
- At each iteration t , residuals $e_i^{(t-1)}$ and associated weights $w_i^{(t-1)}$ should be calculated.
- The new coefficient estimates \mathbf{b}^t is calculated with the following equation

$$\mathbf{b}^t = [X' W^{(t-1)} X]^{-1} X' W^{(t-1)} y, \quad (4.9)$$

where X and $W^{(t-1)}$ are model and current weight matrices, respectively.

The last two steps are iterated until the estimated coefficients converge. Covariance matrix of \mathbf{b} is

$$\Omega(b) = \frac{E(\psi^2)}{[E(\psi')]^2} (X' X)^{-1}. \quad (4.10)$$

To get an estimated asymptotic covariance matrix $\Omega(b)$, we use $\sum [\psi(e_i)]^2$ for estimation of the nominator and $[\sum \psi'(e_i)/n]^2$ for estimation of the denominator.

In the function of **robustfit** on MATLAB, some estimator methods are used to estimate coefficient \mathbf{b} such as 'andrews', 'bisquare', 'cauchy', 'huber', which have different objective and weight functions, ρ and w . They are the piecewise functions with boundaries. This boundaries are named as tuning constants. Residuals e_i are used in the objective and weighted function and that boundaries are applied on that residual

values. Tuning constants are determined to work in rationally high efficiency in normal cases. All of the methods have determined tuning constants as default but it can be modifiable. We try all possible methods with different tuning constants. These results are compared and illustrated in the following sections.

4.1.4 Neural Network Model

A neural network, or more sleekly called as an 'artificial' neural network, is explained by the inventor of one of the first neurocomputers [9]

”...a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs.”

Neural networks are modeled as inspired from a neuronal structure of mamalian cerebral cortex but on much smaller scales. They are setup as an interdisciplinary issue of neuroscience, mathematics, physical science and engineering and psychology. Their techniques are developed continually with the developments of personal computers [6].

An artificial neural network is a group of connected artificial neurons that works on a mathematical or computational model for data processing depending on connected systems. It is intended for solving problems without directly constituting a mathematical, biological and physical model. Because of this unique properties, neural network systems are used to imitate the complex model such as non-linear computation and find the complicated relation between inputs and output. The advantages of using neural network is that it gives the estimates very efficiently for forecasts of the output without knowing the model, the real structure and the link between inputs and output. It finds the interconnections between the inputs and the output and determine the synaptic weights of the relations inside the network to get the closest result of the output that is trained with the experience in the past [23, 13] .

4.1.5 Mathematical Model of A Single Neuron

Let $P = [P_1, P_2, P_3, \dots, P_N]$ be the input vector and $W = [W_{11}, W_{12}, \dots, W_{1N}]$ be the weights vector such that W_{1i} for each $i = 1, \dots, N$, is the weight of individual input element of P_i .

Let K be a single bias. The output O can be calculated using a transfer function f as

$$O = f \left(\sum_{i=1}^N W_{1i} P_i + b \right). \quad (4.11)$$

There are three main transfer functions used in the literature:

a) The first one is the threshold function

$$f(v) = \begin{cases} 1 & \text{if } v \geq 0, \\ 0 & \text{if } v < 0. \end{cases} \quad (4.12)$$

b) The second one is the piecewise-linear function

$$f(v) = \begin{cases} 1 & \text{if } v \geq \frac{1}{2}, \\ v & \text{if } -\frac{1}{2} > v > \frac{1}{2}, \\ 0 & \text{if } v < -\frac{1}{2}. \end{cases} \quad (4.13)$$

c) The third one is the Sigmoid function

$$f(v) = \frac{1}{1 + e^{-av}}, \quad (4.14)$$

where a is a slope parameter of the sigmoid function. Moreover, there is a modified version of Sigmoid function called hyperbolic tangent function such that

$$f(v) = \frac{1 - e^{-v}}{1 + e^{-v}} = \tanh\left(\frac{v}{2}\right). \quad (4.15)$$

4.1.6 Mathematical Model of Multiple Neuron

The neural network process in Figure 4.2 shows the model that we use in the regression of the hourly electrical load. That model is named as non-linear auto regressive model with external input. It starts from the inputs and continue with the hidden layer, output layer and the output. Each input is connected to hidden layer. Moreover, each hidden neuron in the hidden layer is also connected to every output layer. In our model we have 15 inputs for each observation and get 1 output. In addition, since we use closed neural network we get the calculated output as the extra input that help us to improve our solutions with the prediction of the past values of output.

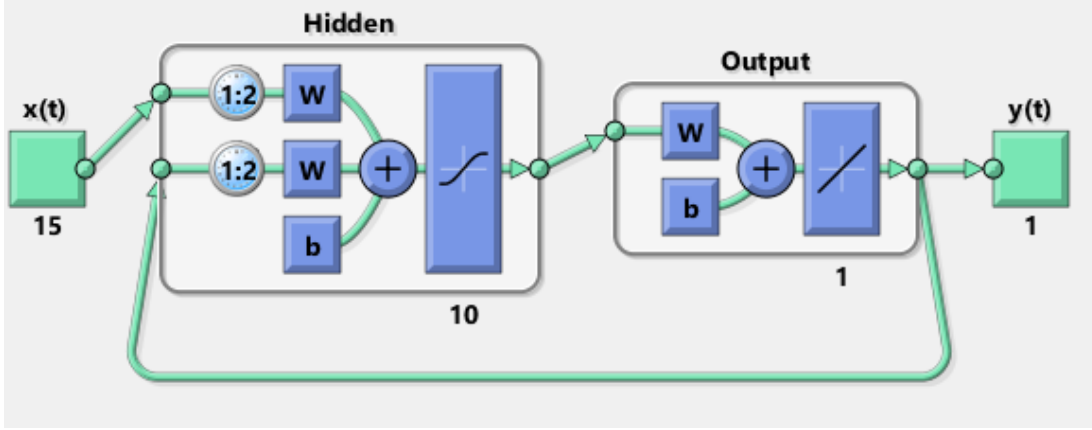


Figure 4.2: Non linear autoregressive neural network model with external input

Signals are passing through the input towards the hidden layer. They are multiplied by their corresponding synaptic weights. When multiplied signals are summed and transformed by a transfer function, the output is reached.

Let i , j and k denote i -th input, j -th neuron of hidden layer and k -th neuron of output layer, respectively. Let y_i be an input signal. Let w_{ij} be a synaptic weight between i -th input and j -th hidden neuron. Let h_j be the summed signal at j -th hidden neuron and $hf_j(.)$ be the transfer function at j -th hidden neuron.

h_j can be calculated as

$$h_j(N) = \sum_{i=1}^N w_{ji}(N)y_i(N), \quad (4.16)$$

where N is the number of input signals. The initiated sum through transfer function $hf_j(.)$ is written as

$$hy_j = hf_j(h_j(N)) + b_j, \quad (4.17)$$

where b_j is a threshold value at j -th hidden neuron.

Let w_{kj} be a synaptic weight between j -th hidden neuron and k -th output neuron.

Let O_k be the summed signal at k -th output neuron and $Of_k(.)$ be the transfer function at k -th output neuron. O_k can be calculated as

$$O_k(M) = \sum_{j=1}^M w_{kj}(M)hy_j(M), \quad (4.18)$$

where M is the number of hidden neurons. The transformed sum Oy_k through transfer function $Of_k(\cdot)$ is written as

$$Oy_k = Of_k(O_k) + b_k, \quad (4.19)$$

where b_k is a threshold value at k -th neuron.

4.2 Data

In this section we briefly explain how we get the historical market data, numerical weather predictions and wind farms information.

4.2.1 Historical Market Data

Historical market data is obtained from the website of EXIST. This data contains total available installed capacities of electricity generator companies in Turkey, Total electrical production, total available installed capacity of wind farms and total real time generation of wind power, the market clearing price and the system marginal price on the TEL [4]. This data is available as hourly basis and can be taken as excel spreadsheet.

4.2.2 Weather Data

Weather data is obtained from European Centre for Medium Range Weather Forecasts (ECMWF) [10] and it is taken from archive of data sets of [10]. It is 70 m height forecasts of the weather of 6 hours intervals and includes temperature, relative humidity, surface pressure and u and v components which show a west to east and south to north flows of the wind, respectively. It is given as 1925 number of grids positioned on Turkey map having coordinates between $36^\circ - 42^\circ$ North latitude and between $26^\circ - 45^\circ$ East longitudes. Note that pressure can not be taken as 70 m height and is obtained from surface.

4.2.3 Wind Farms Data

Wind farms data is obtained from Turkish Wind Energy Association (TWEA) [5] and General Directorate of Renewable Energy [18]. Data includes lists of wind power plants under operation, with the informations of their coordinates and installed capacities.

4.3 Wind Energy Production Forecasting

The wind energy production is important for system operator, market operator and the producers. From the perspective of operators, they can plan the distribution of electricity one day before with a nice forecast. Moreover, from the perspective of producers, they can manage their strategies and prevent themselves from big losses with that predictions. Furthermore, it is important to forecast the wind energy production one day before because MCP is announced a day in advance. Therefore, the relation between the wind energy production and MCP is investigated in this thesis.

Wind energy has a complex relations, since it is a renewable energy source and has many dependant factors in the production. Regression tree model, ensembling of the regression trees model, and robutsfit model tool on the MATLAB programme are used to get results in order to solve this relationship between the dependant factors and total wind energy production. Neural network model is tried to get results but it is not give such a good result to compare with the other regression models. MATLAB code has five main steps that are in order as below:

- 1-Obtain data to our forecast
- 2-Defining predictor data
- 3-Defining data sets and test sets
- 4-Training of model and simulating with test set and get results
- 5-Plotting, comparison and evaluation.

