JUNE 2017

Ph.D in Industrial Engineering

YUNUS EROĞLU

UNIVERSITY OF GAZİANTEP

GRADUATE SCHOOL OF

NATURAL & APPLIED SCIENCES

CONDITION MONITORING AND CONTROL TOOLS FOR WIND ENERGY SYSTEMS USING SCADA DATA

Ph.D THESIS

IN

INDUSTRIAL ENGINEERING

BY

YUNUS EROĞLU

JUNE 2017

Condition Monitoring and Control Tools for Wind Energy Systems Using SCADA Data

Ph.D Thesis

in

Industrial Engineering

University of Gaziantep

Supervisor

Prof. Dr. Serap ULUSAM SEÇKİNER

by

Yunus EROĞLU

June 2017

REPUBLIC OF TURKEY
UNIVERSITY OF GAZİANTEP
GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES
INDUSTRIAL ENGINEERING

Name of the thesis: Condition Monitoring and Control Tools for Wind Energy Systems Using SCADA Data

Name of the student: Yunus EROĞLU

Exam date: 14/06/2017

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Ahmet Necmeddin YAZICI

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Serap ULUSAM SEÇKİNER

Head of Department

This is to certify that we have read this thesis and that in our consensus opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Serap ULUSAM SEÇKİNER

Supervisor

Examining Committee Members	Signature
Prof.Dr. Türkay DERELİ	....
Prof.Dr. Serap ULUSAM SEÇKİNER	....
Assoc.Prof.Dr. Mehmet İshak YÜCE	....
Assoc. Prof. Dr. Mehmet ATAK	....
Assist. Prof. Dr. Serkan ALTUNTAŞ	....

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.

Yunus EROĞLU

ABSTRACT

CONDITION MONITORING AND CONTROL TOOLS FOR WIND ENERGY SYSTEMS USING SCADA DATA

EROĞLU, YUNUS
Ph.D. in Industrial Engineering
Supervisor: Prof. Dr. Serap ULUSAM SEÇKİNER
June 2017
222 pages

Technological developments in wind energy have reduced investment and operating costs. For this reason, wind farms have become more popular around the world. Increasing the share of wind energy in the market has led to the need for easy, inexpensive and effective monitoring and control approaches. In this thesis, various monitoring and control tools are proposed which are cheap and easy to implement in wind farms using existing system data. The first tool is focused on analyzing available data to have a better understand system behavior. Statistical analysis and clustering methods are proposed in this regard. It is necessary to prove that the wind farm operates efficiently and under control. For this reason, the second developed tool is on performance measurements recommended to monitor power generation efficiency. Data Envelopment Analysis, Malmquist Index Approach, and Stochastic Frontier Analysis are proposed to measure power production efficiencies of wind turbines. Forecasting methods have become more important as the wind energy market has increased. Therefore, simple forecasting methods are presented to show the use of available data as the third tool. Also, there is always error in each forecast, so the Particle Filtering approach is suggested to reduce errors. As the last tool, a new training algorithm is developed for multilayer perceptron artificial neural networks which called Antrain ANN. Results are showed that the proposed novel approach can compete with current algorithms.

Key Words: Wind Energy, Monitoring and Control, Performance Analysis, Particle Filtering, Artificial Neural Networks

ÖZET

SCADA VERİLERİNİ KULLANAN RÜZGAR ENERJİ SİSTEMLERİ İÇİN DURUM İZLEME VE KONTROL ARAÇLARI

EROĞLU, YUNUS
Doktora Tezi, Endüstri Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Serap ULUSAM SEÇKİNER
Haziran 2017
222 sayfa

Rüzgar enerjisindeki teknolojik gelişmeler yatırım ve işletme maliyetlerini azaltmıştır. Bu nedenle, rüzgar çiftlikleri dünyada daha popüler hale gelmiştir. Pazardaki rüzgar enerjisi payının artması, kolay, ucuz ve etkili izleme ve kontrol yaklaşımlarına ihtiyaç duyulmasına neden olmuştur. Bu tezde mevcut sistem verilerini kullanarak rüzgar çiftliklerinde ucuz ve kolay uygulanabilir çeşitli izleme ve kontrol araçları geliştirilmiştir. İlk araç, sistem davranışını daha iyi anlamak için mevcut verileri analiz etmeye odaklanmıştır. Bu bağlamda istatistiksel analizler ve kümeleme yöntemleri sunulmuştur. Rüzgar çiftliğinin etkin ve kontrol altında çalıştığının kanıtlanması gerekmektedir. Bu sebeple, geliştirilen ikinci araç güç üretim verimliliğini izlemek için önerilen performans ölçümleri üzerinedir. Veri Zarflama Analizi, Malmquist İndeks Yaklaşımı ve Stokastik Sınır Analizi, rüzgar türbinlerinin güç üretim verimliliklerini ölçmek için önerilmiştir. Rüzgar enerji pazarı geliştikçe tahmin yöntemleri de önem kazanmıştır. Bu nedenle, mevcut verilerin kullanımını göstermek için basit tahmin yöntemleri üçüncü araç olarak sunulmuştur. Her tahminin bir hatası vardır, bu yüzden Parçacık Filtreleme yaklaşımı hataları azaltmak için geliştirilmiştir. Son araç olarak, Antrain ANN olarak adlandırılan, çok katmanlı algılayıcı yapay sinir ağları için yeni bir eğitim algoritması geliştirilmiştir. Sonuçlar, önerilen yeni yaklaşımın mevcut algoritmalarla rekabet edebileceğini göstermiştir.

Anahtar Kelimeler: Rüzgar Enerjisi, İzleme ve Kontrol, Performance Analizi, Parçacık Filtreleme, Yapay Sinir Ağları

To My Parents

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor Prof. Dr. Serap ULUSAM SEÇKİNER for the continuous support of my study and related research, for her advice, patience, motivation, immense knowledge and insightful comments. I am thankful to Prof. Dr. Türkay DERELİ and Assoc. Prof. Dr. Mehmet İshak YÜCE for their insightful comments as members of my thesis progressive committee. I am also grateful to Assoc. Prof. Dr. Mehmet ATAK and Assist. Prof. Dr. Serkan ALTUNTAŞ for their contributions.

I would like to thank the authorities of the company for whom I have received the wind farm data used in this study and cannot give a name because of the Confidentiality Agreement.

I am grateful to Prof. Dr. Istvan ERLICH for his acceptance as a visiting researcher on my PhD topic at the Department of Energy and Power Systems of Duisburg - Essen University. I want to express my appreciation to The Turkish Scientific and Technical Research Council for giving me an opportunity to this research visit for six months with a scholarship program called TUBİTAK – BIDEB 2214 A.

I would like to express my special thanks to Özgür DEMİREZER who gave me the opportunity to attend a valuable project titled “Renewable Cooperation and Sustainable Partnerships Between Green Cities” which is co-founded by the European Union and the Republic of Turkey.

I also would like to thank Res. Asst. Selen YÜCESOY, Harika AKALIN, Fatma CEYLAN, Samet Kürşat BAŞOL, and all my friends at the Department of Industrial Engineering of Gaziantep University for their unconditional support.

Last but not least, I would like to thank my wife, Döndü, and daughter, Elif Ceyda, for their love, continuous support, patience, and encouragement during all phases of my thesis.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ABREVIATIONS

ACO: Ant colony optimization

ACO – BA: Ant colony optimization based algorithm

ACORSES: Ant colony optimization with reduced search space

ACO – NPU: New pheromone updated Ant Colony Optimization

AI : artificial intelligence

ANN : Artificial Neural Networks

ANOVA : Analysis of variance

ARSET: Adaptive random search technique

AWEA : American Wind Energy Association

BCC: Banker, Charnes and Cooper model

bn : Billion

CCR: Charnes, Cooper and Rhodes model

CM : Condition monitoring

DEA : Data Envelopment Analysis

df : Document frequency

DFIG : Doubly-Fed Induction Generator

DRASET: Dynamic random search technique

EC: Efficiency change

EPA : Environmental Protection Agency of United States

EU : European Union

EU-27 : 27 countries of European Union

EUR : Euro

EWEA : European Wind Energy Association

FDS : Fault detection system

GDP : Gross Domestic Production

GW : Giga Watt

GWEC : Global Wind Energy Council

GWh : Giga Watt hour

HAWT : Horizontal axis wind turbine

HRO: Heuristic random optimization

ICS : Industrial Control Systems

idf : Inverse document frequency

IEA : International Energy Agency

IGARSET: Improved genetic algorithm by random search technique

IPCC : Intergovernmental Panel on Climate Change

KE : Kinetic Energy

kVAr : kilo Volt Amper reactive

kW : kilo Watt

LAN : Local Area Network

MACO: Modified ant colony optimization

MAD: Mean absolute deviation

MAPE: Mean absolute percentage error

MEF: Mean error forecast

MI: Malmquist Index

MLP: Multilayer perceptron

MW : Mega Watt

m/s : Meter per second

NASA : National Aeronautics and Space Administration, United States of America

OECD : Organization for Economic Cooperation and Development

PEC: Pure Efficiency Change

PF : Particle Filtering

RPM: Revolution per minute

SCADA : Supervisory Control and Data Acquisition

SEC: Scale Efficiency Change

SES: Simple Exponential Smoothing

SFA : Stochastic Frontier Analysis

SLR: Simple linear regression

SMA: Simple moving averages

SMC: Sequential Monte Carlo

SSE: Sum of squared errors

SZGA: Successive zooming method

TEC: Technical Efficiency Change

TC: Technological Change

TUREB : Turkish Wind Energy Association

VAWT : Vertical axis wind turbine

VAr : Volt-Ampere reactive

WAN : Wide Area Network

WEC : World Energy Council / wind energy converter

wf : Word frequency

wrf : Word raw frequency

WMA: Weighted moving averages

CHAPTER 1

INTRODUCTION

The world population is increasing and the more people, the more needs. Thus, industrialization is inevitable. Energy has begun a common need for modern economies towards these developments. There are many ways to supply energy demand such as; generating energy from fossil fuels, generating energy from nuclear plants, generating energy through renewable energy source, and even countries buy direct electricity from other neighbors. Meanwhile, policy makers realize that energy demand of the country has to be supplied in a secure and clear way because they concern over dramatic climate change and energy supply security.

Especially after Gulf of Mexico oil spill (2010) and Fukushima disaster in Japan (2011), renewable energy resources have begun to have an utmost importance in the world. Renewable energy resources are local and clean energy production alternatives such as solar, bio-mass, wind, hydro, and geothermal energy. Thus, decision makers increase the interest in renewable energy systems and prepare energy policies, which are advantageous on renewable resources, to supply their energy demand. Also, because renewable energy sources are local resources, they have become important instruments for energy policies of countries on the way of secure energy production.

The wind energy is one of the most common used renewable energy resources worldwide. It has also proved itself over the last 20 years. The wind energy technology increases, the investment and operation costs decreases. Therefore, wind farms have become more popular on all over the world. Also, wind energy an important role in the production of local and cleaner energy in Turkey.

Wind Turbine technology is a rapidly developing sector of industry. Many turbine manufacturers such as Enercon, Samsung, Vestas are in a competition with each other to present a larger capacity turbine. One of the market leaders of wind turbine manufacturing sector, MHI Vestas, manufactured the largest offshore wind turbine prototype in the world (V164-8.0 MW). It was tested in 2014 and produced 192,000 kWh in a 24 hour period which was enough to power approximately 13,500 Danish households which demonstrates the full capability of the world's most powerful wind turbine [1]. Because the fact that wind conditions of offshore areas are more profitable than onshore, manufacturers and investors select the offshore environment. On the other hand, the price of operating environment and reduced accessibility are still barriers of offshore wind energy sector [2]. All these improvements of wind turbine technology shows the need of systematic, dynamic, and smart control strategies for spreading wind farms. The main contribution of this thesis is focused on constructing monitoring, control, and management tools using SCADA (Supervisory Control and Data Acquisition) data gathered from system of an installed wind turbine.

This chapter briefly explains the need of local energy resources by considering energy security and climate change. Then, energy market on the world and Turkey are investigated separately. Outline of the thesis is given in detailed and finally, main contributions are presented at the end of this chapter.

1.1 Energy Security

Energy security is important and current issue for especially modern economies [3–5]. Also, energy policy makers mainly focus on energy security. However, the definition of it has not been clearly defined [6–9]. This is an important obstacle for policy makers to measure and to balance against other policy objectives. Meanwhile, energy security is generally discussed through energy imports [10]. There are many studies which tries to explain the definition and components of Energy Security [6–15].

Kruyt et. al. (2009) presented an overview of available indicators for long-term energy security and provided four dimensions of energy security such as availability, accessibility, affordability and acceptability of energy [6].

Winzer (2012) reviewed the multitude of definitions of energy security and characterized according to the sources of risk, the scope of the impacts, and the severity filters in the form of the speed, size, sustention, spread, singularity and sureness of impacts [7]. It was illustrated how the selection of conceptual boundaries along these dimensions determines the outcome by using a stylized case study for three European countries. Thus, the definition of energy security could be defined as the continuity of energy supplies relative to demand.

A detailed and up-to-date survey of energy security was reported by Ang et.al. (2015) and they gave the definitions of energy security, changes in the themes of these definitions, energy security indexes, specific focused areas and methodological issues in the construction of these indexes, and energy security in the wider context of national energy policy [8]. According to their analysis, the definition of energy security is contextual and dynamic in nature. However, they still show that there is no consensus on a widely accepted definition of energy security.

Jonsson et.al. (2015) screened and scoped out a more comprehensive suite of energy security aspects to be considered when assessing low-carbon energy scenarios and apply it using the EU Energy Roadmap as an example [9]. They identified and discussed availability and affordability issues and geopolitical security aspects.

Hughes (2009) presented a methodology that could be used to explain energy security to the general public, policy analysts, and politicians by introducing the "four R's of energy security" as listed below [11];

"Review": understanding the problem, "Reduce": using less energy, "Replace": shifting to secure sources, and "Restrict": limiting new demand to secure sources.

A literature survey on energy security presented by concluding that energy security is composed of availability, affordability, efficiency, and environmental stewardship by Sovacool and Brown (2010) [12]. They analyzed the relative energy security performance of 22 countries from 1970 to 2007. They also analyzed in greater detail following four countries: Denmark which has the strongest energy security, Japan which has the most improved energy security, United States which has weak and stagnant energy security, and Spain which deteriorates energy security. Their study concluded by offering implications for public policy. The global and geopolitical dimensions of the future international energy security and its implications for Europe and the EU-27 were analyzed by another study [13]. Müftüler – Baç and Başkan (2011) asserted that Turkey would play an increasingly critical role for the future of energy security in Europe. They anticipated that Turkey was going to become a major energy hub for Europe in the near future because the most of the pipelines from the energy producing countries in the East to the energy consumer countries in the West would pass through Turkish territory [14].

Brown et al. (2014) contributed the correlation of energy policy by multidimensional concept of energy security and empirical performance over forty years [15]. They concluded after an analysis of 22 countries in the Organization for Economic Cooperation and Development (OECD) between 1970 and 2010 that many industrialized countries had made limited progress toward the goal of achieving secure, reliable and affordable supplies of energy while also transitioning to a low-carbon energy system. Their study offers the implications for energy policy more broadly and by providing empirical evidence that availability, affordability, energy efficiency, and environmental stewardship dimensions envelop the key strategic components of energy security.

By the way, the basic definition of energy security is given as “the uninterrupted availability of energy sources at an affordable price” by the International Energy Agency (IEA) [16]. It can be summarized as relationship between the availability of natural resources for energy consumption and national security. Also, cheap energy is essential for all modern economies to continue their competitive functions in production and manufacturing.

Energy security also has big importance role for the European Union (EU) because the EU imports more than half of all the energy it consumes [17]. The European Commission prepared its Energy Security Strategy in May 2014 aiming to ensure a stable and abundant supply of energy for European citizens and the economy. Figure 1.1 shows the energy systems approach [18]. If you want to secure your energy supply, renewable energy resources would be good options existing over wide geographical areas, in contrast to other energy sources, which are concentrated in a limited number of countries [19]. Because of being local sources, any type of renewable energy resource has many opportunities by considering its distribution loses, energy security, and economic benefits.

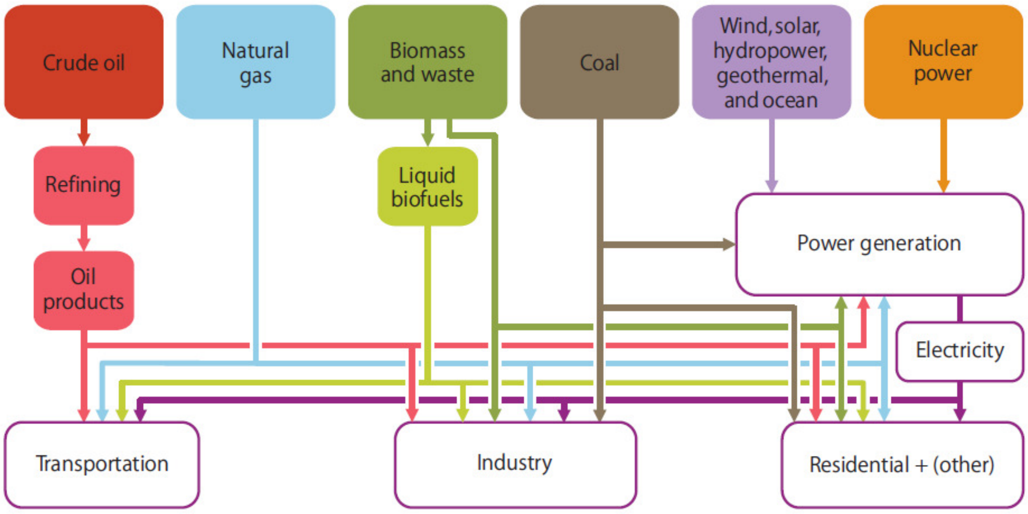

Figure 1. 1 Energy systems approach [18]

1.2 Climate Change

In spite of the fact that the Earth's climate has changed throughout history (Figure 1.2), there is a common sentiment that it is affected by industrialization in the way of our electricity generation [20–23]. Meanwhile, the increasing levels of CO_2 and other greenhouse gases, caused by dramatically increasing industrialization, have strong correlation with climate changes. Since the Industrial Revolution began in the 1700s after the invention of steam engine, a significant amount of greenhouse gases had been added into the atmosphere. With this invention, coal was being extracted from the ground at a previously unseen rate. Industrialization speed was also accelerated by steam engine. Burning fossil fuels to generate electricity, heat and cool buildings, and power vehicles are the main sources of green gases [24].

It can be simply observable from historical weather data that temperatures are rising according to past seasons, even snow and rainfall patterns occurs in a shifted month period. Figure 1.3 gives the yearly anomaly of weather temperature of the world. Especially after 1950s the anomaly is continuously increasing.

The definition of Climate Change is given as "any substantial change in measures of climate (such as temperature or precipitation) lasting for an extended period (decades or longer)" by the Environmental Protection Agency (EPA) of United States [24].

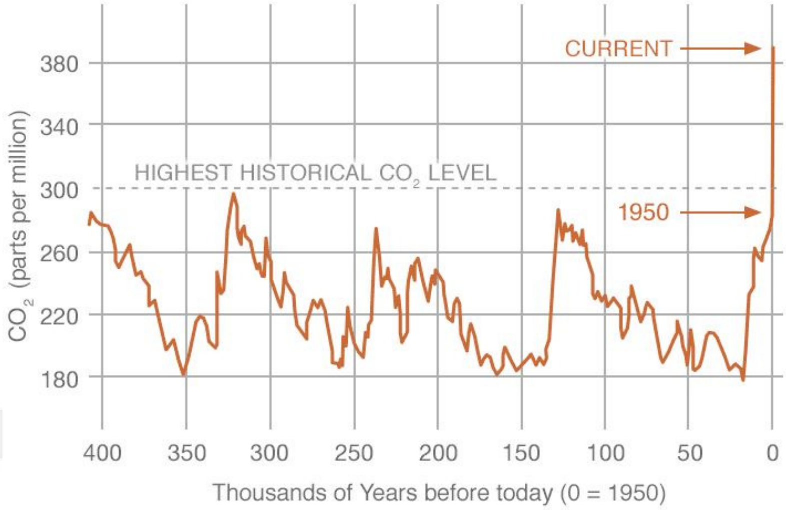

Figure 1. 2 CO_2 level of the world [23]

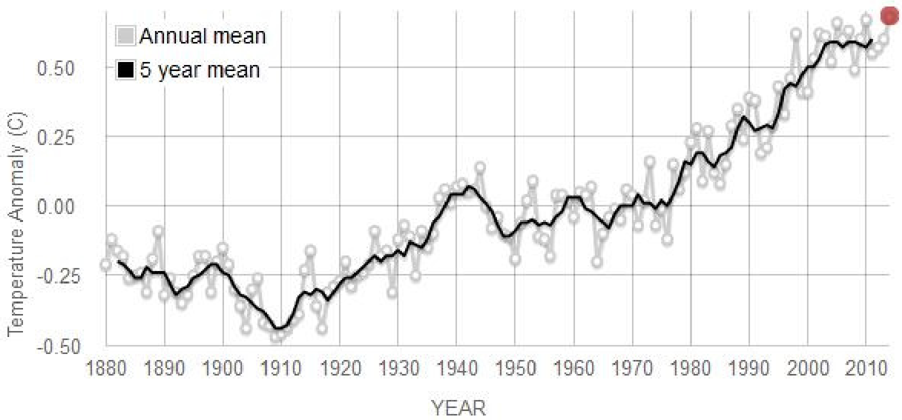

Figure 1. 3 Temperature anomaly of the world [23]

The Intergovernmental Panel on Climate Change (IPCC) prepared a report on Climate Change Synthesis [25] and it was concluded by a group of 1,300 independent scientific experts from all over the world that there were more than 90 percent probability that human activities over the past 250 years had warmed the Earth. Also, it was reported that atmospheric carbon dioxide levels had risen from 280 parts per million to 379 parts per million in the last 150 years by the industrial activities.

There are many negative effects of these changes such as rising of sea levels and melting of glaciers. Unfortunately, this changes the life cycle of plant and animal and can bring about fundamental disruptions in ecosystems. The worst case is that as climate change affects plant and animal populations – communities – and - biodiversity, society and traditional ways of life for certain communities such as where people can live, what kinds of crops are most viable, and what kinds of businesses can thrive in certain areas are also negatively affected [24].

Fossil fuels also can be considered as local energy sources and they may help countries to secure their energy demand as mentioned in Section 1.1. Besides the fact that fossil fuels cause the climate change, they are limited sources at the same time. Thus, renewable energy sources are major components of energy supply – demand equilibrium of modern economies which they produce needed energy in a secure and a clean way.

1.3 Energy Market

Energy is a complex market all over the world. This section summarizes the current energy reports [26–35] and identifies the projections for the energy share.

1.3.1 General situation

The energy market can be thought as the pillars of development, growth and competitiveness in modern economies. European Commission prepared a report on energy Prices and Cost [32]. The report says that while fossil fuel prices are expected to continue to rise and to drive energy costs, electricity costs are expected stabilize and then slightly decrease as fossil fuels are replaced by renewable energy beyond 2020. By the way, Europe's energy sector was discussed as it was in the midst of a major transformation. It was also reported that the gas and electricity market were moving through competitive private companies. Another significant observation was that wind and solar power in particular were strongly growing day by day. Thus, the electricity generation was being decarbonized [28, 32].

There is a report on “The World Energy Issues Monitor” prepared by World Energy Council (WEC) which was based on an annual survey [33]. Macroeconomic risks, technology, geopolitics, energy vision, and business environment were discussed in 40 issues. Over 1000 energy ministers, CEOs, and leaders from over eighty countries completed the survey. The survey results can be concluded that the experts remain more concerned about commodity price volatility, global recession and climate, electric storage, smart innovation on regional interconnection, and the need on new business models.

The WEC thinks that the increasing share of wind and solar power in the electricity mix causes great unpredictability of supply. Therefore, the need for new electricity market design is discussed in their study [33]. Also, The WEC wants to balance action-oriented framework "Energy Trilemma" (Figure 1.4) by their monitoring survey. Energy Trilemma consists of three main issues, Energy Security – Energy Equity – Environmental Sustainability, to be balanced. The World Energy Issues Monitor 2015 can be used as a tool to prepare new energy policies by considering energy Trilemma.

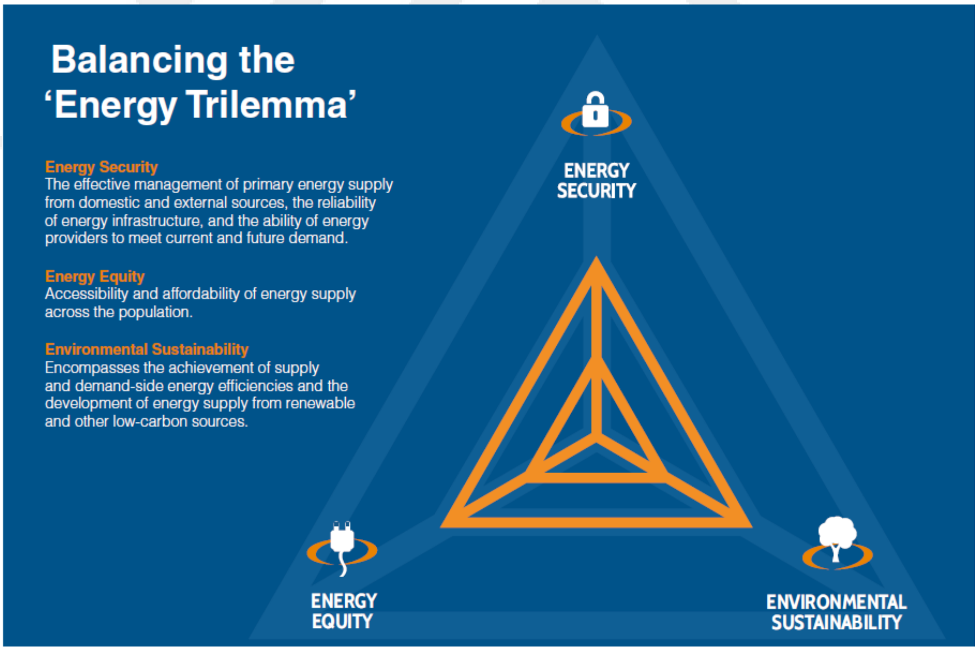

Figure 1. 4 The Energy Trilemma [27]

Figure 1.5 shows the results of the WEC survey on energy sector for the year 2015. Macro-economic factors, such as energy poverty – cyber threats – terrorism – talent – energy prices – commodity prices – capital markets – etc., have continues trend.

Energy and commodity prices, climate framework and electric storage are the most critical uncertainties in 2015 [27].

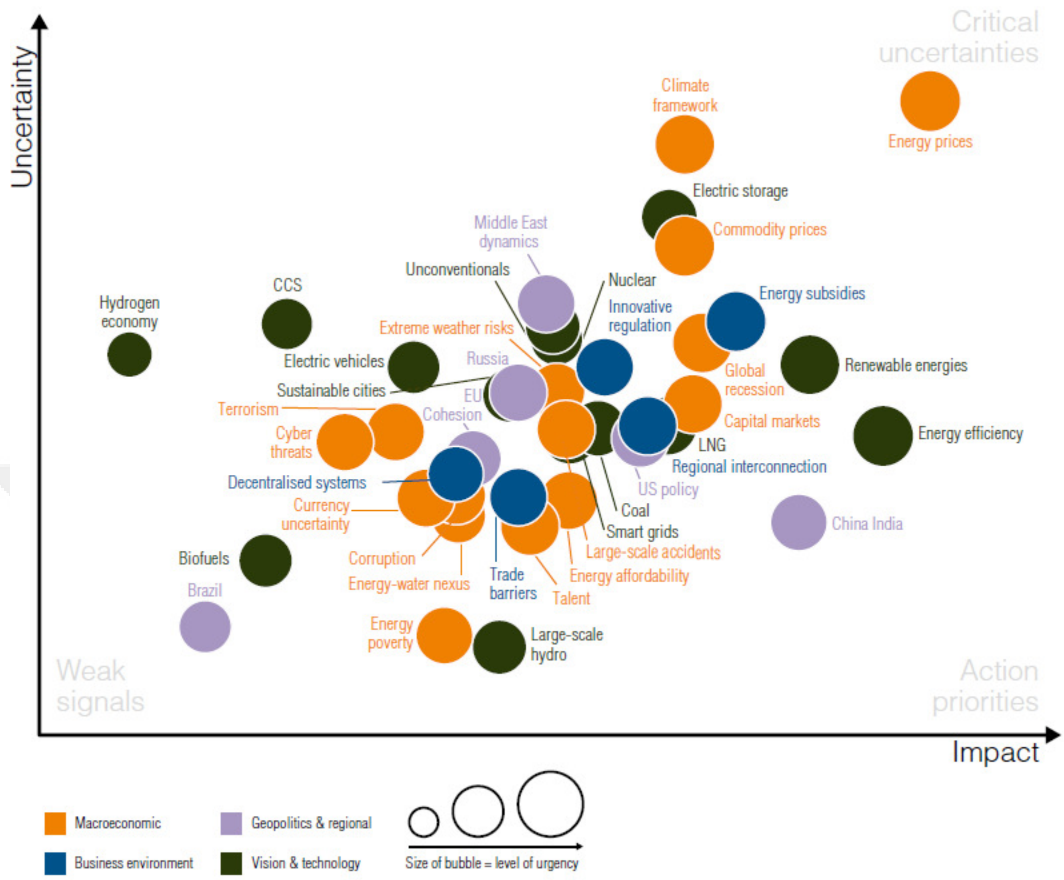

Figure 1. 5 Forty issues and their perceived impact, uncertainty, and urgency for global energy leaders and experts globally [27]

Technology is rapidly developing, thus, the energy systems become modernized, automated and interconnected to each other. Industrial Control Systems (ICS) and SCADA are common today's energy plants. Besides, cyber threats must be considered for the coming years [27]. On the other hand, the current global uncertainties are showed in Figure 1.6 [33]. While the hydrogen economy, nuclear, sustainable cities, etc. are less urgent subjects, still cyber – threats, commodity prices, climate framework, energy efficiency, renewable energy, market design, terrorism, etc. are more urgent subjects according to energy experts.

The British Petroleum Company presented "Energy Outlook 2016" [28]. The results of the report shows that the balance of energy demand shifts and shale gas, ultradeep-water oil, tight oil, or renewables are emerging new energy sources. In addition to this, energy production and consumption levels are affected by many disruptions, from wars to extreme weather. Therefore, energy security and climate change affected the new energy policies. On the other side, the fossil fuels are still dominant in the market with a share of 60 %. Nevertheless, renewables increasing the share of the market. Figure 1.7 shows the primary energy shares up-to-day and future projections [28].

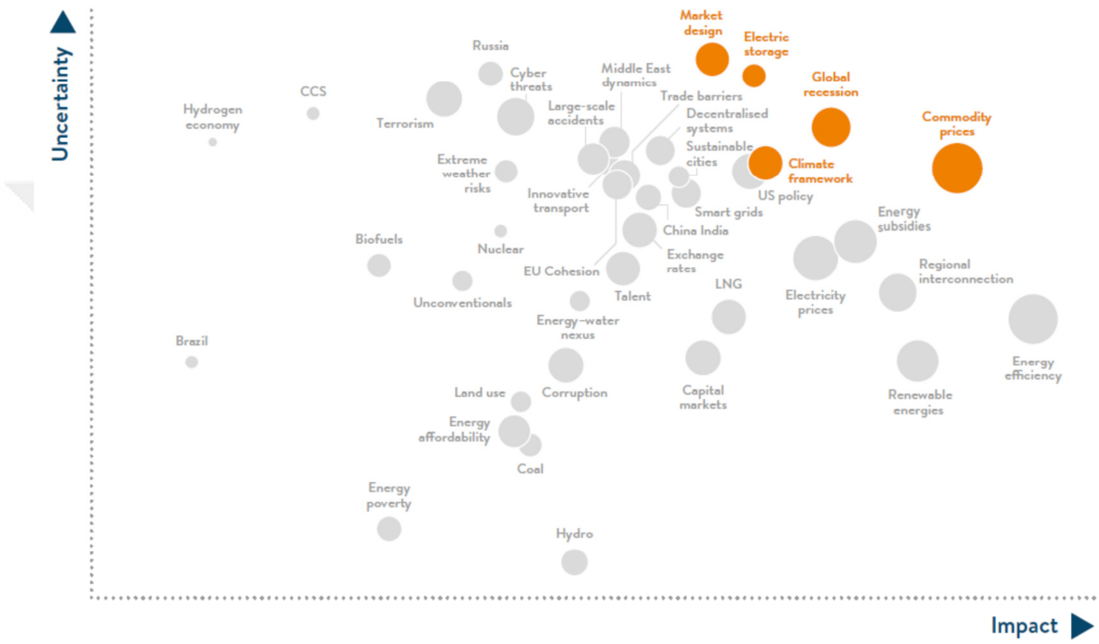

Figure 1. 6 Global critical uncertainties [33]

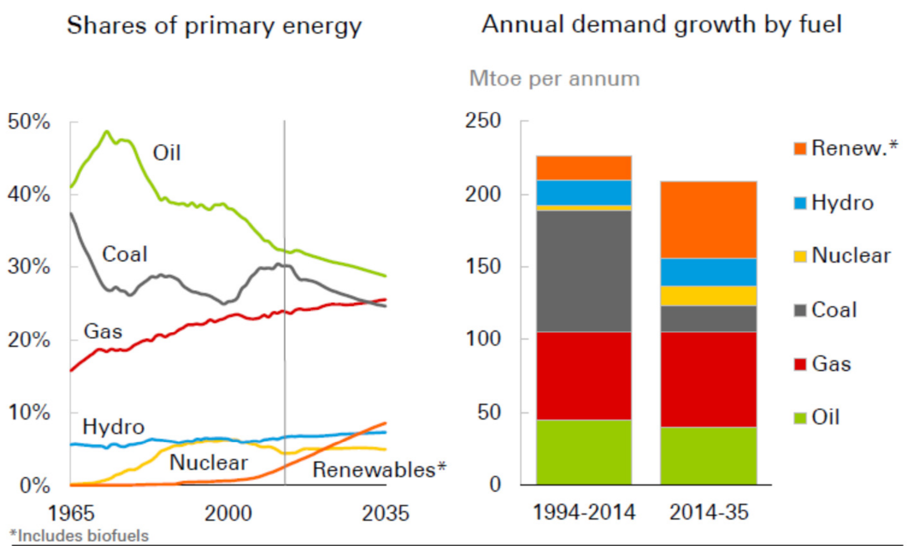

Figure 1. 7 Energy shares versus years [28]

The shift on the share of power generation can be seen in Figure 1.8 [28]. It can be easily observable from the figure that the largest shifts are the increase in the renewables share and the decline in the oil share. The British Petroleum Company projects the outcome by 2035 as more balanced and diversified portfolio of fuels for power generation in their report.

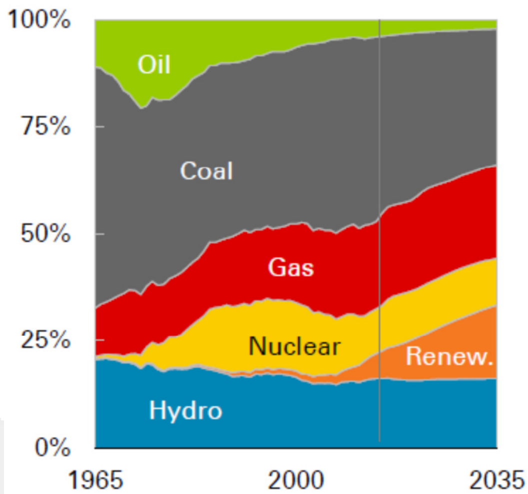

Figure 1. 8 Primary input shares to the power [28]

These general energy market situations guided us to study on renewable energy systems. We focused on wind energy systems which is the most rapid growing renewable energy market all over the world. The next section summarizes the wind energy market.

1.3.2 Wind energy on the World

Wind energy market is growing rapidly and many associations presents reports to follow the dynamics of the market [26, 29, 36]. Global Wind Energy Council (GWEC) presented its latest wind energy report for the year 2016 in April 2016. According to their report, 2015 had many records such as more than 63 GW installed in a single year (22% increase), wind power supplied more new power generation than any other technology in 2015, China installed 30.8 GW of new capacity and invest more than USD 100 billion in renewables in a single year, and Germany installed 6 GW of new capacity.

They considered that wind energy growth rate was much more than projected which could be a reason of decrease in the price of oil and concerns about climate change. Besides the fact that many people believe the effect of oil prices on the wind energy market, the GWEC regrets to compete with oil and considers that there is no evidence of this effect. By the way, wind energy is able to compete successfully on prices against other sources.

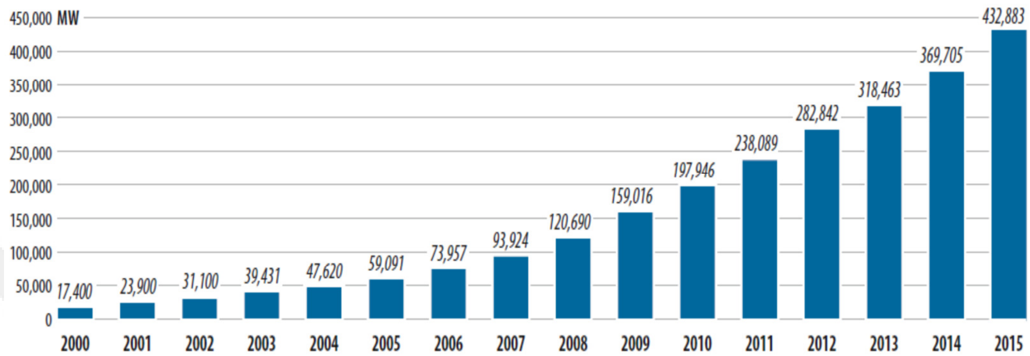

Figure 1. 9 Global cumulative installed wind energy capacity in MW between the years 1997 and 2015 [29]

The wind energy market is also a chance to the modern economies on the way of fighting Energy Security, cost stability, and air pollution [29].

Wind energy market is a global industry by nearly 100 countries. Figure 1.9 shows the global cumulative installed wind energy capacity between the years 1997 and 2015.

On the other hand, there is another valuable annually wind energy report that European Wind energy Association (EWEA) presented wind energy statistics of European Region in February 2015 [26]. They concluded that between EUR 13.1bn and EUR 18.7bn was invested to European wind energy market with a total installations of 11791,4 MW in 2014 with a 3.8% annual growth rate. Wind energy market had the highest rate for new installations in 2014 with 43.7% of total 2014 power capacity installations.

1.4 Energy Market in Turkey

1.4.1 General energy situation of Turkey

Turkey has one of the fastest growing energy markets in the world with a rapidly growing industrialization. Because the energy demand of Turkey increases, it is expected that Turkey will become one of the most dynamic energy economies of the world [29, 37]. Thus, Energy security is an important issue for especially Turkey. The energy policies of Turkey have priorities to the private sector for financing energy investments. Republic of Turkey Ministry of Foreign Affairs [37] explains their primary aim is to realize its own energy security and targets following objectives;

- Energy supply routes and energy source countries have to be diversified,
- The share of renewables has to be increased and the nuclear energy has to be considered in the energy mix,
- Energy efficiency has to be increased,
- Europe’s energy security has to be contributed by Turkey’s energy policies.

Electricity Market Regularity Authority of Turkey presented an Energy Market report in 2011 [38]. According to that report; policies are required to be performed to development of private sector, electricity energy sector has to become the most attractive sector for investors by carrying out related legislative regulations and applications, and production, transmission, wholesale, organized wholesale markets have to be considered as the activities of the main sector.

Turkish Electricity Transmission Company presented a five year energy projection for Turkey. They firstly analyzed the current energy demand, consumption and production situation. Then, they projected the energy market of Turkey up to year 2018.

Table 1.1 gives the Peak Power Demand and Energy Consumption values for the years 2004 and 2013. The peak power has generally increased yearly and energy consumption does. The electricity consumption of Turkey has increased 5,2% for the year 2012 and 2,5% for the year 2013. Except the years 2009 and 2013, energy consumption generally has increased above 4% [39].

Table 1. 1 Peak Power Demand and Energy Consumption of Turkey between 2004 and 2013 [39]

Year	Peak Power Demand (MW)	Energy Consumption (GWh)	Minimum Load (MW)	Minimum Load / Peak Load (%)
2004	23485	150018	8888	38
2005	25174	160794	10120	40
2006	27954	174637	10545	38
2007	29249	190000	11100	38
2008	30517	198085	10409	34
2009	29870	194079	11123	37
2010	33392	210434	13513	40
2011	36122	230306	14822	41
2012	39045	242370	13922	36
2013	38274	248324	14800	39

The projections of energy demand of Turkey are given in Table 1.2. These forecasts depend on the base level of previous Peak Power Demand and Energy Consumption data by considering energy policies of Turkey. The Peak Demand growth rate is over 4% and it grows up to 5,7 % then drops to 5,1% at the end of the projection. The Energy Consumption also tracks the similar growth rate with the Peak Demand [39].

Table 1. 2 Peak Demand and Energy Demand Base Level Projections of Turkey [39]

Year	Peak Demand (MW)	Energy Demand (GWh)
2014	40000	256700
2015	41850	271450
2016	44260	287310
2017	46640	302750
2018	49290	319980
2019	52110	338270
2020	55060	357430
2021	57940	376150
2022	60930	395540
2023	64040	415680

Energy production mix of Turkey is given in Figure 1.10 for the year 2013 and it does not represent the total energy demand. Thermal reactors have a big share of the energy production mix with coal, oil and gas reactors. On the other hands, Renewable Energy resources have great importance in the energy mix of Turkey to improve its energy security, because it has limited oil and gas reserves. Global Wind Energy Council concludes that more renewable energy investments are needed to match rapidly growing energy demand and to achieve the 2023 projections of Turkey (renewable energy sources have to provide 30% of Turkey's electricity) [29].

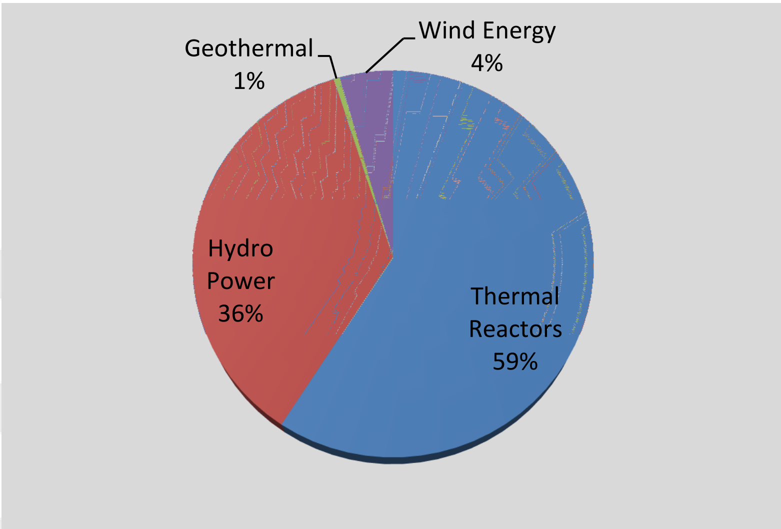

Figure 1. 10 Installed Power Capacity Share of Turkey for the Year 2013 [39]

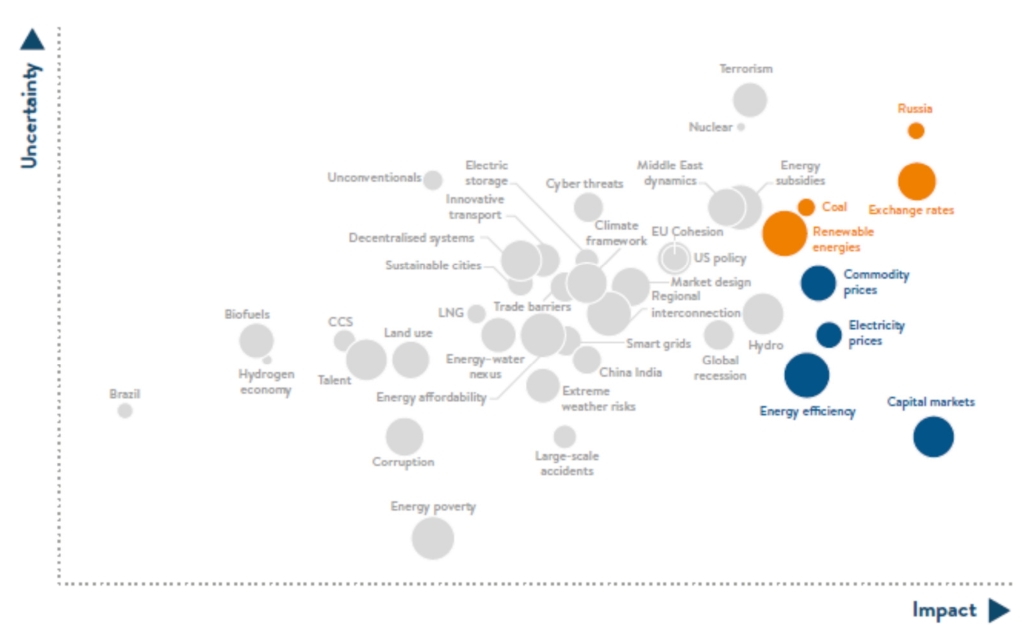

Figure 1. 11 Critical uncertainties of Turkey

There is a summary key points of World Energy Issue Monitor survey of Tukey section in Figure 1.11 [33]. The geopolitical characteristics of Turkey are key determinant for energy leaders. Russia is at the epicenter of the country's energy concerns. Turkey imports 75% of the country's primary energy demand (93% consumed oil, 98% of consumed natural gas). Besides, Turkey uses natural gas to generate almost 50% of its electricity. Therefore, Russia indirectly influences 25% of the electricity generation of Turkey.