As it is explained above, production data is taken from the Internet as an Excel spreadsheet. The data needs to be examined in order to operationalize by dealing with the problems such as outliers, blanks and errors. After this, production data is adapted to be processed MATLAB programme. Weather data is taken from ECMWF website, which is explained in Appendix A. It is in .grib extension and to read this data in windows, WGRIB programme is used to transform it to .txt version to read by MATLAB. Our related weather data with the wind energy are the relative humidity, temperature, u and v components of the wind and the pressure. Weather forecasting data is taken from 70m. height excepting pressure. It is taken from the surface since that can not be taken from 70m. height. While we are using u and v components of the wind we turned this variables to wind speed vector and angle of the wind. It is also explained shortly in Appendix A.

An important point is that our weather forecasts were obtained at six hours, while our wind data was obtained on an hourly basis. Therefore we adapt our wind energy production data as the six-hour periods. We take average hourly production for these periods and get this data as our output variable.

As the second part of code, we define a set to predictor data about using them in regression models. As a predictor data, our weather data is taken in this set, which are wind speed vector, angle of the wind, temperature, relative humidity of the 70m.

height and surface pressure as mentioned before. Weather data is taken for the wind farms on the related grid. As we mentioned before, our weather data has totally 1925 grid points. Only 92 of these grids has one or more wind farms inside their fields. These results are taken according to the grid where the plant is located. The weather forecasts of that grids are taken as inputs with 5 branches for each. Moreover, wind farms total installed capacities for the next day, is taken in the predictor set. So we have 461 input columns for the predictor matrix, that is $92 \times 5 = 460$ input columns for weather forecasts and 1 input column for installed capacities of wind farms.

Third section includes creating training and test sets. We have the average for six-hour periods data processed for 2016 and 2017 years, which has totally 2924 rows of data. Training data includes 23 months six-hour periodic data from 01 January 2016 to 30 November 2017 totally 16800 rows. Moreover, test set has the daily averages in the December 2017 which is 4 rows of data for a day. It is important that our test set is not trained in the model and while testing it is used in the model for the first time. In other words, our test set is out of sample that we trained.

For the fourth section some of the regression models are used for training. Then we simulate these models on the test set and get the predicted values. The models that we use are regression trees model, ensembling of regression trees model and robust regression model as we mentioned before.

Regression tree model is used with **fitrtree** fuction of MATLAB. While we construct ensembling model we use **fitensemble** function with 100 regression tree and we use least square boosting algorthim to minimize residuals. Lastly robust regression model is applied with **robustfit** function. As we explain in our model, there is different weight function with different tuning constants in that model. We applied all of them and we get the best estimator function to compare with the regression tree model and the ensemble model. Best weight function that we applied for that sample is the 'talwar' function with the default tuning constant. While we are calculating the estimates we use **predict** function for regression tree model and the ensemble model. Moreover, we calculate the robust regression result, we multiply the predictor matrix with the coefficient **b** that we determined by **robustfit** function.

At the end of the compiling session residuals are calculated as

$$r = Y - PredY, \quad (4.20)$$

where r is the residuals between the real values Y and forecasted values $PredY$.

Furthermore, Mean Absolute Percent Error $Mape$ is obtained as

$$Mape = \frac{\sum_{t=1}^n \frac{|r_t|}{Y_t}}{n} \times 100, \quad (4.21)$$

where $Mape$ is the mean absolute percent error, n is the number of all values and r_t is the t -th residual value. $Mape$ is the most common measure of the forecast error and it

shows the quality of the predictions.

Moreover Mae is the mean absolute error, which is calculated as in Equation (4.22) shows the amount of the error.

$$Mae = \frac{\sum_{t=1}^n |r_t|}{n}. \quad (4.22)$$

Table 4.1: Forecasted amounts of wind energy production for 1 December 2017

	REAL WIND ENERGY PRODUCTION	ENSEMBLE MODEL	REGRESSION TREE MODEL	ROBUST REGRESSION MODEL
'01-Dec-2017 00-06'	4428.4	4181.3	4025.1	4405.7
'01-Dec-2017 06-12'	4272.7	4048.8	3397.7	4184.2
'01-Dec-2017 12-18'	4066.6	3754.7	3397.7	3346.8
'01-Dec-2017 18-24'	4076.1	4244.8	3397.7	3655.4

Table 4.1 shows the real and the forecasted amounts for the the six-hour averages of the wind energy production in 1 December 2017. It is also showed as graphic in Figure 4.3. As can be seen from the table and the figure all of the three models follow the real values.

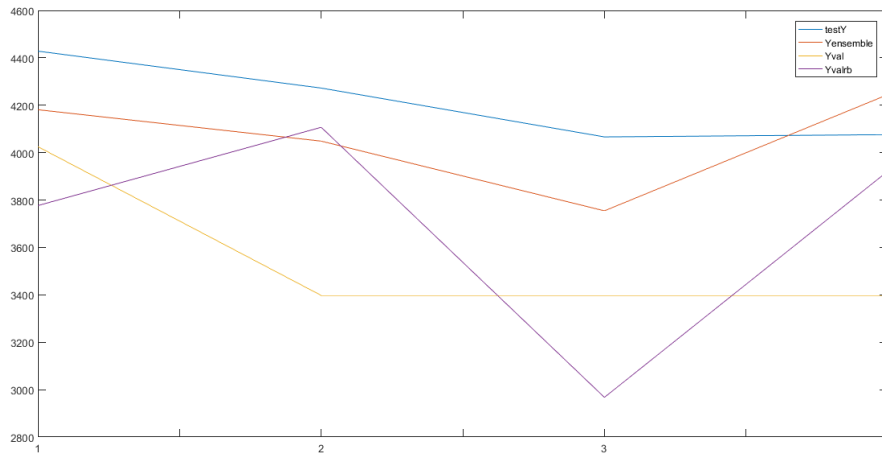


Figure 4.3: Wind Production Data and Model Prediction

In order to compare performances of the predictors we need to look at the $Mape$ of the forecasts. $Mape$ is the most common measure of the forecast error and it shows the quality of the predictions. As shown in Table 4.2 best regression model of the wind energy forecasting is the ensembling of regression trees model from the perspective of $Mape$ values. $Mape$ for this model is calculated as 5.66% and Mae is calculated as 237.89 MWh.

Table 4.2: *Mape* and *Mae* values of regression models for 1 December 2017

	Mape (%)	Mae (MWh)
ENSEMBLE MODEL	5.66	237.89
REGRESSION TREE MODEL	15.67	656.37
ROBUST REGRESSION MODEL	7.65	312.95

We can easily say that ensembling regression trees model gives the best prediction values, when comparing with the other models that we use. We do not compare this values only with *Mape* and *Mae* values of one day. We also look at the errors of the training set and all of the other days in December 2017. In these studies we get similar results with that ensembling regression trees model which has more accurate forecasts except one or two days of outliers. These outliers arise from not finding a similar solution in the training sample for this case or forecasting errors of the numerical weather predictions. In Turkey, average mean absolute percent errors of the estimates of the wind power plants is in between 10 – 11% [23, 21, 20]. Therefore, it is clear that our prediction is good because it reduces differences caused by all of the prediction errors, imbalances due to power plants malfunctions and bidding strategies that wind farms have made to earn more. In this way, a producer has a reference predictor value about the prices of tomorrow a few hours before all of the transactions are done. Moreover these results show the determination of the estimates and how few errors are made per hour. Thus, from the perspective of the system operator possible imbalances and system shutdowns can be predicted and necessary actions can be taken a day in advance.

4.4 Electricity Load Forecasting

Electricity load forecasting is the most important issue in electricity markets that gives us the estimated electrical load reference value of the distribution of production resources while showing the total load on the electrical system in Turkey. Our model is shaped according to the social changes in the flow of life, using the non-linear relationship between the predictor data and TEL of the past. The days of the week and the hours of each day were grouped separately to catch social changes.

As shown in Figure 4.4 while electricity consumption at night is severely down, the consumption reaches its highest level during daytime. Similarly, there are significant differences in the use of electricity between working days and holidays. The reduction in electricity use is not as much as on Sunday because Saturday's weekend is not a weekend break for the entire community. For this reason, the days are not only divided into working days and holidays, but each day is also classified as the day of the week.

Similarly, in religious holidays and official holidays, electricity consumption level decreasing to the same level as in the weekend holidays. Especially during the religious holidays electricity consumption level decrease lowest level of the year. Therefore, these holidays are grouped as the weekend holidays.

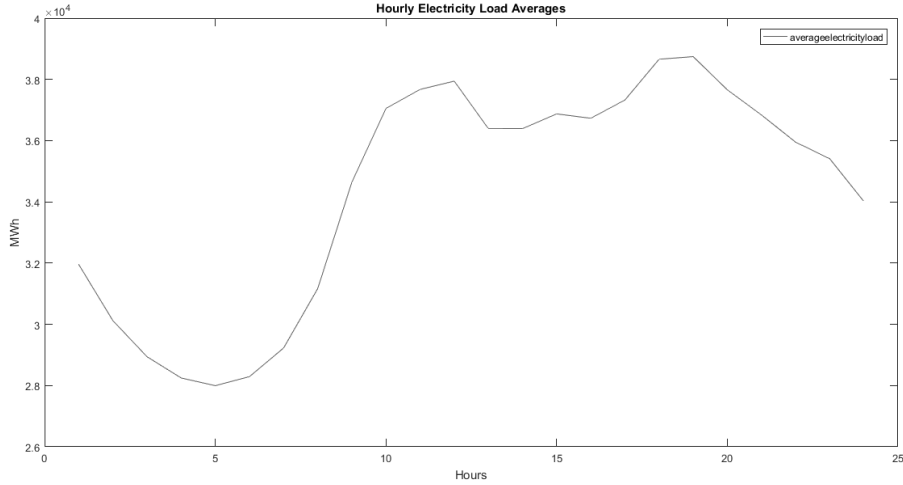


Figure 4.4: Hourly electricity load averages in December 2017

Totally data of two years is used for training and testing part. They are not similar to each other but they can carry each other seasonal information. As a result of this, 365 days are divided to angles as 360 degree in a year so that each day of a year has a unique angle. Moreover, to look at seasonality of years degree of the each day are turned into radians and these radians are used to find the sine and cosine values of each day. Thus, it was concluded that the start and end days of the year should have very close values. These values are also used as a predictor data and added to the predictor matrix of our model.