Although the coal is the base for national energy production, there are projects on hydro-energy. Renewable energy has an uncertainty which causes remarkable concern. There is a huge unutilized wind, solar, and hydro energy potential.

Renewable energy has the utmost priority in the plans of the Minister of Energy and Natural Resources. Because of the energy investments in Turkey are financed international loans; volatile exchange rates are another critical issue. So that, energy investors have to find alternative financing sources. By the way, energy efficiency remains a critical issue because Turkey's primary energy intensity still lags behind the OECD countries.

1.4.2 Wind energy in Turkey

Turkey has given much importance to wind energy in recent years. The Turkish Wind Energy Association (TUREB) is an important organization who creates awareness about wind energy, creates unity of wind energy investors, and makes sectoral studies by creating statistics and market reports. gives the current wind energy statistics in a current report up to date January 2017 [34]. Figure 1.12 gives the wind energy investments of Turkey for the years 2007 – 2016. There are 152 wind farms with a 6106,05 MW total wind energy installed capacity. The increase of new installed capacity is bigger than previous years with a 1387,75 MW.

Doloitte Danışmanlık A.Ş. and Elia Grid International prepared a Turkish Wind Energy Association Wind Energy and Interaction Report for Turkey (2016) [35]. The report gives a projection of 1000 MW of new installed capacity for each year up to 2035 and summarizes its interaction of Turkey's economic dynamics. If there is at least 1000 MW new wind capacity installation every year; total capacity will reach at least 25 GW in 2035, there is increase on employment (there would be an employment multiplier factor of between 2,7 and 3,4 per MW), there is increase on Gross Domestic Production (GDP) (there would be an extra addition on GDP between USD 374000 and 423000), there is decrease on Current Account Deficit (with a total of USD 12,9 billion), and the environment would be protected (with an equal tons of 279,6 million carbon emission).

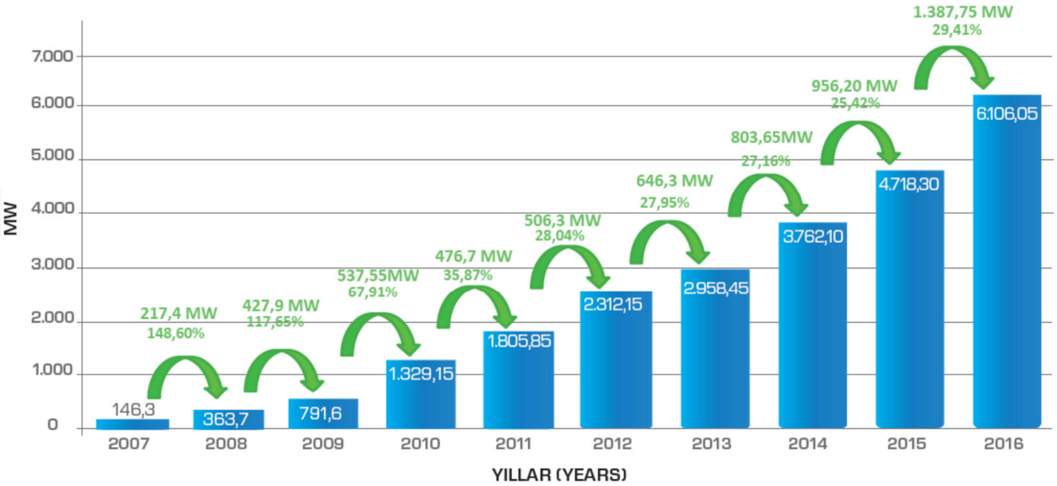

Figure 1. 12 Turkish wind energy market for the years 2007 – 2016 [34]

The wind energy potential of Turkey is summarized in Figure 1.13 as a topologic map. It can be easily concluded that the best wind resources are located near Balıkesir, Çanakkale, Hatay, İzmir, and İstanbul.

Turkey has many advantageously Renewable Energy polices for the investors. Turkey's Renewable Energy Law says that the feed-in tariff is set at EUR 6,5 cent/kWh for wind power, for a period of ten years, also, if the locally manufactured components is used in the investment, an extra bonus of up to EUR 3,3 cent is presented for up to five years (The Renewable Energy Law Number 5346 dated 18th May 2005).

Wind energy investors are able to sell produced power to the national power pool or they can even engage in bilateral power purchase agreements. Additionally, they also have 85% discount for the right of easement on State owned land for transportation and transmission, which will be applied during the first ten years after the establishment of the wind farm, for the investments that begin operations before 31 December 2020.

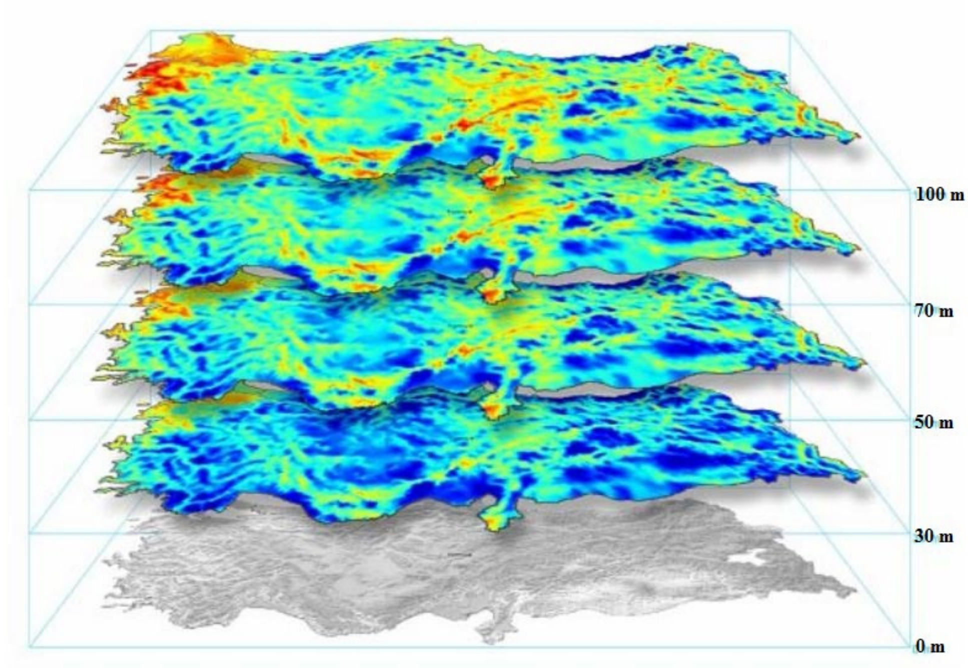

Figure 1. 13 Wind energy topology of Turkey at 30m – 50m – 70m – 100m heights

These positive policies encourage the investors to play a role in wind energy market of Turkey. The market is mostly dominated by local developers in Turkey. Polat Energy (560,8 MW), Demirer Holding (408,9 MW) and Guriş (404 MW) are leading investments followed by Bilgin Energy (371,5 MW) and Borusan EnBW (331,7 MW) [34].

In spite the fact that these advantageously renewable energy law helps the rapid growth of the wind energy market and there are some issues still needing to be mentioned such as; immature electricity and gas market have to be handled to predict the market prices, technical difficulties in transmission have to be solved, continuous and predictable grid connection capacity allocation have to be handled, and long administrative procedures have to be shortened [35].

The projections of the market is given by the Turkish Wind Energy Association such that Turkey will have to install a new capacity of 1,000 MW for each year to reach the 2035 goals [35]. By considering the projections, Turkey is one of the biggest on-shore wind markets in Europe with its wind power projects for the next years.

1.5 Structure of the Thesis

The structure of the thesis is constructed to reflect the direction, progress and results of research since September 2011. The eight chapters of the thesis present different areas of research and summarized in the Figure 1.14.

Thesis starts with the introduction part, Chapter 1, which has briefly discussed the basis for interest in wind power and introduced the reader to general energy market all over the world. This chapter provides the motivation of the related energy sector, explains the reasons of selecting wind energy area by mentioning the energy security and climate change threats, and it is reinforced with actual energy reports.

Chapter 2 is constructed as a technical background part. Since this thesis is presented in Industrial Engineering area, the technical information is a common need to understand related wind farm place. Therefore, an overview on wind energy concepts is presented here. Also, SCADA is introduced in this chapter. It is the main data source of this thesis and a typical SCADA data system is explained in Chapter 2. The objective of this chapter is to provide the unfamiliar reader with a comprehensive technical overview of the wind energy terminology. Thus, this chapter is suggested for the readers who think that he/she is far away from wind energy and SCADA system concepts.

This thesis is differently planned from many other dissertations in terms of defining the main problem in a research field. Generally, research problem is defined at beginning of the doctoral study, then researcher studies on giving solution ideas for the problem. In this thesis, the main research field was defined at the beginning. Then, the main problem was set forth through a systematic and intelligent literature review using one of the Industrial Engineering tool, text mining, as a new literature review tool which can be used as a neoteric approach to find the hot topics of general literature. Therefore, Chapter 3 presents the methodology of defining the trend topics of wind energy literature.

Chapter 4 presents main definition of the problem. Chapter 3 gives the clue on the importance of condition monitoring, control algorithms and using smart managerial tools for the wind energy systems. Thus, the main problem was defined to contribute new and smart managerial tools for optimal monitoring control strategies using SCADA system of the wind farms. Then, the focus of the chapter moves from background literature in the field.

Chapter 5 provides a statistical analysis tool of the SCADA data. In this chapter, a new clustering approach was presented for the wind farms to identify similarities and differences between wind turbines in a single wind farm. This approach allows the managerial team to identify characteristic behavior of each wind turbine in the farm. Thus, they can apply individual managerial approach for the turbines or turbine clusters. For example, the managerial team is able to prepare a preemptive maintenance plan using similarities of operational behavior of the wind turbines.

The performance monitoring of wind turbines is lack of literature. As the literature is carefully reviewed, it can be concluded that cheaper, important, and easy-implemented strategies for performance monitoring and control of wind turbines have many advantages (to define low performance turbine, to make a strategic maintenance plan, to prevent possible faults) to fill the gap in the literature. Thus, a Data Envelopment Analysis (DEA), Malmquist Index (MI), and a Stochastic Frontier Analysis (SFA) approaches were proposed to monitor current technical power production performances of the wind turbines. The studies were given in Chapter 6.

Addition to these approaches, a Particle Filtering (PF) algorithm was proposed as an alternative smart and cheap tool by using only SCADA data to fine-tune the power production forecast. Contributed heuristic algorithm was presented in the Chapter 7.

A new Training Algorithm for Artificial Neural Networks (ANN) is developed and presented in Chapter 8. First of all, a continuous ant colony algorithm with a new pheromone updating methodology (ACO-NPU) was presented to solve global minimization problems. Secondly, ACO-NPU was adopted to train ANNs which is called *Antrain ANN*. Finally it was used to model different case studies including wind speed, wind power, and wind turbine fault forecast.

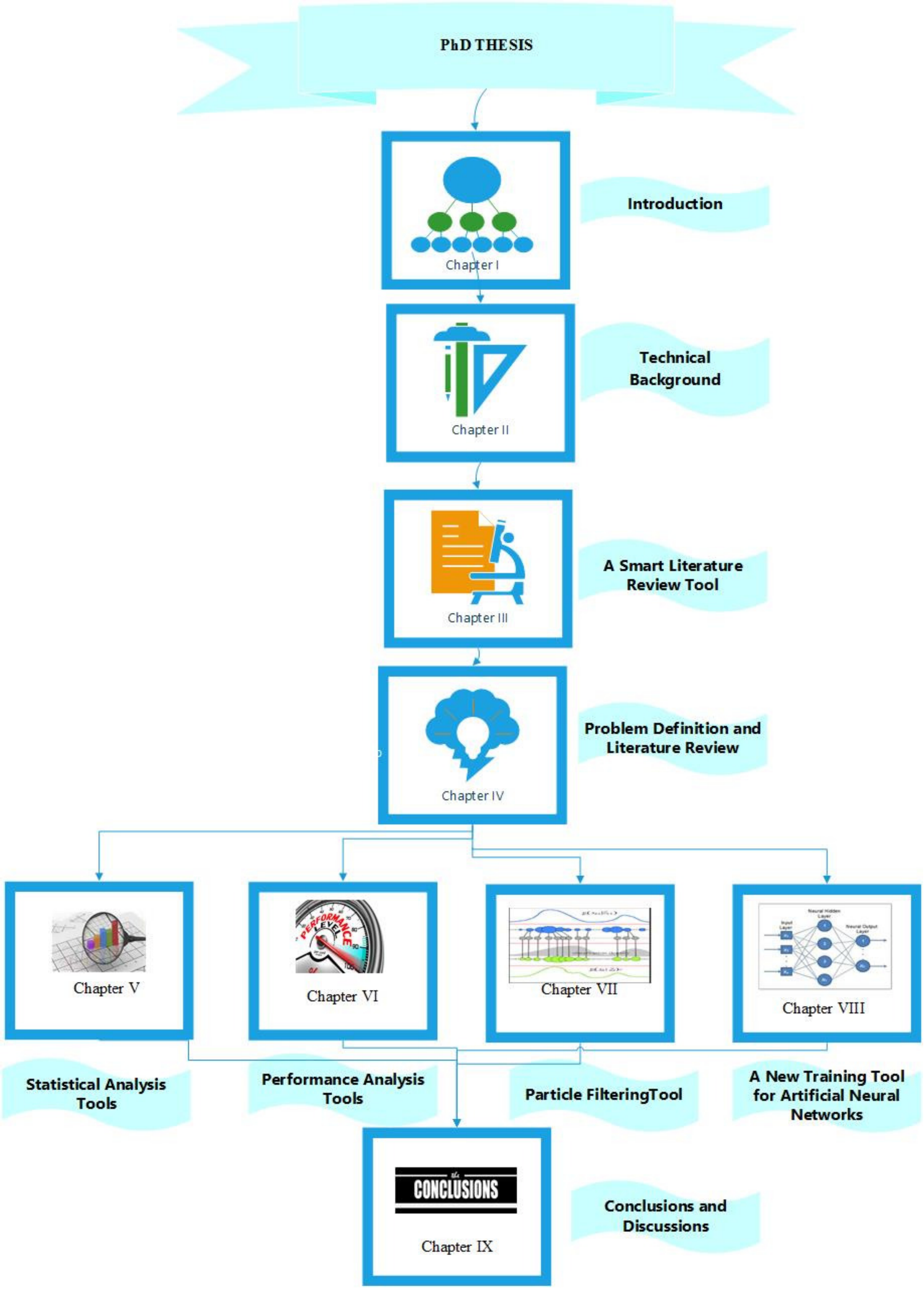

Figure 1. 14 Main Structure of the Thesis

Finally, Chapter 9 gives the conclusions and discussions of the contributed managerial tools with recommendations for further studies.

1. 6 Original Contribution of the Thesis

The main subject of this thesis was constructed after systematically analyzing of all possibly related articles by Text Mining approach. Text mining helps the researchers on finding the possible gaps and new trends and hot topics on a defined general area. Due to the applied process is a strong and noetic approach in Industrial Engineering area, this procedure can be handled as the first contribution of this dissertation. The study firstly titled as “Text Mining For Systematic Literature Review on Wind Energy” and presented in an international conferences, The Second European Workshop on Renewable Energy Systems [40]. Then, it is re-studied by considering the discussions on the conferences and prepared and published as a journal article [41].

The results of the systematic and intelligent review of all wind energy related articles in the Thomson Reuters Database Web of Knowledge shows that the control is the hottest topic in wind energy. Then, after a deep literature on control algorithms and methods on wind energy sector, it was seen that the common problem to control a wind farm/turbine could be handled by Condition Monitoring systems. Current literature in Chapter 4 shows that condition monitoring systems have become a commonplace feature of wind turbines. Actually, all new turbines have some form of condition monitoring systems which have advanced significantly over recent years. However, many of these are based on the basic Fourier transform as applied in industries where stationary conditions are standard and they are not especially designed for wind turbines. Therefore many commercially available condition monitoring systems are expensive and they still need a large degree of manual analysis to interpret results.

This PhD thesis is inspired by the perceived lack of appropriate dynamic, smart and cheaper solutions to control and monitor for wind energy systems. In this regard, this thesis proposes different tools to help managers on condition monitoring area of the wind energy market by using current SCADA data.

Developing smart condition monitoring system approaches within the domains of wind turbine power generation in a cheaper way are the main objectives of this thesis. The thesis aims to exploit system data which is already collected as part of SCADA systems, which are installed as standard on most modern wind turbines.

The all data utilized was collected from real-world operational wind farm in Turkey. The names, addresses and turbine brands and models of the considered wind farm have not been disclosed due to the Confidentiality Agreement with the company authorities. The developed solutions have demonstrated capabilities and applicability to real world condition monitoring applications. Therefore, another contribution is demonstration of how existing SCADA systems could be used to monitor large wind turbines.

The original contribution of this thesis is the development of an efficient method for the automation of Condition Monitoring systems which is based on the progressive process of algorithm development and testing through the use of industrial information and knowledge.

The best way to manage and control on organizations is going on the way of understanding the best of its data. The SCADA data is a big data and need to be re-organized and re-structured to be totally used to gather meaningful managerial information. Thus, the first tool is Statistically Analyzing of the SCADA data. So that, one can identify the main characteristics of the wind turbines. In the current literature, generally wind farm is investigated as a whole. What about the individual wind turbines or clustered ones? It is known that investigate all individual wind turbines is very expensive procedure but also thinking as a whole may also be wrong. So that, a clustering of wind turbines in a single site using SCADA data was contributed in this thesis and presented in an international conference [42]. It is also planned to publish the same study in an international journal after improving.

New performance monitoring tools are other crucial contributions of this thesis. In the wind energy researches, performance is generally thought as in the mechanical perspective. On the other hand, there are many tools which can measure and monitor economic performance of the organizations. DEA is only one of them which only considers economic borders. Thus, a DEA was contributed to monitor wind turbines individual production efficiency. On the other hand, because of the nature of DEA, it is lack of measuring time periodic performance monitoring and parametric effects of performance. Thus, a Malmquist Index approach was also applied to monitor time dependent performances of the wind turbines and published in a scientific journal [43]. Also, a SFA was contributed to identify parametric effects on wind turbine power production efficiency. It was presented at an international conference [44] and its improved version is accepted in a journal [45].

Power forecasting in the wind systems is an important issue especially when it is obligated to give forecasted data to a Central Wind Energy Monitoring and Management System as in Turkey. In the current literature, there are many wind power forecasting methodologies/algorithms. In this thesis, a Particle Filtering tool was developed to fine-tune the errors of power forecasts and presented in an international conference [46] as condition monitoring algorithm to track and forecast the wind power generation. The notable feature of this algorithm is that, it uses only SCADA data to track generated power and able to identify the anomalies in the wind speed and generated power curve.

Finally, a novel training algorithm for ANN, Antrain ANN tool, was introduced. First of all, ACO-NPU was developed to optimize global minimization problems and presented in a journal [47]. Then, it was modified to train ANNs. Finally, it was applied to model different types of case problems. The best part of the Antrain ANN algorithm is that it can be applied to all kind problems in general including wind speed, wind power, and turbine faults forecast.

The proposed tools in this thesis are totally novel to the literature. They have been presented in several conferences and sent to publish in several international journals. The developed tools can easily be used in industries for condition monitoring as smart managerial tools.

1.6 Conclusion

The main reasons for selecting wind energy sector were summarized in this introduction part of the thesis. Firstly, the need and importance of energy introduced. Then, as parts of threats on energy field, energy security and climate change were explained briefly. Because energy statistics are generally reported annually, it is necessary to follow organizations ant to keep up to date on the energy and wind energy market, you have to follow listed organizations below in order to stay up to date.

- European Wind Energy Association (EWEA),
- Global Wind Energy Council (GWEC),
- International Energy Agency (IEA),
- Turkish Wind Energy Association (TUREB),
- World Energy Council (WEC),
- World Wind Energy Association (WWEA).

The general energy market on the world was mentioned with current energy reports. It can be easily seen from general market reports that after Paris Agreement (The CoP21 Agreement), renewable and low -carbon emission energy sources have gained more importance than before. Wind energy market was also introduced in detailed as a sub part of the energy market. Then, Turkey's energy market was introduced with a separate part of wind energy market.

The main structure of this dissertation was also given here with the reasons for the readers. Finally, the original contribution of this thesis was mentioned. In the next chapter, the technical background about wind energy and SCADA systems is presented.

CHAPTER 2

TECHNICAL BACKGROUND

Wind energy may be somehow indirect subject for audience of this thesis. Therefore, the technical background of wind energy was mentioned in this section. This section does not contain detailed wind related definitions, formulations and issues, but it gives a brief summary on wind energy concepts to help unfamiliar readers. Thus, the main purpose of this chapter is to give the core definitions and statements of wind, wind energy, use of wind, wind turbines, and SCADA systems.