As in the wind energy production, total energy production is effected by weather conditions. Different level of the temperature affects the electricity consumption of the industry. Also different weather conditions influence people directly. Air conditioners and infrared heaters are used more in hot and cold temperatures so that electricity demand is directly affected from this conditions. Therefore, surface temperature, surface relative humidity, surface pressure and surface dewpoint are taken as a predictor data and added to the predictor matrix. All of these weather data is taken from 2 m height and weighted averages of this data is calculated with the usage percentages of the cities (see Appendix A) .

Finally, historical production data is taken from the market, which is hourly total installed capacity of Turkey, real time production at the same hour of the same day of the previous week, real time production at the same hour of the previous day, and average 24 hour production of the previous day. These are taken into predictor set to benefit from statistics of the past.

To summarize our predictor set has 15 input and our target has 1 output. Inputs are average surface dewpoint temperature, average surface pressure, average surface temperature, average relative humidity, total installed capacity, day of year, holidays, sine value of the day, cosine value of the day, determined total production amount of tomorrow, day of week, working day or not determiner, same hour load of previous week, same hour load of previous day and average load of the previous 24 hours. In addition

the output is hourly total electricity load of tomorrow.

Similar to the the wind energy production forecasting, training and test sets are created to use. As explained before, database includes totally 17544 rows of data. Training data has 16800 rows of data that is uploaded as hourly from 01 January 2016 to 30 November 2017. Furthermore, test set has 744 rows of hourly data that starts from 01 December 2017 and ends at 31 December 2017.

Regression tree model, ensembling of regression trees model, robust regression model and the neural network tool are used for training data set and then calculate forecasted values of the electricity production by simulating the predictor matrix of the test set. While we are using regression tree model, trees are constructed by **fitrtree** function and the result are evaluated with **predict** function.

The second model that we use is the ensemble of regression trees, that is applied with the composition of 100 trees. We use **fitensemble** function and we use the least squares boosting as the ensemble-aggregation algorithm as we mentioned before. It aims to decrease mean-square error by fitting new weights for each step. We can increase the number of trees that we composed. While it diminishes the error of the training set, for some cases it increase the error of the test set. Moreover, it increases compilation time of training session.

The robust regression model is applied with **robustfit** function. This function defines coefficient **b** for the predictor inputs and we multiply this **b** with the predictor matrix to get the predictions. As we explained in mathematical structure of model, there is different objective and weight function used on the residual values. We applied all of these functions, compare them with their *Mape* values and we evaluated results of the function that gave the best estimates. Moreover tuning constants are used in default for this model.

Our last model is the neural network model. It is applied with **ntstool** which is the neural network time series tool. The model that we use in this tool is nonlinear autoregressive with external input (NARX). Delays of that function are used as default(1:2). We use 10 hidden neuron in the hidden layer. Hidden neurons in this model are using hyperbolic tangent function, which is the modified version of sigmoid function. While we are putting predictors and the target variable to the neural network model, we prepare this inputs and the output with **preparets** function. After that, the model is trained and simulated with test set.

After using these models, performances of them are evaluated similar as in the previous part. Residuals and *Mape* are calculated with the Equations 4.20 and 4.21, respectively. *Mape* and *Mae* values are tabulated in Table 4.3. As seen from this table, when evaluated over *Mape* values best predictor model for electricity load forecasting is the neural network model. The mean absolute percent error of the total electricity production is calculated as 1.38% and the mean absolute error is calculated as 475.20 MWh. In other words, the estimates with neural network model are accurate with an error of 1.38%.

Residuals of forecasts with neural network model are shown in Figure 4.5. Predictions

Table 4.3: *Mape* and *Mae* values for electricity load forecasting in December 2017

	Mape (%)	Mae (MWh)
ENSEMBLE MODEL	2.76	942.73
REGRESSION TREE MODEL	2.38	814.40
ROBUST REGRESSION MODEL	1.64	508.77
NEURAL NETWORK MODEL	1.38	475.20

show accuracy in 2000MWh difference level. That level is achieved only four times in the month. These outliers can sometimes happen due to the sudden demand for the electricity, depending on unexpected weather conditions or the increasing in the power consumption of the big corporations.

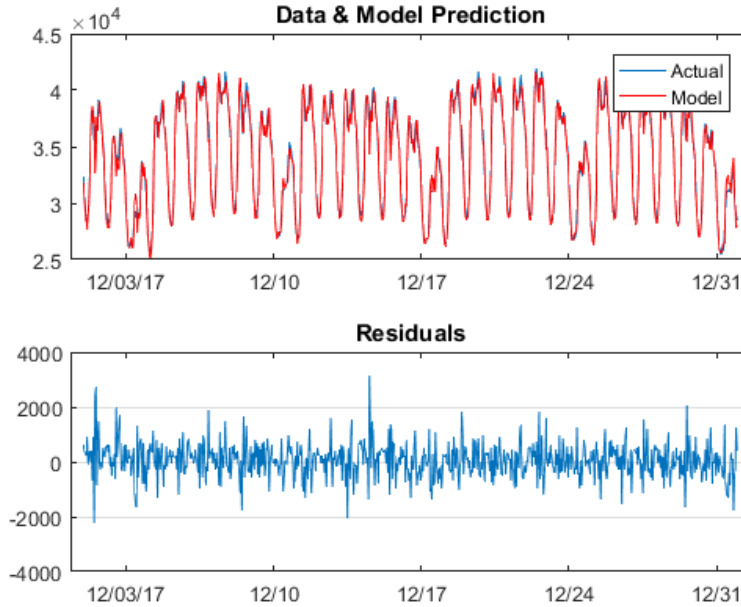


Figure 4.5: Real Time Total electricity production and Model Prediction with neural network

Figure 4.6 shows the outliers as "+" from out of prediction box. Each hour contains 31 data of each day of December 2017. In other words, there are 744 tested data in total. Out of this, 21 of them are the outlier and the rest remains in the prediction box. The most erroneous hours are 1 am and 9 am, which are the changing hours of the decrease trend and the increase trend in energy production, respectively as seen in Figure 4.4 that is plot of average hourly electricity load in December 2017, i.e., the model does not catch perfectly the sudden changes in electricity usage. On the other hand, the rest is obtained with success.

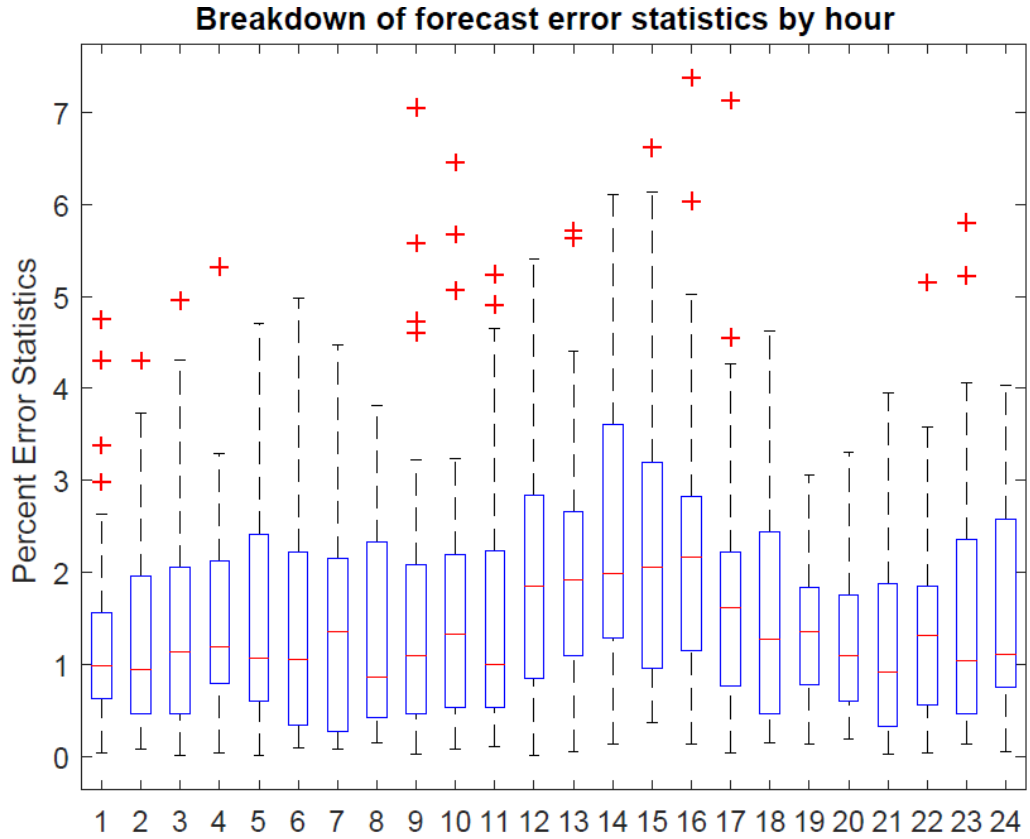


Figure 4.6: Boxplot of the Electricity Load Predictions

4.5 Market Clearing Price(MCP) Forecasting

MCP is the most difficult one to predict when compared with the other two. As estimating the trends and demands of people in electricity load and estimating the weather condition about wind energy production, it is necessary to find a forecasting model in price estimation that includes production forecasts and proposals of producers and consumers in addition to the demand of people.

As it is explained in Chapter 2 MCP occurs at the point where the supply surplus and the demand surplus intersect. Therefore, our model should include supply surplus and demand surplus in it. Then, rest of the model is completed by regression models in MATLAB.

Similar to the our previous forecasts, we define a predictor function that takes out the predictor matrix in order to train the predictor in the programme and test it.

As we explained in wind energy forecast, we have weather forecasts with 6 hour intervals. In order to adapt the other predictors and price, we calculate averages for six-hour periods. Therefore, we have 4 period for a day, these are 00 : 00 – 06 : 00, 06 : 00 – 12 : 00, 12 : 00 – 18 : 00 and 18 : 00 – 24 : 00.

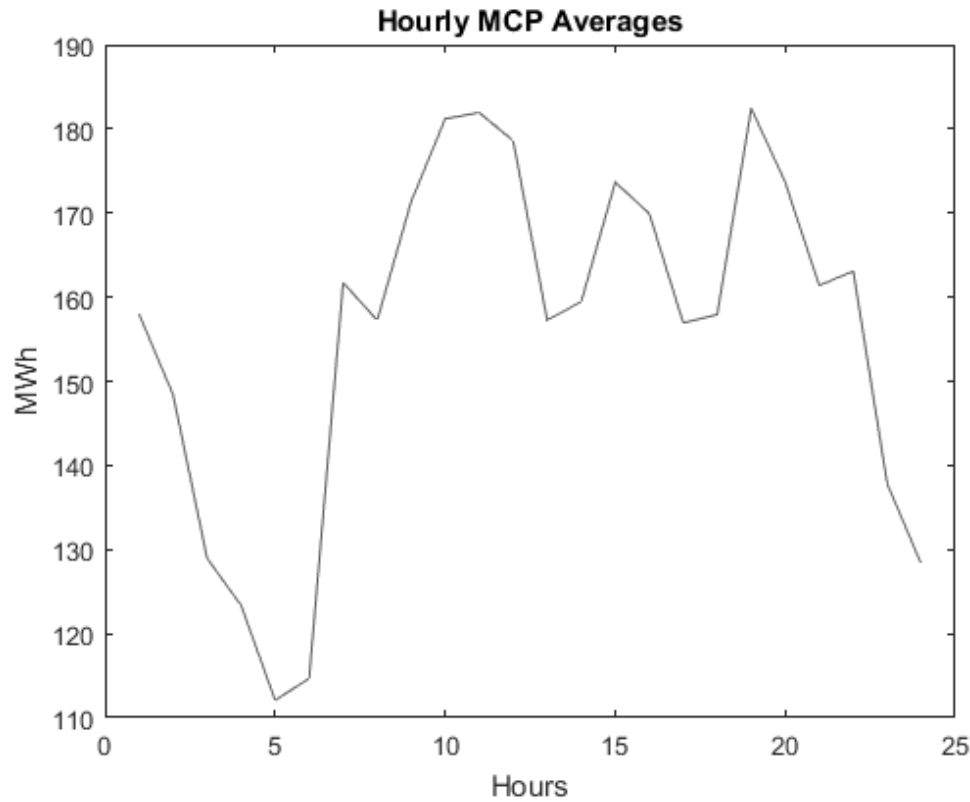


Figure 4.7: Hourly Average MCPs of December 2017

As seen in Figure 4.7 average MCP goes down at night because of the excess demand and the night hours not being the collective working hours of the community. Price is at the bottom level in between 05 : 00 – 06 : 00 and reaches its peak in working hours of the community. Moreover, MCP graph is partially in line with the distribution of the averages over the electricity load. These indicators give us two guiding results. Firstly, we should classify price results as periods because prices vary from hour to hour. Secondly, we should include the predictors that we used in the electric charge, while forecasting MCP.

We have a very large predictor matrix that includes the data that is elaborated below. Timezone is used since we grouped as six-hour periods. Furthermore, because working and not-working hours affects the price, we need to classify our data daily. Thus, we add day of week, and group our data as working day and holiday.

Moreover, since price is affected from demand and supply, we add all of the historical market data, total installed capacity of Turkey and real time total production of Turkey. Also, we add total installed capacity about wind production in Turkey, final daily production program of wind energy, and real time wind energy production in the predictor matrix. It is important to use data because we need to see the effect of that predictors to our forecast. We are trying to capture the balance about the ratio of the total electric charge and the actual wind output and the amount given as the forecasting plan.

Moreover, historical statistics of the electricity load, MCP and SUMP are added to the predictor matrix. Real time production at the same hour of the same day of the previous week, real time production at the same hour of the previous day, and average 24 hour production of the previous day, MCP at the same hour of the same day of the previous week, MCP at the same hour of the previous day, and average 24 hour MCP of the previous day, SUMP at the same hour of the same day of the previous week, SUMP at the same hour of the previous day, and average 24 hour SUMP of the previous day are inserted to predictor matrix.

The weather forecast was added to the forecasting matrix as it was in the previous forecasts in order to be able to feel more of the effect of the wind energy and the renewable energy source in the estimation. Surface dew point, surface pressure, surface temperature and relative humidity forecasts, which is taken from 2 m height, are included to the predictor matrix.

To summarize, total 578 columns of input are included in the system to use on the regression models prediction matrix. 560 of them are the weather forecasts in four point. The other 18 inputs are total electrical production, period of time, holidays, determined total production amount of tomorrow, day of week, working day or not determiner, determined total wind energy production amount of tomorrow, total installed capacity, total installed capacity of wind, previous week same hour load, MCP, SMP, previous day same hour load, MCP, SMP and previous 24 hour average load, MCP, SMP.

Similar to the previous forecasts, we split data to two group as training and test sets. Training set has 2800 rows of data that is uploaded as averages of six-hour periods from 01 January 2016 to 30 November 2017. Furthermore, test set has 124 rows of six-hour periodic data that starts from 01 December 2017 and ends at 31 December 2017. We examine our test set day by day.

We try five different regression models for this case. We try regression tree model, ensemble regression trees model, robustfit model, neural network fitting model(nftool) and the neural network time series model(ntstool). While checking the results of the ntstool we get the extraordinary solutions same as the wind energy forecasting. Neural network time series non linear autoregressive with external input model (NARX) gave irrelevant results with the real values. Sometimes residuals are more than twice of the real values, so that we do not compare this model with the others. We use regression tree model, ensemble of regression trees model and robust fit model similar to the wind energy forecasting model. So we will not repeat their details again in this section. While we compile our programme with neural network fitting tool, we use ten hidden neuron in our hidden layer, Levenberg-Marquardt backpropagation algorithm for training. It completes its operation in 12 iteration and gives the weights for the inputs. Then we simulate them with test sample of predictors.

Table 4.4 shows *Mape* and *Mae* values of price forecasts for 1 December 2017. Robust regression model is best model with 8.42% *Mape* and the 12.78 TL *Mae* value.

Table 4.4: Mape and Mae values of regression models for MCP in 1 December 2017

	Mape (%)	Mae (TL)
ENSEMBLE MODEL	9.93	14.72
REGRESSION TREE MODEL	12.60	19.18
ROBUST REGRESSION MODEL	8.42	12.78
NEURAL NETWORK MODEL	16.97	27.27

As it is shown in Figure 4.8 prediction does not have an identical forecasts of the real time prices, but it has similar fracture and gives a good information about the price of tomorrow except neural network model. In this study, an important issue is detected that, while working with the large input set with so many inputs, neural network models slog away. On the other hand, when number of inputs are small as in the case of forecasting electricity load, neural network gives more accurate results than the other regression models.

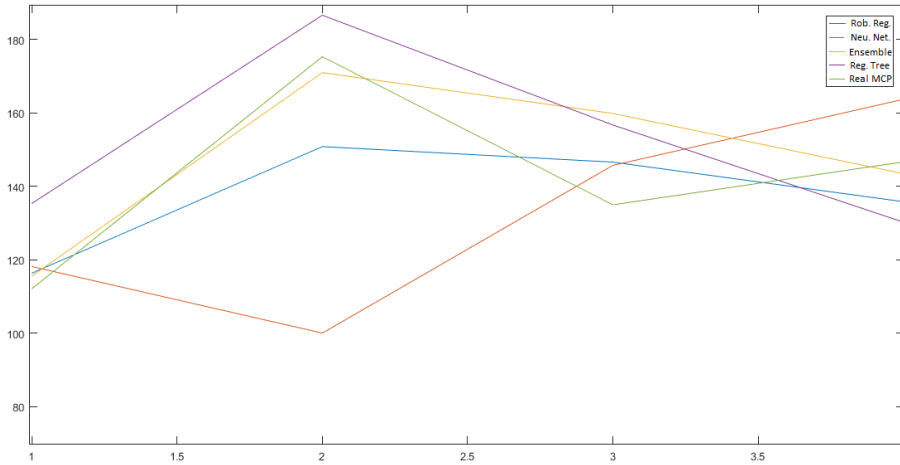


Figure 4.8: Real Time MCP and Forecasted Prices

Figure 4.9 shows the comparison of the real MCP values and the forecasted values with the robust regression model in the first 10 days of December. As can be seen in the figure, when the price goes in the normal flow, the estimates show high accuracy as parallel to the change in predictors. In the event of a sudden loss of power or a sudden demand per hour in a large producer, it is very difficult for the estimates to catch up because the price change is sudden and the estimates must be calculated before the bids of the producers and consumers are shown.