2.1 What Is Wind

Since the air is cooled or warmed by the sun, there are high and low pressures in the world's atmosphere. The free air mostly moves from high to low pressure, which is basically called "wind". The greater the difference in pressure, the faster the air flows.

The wind type is generally classified by its speed, its spatial scale, the types of forces that cause it, the regions in which it occurs, and its effect. There are many forms of winds; thunderstorm, local breezes, and global winds [48].

The cycle of sea breeze and land breeze can define local winds in coasts. The breezes between mountain and valley are able to dominate local winds in variable terrain areas. Figure 2.1 shows the circulations of air for sea breezes and mountain-valley winds at night and day. On the left side, Sea breezes are illustrated. As they have different rates of heating and cooling, the different heat capacities of sea and land cause Sea breezes. Meanwhile, when cool mountain air warms up in the morning and, it begins to rise, at the night, vice versa [48].

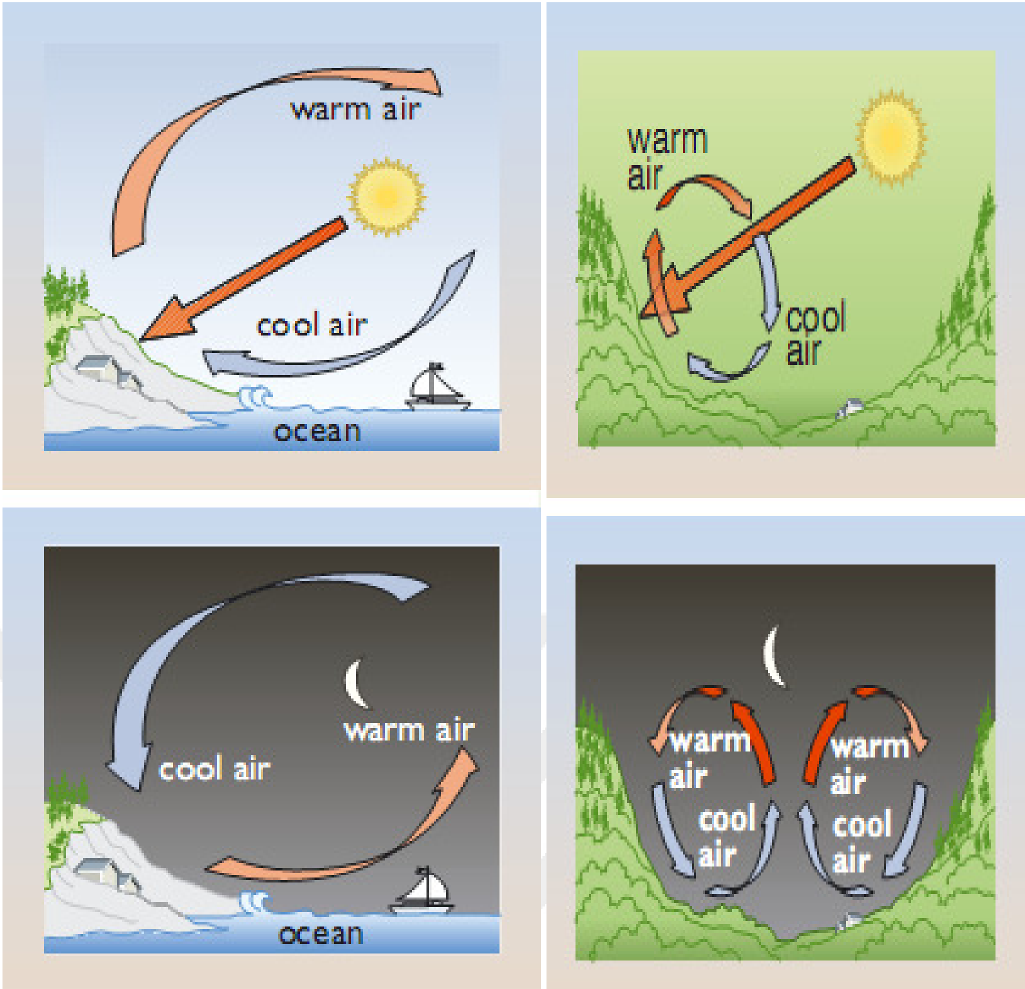

Figure 2. 1 Wind circulations at night and day [48]

The wind is commonly described with the speed and direction of blowing. Also, the speed of it, the density of the gas (air), and the energy are the various aspects of wind. Wind is a common need for humanity and the wild life in some issues such as shaping landforms, moving dost from large deserts, affecting the spread of wildfires, and having a negative impact on livestock when combined with temperatures [48].

2.2 Energy And Power In The Wind

Wind energy has commonly been used for thousands of years for milling grain, pumping water and other mechanical power applications [48–53]. Today's technology allows us to use wind in generating electricity by wind turbines. The energy contained in the wind is its kinetic energy and can be evaluated by following Equations (2.1 – 2.4);

$$KE = 0{,}5mV^2 \qquad \text{Equation 2.1}$$

where *KE* stands for kinetic energy, *m* is mass in kilograms and *V* is velocity in meters per second.

In wind energy, *m* has to be taken as the mass of air per second in Equation 2.1 and can be evaluated by following Equation 2.2;

$m_{air}= \rho V_{air}$ Equation 2.2

where m_{air} stands for mass of air per second, ρ represents air density, and V_{air} is the volume of air flowing per second and can be evaluated by the Equation 2.3;

$V_{air} = A_r L$ Equation 2.3

where *L* stands for length of cylinder of air flowing per second and can also be represented by the velocity of the air per second *V*, A_r represents the area.

Thus, kinetic energy per second on the wind can be calculated by the following Equation 2.4;

$KE_w = 0{,}5\ \rho\ A_r L V^2$ (joules) Equation 2.4

where ρ is in kilograms per cubic meter, A_r represents the area of blowing air in square meters, and *V* is in meters per second.

Energy per unit of time is equal to power, and then the power (P) in the wind is given by following equation 2.5,

$P = 0{,}5\ \rho\ A_r\ V^3$ (watts) Equation 2.5

Figure 2.2 gives an example and represents the energy creation of the wind through a wind turbine. In this example, it is assumed that the wind turbine blades rotate in a circular area A_r = 100 meter squared, wind speed V=10 meter per second, air density is 1,29 kg per meter cube, then we have 64500 watts (joules per second).

The wind energy is used to generate electricity firstly in the late 19th century. It was still limited to small scale by the middle of the 20th century. The total power from was less than 1000 MW in 1980s [50]. Then, wind technology has improved significantly day by day [54]. Modern turbine technology presented 2000 MW turbines in 2000 and today, a single wind turbine can produce up to 8 MW electrical powers.

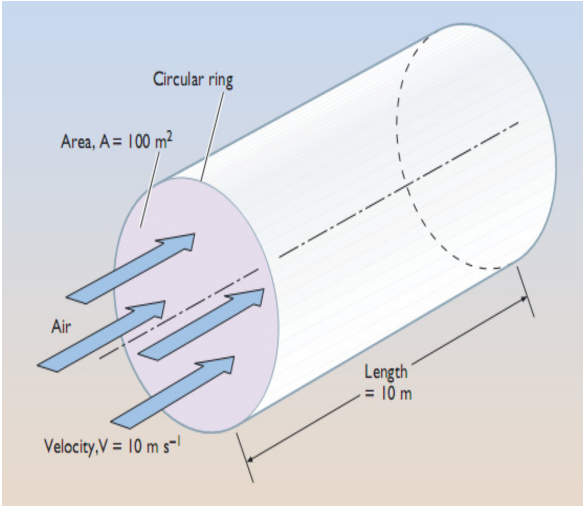

Figure 2. 2 Wind Energy Creation [48]

Wind energy is by far the fastest growing renewable energy source as mentioned in the Chapter 1. It is free, clean and can be thought as endless. Furthermore, the cost of the electricity produced by wind turbines is nearly fixed as management costs once the plant has been built and it has already reached the point where the cost of the of electricity produced by some of the conventional, fossil based power plants [49, 55]. Now, the technology is continually being improved to make it both cheaper and more reliable, so it can be expected that wind energy will become even more economically competitive over the coming decades. Thus, Wind Farms, which are constructed to generate more electricity in a single site and have more than one wind turbines, are grown up all over the world.

2.3 Wind Turbines

The term turbine is used for rotating mechanical devices that extracts energy from a fluid flow. A turbine is a turbomachine with at least one moving part called a rotor assembly, which is a shaft or drum with blades attached. A turbine converts the energy of moving fluid on the blades so that a rotational energy occurs [56]. Thus, a wind turbine can be considered as a device that converts kinetic energy from the wind into electrical power.

Beside the first known wind machine is Hero of Alexandria in history [57], the first known practical windmills were built in Sistan (eastern Iran) at the 7th century which vertical axis with long vertical drive shafts and rectangular blades [58]. Europe met with windmills during the Middle Ages with the first historical records of 11th or 12th centuries in England and German crusaders had taken their windmill engineering skills from Syria around 1190 [51, 59, 60]. German windmills were in use to drain areas of the Rhine delta by the 14th century and then windmills became an important tool. Generally, first windmills of Europe were used for grinding, but some of them were especially used to pump river water to the land located below sea level in the Netherlands [51]. The engineering skills of the windmills had been depended on trial and error principals between the 18th and 19th centuries, then theories were developed which were able to introduce new designs by improving the efficiency of energy conversion. In the middle of the 18th century there were many windmills in Denmark, England, Germany and the Netherlands.

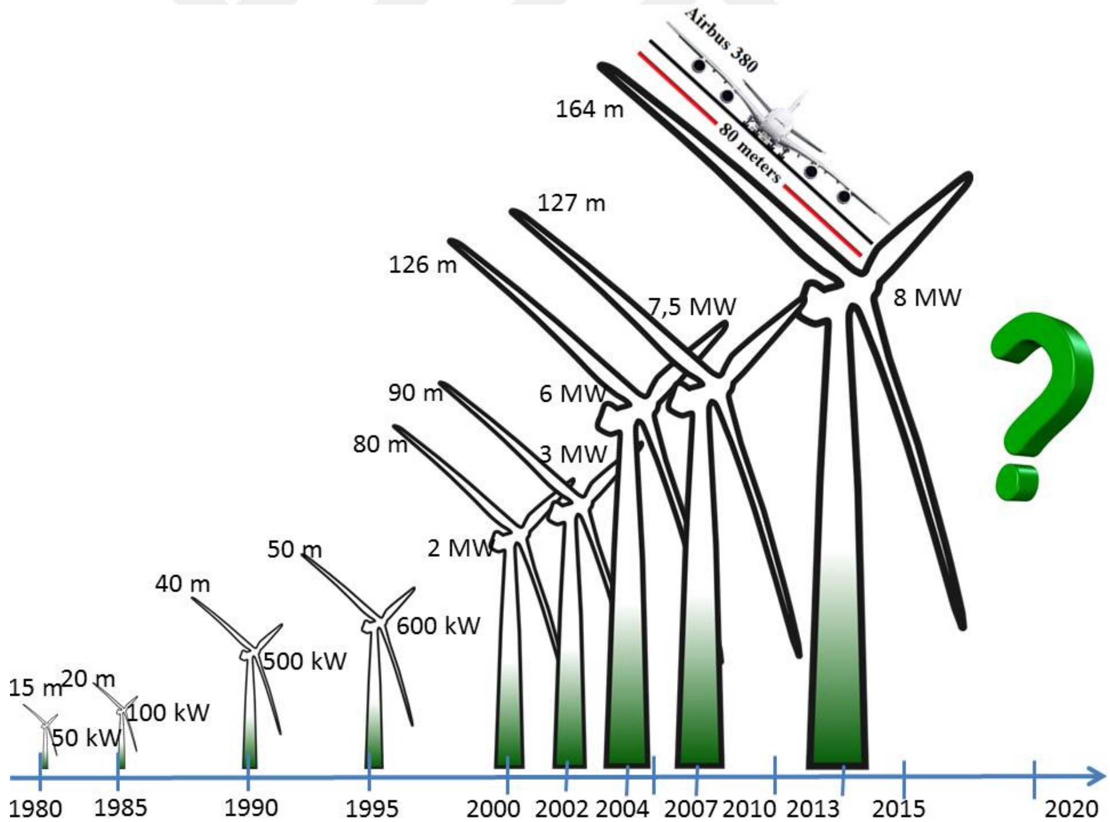

Figure 2. 3 Evaluation of modern wind turbines [43]

The wind turbines were firstly used to generate electricity in the late 19th century in the United States [49, 51, 61]. Charles Brush invented one of the first examples of wind turbines in Cleveland, Ohio and it operated for 12 years, from 1888 to 1900 supplying the needs of his mansion [51]. It was multi-vane type with 144 blades, a diameter of 18.3 meters, and the hub height was 16.8 meters [61]. Nowadays, wind turbine technology is one of the most rapidly growing renewable energy markets. Figure 2.3 summarizes the last improvements after the 1980s [43]. The largest capacitated wind turbine is manufactured by MHI Vestas Offshore in 2013 and it is as wide as double wingspan of Airbus A380 with a 164-meter diameter [1].

Modern wind turbines are generally categorized by their rotation blades across the wind. There are two categories, horizontal axis and vertical axis wind turbines. Figure 2.4 simply demonstrates the working principles of these two types and following sections give detailed information.

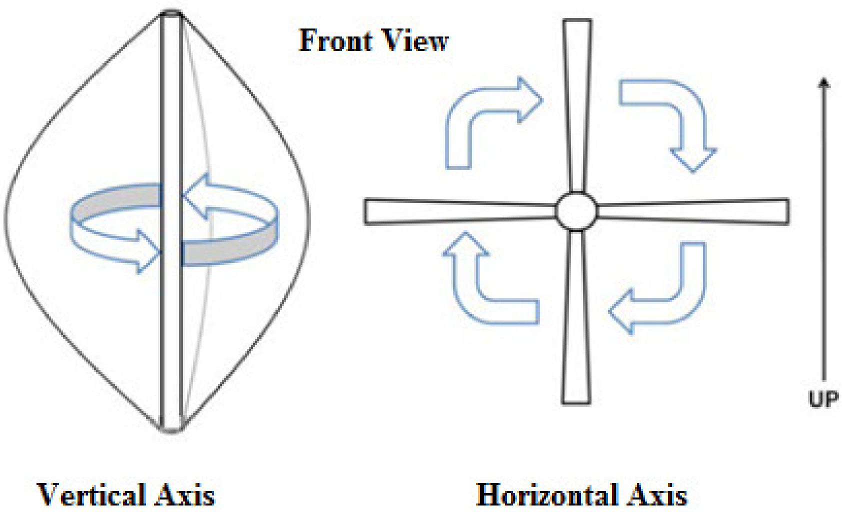

Figure 2. 4 Working principles of Horizontal and Vertical axis wind turbines [50]

2.3.1 Horizontal Axis Wind Turbines

The main characteristic of a horizontal axis wind turbine (HAWT) is that the blades rotate around a horizontal axis (Figure 2.5). They have the main rotor shaft and electrical generator at the top of a tower, and must be pointed into the wind. They also have a gearbox turning the slow rotation of the blades into a quicker rotation. The turbine is mostly positioned upwind of its supporting tower and the blades are placed a considerable distance in front of the tower [48].

They firstly appeared in Europe in the middle ages. They could be used for pumping water, sawing wood, and grinding grain [49]. Then, downwind HAWTs have been built. In spite of the fact that downwind turbines have a turbulence problem, additional mechanism to keep them in the same line with the wind is not a common need. Also, the blades can be allowed to bend which reduces their swept area in high winds and thus their wind resistance. Nevertheless, the most of HAWTs are designed to upwind to reduce turbulence [48].

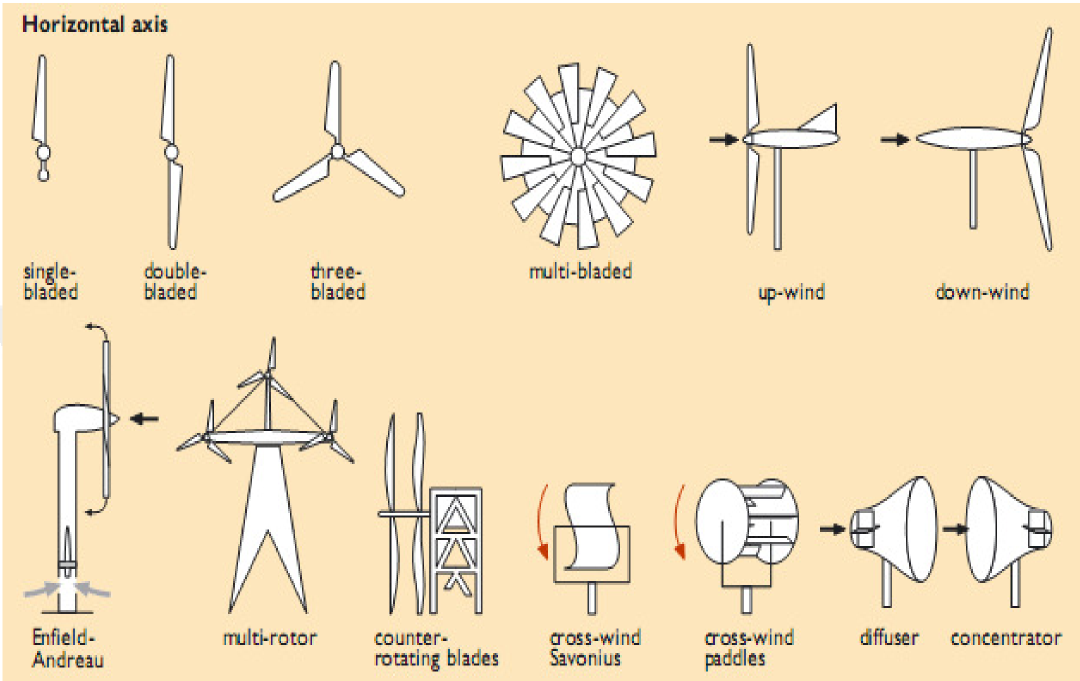

Figure 2. 5 Horizontal axis wind turbines [48]

HAWTs are the most common turbines in use all over the world today. Figure 2.5 summarizes the types of HAWTs and Figure 2.6 gives common used HAWT types. Although many types of HAWTs exist commercially, three-bladed ones are mostly used in wind energy market.

Main advantages of three bladed turbines are as follows; tip speed is quite high (over 320 km/h), efficiency is high, and torque ripple is low. They have light gray colored blades with lengths between 20 and 40 meters (or more) and rotation is between 10 and 22 rpm. Tower is generally tubular steel (60 to 90 meters hub heights). They have a gear box to step up the speed of the generator. Some of them operate at constant speed, others at variable-speed. They are protected to avoid damage at high wind speeds with equipment [48].

Figure 2.3 shows the evaluation of modern three bladed turbines. Especially large sized turbines, such as 5 MW, 6 MW or 8 MW, are designed for offshore wind energy. The future of the three bladed HAWTs has a big growing potential for the offshore.

Figure 2. 6 Common used horizontal axis wind turbines [48]

2.3.2 Vertical Axis Wind Turbines

When the blades rotate around a vertical axis, it is called a vertical axis turbine. The main rotor shaft is arranged vertically in vertical-axis wind turbines (VAWTs). Generally, its visual appearance has been likened to an eggbeater [48]. Wind was firstly used to drive vertical axis wind turbines at 10th century and early ones had a simple design which were particularly susceptible to damage in high winds [49].

The main advantage of the VAWTs is that they do not need to be pointed into the wind to be effective. It is difficulty to model the wind flow accurately and the their rotor have challenges to being analyzed and designed prior to fabricating a prototype [48]. Because of these disadvantages of VAWTs, they are not used as common as HAWTs. Figure 2.7 gives many types of vertical axis wind turbines.

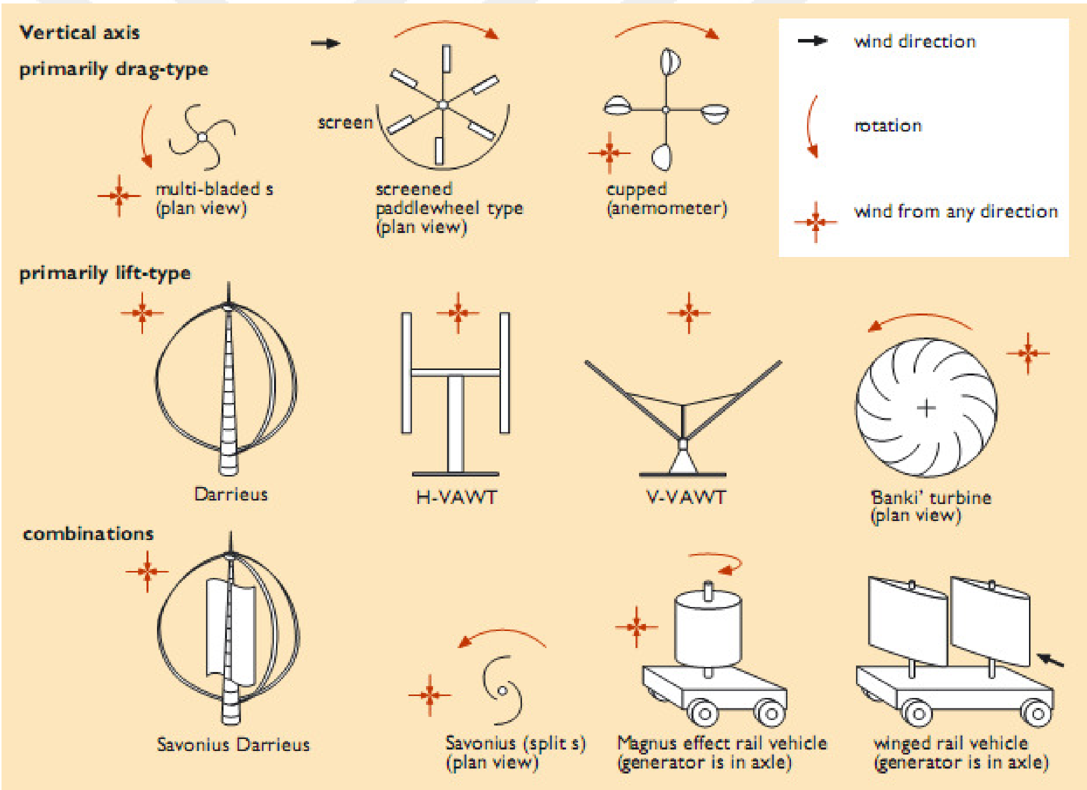

Figure 2. 7 Vertical axis wind turbines [48]

2.4 Pros And Cons Of Wind Energy

Every system has advantages and disadvantages, so do wind energy systems. Wind energy is clean because it generates energy by the naturally flowing wind. It does neither pollute the air like coal or natural gas power plants nor cause acid rain or greenhouse gases by producing harmful atmospheric emissions. Also, it does not leave behind dangerous residues as do nuclear energy systems. It is a local energy source in many countries all over the world. It is competitive with its low priced technologies compared to other renewable energy sources available today. It has lower decommissioning costs than those of many other types of energy systems especially compared to fossil fueled systems and nuclear generators. It does not require as whole large areas, it can be built even on farms or ranches. Besides, farmers and ranchers can continue to use their land because the wind turbines need only a small fraction of the area. These areas can also be rented from farmer or rancher to use a small area. Therefore, one land can be used for both farming and wind energy. Those abovementioned issues are the advantages that make wind energy a competitive alternative [51, 59].