There is an outlier that does not reflect the true trend and affects the stability of this model. This outlier can be removed by checking the forecast with the other regression models to obtain a better estimate. Other models can give accurate solutions for these points. While doing day by day estimation, we crosscheck the predictions and do not use the abnormal solutions. In this way outliers in the forecast of the test set are cancelled out.

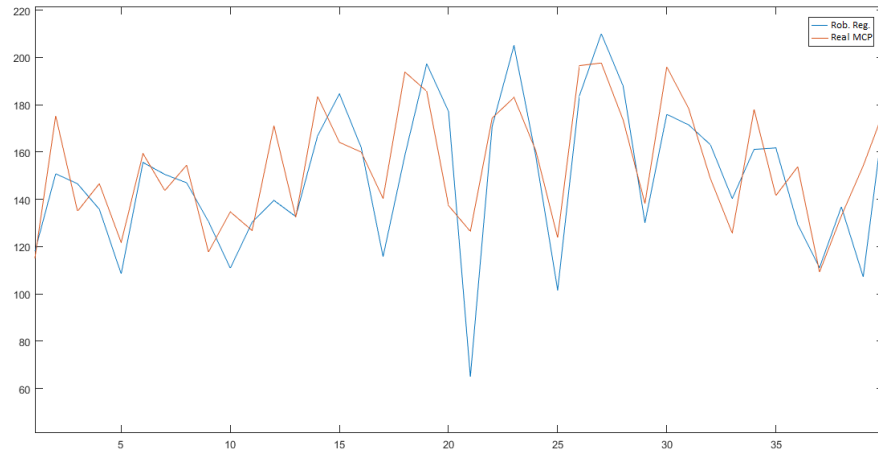


Figure 4.9: Plot of Real and forecasted MCP with robust fit model in first 10 days of December 2017

To summarize, the market player has to know MCP of tomorrow which is the reference value of the market, the total electricity to be produced tomorrow in order to get an estimate of how much the company will earn and the total transaction volume of market. Results of our model show that model gives very good reference values with a short calculation time and the wind energy affects the volatility of MCP due to the fact that it is a too volatile energy source and forecasting of wind energy production needs some times to enhance itself.

CHAPTER 5

CONCLUSION

In this thesis, we predict average wind energy production amount for six-hour periods with the numerical weather predictions and the installed capacity of the wind turbines. While we are evaluating predictions, best model that we observe is ensembling of regression trees model. This model diminishes mean square error for the training sample to very small values. However, because of working with out-of-sample test set, for some cases it may give irrelevant predictions. This is because the model can not find similar case in the training sample. Wind energy production is developing year by year and new wind turbines are added with every new month. Wind turbines that are taken into maintenance, or newly commissioned to the system directly affects the production amount, but there is not a predictor variable for that cases in our model. We use numerical weather predictions on grid base and we can not change this grids regarding the active or inactive position. These limitations are increasing the error values. To overcome these problems production values for every wind turbine needs to be forecasted in its special conditions. On the other hand, the forecast values that we achieved are useful for the prediction of MCP. Deviations lose their effect together with other predictors used in the model.

As a system operator, it is important to predict the amount of total electricity load in order to reduce the imbalance costs and to avoid possible foreseeable imbalances. On the other hand, as an electricity generator company, total electricity load needs to be forecasted to make a strategy about selling its power on the most profitable market. For these purposes, hourly total electrical load of the next day is forecasted with four different regression models. The best regression model that we use is achieved by neural network time series tool. Forecast results in that model are 1.38% mean absolute percentage error and about 475MWh mean absolute error. While we get the predicted values, we do not have an information system about power failures and power cuts happens at the past. We know these failures and cuts affect the usage levels and we could not have information about them. If we benefit from this information we can follow the real time fluctuations and improve our forecasts.

Moreover, as an electricity supplier, it is important to minimize the wholesale purchase cost, because this is the factor that determines the profitability of the company. On the retail side, the prices are more or less certain, so the company that wants to gain competitive advantage must reduce what it pays for wholesale electricity sells at retail. As a result of these, market participants and system operator should forecast market clear-

ing price, which is the reference value of the markets. In this direction market clearing price is forecasted with four different regression models. The best model that we adapt is the robust regression model with 8.42% mean absolute error. The results of the estimations show that the market participants can obtain profitable results by making a strategy in balancing power market, in the day ahead market or in the intra-day market with the forecast values obtained. Moreover, a company that will consume electricity at any time of the day can predict the supplying time to minimize the electricity cost and will make the purchase bid accordingly. As it is mentioned in previous paragraph, real-time information affects the price levels. However, this indicates that predictions should be evaluated on a case-by-case basis for every day and the estimates should be adapted to real-time changes. Furthermore, we know numerical weather predictions that we get has six hours intervals and we had to adapt our data to this situation so that average price levels were calculated for six-hour periods. This increases the residuals and we can not take this under the control, because weather predictions taken for a certain time are not suitable directly to average MCP values. This problem can be solved with appropriate weather data, which can be hourly weather forecasts or average weather forecasts for six-hour periods.

For the future work, the relationship between the estimates and the results of the numerical weather predictions can be explored and the best predictive data can be used on that system. Different mathematical or statistical forecasting models can be used, such as Monte Carlo simulations, time series models in order to diversify forecasting models and to make comparisons between various models. Furthermore, number of predictor values used in regression models can be augmented to improve the quality of forecasts. Some other predictor values such as fuel prices can be added to decrease the error of forecasts. To conclude, with the addition of new models and data, this topic can be expanded, developed and more effective results can be achieved.

REFERENCES

- [1] Şekkeli, Mustafa and Yıldız, Ceyhun and Karık, Fatih and Sözen, Adnan, Türkiye Elektrik Piyasasında Rüzgar Enerjisi, Gazi Mühendislik Bilimleri Dergisi, 1(2), pp. 253–264, 2016.
- [2] Yenilenebilir Enerji Genel Müdürlüğü, Yenilenebilir Enerji Kaynaklarının Belgelendirilmesi ve Desteklenmesine ilişkin yönetmelik, http://www.yegm.gov.tr/yenilenebilir/document/EP_YEK_BelgeDestekY.doc, 2011, [Online; accessed 02.07.2018].
- [3] Türkiye Elektrik İletim A.Ş., Türkiye Kurulu Güç ve Üretiminin Yıllar İtibariyle Gelişimi, <http://www.teias.gov.tr/sites/default/files/2017-10/21.docx>, 2017, [Online; accessed 02.07.2018].
- [4] Energy Exchange Istanbul, Transparency Platform, <https://seffaflik.epias.com.tr/transparency/>, 2018, [Online; accessed 02.07.2018].
- [5] Turkish Wind Energy Association, Türkiye'deki Rüzgar Enerjisi Santralleri Durumu, <https://www.tureb.com.tr/bilgi-bankasi/turkiye-res-durumu>, 2018, [Online; accessed 02.07.2018].
- [6] M. H. Beale, M. T. Hagan, and H. B. Demuth, *Neural network toolbox user's guide*, in *R2012a, The MathWorks, Inc., 3 Apple Hill Drive Natick, MA 01760-2098, www.mathworks.com*, 2012.
- [7] R. Behçet, H. Gül, H. Oral, and F. Oral, Rüzgâr Enerjisi Potansiyeli Bakımından Malatya İlinin Doğu Anadolu Bölgesindeki Yeri, Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 3(1), pp. 65–73, 2014.
- [8] F. D. Bianchi, R. J. Mantz, and H. De Battista, *The Wind and Wind Turbines*, Springer, 2007.
- [9] M. Caudill, Neural Networks Primer, part I, AI expert, 2(12), pp. 46–52, 1987.
- [10] European Centre for Medium-Range Weather Forecasts, <https://www.ecmwf.int/>, 2018, [Online; accessed 02.07.2018].
- [11] Resmi Gazete, Elektrik Piyasası Kanunu no: 4628, 2001.
- [12] Resmi Gazete, Elektrik Piyasası Kanunu no: 6446, 2013.
- [13] K. Gurney, *An Introduction to Neural Networks*, CRC press, 2014.
- [14] Global Wind Energy Council, Global Wind Statistics 2017, http://gwec.net/wp-content/uploads/vip/GWEC_PRstats2017_EN-003_FINAL.pdf, 2017, [Online; accessed 02.07.2018].

- [15] J. F. Manwell, J. G. McGowan, and A. L. Rogers, *Wind Energy Explained: Theory, Design and Application*, John Wiley & Sons, 2010.
- [16] A. W. Manyonge, R. Ochieng, F. Onyango, and J. Shichikha, Mathematical Modelling of Wind Turbine in a Wind Energy Conversion System: Power Coefficient Analysis, *Applied Mathematical Sciences*, 6(91), pp. 4527–4536, 2012.
- [17] Yenilenebilir Enerji Genel Müdürlüğü, Türkiye Rüzgar Enerjisi Potansiyel Atlası, http://www.yegm.gov.tr/YEKrepa/REPA-duyuru_01.html, [Online; accessed 02.07.2018]”.
- [18] Yenilenebilir Enerji Genel Müdürlüğü, <http://www.yegm.gov.tr/yenilenebilir/ruzgar.aspx>, 2018, [Online; accessed 02.07.2018].
- [19] A. Özdamar, Dünyada ve Türkiyede Rüzgar Enerjisinden Yararlanılması Üzerine Bir Araştırma, *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 6(2), pp. 133–145, 2000.
- [20] M. A. Özpınar, *Yenilenebilir Enerji Santrallerinde Yapay Sinir Ağları Yöntemiyle Enerji Üretiminin Modellenmesi ve Planlanması*, Ph.D. thesis, YTÜ Fen Bilimleri Enstitüsü, 2007.
- [21] Ü. Şenol and Z. Musayev, Rüzgar Enerjisinden Elektrik Üretiminin Yapay Sinir Ağları ile Tahmini, *Bilge International Journal of Science and Technology Research*, pp. 23–31, 2017.
- [22] D. A. Spera, *Introduction to Modern Wind Turbines*, ASME Press, 1994.
- [23] X. Yan and N. Chowdhury, Electricity Market Clearing Price Forecasting in a Deregulated Electricity Market, in *Probabilistic Methods Applied to Power Systems (PMAPS), 2010 IEEE 11th International Conference*, pp. 17–30, IEEE, 2010.

APPENDIX A

Weather Data Acquisition and Processing

Weather data is important for quality and determination of our forecast. Hourly weather predictions are needed to implement our forecasting model. We choose the European Centre for Medium-Range Weather Forecasts (ECMWF) data, which is well accepted, reliable, and accurate for weather forecasting. ECMWF data is freely available and is disseminated to the member states. Relative humidity, dew point, pressure and temperature data are taken from 2m height for electricity load forecasting. For wind energy forecasting, we take these data from 70m height which is related to energy production of wind turbines. The data is taken between 12 hours or 6 hours, and we chose the data that was taken with 6 hours gaps. When the time is 00:00, we have the forecast of 00:00, 06:00, 12:00, 18:00 and 24:00. Altogether 5 different data are taken at 5 times each day.

Data can be downloaded in *.grib* extension in order to be workable in windows and converted to test data *.txt* which is readable in MATLAB. We WGRIB programme for data transformation which is used manipulate and decode GRIB data files for weather forecasting.

```
wgrib -v -text -d all file.grb -o file.txt
```

As a result of this we obtain a matrix *.txt* extension with $77 \times 25 = 1925$ elements, where single column contains data for each hour and for each variable. This data is read in Matlab with **dlmread** and **vec2mat** functions and processed with the codes in section A.1. The codes can read the *.txt* data, distinguish different data from the same file and merge the data which taken for 2016 and 2017 separately for each observation.

We have totally 3655 rows and 1925 columns of data for each variable of relative humidity of 10 m height, relative humidity of 70m height, surface pressure, 2m dewpoint temperature, 24m temperature, 70m temperature, 70m height *u* component of wind, 70m height *v* component of wind. 3655 rows consists of 731 days of the 2 years multiplied with the 5 observation for each day, taken at the six hour intervals. 1925 columns of data consists of the matrix sited to the coordinates of the Republic of Turkey, located between 26 – 45 east meridian, 36 – 42 northern latitudes. This 1925 observation for each variable and for each time is got 1925 grids, ordered from 1 to 1925 as showing in the figure below:

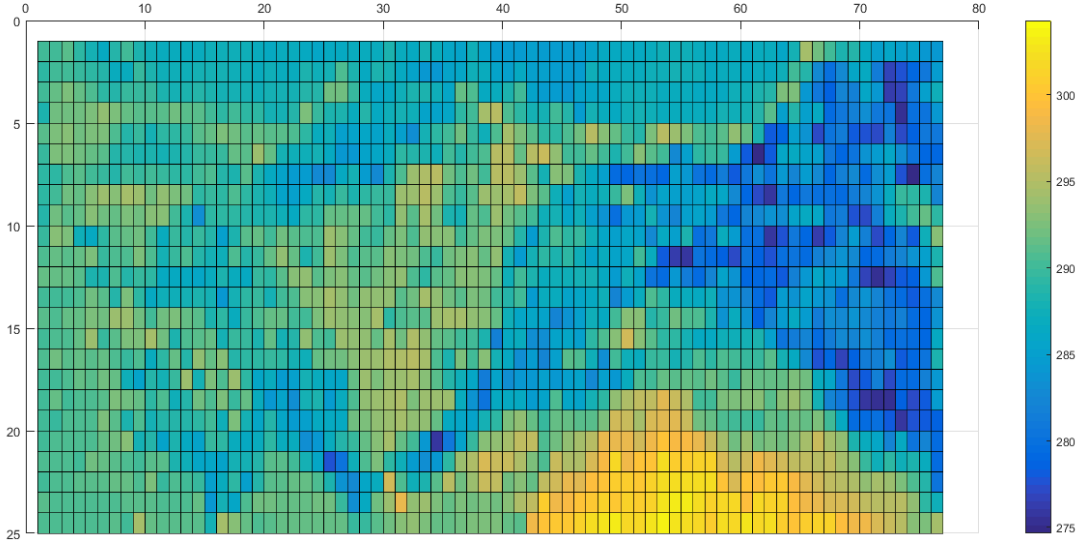


Figure A.1: Republic of Turkey surface temperature forecast map with grids resolution on 14.04.2016 at 00:00

As seen in Figure A.1, the numbers on the graph indicate grid coordinates. The grid at the top left is the grid number 1, the grid at the top right is 77, the grid at the bottom left is 1849 and the grid at the bottom right is the grid number 1925. We evaluate these grids by means of the number of the grid number that the center falls with its coordinates, while appointing the grids and calculating the averages.

Linear interpolation was used to fill the empty data gaps in order to adapt them to hourly electricity data. Moreover, we have duplication about the time 00:00. Because we have the temperature estimate and instantaneous temperature value after 24 hours, taken from a day before, at the same time. We do not need instantaneous temperature values, so that we extract these rows from our data and then interpolate our data for the forecasting of electricity load. We do these with codes written in section A.2. After that, we will take the hourly averages of the data. Taking the averages of the data is done for the electricity usage and it is affected from density and the location of the cities.

Total Electricity load is related to the weather conditions of the place where it is used, as we explained in previous chapters. Thus, weather prediction of the place where electricity is consumed more should affect the average of the numerical weather predictions. Therefore, we collect a list of cities' electricity usage percentages with their city centers coordinates. We use the usage percentages as their weights and find the grid numbers that the city centers coordinates fall. While finding the grid numbers, we use the codes written in section A.4. We find the numerical weather prediction weighted averages of cities related to their electricity usage percentages with the codes written in section A.5.

As a result we have the weighted average numerical weather prediction of relative humidity from 2 m height, dewpoint from 2 m height, surface pressure and temperature from 2 m height. We use this data directly on predictor matrix of electricity load forecasting. In MCP forecasting case, we want the model feels the effect of weather forecast more. As a result of this we bring the grids where the city centers located in. We determine these grids and take their weather forecast results as input directly. Furthermore, we do not use interpolation to adapt weather forecasts to hourly price data. On the contrary, we constrict our price data about adapting to our weather forecasts. In order to do this we take averages of MCP for six-hour periods.

Wind energy is produced in wind plants and it is affected from the coordinates of the wind plants with respect to its production capacity and production capacity is directly dependent to its installed capacity. We collect the list of all wind plants of Turkey that is updated twice per year. So we have four list in total for 2016 and 2017 and they have all the details from installed capacity to their coordinates.

To find the grid numbers of wind plants we use codes in section A.3 at the end of this chapter. After that we take the grid, which has at least one wind farm in its field. We take the weather forecasts of that grids as an input. So we do not need to work with weights and the installed capacities of the wind plants. We only determine the grids which has wind farms. Then we use the weather forecasts of the grids directly in our predictor matrix. In order to use u and v components of the wind we need to transform these components to wind velocity vector v_h and wind angle ϕ_v : following equations to calculate them as

$$\phi_v = \text{atan2}(u, v), \quad |v_h| = \sqrt{u^2 + v^2}.$$

We use v_h and ϕ_v in the predictor matrix for forecasting of wind energy production.

A.1 NWP data codes

```
M1 = dlmread('dewtempre2016.txt');
%10573740 row of data
M1(:,2) = [];
%delete second column
dewtempre2016 = vec2mat(M1,1926);
%every step give 1926 grid;
%then delete the first column
dewtempre2016(:,1) = [];
%we have 5490X1925 sized column
%dewtempre(3k-2) for k from 1 to 1830 gives Surface...
... pressure
for i=(1:1925)
for k=(1:1830)
surpre2016(k,i)=dewtempre2016((3*k-2),i);
```

```

end
end
%dewtemp(3k-1) for k from 1 to 1830 gives 2 metre...
... temperature
for i=(1:1925)
for k=(1:1830)
surtemp2016(k,i)=dewtemp2016((3*k-1),i);
end
end
%dewtemp(3k) for k from 1 to 1830 gives 2 metre ...
...dewpoint temperature
for i=(1:1925)
for k=(1:1830)
surdewp2016(k,i)=dewtemp2016((3*k),i);
end
end
M2 = dlmread('dewtemp2017.txt');
%10544850 row of data
M2(:,2) = [];
%delete second column
dewtemp2017 = vec2mat(M2,1926);
%every step give 1926 grid;
dewtemp2017(:,1) = [];
%then delete the first column
%we have 5475X1925 sized column
%dewtemp(3k-2) for k from 1 to 1825 gives Surface...
... pressure
for i=(1:1925)
for k=(1:1825)
surpre2017(k,i)=dewtemp2016((3*k-2),i);
end
end
%dewtemp(3k-1) for k from 1 to 1825 gives 2 metre...
... temperature
for i=(1:1925)
for k=(1:1825)
surtemp2017(k,i)=dewtemp2016((3*k-1),i);
end
end
%dewtemp(3k) for k from 1 to 1825 gives 2 metre ...
...dewpoint temperature
for i=(1:1925)
for k=(1:1825)
surdewp2017(k,i)=dewtemp2016((3*k),i);
end
end