Of course wind energy systems have some disadvantages as many systems have. Besides the fact that it can compete with other renewable energy systems' cost, it also has to compete conventional power generation systems. It requires a higher initial investment than fossil-fueled systems. The wind is intermittent and it does not always blow when it is needed. It cannot be stored. Good wind sites are often located in remote locations, far from places that the electricity is needed. The rotor blades produce noise. People concern about visual impacts of wind turbines aesthetically. They threaten to the birds that fly near the wind farms [51, 59]. By the way, most of those disadvantages can be resolved or greatly reduced through technological development.

2.5 Supervisory Control And Data Acquisition (SCADA) Systems

Supervisory Control and Data Acquisition (SCADA) is a tool which collects and gives very useful data about the system integrated. It was first used in the 1960s [62]. It is generally used in huge industrial facilities and infrastructures such as oil refineries, power generation plants, water/sewage treatment plants, petrochemical plants, and chemical plants [63]. A SCADA system gathers information, such as where a leak on a pipeline has occurred. Also is able to determine if the leak is critical or not by carrying out necessary analysis and control tools. It can be a very simple or more complex in the manner of the controlled and monitored facilities.

SCADA systems have generally used hybrid systems of radio and direct wired connections. Today's increasing security demands leads increasing use of satellite-based systems [64]. A SCADA system generally consists of remote terminal units (RTUs), programmable logic controllers (PLCs), a telemetry system, a data acquisition server, a human-machine interface, a historian, a supervisory system, communication infrastructure.

The most common used general purposed SCADA software are as follows: WinCC, Citect, ICONICS, iFIX, Indusoft, Entivity Studio.

Generally the SCADA system can be classified by two groups such as Energy SCADA systems and Factory-Plant SCADA system. The SCADA system can be used for monitoring and control of very huge industrial and critical systems such as: Nuclear Plants, Electricity Plants, Water Supply Systems, Pipeline Systems, Traffic Control Systems, Automotive Industries, Naturel Gas Systems, Railway Systems, and so on [65].

There are many advantages of the SCADA system. It allows to monitor the system whenever you want and avoids from time and work-hour wasting. It records whole data of the production system. It alarms whenever an abnormal situation is occurred. It records all system interventions. Additionally, all recorded data can be achievable and statistical reports can be generated whenever you want. The system can be monitored from only one main computer and also from a network or even on any mobile device. SCADA systems have evolved through four generations [66]:

I. Monolithic – it was firstly designed in the concept of computing in general centered on "mainframe" systems. These generations were standalone systems with virtually no connectivity to other systems.
II. Distributed – Local Area Network (LAN) technology is used to distribute processing data across multiple systems in this generation.
III. Networked – Wide Area Network (WAN) protocols are the main differences of Networked generation of SCADA systems to communicate between the master station and other communication equipment.

IV. Internet of Things – This will be the next generation for the SCADA systems which will be the smartest one by improvement on internet on things concepts.

Despite the fact that SCADA systems was initially used to simply monitoring and control of turbines, the system serves beyond the basic needs of reducing downtime and increasing availability [67]. Today, SCADA acts as a ‘nerve center’ for the wind farms [68].

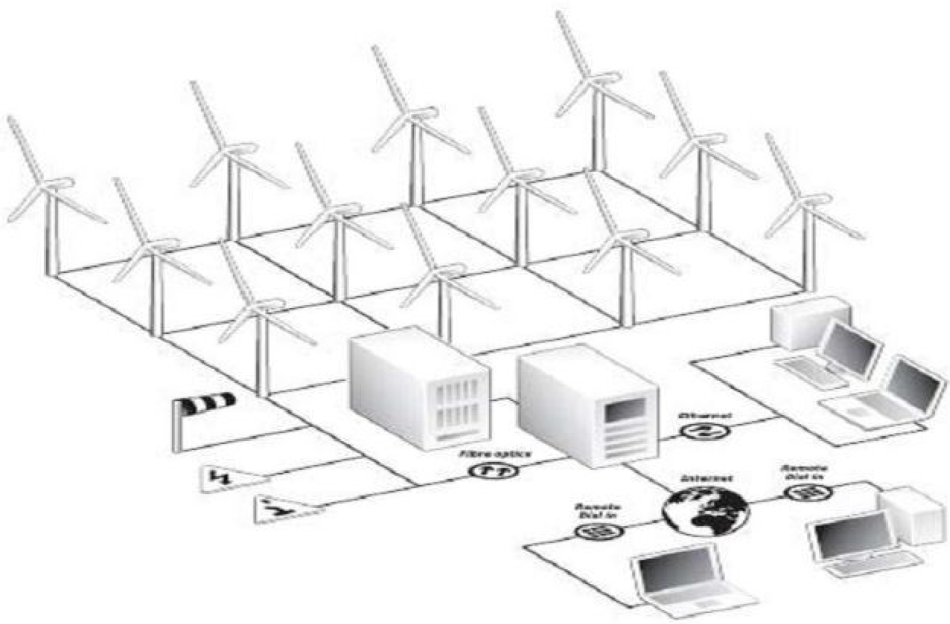

Figure 2. 8 Structure of a SCADA system in a wind farm [68]

Figure 2. 8 demonstrates the main structure of a SCADA system of a wind farm. SCADA systems provide you to control and acquire data from your wind farm. Therefore you can operate the whole wind power plant. The individual turbines, the substation, and meteorological stations are connected to each other by a central computer. Thus, the operator control and monitor the all individual wind turbines and also the wind farm as a whole by this SCADA system.

Despite independent suppliers offer commercial SCADA systems, the turbine suppliers generally provide their own systems for contractual simplicity. Thus, the system reports identical data and have identical analysis formats even there are different wind turbine models and types in the wind energy project. Also, turbine suppliers guarantee the transparency of calculation of availability and other possible warranty issues by providing their own systems. Table 2.1 gives a selected list of current SCADA systems.

Table 2. 1 A selected list of current SCADA products

Company	Product
Alstom	WindAccess™
ARC Informatique	PcVue SCADA
Bachmann Electronic	Wind Power SCADA
DEIF-Wind Power Technology	Scada Systems
DNV GL	WindHelm
Emerson Process Management	High Voltage SCADA
ENERCON	ENERCON SCADA
InduSoft	InduSoft Web Studio
Siemens	Wind Farm Management SCADA
Vestas	VestasOnline® SCADA Systems
WICOtech	SCADA software

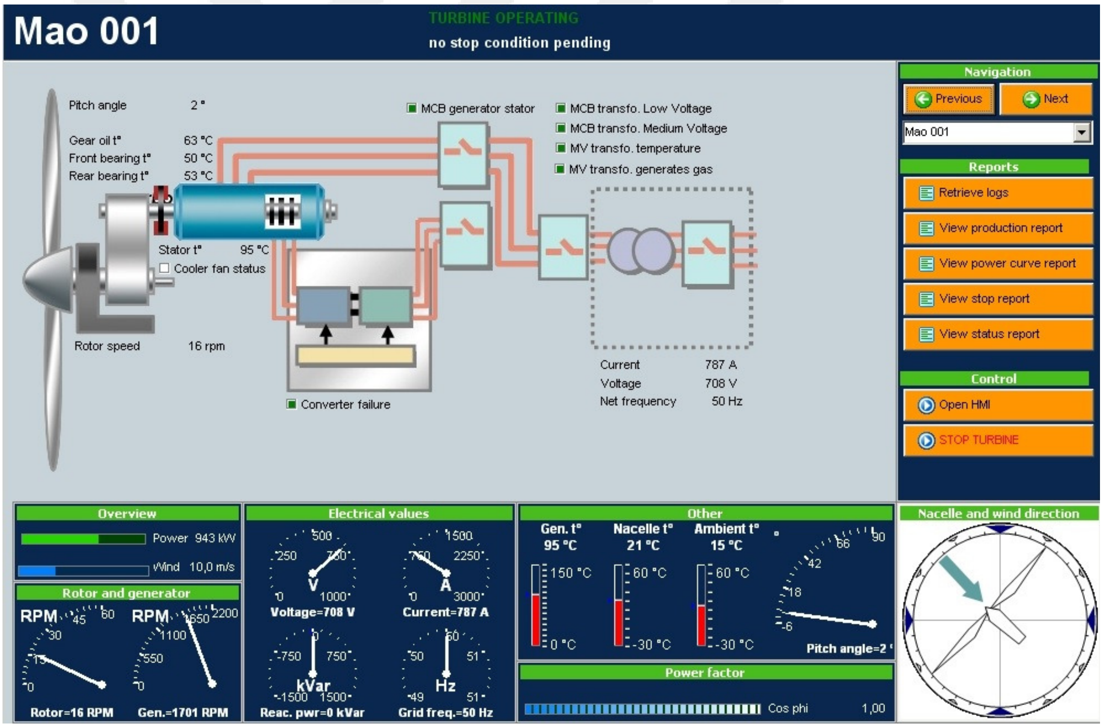

Figure 2. 9 A simple SCADA Interface of a wind turbine system [69]

A screenshot of a typical wind turbine SCADA system interface is given in Figure 2.9. It is a view from DEIF – Wind Power Technology's SCADA system [69] and summarizes a currently working wind turbine's general status. This interface directly gives data related temperatures such as gear oil temperature, front and rear bearing temperatures, stator temperatures, generator temperature, nacelle temperature, and ambient temperature. Also one can directly observe pitch angle, rotor speed, current, voltage, net frequency, power, reactive power, wind speed, rotor rpm, generator rpm, nacelle direction, and wind direction data.

Table 2. 2 A sample of SCADA data signal for a wind turbine with its description, location of sensor, and unit [73]

Name of variable	Unit	Sensor location	Short description
Spinner temp.	◦C	1	Spinner temperature (in hub housing)
Hub controller temp.	◦C	1	Pitch controller temperature
Pitch angle	◦	1	Blade pitch angle
Hydraulic oil temp.	◦C	10	Hydraulic oil temperature
Rotor speed	rpm	2	Rotor speed
Gear bearing temp. (HSS)	◦C	3	High speed shaft bearing temperature
Gear oil temp	◦C	3	Gearbox oil temperature
Generator speed	rpm	4	Generator speed
Generator bearing temp.1	◦C	5	Generator bearing temperature gearbox
Generator current ph.1	A	6	Generator current phase 1
Power output	kW	6	Turbine power output
Reactive power	kVAr	6	Turbine reactive power consumption
Grid busbar temp.	◦C	8	Busbar temperature
Nacelle temp.	◦C	7	Temperature in nacelle of the turbine
Wind speed	m/s	9	Wind speed
Wind direction	◦	9	Wind direction
Ambient temp.	◦C	9	Outdoor temperature

In the wind energy systems, more than 120 data are collected and stored by SCADA [70–72]. The system records all the activity with also error signals of the turbines and helps to the operator on determining what corrective action needs to be taken.

Figure 2.10 demonstrates the sensor positions schematically on a wind turbine and its sub systems' positions. Table 2.2 gives a sample of SCADA data signal for a wind turbine with its description by giving its unit and the location of sensor on the wind turbine [73].

These data can be classified by three groups:

I. Controllable Data: Blade pitch angle, yaw angle, rotor torque

II. Un-Controllable Data: Wind speed, wind direction, ambient temperature, etc

III. Response Data: power output, gearbox torque.

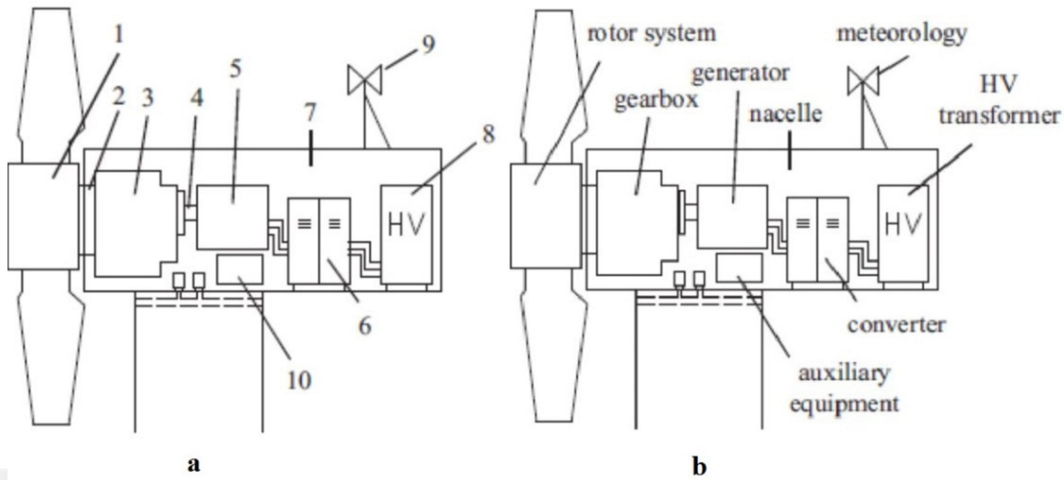

Figure 2. 10 a) Wind turbine sensor positions b) Wind turbine components / subsystems [73]

2.6 Conclusions

A technical background about wind energy and SCADA systems were presented to ones who is unfamiliar with this subject. The definition of the wind, the wind types on the world, and its historically usage areas were given briefly. The energy potential of the wind was given with formulas. A short history on using and improvement of wind turbines was given. Horizontal axis wind turbines and vertical axis wind turbines were introduced. Modern wind turbine evolutions, which commonly used in the market, were illustrated with the dimensions versus years. Then advantages and disadvantages of the wind energy were given to discuss it’s widely usage on the world. After all these technical background on wind energy, it is hoped that the readers of this dissertation could easily understand the mentioned technical terms.

The main used data in this study is gathered from a real SCADA system of a wind farm. Therefore, the basic concepts of the general SCADA systems were introduced. And also, a wind farm’s SCADA system was presented with its requirements, benefits, and main frames.

CHAPTER 3

A SYSTEMATIC AND SMART LITERATURE REVIEW TOOL: CASE STUDY IN WIND ENERGY

This chapter describes how text mining solutions can be used to enhance the wind energy research tracking of systematic review procedure. The first results of this study was presented at The Second European Workshop on Renewable Energy Systems – EWRES [40]. Then, the expanded version of this study was sent to an international journal to publish final results [41]. Thus, this chapter has great similarities with the journal paper.

In this study, the most common used Web of Knowledge database was selected to collect wind energy researches. All of the available abstracts of articles and the proceedings relevant to our study were collected and arranged for analysis. Each paper or proceeding; title, abstract, publication name, host of publication, and publication date were extracted from the web of knowledge database. To include only relevant studies; irrelevant editorial notes, research notes, patents, and reviews were removed from our data collection. The collected data were analyzed by text mining tools. The proposed systematic review provides sufficient guidance to generate flexible and robust tracking.

3.1 Introduction

In order to make use of the full potential of wind energy, it will be of crucial importance to find out relevant researches and its related frameworks, policies, future challenges and prospects worldwide. The projection of problems on wind energy sector can be traced by analyzing the current literature. However, there are many numbers of studies related with wind energy. Special consideration has to be given to the deployment of wind energy in the World. There has been no systematic analysis of the academic researches to date; neither of wind energy all over the world, nor of mining and analytics of large datasets within relevant literature extracting valuable future challenges and prospects in area as wind energy.

Thomson Routers-ISI Web of Science database gives the statistics on wind energy oriented publications between the years 1980 and 2013 as seen in Figure 3.1. It can easily be observed that publications suddenly increase especially after the year 2000. This also supports the fact that importance on wind energy rises year by year. Figure 3.2 shows the citation numbers of wind energy oriented studies. These citation numbers also proofs exponentially growing of wind energy publications' importance of each year.

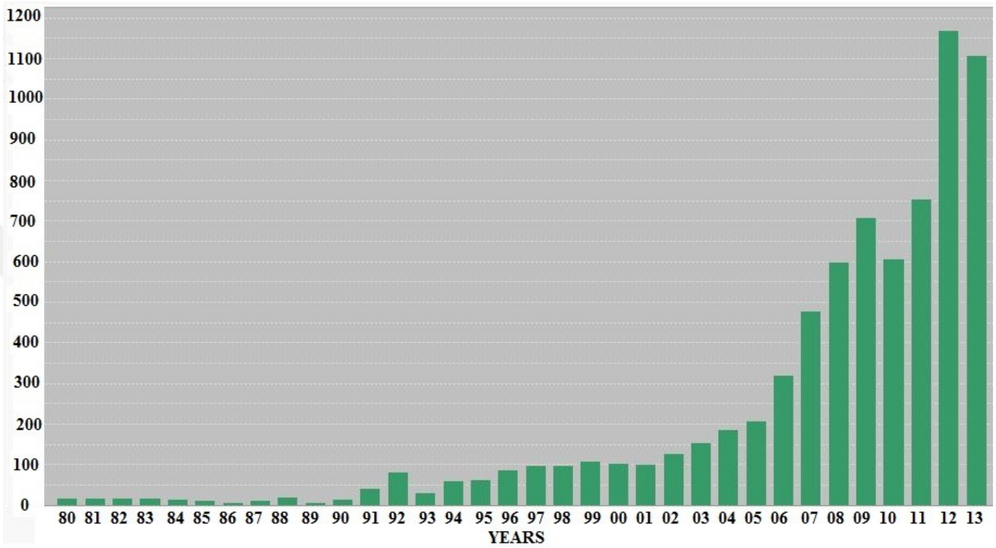

Figure 3. 1 Published items per year (Thomson Routers – ISI Web of Science, 2014)

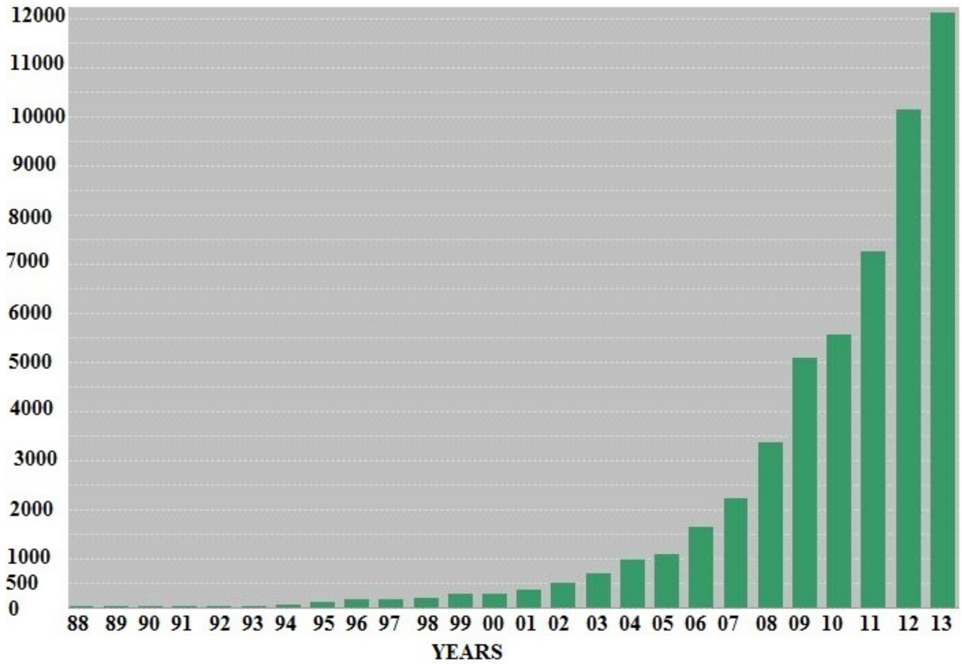

Figure 3. 2 Citations per year (Thomson Routers – ISI Web of Science, 2014)

Table 3.1 gives the country/territory based publication counts until the end of 2013. There are extra three categories in this table such as group A countries which have between 100 and 50 publications, group B countries which have between 50 and 10 publications, and group C countries which have less than 10 publications. The list of these three groups is given in Table 3.2. The USA has the most wind energy publication with 1353 studies followed by Peoples Republic of China with 783 publications.

Table 3. 1 Country/Territory rankings on number of publication all over the world until the end of the year 2013

Country/Territory	**Publication numbers**	**% of 7145 (percentage of total number of publication)**
USA	1353	18,94
A*	980	13,72
Peoples R China	783	10,96
B**	665	9,31
Canada	446	6,24
Germany	423	5,92
England	359	5,02
India	356	4,98
Spain	295	4,13
C***	285	3,99
Turkey	259	3,62
Japan	209	2,93
Denmark	198	2,77
France	188	2,63
Italy	162	2,27
Australia	160	2,24
Greece	149	2,09
Netherlands	145	2,03
Scotland	114	1,60
Iran	111	1,55
Romania	105	1,47
Portugal	102	1,43
Egypt	100	1,40

*A = countries which have number of publications between 100 and 50

**B = countries which have number of publications between 50 and 10

***C = countries which have number of publications less than 10

Massive growth on wind energy literature is the main reason for needing a systematic approach to review wind energy publications. Currently, reviewing procedure is performed mostly manually, thus, it has many problems such as increase of textual information day by day that reading them and summarizing all is impossible [74].