```

```

M3 = dlmread('uvtem702016.txt');
%10573740 row of data
M3(:,2) = [];
%delete second column
uvtem702016 = vec2mat(M3,1926);
%every step give 1926 grid;
uvtem702016(:,1) = [];
%then delete the first column
%we have 5490X1925 sized column
%dewtempre(3k-2) for k from 1 to 1830 gives Surface...
... pressure
for i=(1:1925)
for k=(1:1830)
temp702016(k,i)=uvtem702016((3*k-2),i);
end
end
%dewtempre(3k-1) for k from 1 to 1830 gives 2 metre...
... temperature
for i=(1:1925)
for k=(1:1830)
windu702016(k,i)=uvtem702016((3*k-1),i);
end
end
%dewtempre(3k) for k from 1 to 1830 gives 2 metre...
... dewpoint temperature
for i=(1:1925)
for k=(1:1830)
windv702016(k,i)=uvtem702016((3*k),i);
end
end

M4 = dlmread('uvtem702017.txt');
%10544850 row of data
M4(:,2) = [];
%delete second column
uvtem702017 = vec2mat(M4,1926);
%every step give 1926 grid;
uvtem702017(:,1) = [];
%then delete the first column
%we have 5490X1925 sized column
%dewtempre(3k-2) for k from 1 to 1825 gives Surface...
... pressure
for i=(1:1925)
for k=(1:1825)
temp702017(k,i)=uvtem702017((3*k-2),i);
end
end

```

```

%dewtempre(3k-1) for k from 1 to 1825 gives 2 metre...
... temperature
for i=(1:1925)
for k=(1:1825)
windu702017(k,i)=uvtem702017((3*k-1),i);
end
end
%dewtempre(3k) for k from 1 to 1825 gives 2 metre...
... dewpoint temperature
for i=(1:1925)
for k=(1:1825)
windv702017(k,i)=uvtem702017((3*k),i);
end
end
%2016 relative humidity of 70 and 10 metres
M5 = dlmread('relhum2016.txt');
%7049160 row of data
M5(:,2) = [];
%delete second column
relhum2016 = vec2mat(M5,1926);
%every step give 1926 grid;
relhum2016(:,1) = [];
%then delete the first column
%we have 3660X1925 sized column
%relhum2016(2k-1) for k from 1 to 1830 gives relativehum...
... of 10 metre
for i=(1:1925)
for k=(1:1830)
relhum102016(k,i)=relhum2016((2*k-1),i);
end
end
%relhum2016(2k) for k from 1 to 1830 gives relativehum...
... of 70 metre
for i=(1:1925)
for k=(1:1830)
relhum702016(k,i)=relhum2016((2*k),i);
end
end
%2017 relative humidity of 70 and 10 metres
M6 = dlmread('relhum2017.txt');
%7049160 row of data
M6(:,2) = [];
%delete second column
relhum2017 = vec2mat(M6,1926);
%every step give 1926 grid;
relhum2017(:,1) = [];
%then delete the first column

```

```

%we have 3660X1925 sized column
%relhum2016(2k-1) for k from 1 to 1825 gives relativehum...
... of 10 metre
for i=(1:1925)
for k=(1:1825)
relhum102017(k,i)=relhum2017((2*k-1),i);
end
end
%relhum2016(2k) for k from 1 to 1825 gives relativehum...
... of 70 metre
for i=(1:1925)
for k=(1:1825)
relhum702017(k,i)=relhum2017((2*k),i);
end
end

%Merging datas of 2016 and 2017
for k=(1:1830)
surpre(k,:)=surpre2016(k,:);
end
for k=(1:1825)
surpre(k+1830,:)=surpre2017(k,:);
end
for k=(1:1830)
surtemp(k,:)=surtemp2016(k,:);
end
for k=(1:1825)
surtemp(k+1830,:)=surtemp2017(k,:);
end
for k=(1:1830)
surdewp(k,:)=surdewp2016(k,:);
end
for k=(1:1825)
surdewp(k+1830,:)=surdewp2017(k,:);
end
for k=(1:1830)
temp(k,:)=temp702016(k,:);
end
for k=(1:1825)
temp(k+1830,:)=temp702017(k,:);
end
for k=(1:1830)
windu(k,:)=windu702016(k,:);
end
for k=(1:1825)
windu(k+1830,:)=windu702017(k,:);
end

```

```

for k=(1:1830)
windv(k,:)=windv702016(k,:);
end
for k=(1:1825)
windv(k+1830,:)=windv702017(k,:);
end
for k=(1:1830)
relhum10(k,:)=relhum102016(k,:);
end
for k=(1:1825)
relhum10(k+1830,:)=relhum102017(k,:);
end
for k=(1:1830)
relhum70(k,:)=relhum702016(k,:);
end
for k=(1:1825)
relhum70(k+1830,:)=relhum702017(k,:);
end

```

A.2 Expanding and adapting dataset of NWP

```

windv=abs(windv);
windu=abs(windu);
for k=2:731
surtemp(5*k-4,:)=zeros;
end
for k=2:731
surtemp((5*k-4)-k+2,:)=[];
end
for k=1:2924
surtempall(6*k-5,:)=surtemp(k,:);
end
for k=1:2923
surtempall(6*k-4,:)=((5*surtempall(6*k-5,...
...+surtempall(6*k+1,...))/6;
surtempall(6*k-3,:)=((4*surtempall(6*k-5,...
...+(2*surtempall(6*k+1,...))/6;
surtempall(6*k-2,:)=((3*surtempall(6*k-5,...
...+(3*surtempall(6*k+1,...))/6;
surtempall(6*k-1,:)=((2*surtempall(6*k-5,...
...+(4*surtempall(6*k+1,...))/6;
surtempall(6*k,:)=(surtempall(6*k-5,...
...+(5*surtempall(6*k+1,...))/6;
end
surtempall(17540:17544,:)=surtempall(17516:17520,...);
surtemp=surtempall;

```



```

for k=2:731
surdewp(5*k-4,:)=zeros;
end
for k=2:731
surdewp((5*k-4)-k+2,:)=[];
end
for k=1:2924
surdewpall(6*k-5,:)=surdewp(k,:);
end
for k=1:2923
surdewpall(6*k-4,:)=((5*surdewpall(6*k-5,:))...
...+surdewpall(6*k+1,:))/6;
surdewpall(6*k-3,:)=((4*surdewpall(6*k-5,:))...
...+(2*surdewpall(6*k+1,:)))/6;
surdewpall(6*k-2,:)=((3*surdewpall(6*k-5,:))...
...+(3*surdewpall(6*k+1,:)))/6;
surdewpall(6*k-1,:)=((2*surdewpall(6*k-5,:))...
...+(4*surdewpall(6*k+1,:)))/6;
surdewpall(6*k,:)=(surdewpall(6*k-5,:)+...
...(5*surdewpall(6*k+1,:)))/6;
end
surdewpall(17540:17544,:)=surdewpall(17516:17520,:);
surdewp=surdewpall;
for k=2:731
surpre(5*k-4,:)=zeros;
end
for k=2:731
surpre((5*k-4)-k+2,:)=[];
end
for k=1:2924
surpreall(6*k-5,:)=surpre(k,:);
end
for k=1:2923
surpreall(6*k-4,:)=((5*surpreall(6*k-5,:))...
...+surpreall(6*k+1,:))/6;
surpreall(6*k-3,:)=((4*surpreall(6*k-5,:))...
...+(2*surpreall(6*k+1,:)))/6;
surpreall(6*k-2,:)=((3*surpreall(6*k-5,:))...
...+(3*surpreall(6*k+1,:)))/6;
surpreall(6*k-1,:)=((2*surpreall(6*k-5,:))...
...+(4*surpreall(6*k+1,:)))/6;
surpreall(6*k,:)=(surpreall(6*k-5,:)+...
...+(5*surpreall(6*k+1,:)))/6;
end
surpreall(17540:17544,:)=surpreall(17516:17520,:);
surpre=surpreall;
for k=2:731

```

```

temp(5*k-4,:)=zeros;
end
for k=2:731
temp((5*k-4)-k+2,:)=[];
end
for k=1:2924
tempall(6*k-5,:)=temp(k,:);
end
for k=1:2923
tempall(6*k-4,:)=(5*tempall(6*k-5,:))...
...+tempall(6*k+1,:))/6;
tempall(6*k-3,:)=(4*tempall(6*k-5,:))...
...+(2*tempall(6*k+1,:)))/6;
tempall(6*k-2,:)=(3*tempall(6*k-5,:))...
...+(3*tempall(6*k+1,:)))/6;
tempall(6*k-1,:)=(2*tempall(6*k-5,:))...
...+(4*tempall(6*k+1,:)))/6;
tempall(6*k,:)=(tempall(6*k-5,:))...
...+(5*tempall(6*k+1,:)))/6;
end
tempall(17540:17544,:)=tempall(17516:17520,:);
temp=tempall;
for k=2:731
windu(5*k-4,:)=zeros;
end
for k=2:731
windu((5*k-4)-k+2,:)=[];
end
for k=1:2924
winduall(6*k-5,:)=windu(k,:);
end
for k=1:2923
winduall(6*k-4,:)=(5*winduall(6*k-5,:))...
...+winduall(6*k+1,:))/6;
winduall(6*k-3,:)=(4*winduall(6*k-5,:))...
...+(2*winduall(6*k+1,:)))/6;
winduall(6*k-2,:)=(3*winduall(6*k-5,:))...
...+(3*winduall(6*k+1,:)))/6;
winduall(6*k-1,:)=(2*winduall(6*k-5,:))...
...+(4*winduall(6*k+1,:)))/6;
winduall(6*k,:)=(winduall(6*k-5,:))...
...+(5*winduall(6*k+1,:)))/6;
end
winduall(17540:17544,:)=winduall(17516:17520,:);
windu=winduall;
for k=2:731
windv(5*k-4,:)=zeros;