Table 3. 2 List of Countries/Territories which have less than 100 publications until the end of the year 2013

Countries/Territories			
Group A 50=<X<100	**Group B 10=<X<50**		**Group C X<10**
South Korea	Finland	Lebanon	Others
Taiwan	Malaysia	North Ireland	
Brazil	Russia	Lithuania	
Norway	Switzerland	U Arab Emirates	
Sweden	New Zealand	Czech Republic	
Belgium	Poland	Hungary	
Argentina	Tunisia	Bangladesh	
Algeria	Jordan	Estonia	
Ireland	Thailand	Oman	
Saudi Arabia	Austria	Wales	
Chile	Croatia	Colombia	
Mexico	Morocco	Serbia	
South Africa	Nigeria		
Singapore	Israel		

If a reviewer wants to search a literature, she/he has to spend more time to ensure having not missed any relevant studies. She/he even has to download and manually screen the titles, abstracts or full texts of potential research topic. So that, this is the most time-consuming part of the reviewing process [75].

In this chapter, a systematic and intelligent literature review technique is contributed to analyze wind energy researches and to define hot topics in related research field. A detailed literature review on used technique, text mining, is given in the next section. Methodology is explained in third section. Then, results are discussed. Finally, concluding remarks are given.

3.2 Literature Review

Text mining is a very useful methodology to develop a query for comprehensive information retrieval from published literature [76, 77]. It is also an important and fascinating area of modern analytics with its extremely iterative process [78]. Text mining is about inferring structure from sequences representing natural language text and may be defined as the process of extracting information from unstructured text data [79]. It can be used to identify technical themes in large unstructured text databases [80–85]. Patents are another kind of scientific document types which can be used to discover technology trends [85–89].

Text mining can also be used to enhance information retrieval [90, 91], to discuss potential discovery and innovation based on merging linkages between different works areas [92, 93], to uncover unexpected asymmetries from the literature [94, 95], to estimate global levels of effort in science and technology sub-disciplines [96–98], to help authors for increasing the citation of their works [76, 96, 97, 99], and to track myriad research impacts across time and application areas. An agent based text mining approach was investigated on financial news to assist the investors in deciding to buy and to sell stocks in Taiwan market [100]. That study gives a good example of gathering forecasts from financial newspapers texts in the field of agent based data mining.

However social media data are usually large – noisy – and unstructured, they can also be analyzed by using text mining to extract meaningful information about a subject [101]. As being a social area, text mining could be a powerful technique to yield interesting finding on human behavior and human interaction [101–104]. A competitive analysis in the pizza industry was studied by text mining using Facebook and Twitter comments of customers as data base [105].

With the large number of research articles in the literature, a manual review of the works would be tedious and time consuming [74, 78]. Thus, text mining can be used as a powerful tool to analyze current studies on a defined field. There are several studies on text mining application for literature analyze [74, 81, 85, 88, 92, 106–121].

Delen and Crossland [113] were investigate a survey analysis on management information field. They used text mining to identify trends on the topics and classify works which was presented in three major journals in the field of management information systems. Kostoff et al. [117] presented a literature related discovery method. They used links of different concepts had not been linked before in order to produce novel, interesting, plausible, and intelligible knowledge. The evolution of literature related discovery method was discussed in a recent study by Kostoff [119]. Ananiadou et al. [74] proposed a supporting systematic reviews using text mining methodology. They described how a systematic literature review could be produced by using text mining. Emerging topic detection is another field which considers literature review by text mining. Tu and Seng [120] presented a novel strategy to detect emerging topics in literature and investigated novelty index and published volume index.

There are a few works on energy related text mining studies in the literature. Kostoff et al. [98] have prepared a report on science and technology for electric power sources by using database tomography. They suggested the combination of text mining and bibliometrics to extract useful information from large volumes of technical text. They also studied on deriving technical intelligence from a Power Sources databases derived from the Science Citation Index [109]. Yoon [122] investigated a diffusion approach to detect weak signals for long-term business opportunities for solar – cells by using text mining. On the wind energy area, there is only a new study in the literature which aims to generate a domain ontology for wind energy field [121].

3.3 Methodology

In this study, text mining was used to investigate a systematic review on wind energy researches by analyzing huge number of papers, which consist of unstructured text data. It will be discussed in more detail that the application of text mining techniques for improving the search and screening strategy of systematic reviewing for the domain of wind energy issues. Following sections gives main usage areas of text mining and steps of this study.

Text mining is a burgeoning field, which has the process of turning unstructured text data into structured numerical data to collect meaningful information from naturel language text. The proposed systematic review follows stages at below:

i. Database Selection: Extensive wind energy searches are carried out in order to locate as much relevant research as possible according to a query. These searches include electronic databases to collect and store published literature. Thomson

Reuters – Web of Science database is a wide-used most common electronic database by researchers and academicians.

ii. Data Selection: A query-based mining are applied for the topics including "Wind Energy" to select relevant studies. Because the aim is to highlight key evidence, the scope of search is reduced by collecting only relevant documents. By the way, reviews, editorials, news, patents, books, case results are eliminated from the search results; papers and conference proceedings are only selected study types. This narrows the collected data to prepare a specific review.

iii. Data Preparation: Selected data is categorized by their authors' country/territories and stored in an Excel Sheet. The sheet contains following data about studies; Publication Type: Journal or Proceeding, Authors' Name, Title, Source: Published Journal or Conference, Publication Year, Authors' country/territories, Abstract, and Citation Number.

Studies having more than one authors, who are from different country/territory, may occur more than one times. These studies are categorized as Multinational in Country/Territory column. This correlates evidence from a plethora of resources and summarizes the results and creates meaningful contents in wind energy literatures.

iv. *Analyzing*: Text mining is used on prepared data.

v. *Extracting Knowledge*: Results of Text Mining gives meaningful statistics and data about collected studies. Also, clustering analyses and ANOVA (analyses of variances) helps to extract useful knowledge.

Data gathering and preparation processes are detailed given the next section. Then, text mining method is explained step by step.

3.3.1 Data gathering

In this study, Thomson Reuters database is used to collect wind energy related studies. All of the available abstracts of journal papers and conference proceedings having topic "wind energy" are collected from the year 1990 to end of 2013. The data-base is constructed with the Publication type, Authors' name, title, Source, Publication year, Authors' country/territory, abstract, and citation numbers of selected papers. Irrelevant editorial notes, research notes, patents, news, and reviews are not selected for data-base to include only relevant studies. Collected data is categorized by their authors' country/territories and stored in an Excel Sheet (Table 3.3).

Table 3. 3 An example part of stored data

Type	Authors	Title	Source	Year	Abstract	Citations	Territory
J	Muljadi, E; Hess, HL; Thomas, K	Zero sequence method for energy recovery from…	IEEE TRANSACTIONS ON ENERGY CONVERSION	2001	An innovative power conversion system to convert energy from a variable-frequency…	6	USA
J	Mays, I	WREC 1996 – The status and prospects for wind energy…	RENEWABLE ENERGY	1996	This paper reviews developments in the Weibull distribution of wind energy in Europe over the past decade,….	1	England

Table 3. 4 Categories of countries/territories according to number of published studies

X>= 100		**Group A 50=<X<100**	
1. USA	11. France	21. South Korea	31. Chile
2. Peoples R Chine	12. Australia	22. Taiwan	32. Mexico
3.Canada	13. Italy	23. Brazil	33South Africa
4. Germany	14. Greece	24. Norway	34. Singapore
6. India	16. Scotland	26. Belgium	
7. Spain	17. Iran	27. Argentina	
8. Turkey	18. Romania	28. Algeria	
9. Japan	19. Egypt	29. Ireland	
10. Denmark	20. Portugal	30. Saudi Arabia	
Group B 10=<X<50			
35. Finland	44. Austria	53. Czech Republic	
36. Malaysia	45. Croatia	54. Hungary	
37. Russia	46. Morocco	55. Bangladesh	
38. Switzerland	47. Nigeria	56. Estonia	
39. New Zealand	48. Israel	57. Oman	
40. Poland	49. Lebanon	58. Wales	
41. Tunisia	50. North Ireland	59. Colombia	
42. Jordan	51. Lithuania	60. Serbia	
43. Thailand	52. U Arab Emirates	**Group C : Others**	

Country/Territories of publications are classified in four groups as number of publications is more than 100, between 100 and 50, between 50 and 10, and less than 10 Studies having more than one authors, who are from different country/territory, may occur more than one times (Table 3.4). The first groups of these are stored with their country/territory name, rest of them stored with their class name in the data sheet. Also, a Multinational category is created for international collaborative publications in Country/Territory column. After collection and preparation of data, data-base is ready for the next step, Text Mining.

3.3.2 Text mining

Text mining is a burgeoning methodology to semi-automatically extract information from unstructured text data and involves imposing structure upon text so that relevant information can be extracted from it [78, 113, 123, 124]. Text mining process of is as follows;

i. Create data-base: Data-base is created in an Excel Sheet and imported to a Statistica Sheet.

ii. Select text variable to be mined: Abstracts of the publications are selected as variable to text mining

iii. Set text mining stemming language parameters: Stemming language is English, because all abstracts are in English. Other parameters are left in default values.

iv. Give the list of stopping words, phrases, and synonyms: Stopping word list is the list of words which do not need to be selected for further analysis. A standard English Stop List consist of pronouns, auxiliaries, articles, prepositions, conjunctions, adverbs, and some common words such as “now, never, between, into, etc.”. This entire word list is unnecessary for an academic literature text mining process.

Table 3. 5 Added extra stop words

1. abstract	5. copyright	9. paper	13. wind
2. address	6. Elsevier	10. reserved	14. wind energy
3. article	7. energy	11. rights	15. wind-energy
4. author	8. ltd	12. study	16. work

Besides, an academic abstract has some common words (Table 3.5) which is also unnecessary such as “abstract, author, study, paper, etc.”. These are also added to stop list.

Table 3. 6 Phrase words

1. converting system	4. variable speed	7. wind power
2. renewable energy	5. Weibull distribution	8. wind turbine
3. three phase	6. wind farm	

The phrases list is the list of phrases which consists joint words (Table 3.6). Some words have similar meanings/synonyms and they have to be considered in only one word (for instance; predict-forecast-estimate, these all three words combined as the word “forecast”).

v. Set the word processing/filtering parameters: All parameters are left in default values.

Table 3. 7 The most frequent words

Word	**Count**	**Number of Documents**	**Word**	**Count**	**Number of Documents**
1. system	10792	4067	6. speed	4496	2081
2. power	9542	3763	7. result	4226	3247
3. generation	7914	3650	8. develop	4064	2432
4. control	6813	2128	9. electric	3958	2157
5. model	5618	2543	10. present	3625	2787

vi. Start indexing words: Push the “Index” button to index/select/filter words to Text mining. There are 7145 documents in the database, 473 words are selected, and 18954 words are not selected. Most frequent ten words are given in Table 3.7. The word “*system*” is the most repetitive word through the whole publications which is used 10792 times in a total of 4062 documents. The second one is “*power”* with 9542 repetitions in 3763 documents, followed by the words “*generation*”, “*control*”, and “*model*”.

vii. *Concept Extraction*: There are four methods to extract concepts in text mining.

a. *Raw Statistics:* Gives only counts of words in all documents
b. *Binary Frequency:* Returns “1” for a word if it occurs in any document else “0”

c. *Logarithmic Frequency:* Computes frequency for a word by using following formula (Equation 3.1);

$$F = 1 + log(\text{wrf}) \;\; for \quad \text{wrf} > 0 \qquad \text{Equation 3.1}$$

where *wrf* stands for word raw frequency. If a word occurs “1” time in document *A* but “3” times in document *B*, then it is not necessarily reasonable to conclude that this word is “3” times as important as descriptor of document *B* as compared to *A*. Thus, logarithmic frequency is useful in these situations.

d. *Inverse Document Frequency:* This is the relative document frequencies (*df*) of different words. A common and very useful transformation that reflects both the specificity of words (document frequencies) as well as the overall frequency of their occurrences (word frequencies) is the so-called inverse document frequency (for the *i'th* word and *j'th* document). The formulation of inverse document frequency (*idf*) is given below (Equation 3.2), where *N* is the total number of documents, *wf* is the word frequency for all documents, *df* is the word frequency for current document;

$$idf(i,j) = \begin{cases} 0 & if\ wf_i = 0 \\ (1 + \log(wf_i)) log \frac{N}{df_i} & if\ wf_i \geq 1 \end{cases} \qquad \text{Equation 3.2}$$

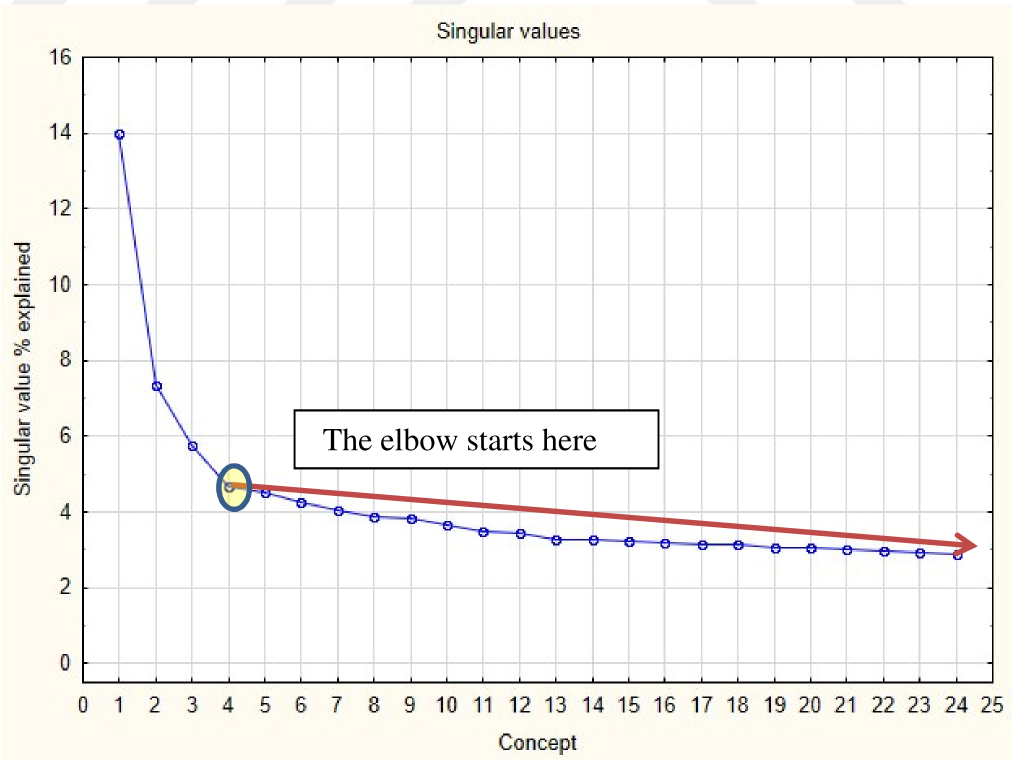

Figure 3. 3 Scree Plot of extracted concepts

In this study, Inverse Document Frequency is used to extract concepts and to determine words' importance. Figure 3.3 gives the Scree Plot of extracted concepts which is used to decide on the number of singular values that are useful and informative, and that should be retained for subsequent analyses. Usually, the number of "informative" dimensions to retain for subsequent analysis is determined by locating the "elbow" in this plot. The points, which are above from elbow of the graph, have more importance than others [78]. Consequently, there are 24 concepts extracted and the first three concepts have more importance than others because the elbow starts from the fourth concept.

Text mining process transforms unstructured text data into structured, analyzable data. All results from text mining process stored in database and further analysis, such as ANOVA and Clustering, are discussed in the next section.

3.4 Results and Discussions

The text mining process gives the importance value of indexed words. The most important ten words are listed in Table 3.8. As it can be seen from the table, the important words are not totally same with the most repetitive words as mention in the Table 3.7. This is the secret of the text mining process. It does not only give the frequent words, it also selects the important ones from frequents.

Table 3. 8 The most important 10 words

Words	Importance	Words	Importance
1. Control	100	6. Wind farm	81,490206
2. Offshore	93,880264	7. Storage	80,12595
3. Solar	86,830387	8. WEC	76,725528
4. Blade	84,827511	9. DFIG	75,564418
5. Voltage	82,696561	10. Forecast	75,176318

According to the inverse document frequency analysis, wind energy related publications are mostly interested in *control* of wind energy systems. Day by day, the need for the energy increases which causing the new installation areas of energy plants. *Offshore* researches for the wind plants have a second importance by the wind energy researchers. Then, *solar* based hybrid studies have a valuable importance in the field. Design problems also have a significant place with *blade* and *WEC* (wind energy converters) related studies. By the incensement of wind energy market all over the world, *Wind farm* is in the top ten lists as another important field for the researchers.

As mentioned in the previous section, twenty-four concepts are extracted after text mining analysis. The most important words for each concept are listed in Table 3.9. It can be seen from the Table that nearly all concepts can be titled according to included important words in their lists. All concepts have caption numbers according to their importance levels and also words are in the importance order as seen in the list.

Results show that the most important concept is related with *System Control Models* including following words: *control*, *model*, *power*, *speed*, *system*, *forecast*, and *electric* consequently. It can be concluded from this concept that many researchers are mostly cared about "*system control*" of the wind energy field. This concept explains the big importance of the system oriented control issues and gives clues on need for development of smart managerial tools.

The second most important concept can be labelled as *Electrical Control Systems* by including the following words: *control*, *voltage*, *convert*, *induct*, *DFIG*, *WEC*, and *reactive*. All-important words listed in this concept are common electrical terms except *control* and *WEC*. These two words combine others in the field of *control* of *wind energy convertors*.

The term *control* has the biggest importance for the first and second concepts, thus *control* issues can be handled as hot and emerging topic for the wind energy.

Data Observation is the third important concept which has to be considered in detailed. One can understand from the important terms of this concept (*speed*, *data*, *observe*, *measure*, *forecast*, *surface*, and *density*) that data observation is still hot topic for the wind energy researchers.

The scree plot (Figure 3.3) tells that remaining concepts have nearly same weights by considering importance level. Solar energy systems, storage systems, policies, offshore, aerodynamics and etc. can be determined as other research fields in the wind energy but they do are not as important as the first three concepts.

Table 3. 9 The most important words of extracted concepts

Concept 1	Concept 2	Concept 3	Concept 4	Concept 5	Concept 6
System Control Models	***Electrical Control Systems***	***Data Observation***	***Storage Systems***	***Solar Systems***	***Electrical Systems***
control	control	Speed	storage	solar	DFIG
model	voltage	Data	hour	magnet	voltage
power	convert	Observe	hybrid	heat	reactive
speed	induct	Measure	battery	fuel	induct
system	DFIG	forecast	distribute	flux	grid
forecast	WEC	Surface	month	temperature	FED
electric	reactive	Density	speed	field	site

Concept 7	Concept 8	Concept 9	Concept 10	Concept 11	Concept 12
Network-Grid Systems	***Renewable Energy Policies***	**Turbine Components**	**Solar Hybrid Systems**	**Aerodynamics**	**Offshore Systems**
network	policy	DFIG	hybrid	Blade	offshore
penetrate	forecast	rotor	solar	Turbulence	control
integrate	track	induct	DFIG	fluctuate	wave
grid	algorithm	blade	battery	renewable energy	water
model	uncertainty	FED	photovoltaic	change	sea
stability	approach	double	induct	wind turbine	strategy
storage	renewable energy	stator	data	source	surface

Concept 13	Concept 14	Concept 15	Concept 16	Concept 17	Concept 18
***NWD**	***NWD**	**Market Researches**	***NWD**	**Electric Distribution Systems**	***NWD**
battery	monitor	price	forecast	distribute	monitor
storage	offshore	market	market	wave	rotor
market	structure	emission	accuracy	load	algorithm
project	solar	load	storage	rotor	reliable
policy	year	frequency	wind power	voltage	cost
solar	industry	forecast	machine	density	optimum
support	blade	voltage	flux	probably	hybrid

Concept 19	Concept 20	Concept 21	Concept 22	Concept 23	Concept 24
***NWD**	***NWD**	***NWD**	**Island Plants**	**Climate Researches**	***NWD**
monitor	WEC	heat	plant	change	wave
fault	machine	storage	island	temperature	machine
measure	reliable	water	station	climate	device
WEC	sea	temperature	area	machine	rate
convert	change	frequency	region	heat	frequency
layer	transmission	air	magnet	DC	communicate
install	wave	impact	water	capacity	hybrid

***Not well determined**

All these concepts are not clearly identified by only looking up its important word lists. Especially after the first ten concept, the important word lists consist of directly irrelevant terms such as in the thirteenth concept: *battery*, *storage*, *market*, *project policy*, *solar*, and *support*. Thus some concepts are labelled as *not well determined (concepts 13, 14, 16, 18, 19, 20, 21, and 24).*

3.4.1 Clustering analysis of important words

Text mining analysis results gave four-hundred-seventy-three indexed words. Another important issue is determining the number of selected important words for further analysis. Therefore, K-means clustering algorithm is applied to word importance values and six clusters are created.

Table 3.10 lists the members of first clusters which have the highest importance levels. As the words are clustered as their importance values, the rank of clusters also represents its importance rank. Furthermore, the rank of words gives the importance level of the term in the cluster.

Table 3. 10 Members of the first Cluster according to word importance level

Cluster 1	
1 control	5 voltage
2 offshore	6 wind farm
3 solar	7 storage
4 blade	

The most important seven words are listed in the first cluster. This cluster clarifies that these words have exactly different importance levels than others. By the way, as the importance value of words decreases, number of members of clusters increases. The terms in the sixth cluster have the least attention of wind energy researchers. By contrast, the first cluster terms; control, offshore, solar, blade, voltage, wind farm, and storage, deserve to be handled in detail.

3.4.2 Anova for text mining results

Anova, the acronym for Analysis of Variance, is developed by R.A. Fisher to analyze the differences between group means and their associated procedures. This statistical procedure gives important knowledge about variants of different groups. In this study, it is used to explain wind energy research trend differences between countries/territories and years. All Anova results have a p value less than 0,01 and vertical bars of Anova graphs denote 0,95 confidence intervals.

Figure 3.4 can light the way for new technologies of wind energy systems. Just doing an Anova analysis on all extracted concepts versus countries and Publication Years, One can conclude on countries research/interest areas and trends of hot topics. In Figure 3.4, only first three concepts are deeply analyzed because these concepts have more importance level than others.

This figure gives the Anova result of most three concepts according to countries/territories and publication years. While Concept 1 has the biggest mean of importance value, it has almost no significant variants according to countries. This means that researches on Electrical Control Systems has nearly same importance for all over the world. The control related studies have a positive trend importance level according to years especially after the year 1998. This can be explained by the improvement on the control equipment technology such as SCADA systems.

Today, there are many types of sensors which can be easily adapted to a wind turbine. This also allows researchers to get many different type of knowledge about working conditions of the system. The more data, the more needs to be controlled.

As Concept 2 is labeled as *Electrical Control Models* related publications, India and Egypt handle *system control models* with the same importance as *electrical systems*. Egypt, India, Iran, France, Japan, Canada, and China give more attention on Concept 2. In contrast, Turkey, Greece, Denmark, England, and Netherlands give less. Especially after the year 2000, this concept has an increasing trend of importance level.

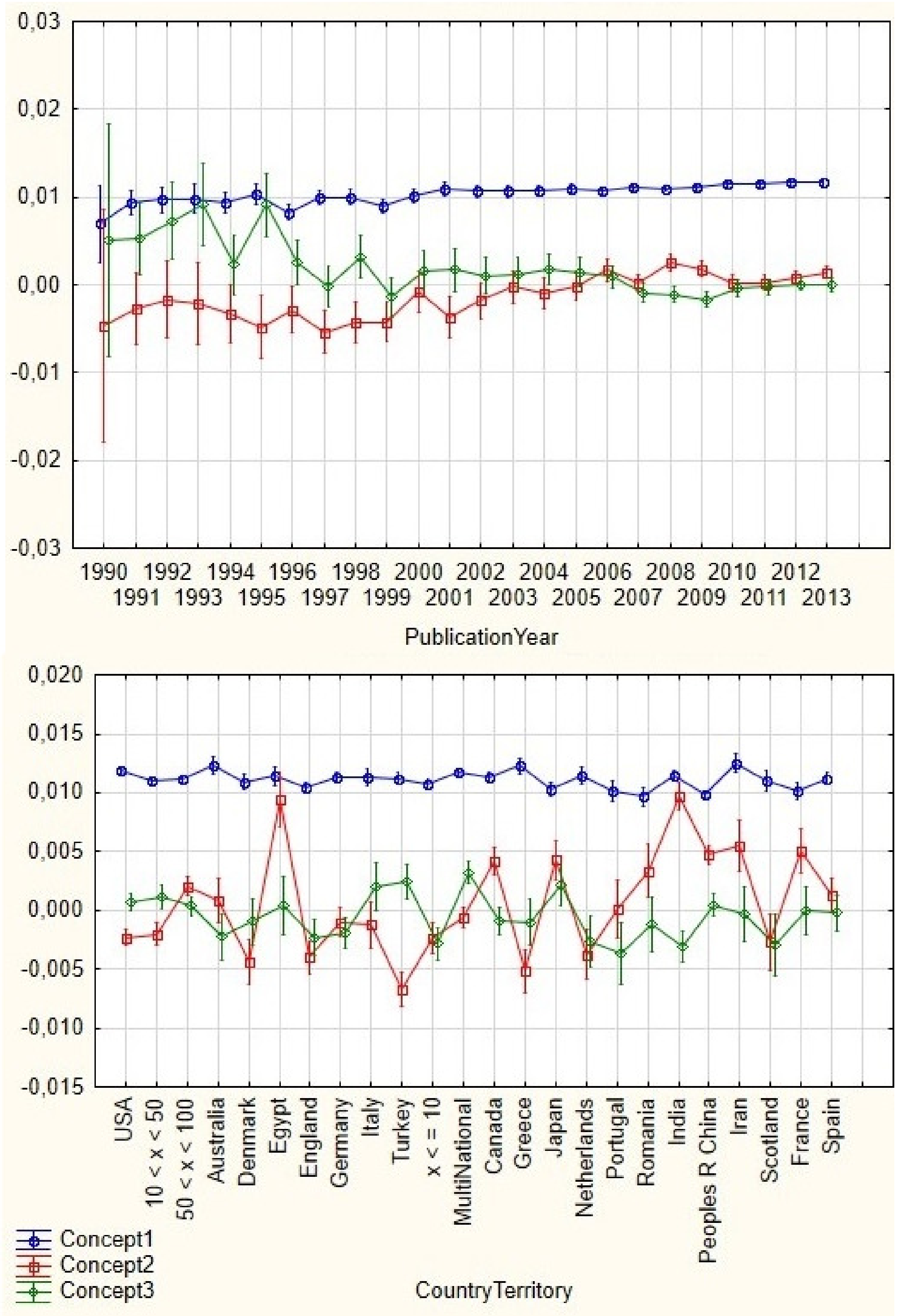

Figure 3. 4 Anova for the most three concepts vs Country/Territory and Publication Year

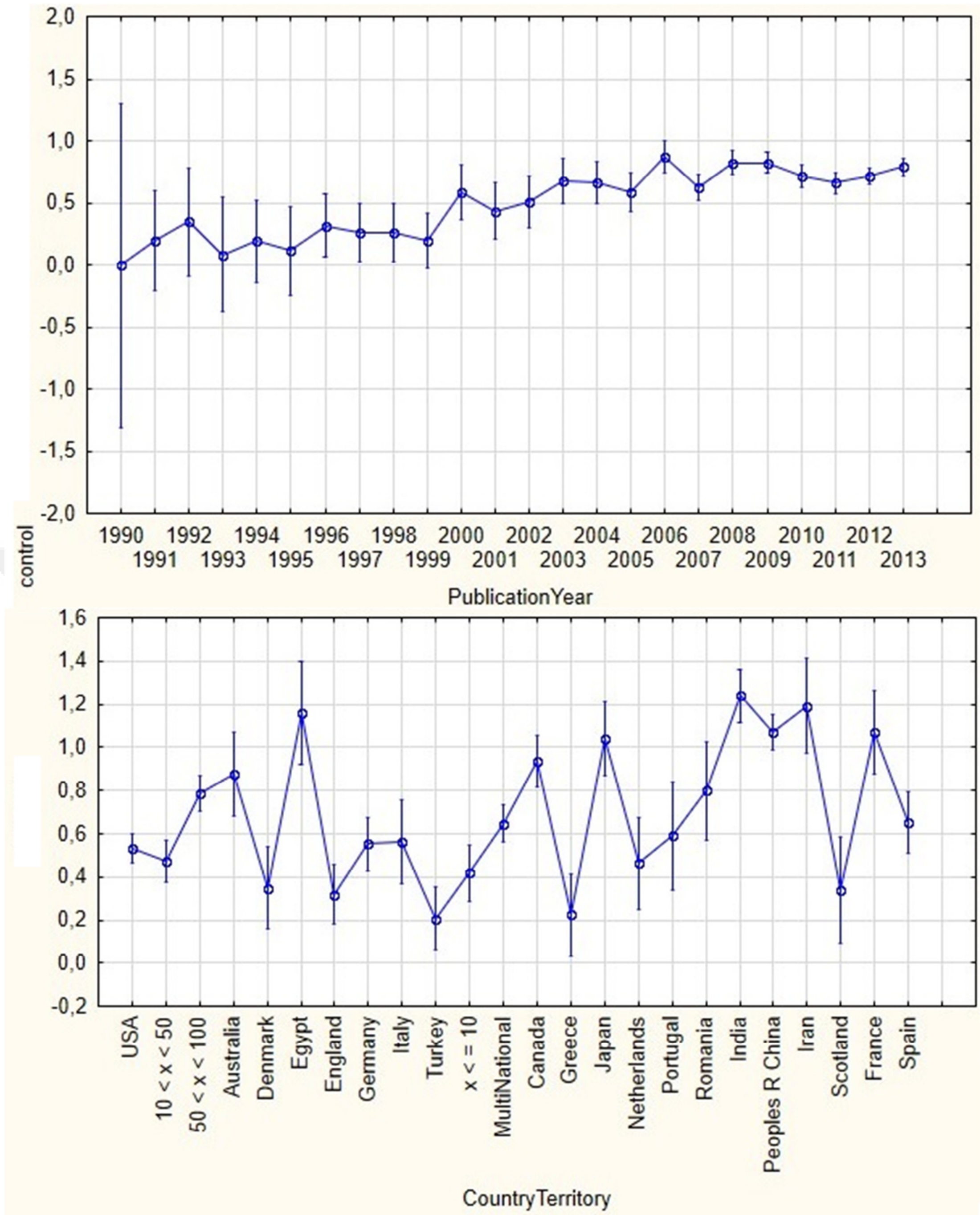

Figure 3. 5 Anova for the word *control* vs Country/Territory and Publication Year

The third concept includes *Data Observation* related studies. While the Concept 1 and 2 have an increasing importance trends, Concept 3 loses its importance year by year especially after 1995 (Figure 3.5). This can be explained by the improvement of the energy market. Today, many countries have wind farms. They are mostly interested in improving productivity, efficiency, and profitability of current wind energy market. Therefore, decrease on the interest of feasibility studies, which includes *Data Observation,* should be considered to be normal. Italy, Turkey, Japan, and Multinational oriented publications are more interested in *Data Observation* studies than other countries/territories.

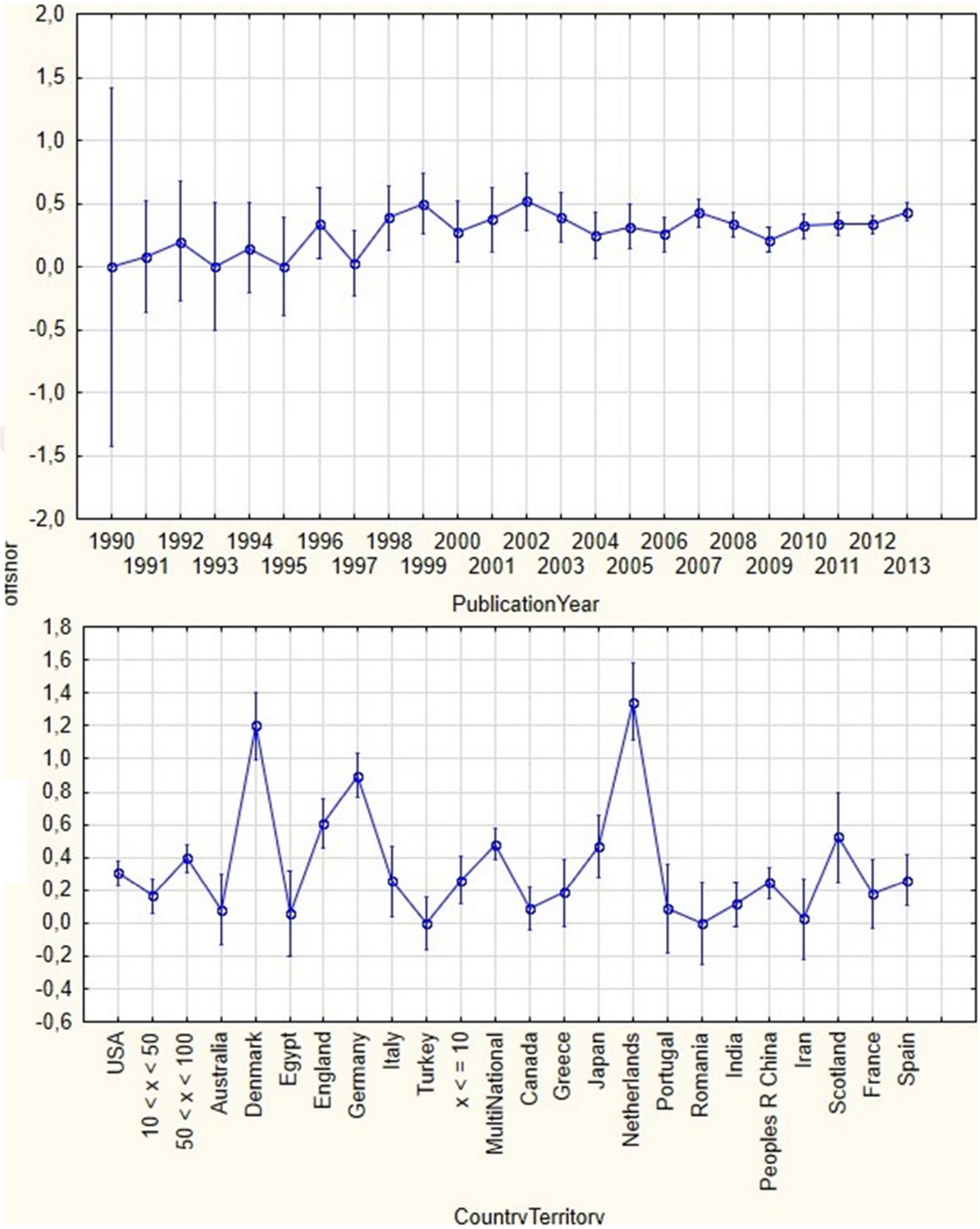

Figure 3. 6 Anova for the word *offshore* vs Country/Territory and Publication Year

The most important word *control* has an increasing yearly trend. India, Iran, China, Egypt, and France are mostly interested in *control* studies. On the other hand; Turkey, Denmark, England, Greece, and Scotland gives less attention to *control* (Figure 3.5). This word occurs in both first and second Concepts in the top of the list.

So, its importance is explicit through all wind energy publications. Technological improvements on equipment of wind energy sector (sensors, blades, brake systems, etc.) lead to increase on *control* focused researches. The more sensitive unites, the more *control*. Thus, it can be forecasted from the positive trend on *control* studies that wind energy technology has still positive trend.

Offshore wind energy plants are preferred to on-land plants because of following advantageous;

- It has better wind speeds available,
- Huge wind turbines can be constructed offshore, so, electricity supplied per turbine is higher than on-land,
- Usually construction is weaker.

Figure 3.6 shows that *offshore* is one of the trending topics of wind energy publications. Countries such as Denmark, Netherlands, and Germany, which have offshore wind energy plants, give more importance on *offshore* studies. This figures out that if one has a technology, he gives more attention on its problems.

It can be understood from the Figure 3.6 that the year 1996 gives the first attention on offshore technologies. Thanks to increasing installed offshore wind farms, it has an increasing trend

Figure 3.7 gives an interesting result that instead the word solar is the third important word; its importance value tragically decreases especially after the year 2005. Meanwhile; Turkey, Multinational publications, Category C countries (having less than 10 publications in total), and Category B countries (having publications between 10 and 50 in total) give more attention to solar related researches.

Blade related publications have a nearly steady importance value year by year except 2005. Japan is the most interested in blade researches (Figure 3.8). Because of having leading technologies all over the world, it is not surprised that Japan and USA are interested on wind turbine blade technology. By the way, aerospace industry is one of the leading sectors of England which explains their interest on blade technology.