```

```

end
for k=2:731
windv((5*k-4)-k+2,:)=[];
end
for k=1:2924
windvall(6*k-5,:)=windv(k,:);
end
for k=1:2923
windvall(6*k-4,:)=(5*windvall(6*k-5,:))...
...+windvall(6*k+1,:))/6;
windvall(6*k-3,:)=(4*windvall(6*k-5,:))...
...+(2*windvall(6*k+1,:)))/6;
windvall(6*k-2,:)=(3*windvall(6*k-5,:))...
...+(3*windvall(6*k+1,:)))/6;
windvall(6*k-1,:)=(2*windvall(6*k-5,:))...
...+(4*windvall(6*k+1,:)))/6;
windvall(6*k,:)=(windvall(6*k-5,:))...
...+(5*windvall(6*k+1,:)))/6;
end
windvall(17540:17544,:)=windvall(17516:17520,:);
windv=windvall;
for k=2:731
relhum10(5*k-4,:)=zeros;
end
for k=2:731
relhum10((5*k-4)-k+2,:)=[];
end
for k=1:2924
relhum10all(6*k-5,:)=relhum10(k,:);
end
for k=1:2923
relhum10all(6*k-4,:)=(5*relhum10all(6*k-5,:))...
...+relhum10all(6*k+1,:))/6;
relhum10all(6*k-3,:)=(4*relhum10all(6*k-5,:))...
...+(2*relhum10all(6*k+1,:)))/6;
relhum10all(6*k-2,:)=(3*relhum10all(6*k-5,:))...
...+(3*relhum10all(6*k+1,:)))/6;
relhum10all(6*k-1,:)=(2*relhum10all(6*k-5,:))...
...+(4*relhum10all(6*k+1,:)))/6;
relhum10all(6*k,:)=(relhum10all(6*k-5,:))...
...+(5*relhum10all(6*k+1,:)))/6;
end
relhum10all(17540:17544,:)=relhum10all(17516:17520,:);
relhum10=relhum10all;
for k=2:731
relhum70(5*k-4,:)=zeros;
end

```

```

for k=2:731
relhum70((5*k-4)-k+2,:)=[];
end
for k=1:2924
relhum70all(6*k-5,:)=relhum70(k,:);
end
for k=1:2923
relhum70all(6*k-4,:)=((5*relhum70all(6*k-5,:))...
...+relhum70all(6*k+1,:))/6;
relhum70all(6*k-3,:)=((4*relhum70all(6*k-5,:))...
...+(2*relhum70all(6*k+1,:)))/6;
relhum70all(6*k-2,:)=((3*relhum70all(6*k-5,:))...
...+(3*relhum70all(6*k+1,:)))/6;
relhum70all(6*k-1,:)=((2*relhum70all(6*k-5,:))...
...+(4*relhum70all(6*k+1,:)))/6;
relhum70all(6*k,:)=(relhum70all(6*k-5,:))...
...+(5*relhum70all(6*k+1,:)))/6;
end
relhum70all(17540:17544,:)=relhum70all(17516:17520,:);
relhum70=relhum70all;

```

A.3 Grid appointment of wind plants

```

for i=1:109
a1(i,1)=wtinfos20161(i,6);
a1(i,2)=wtinfos20161(i,7);
a1(i,3)=wtinfos20161(i,8);
end
for i=1:123
a2(i,1)=wtinfos20162(i,6);
a2(i,2)=wtinfos20162(i,7);
a2(i,3)=wtinfos20162(i,8);
end
for i=1:145
a3(i,1)=wtinfos20171(i,6);
a3(i,2)=wtinfos20171(i,7);
a3(i,3)=wtinfos20171(i,8);
end
for i=1:150
a4(i,1)=wtinfos20172(i,6);
a4(i,2)=wtinfos20172(i,7);
a4(i,3)=wtinfos20172(i,8);
end
a1=table2array(a1);
a2=table2array(a2);

```

```

a3=table2array(a3);
a4=table2array(a4);

for j=1:109
for i=1:77
for k=1:25
if (a1(j,1)>=26+(i-1)*19/77) && (26+i*19/77>a1(j,1))...
... && (42-(k-1)*6/25>=a1(j,2)) && (a1(j,2)>42-k*6/25)
a1(j,4)=grids(k,i);
end
end
end
end
for j=1:123
for i=1:77
for k=1:25
if (a2(j,1)>=26+(i-1)*19/77) && (26+i*19/77>a2(j,1))...
... && (42-(k-1)*6/25>=a2(j,2)) && (a2(j,2)>42-k*6/25)
a2(j,4)=grids(k,i);
end
end
end
end
for j=1:145
for i=1:77
for k=1:25
if (a3(j,1)>=26+(i-1)*19/77) && (26+i*19/77>a3(j,1))...
... && (42-(k-1)*6/25>=a3(j,2)) && (a3(j,2)>42-k*6/25)
a3(j,4)=grids(k,i);
end
end
end
end
for j=1:150
for i=1:77
for k=1:25
if (a4(j,1)>=26+(i-1)*19/77) && (26+i*19/77>a4(j,1))...
... && (42-(k-1)*6/25>=a4(j,2)) && (a4(j,2)>42-k*6/25)
a4(j,4)=grids(k,i);
end
end
end
end

```

A.4 Grid appointment of cities

```
for i=1:81
for k=1:26
c1(i,k)=cityinfos(i,k+2);
end
end
c1=table2array(c1);
for j=1:81
for i=1:77
for k=1:25
if (c1(j,1)>=26+(i-1)*19/77) && (26+i*19/77>c1(j,1)) ...
... && (42-(k-1)*6/25>=c1(j,2)) && (c1(j,2)>42-k*6/25)
c1(j,27)=grids(k,i);
end
end
end
end
```

A.5 Weighted NWP averages of cities

```
for k=1:17544
avgsurtemp(k,1)=0;
avgsurdewp(k,1)=0;
avgsurpre(k,1)=0;
avgrelhum10(k,1)=0;
end
for k=1:744
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,3)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,3)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,3)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,3)*relhum10(k,c1(i,27)));
end
end
for k=745:1440
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,4)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,4)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,4)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,4)*relhum10(k,c1(i,27)));
end
end
for k=1441:2184
for i=1:81
```

```

avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,5)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,5)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,5)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,5)*relhum10(k,c1(i,27)));
end
end
for k=2185:2904
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,6)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,6)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,6)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,6)*relhum10(k,c1(i,27)));
end
end
for k=2905:3648
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,7)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,7)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,7)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,7)*relhum10(k,c1(i,27)));
end
end
for k=3649:4368
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,8)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,8)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,8)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,8)*relhum10(k,c1(i,27)));
end
end
for k=4369:5112
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,9)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,9)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,9)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,9)*relhum10(k,c1(i,27)));
end
end
for k=5113:5856
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,10)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,10)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,10)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,10)*relhum10(k,c1(i,27)));
end
end
for k=5857:6576

```

```

for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,11)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,11)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,11)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,11)*relhum10(k,c1(i,27)));
end
end
for k=6577:7320
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,12)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,12)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,12)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,12)*relhum10(k,c1(i,27)));
end
end
for k=7321:8040
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,13)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,13)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,13)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,13)*relhum10(k,c1(i,27)));
end
end
for k=8041:8784
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,14)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,14)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,14)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,14)*relhum10(k,c1(i,27)));
end
end
for k=8785:9528
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,15)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,15)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,15)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,15)*relhum10(k,c1(i,27)));
end
end
for k=9529:10200
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,16)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,16)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,16)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,16)*relhum10(k,c1(i,27)));
end
end

```



```

for k=10201:10944
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,17)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,17)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,17)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,17)*relhum10(k,c1(i,27)));
end
end
for k=10945:11664
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,18)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,18)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,18)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,18)*relhum10(k,c1(i,27)));
end
end
for k=11665:12408
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,19)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,19)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,19)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,19)*relhum10(k,c1(i,27)));
end
end
for k=12409:13128
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,20)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,20)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,20)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,20)*relhum10(k,c1(i,27)));
end
end
for k=13129:13872
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,21)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,21)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,21)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,21)*relhum10(k,c1(i,27)));
end
end
for k=13873:14616
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,22)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,22)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,22)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,22)*relhum10(k,c1(i,27)));
end
end

```

```

end
for k=14617:15336
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,23)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,23)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,23)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,23)*relhum10(k,c1(i,27)));
end
end
for k=15337:16080
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,24)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,24)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,24)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,24)*relhum10(k,c1(i,27)));
end
end
for k=16081:16800
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,25)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,25)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,25)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,25)*relhum10(k,c1(i,27)));
end
end
for k=16801:17544
for i=1:81
avgsurtemp(k,1)=avgsurtemp(k)+(c1(i,26)*surtemp(k,c1(i,27)));
avgsurdewp(k,1)=avgsurdewp(k)+(c1(i,26)*surdewp(k,c1(i,27)));
avgsurpre(k,1)=avgsurpre(k)+(c1(i,26)*surpre(k,c1(i,27)));
avgrelhum10(k,1)=avgrelhum10(k)+(c1(i,26)*relhum10(k,c1(i,27)));
end
end
avgsurtemp=avgsurtemp/100;
avgsurdewp=avgsurdewp/100;
avgsurpre=avgsurpre/100;
avgsurtemp=avgsurtemp-273.15;
avgsurdewp=avgsurdewp-273.15;

```