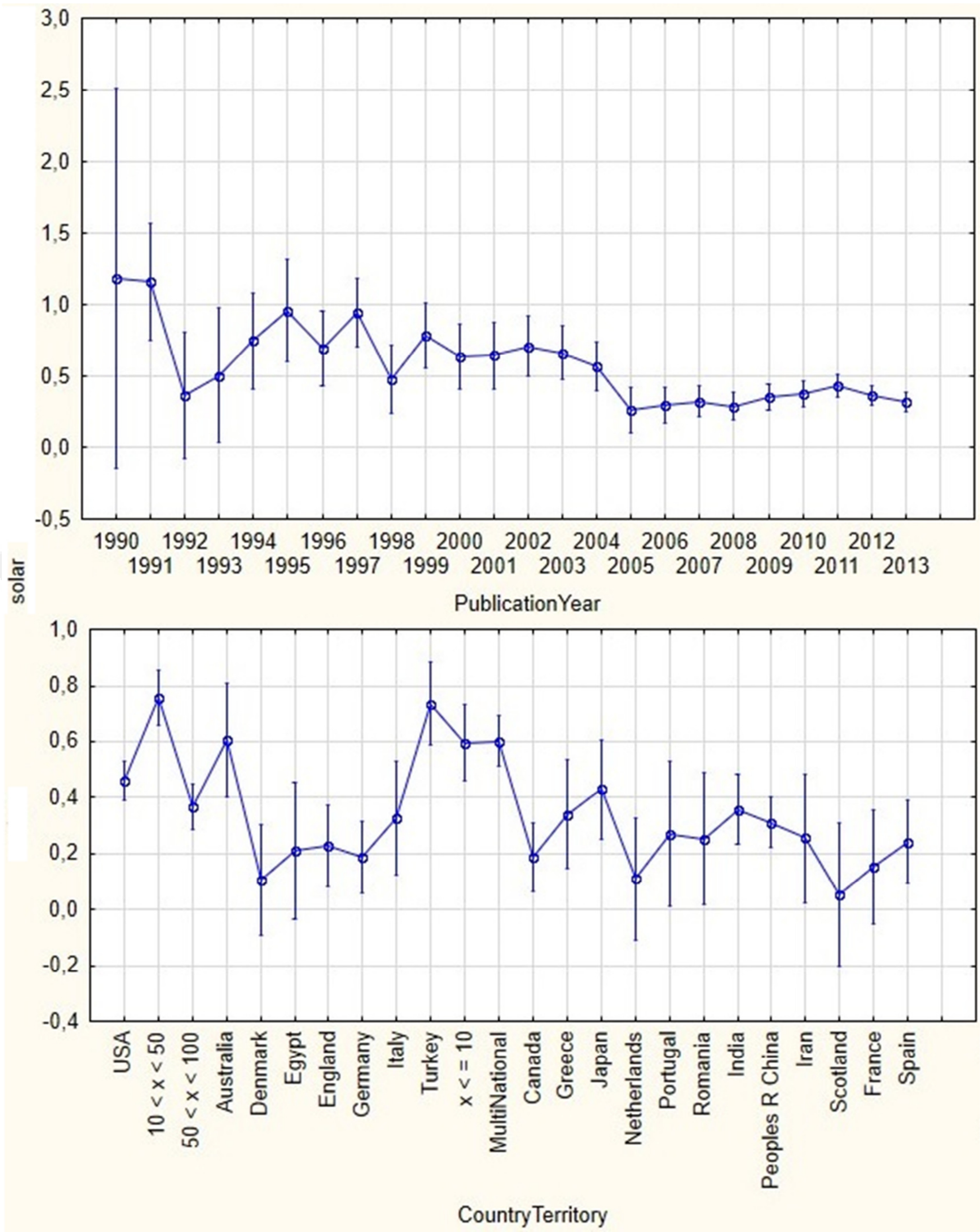

Figure 3. 7 Anova for the word *solar* vs Country/Territory and Publication Year

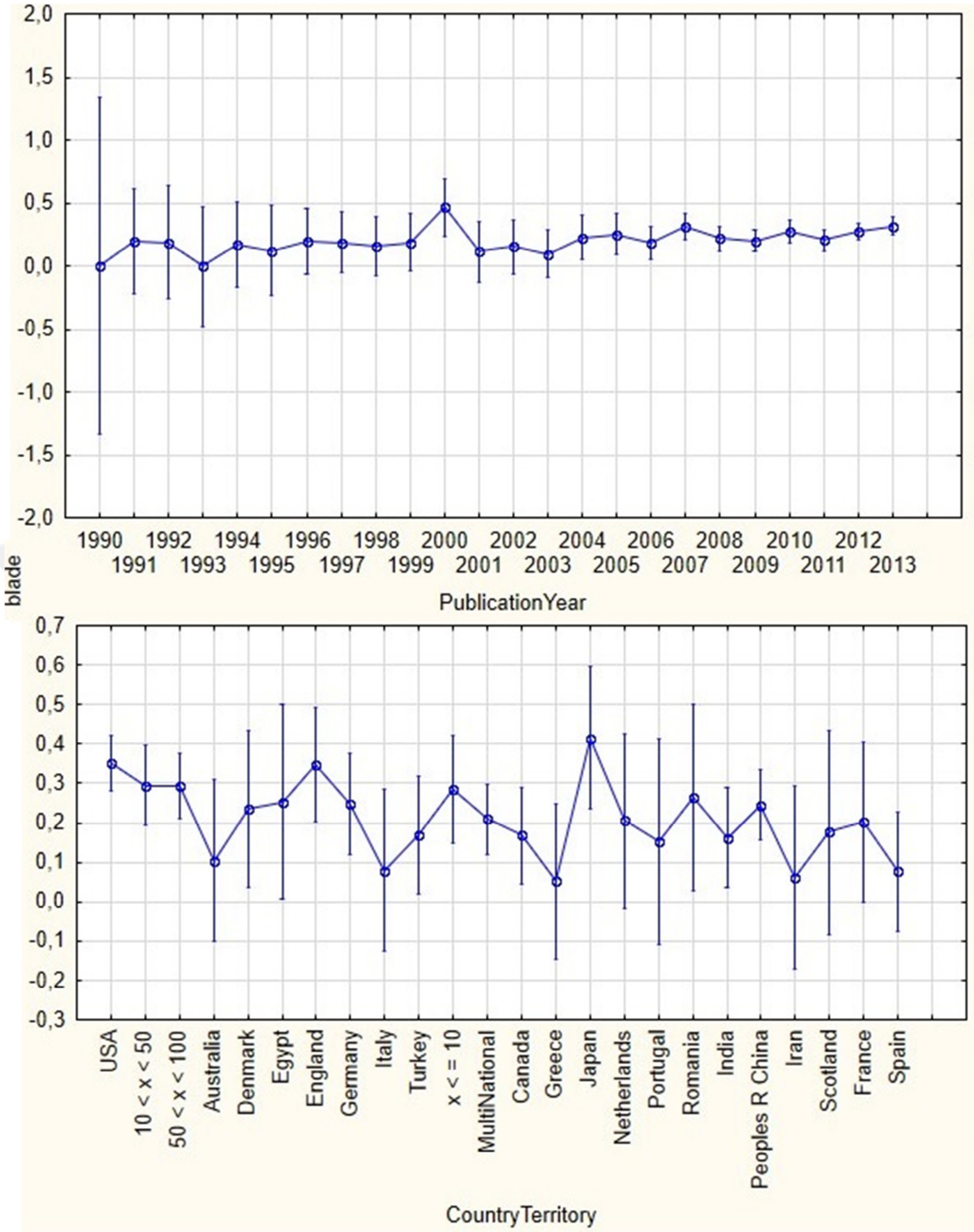

Figure 3. 8 Anova for the word *blade* vs Country/Territory and Publication Year

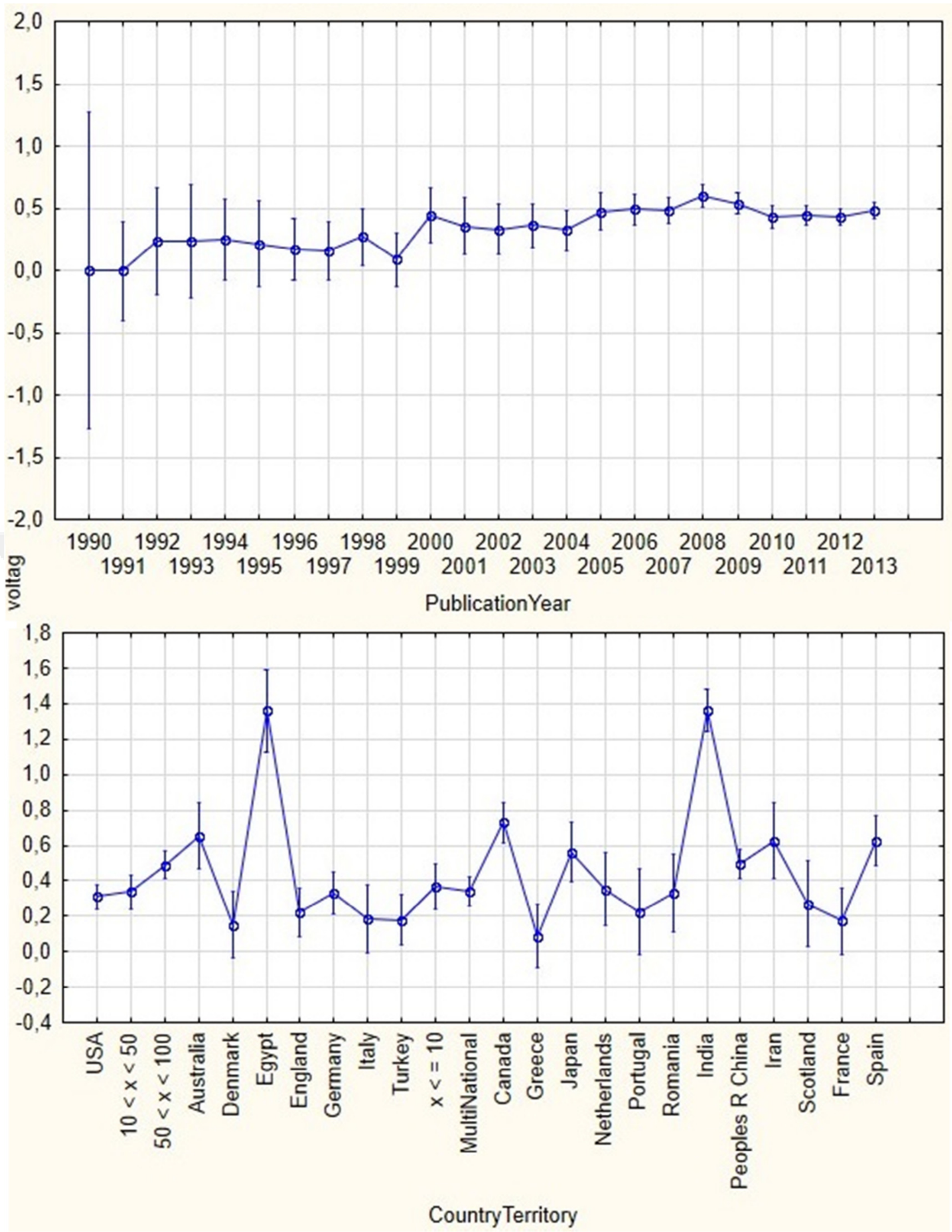

Figure 3. 9 Anova for the word *voltage* vs Country/Territory and Publication Year

Egypt and India have significantly different interest on *voltage* related studies. This can be explained by having incapable or weak electrical transformation and network devices. Because the demand of electricity increases dramatically, this topic has also an increasing importance especially after the year 2000 (Figure 3.9).

Wind turbine technology and needs for renewable energy sources grow yearly. This causes increase on installed *wind farms* all over the world. A *wind farm* is a group of wind turbines in the same location used to produce electric power. A large *wind farm* may consist of several hundred individual wind turbines, but the land between the turbines may be used for agricultural or other purposes. A *wind farm* may also be located offshore. Thus, interest on *wind farm* also increases. Spain and Scotland give more importance on *wind farm* researches (Figure 3.10).

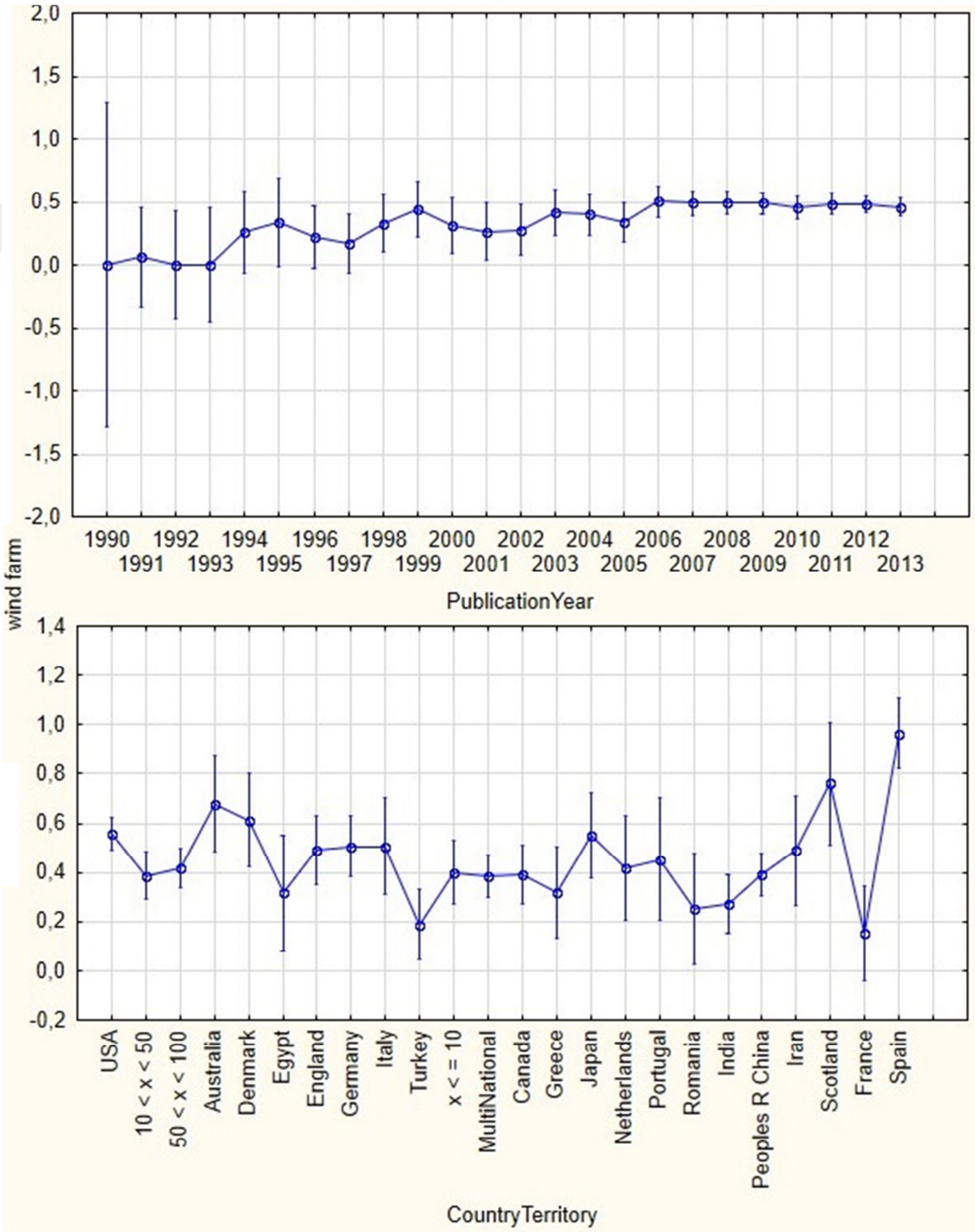

Figure 3. 10 Anova for the word *wind farm* vs Country/Territory and Publication Year

Because today's technology is incapable to *store* it, one of the main problems on renewable energy plants is that once you produce energy, you have to use it. Thus, *storage* researches have increasing trends. While Australia, Canada, and Germany give more attention on *storage* studies; Turkey gives the least (Figure 3.11).

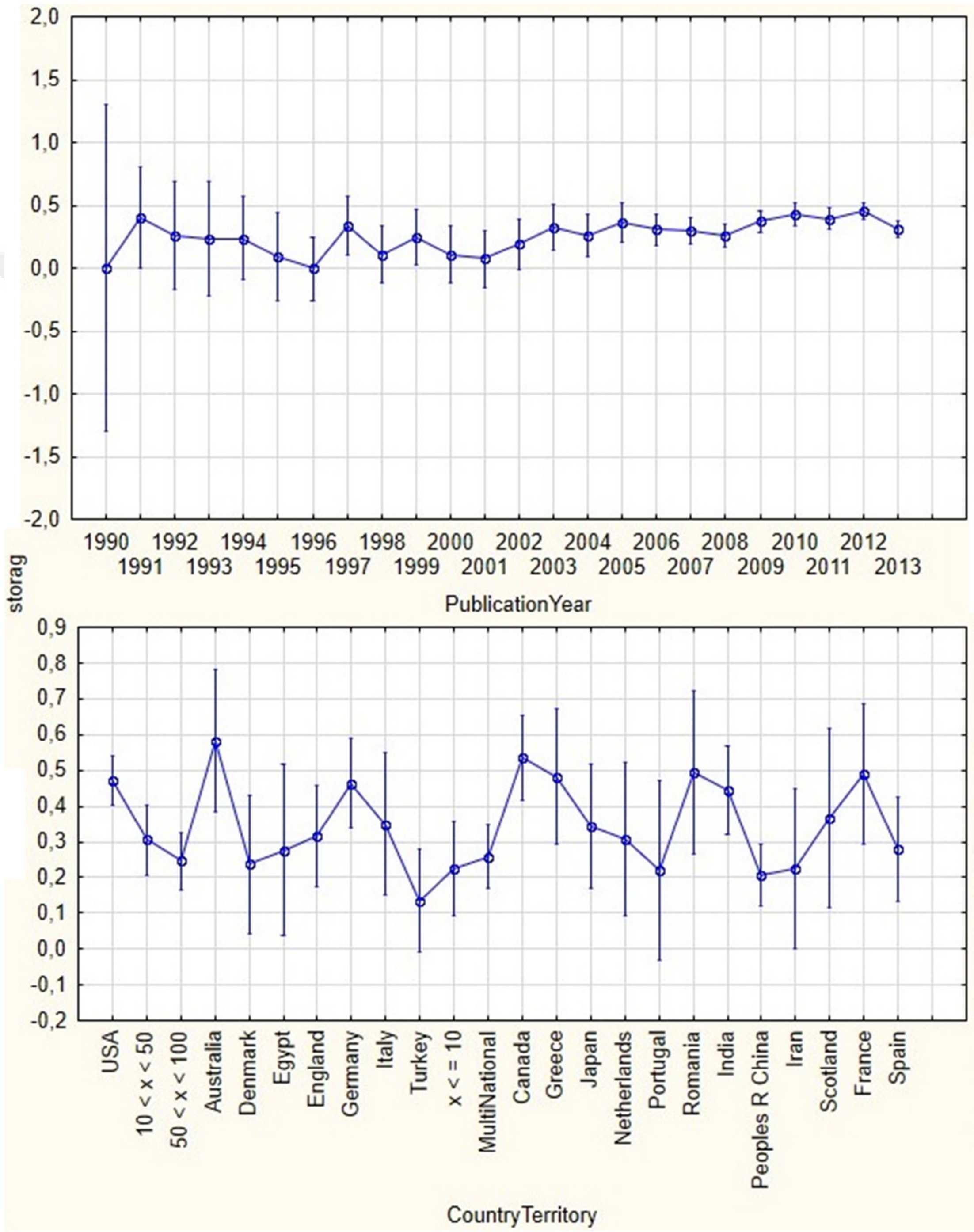

Figure 3. 11 Anova for the word *storage* vs Country/Territory and Publication Year

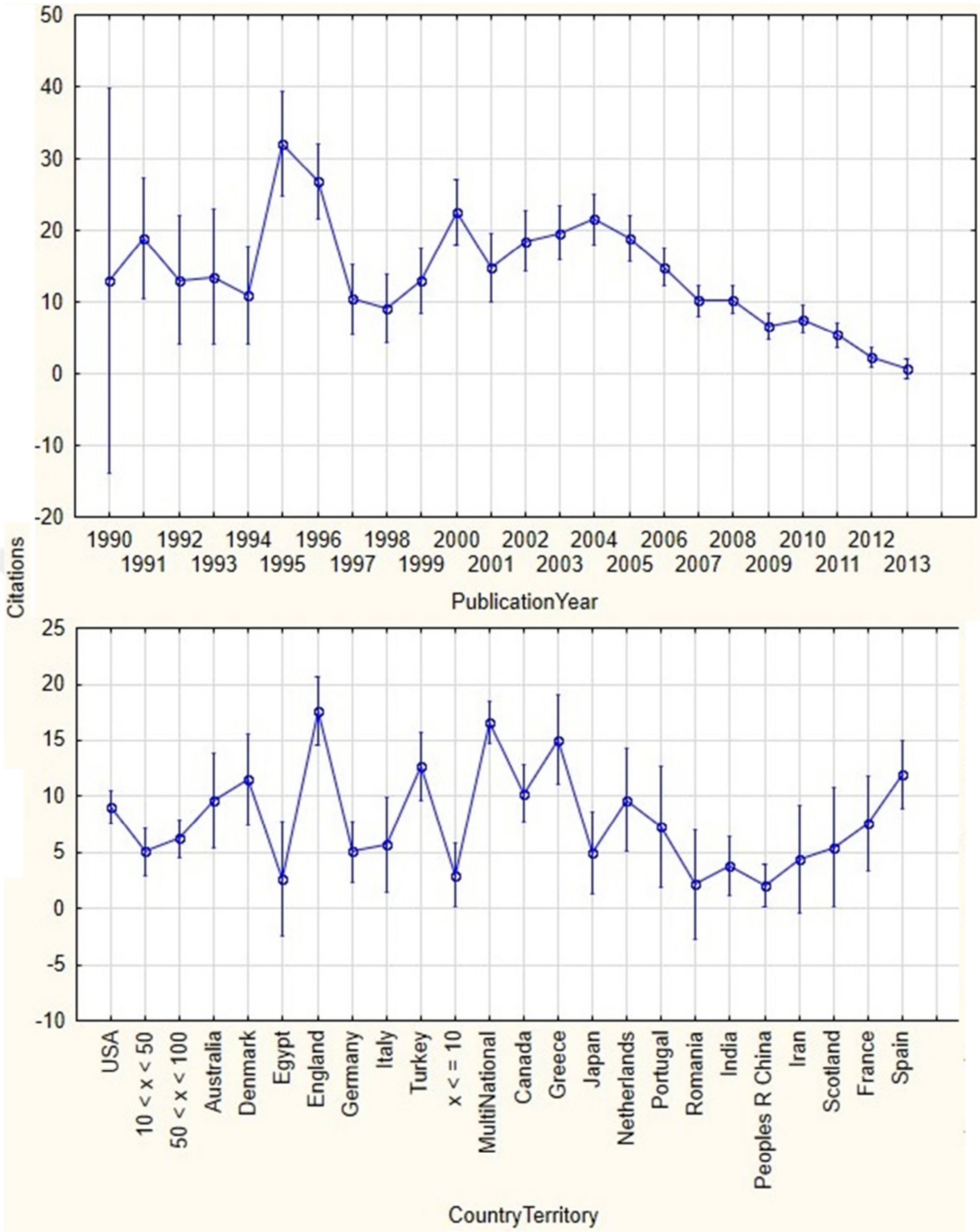

Figure 3. 12 Anova for citation numbers vs Country/Territory and Publication Year

Figure 3.12 tells an interesting story about citation numbers. If the publication is prepared with a collaboration of more than one country, it may have more citations. Also, if the author’s origin is England, Turkey, or Greece, the citation number will be higher. The mean citations of studies published in 1995, 2000, and 2004 are greater than other years. Recently, there is a decrease of mean citations due to increasing number of total published researches.

3.5 Conclusion

This chapter contributes a systematic and intelligent literature review to search hot topics and extract main concepts on wind energy related publications. The wind energy literature is analyzed to extract main concepts. 7415 publications between the publication years of 1990 to end of 2013 is stored from a most common used web based database, Thomson Reuter-ISI Web of Knowledge. Abstracts of stored publications are analyzed by text mining method.

The most frequent words in these publications are *system*, *power*, *generation*, *control*, and *model* respectively. By the way, most important words of this data set are different. *Control* is the most important topic in wind energy systems. Because of increasing installed wind turbines, *control* becomes hot issue for wind energy systems. Technological improvements on wind devices such as new sensors let wind systems to collect more data about the whole plant. These collected data can be used to take the wind energy plant under *control*.

Wind turbine technology groves up day by day. Huge wind turbines (7 – 8 MW capacity, greater than 100 meters rotor diameter) can be manufactured and these are installed off-shores to produce more electricity from a single turbine. Thus, it can be possible to produce more energy in less area. Therefore, the second important word is found as *offshore*. *Solar* energy is another high potential renewable energy source. It is the third important issue, but the trend of its importance decreasing especially after the year 2005.

All these main inferences of text mining can help technological forecasts of wind energy sector. For instances, it can be thought that control strategies still need to improve because trends of researches on it is positive. Vice versa, technologic investments on solar energy to use with wind energy lose its value. May be solar energy can be handled separately.

Inverse document frequency analysis gives the twenty-four concepts of wind energy publications. The most important concepts are related with *System Control Models*, *Electrical Control Systems*, and *Data Observation* respectively. This means that System Approach has significant value in wind energy area.

Following results can be summarized from Anova;

- If a country has offshore wind energy facilities, offshore related studies are important for them (Denmark, Netherlands, and Germany).
- Wind turbine manufacturers are interested in new designs to produce more energy. So that, technologically grown countries such as Japan and USA give more attention on blade studies. Because the blade technology has some similarities with wings of aerospace industry in terms of composite materials, England is also interested on blades.
- The most interesting result is that citation number of a publication can be affected by its authors' origin.
- Collaborative studies have more chance to be cited then others.

Through the light of all results, wind energy systems will need health monitoring strategies and optimization methods to maximize the utilization of installed wind turbines. There are not enough studies which give attention on the performance, productivity and efficiency of wind energy systems. Instead of being a crucial area for wind energy systems, Maintenance Planning subject is also the lack of academic researches.

CHAPTER 4

PROBLEM DEFINITION AND LITERATURE REVIEW

The problem definition of this thesis was clarified after a text mining procedure which was given in detail at Chapter 3. The results of the study show that the control issue is the hottest topic in wind energy studies. The current literature on control of wind energy systems summarizes that condition monitoring algorithms and methodologies are commonly used to control of wind energy plants. Thus, the focus of this thesis is on developing smart control tools by considering condition monitoring. This chapter gives the definition of the problem and a detailed literature review on the control and condition monitoring methodologies for the wind energy.

4.1 Problem Definition

Nowadays, number of wind turbines may be even more than 7000 with a total capacity of about 8 GW as in the world's largest wind farm is Gansu Wind Farm / Gansu – China (Turkey's largest wind farm is Soma Wind Farm / 169 wind turbines / 240 MW capacity). Commercial wind energy systems are generally built on rural areas and even today on the sea which are remote access places. Thus, management of a wind farm has more importance than before. The old turbine may lose their production efficiency more than 30 %. It is obvious that even a slightest failure can cause very significant cost by considering the general characteristics of these plants. Early detection and prevention of potential failures in wind power plants is of great importance for the efficiency, performance, and control of these plants. The prediction of any type of failure of a wind turbine will provide a major contribution to the company management on taking decisions to facilitate the planning of production and maintenance of wind farm.

Effective Condition Monitoring Systems seem to be important needs for wind energy investors to prevent or decrease unwanted turbine failures, to minimize unplanned system failures, and to maximize profitability. Existing systems in wind power plants can be considered as important tool for monitoring the performance of the plant. Sensor technology is rapidly increasing today and many wind turbine sensor data, which contain important information to manage wind farms in a controlled way, are stored in SCADA systems. Not only SCADA data can be used in retrospective analysis, but also they can be used as an instant monitoring tool. Because of this property, they are very suitable to be used in the development of intelligent control and optimization strategies which will enable the wind farm to be managed dynamically. You can found detailed information about SCADA systems and SCADA data in Chapter 2.

Firstly, it is required to determine the methodologies and algorithms that will allow control and monitor of wind farms. So, the next section gives a detailed review of current literature on control and monitoring techniques and algorithms of wind turbines. We want to find answers in the current literature to the following questions;

- Can SCADA data be used to control and monitor a wind farm in order to manage the farm in a smart way?
- What kind of parameters and algorithms can be used for processing and analyzing SCADA data to achieve the mentioned problem?

Then, the gap and opportunities in the literature are concluded as control and monitoring parameters which are discussed in the conclusion part of this section. This thesis focused on analyzing and processing of current SCADA data of a wind farm to contribute optimum management and monitoring tools. To achieve this goal, the SCADA data is analyzed statistically and the characteristic behavior of wind turbines are identified according to different parameters. This allows manager to monitor different or similar behaviors of each turbine by considering much kind of parameters (wind speed, wind direction, oil temperature, bearing temperature, produced power, and etc.) in the farm. By the way, there is a gap on the wind turbine clustering studies in the current literature. So that, a clustering study is also proposed to fill the gap and presented in Chapter 5 with a starting point of statistical analyze. The power production performance is also studied to gain turbine efficiency monitoring capability to the manager with three different tools, Data Envelopment Analysis (DEA), Malmquist Index (MI), and Stochastic Frontier Analysis (SFA), to look from different perspectives of performance monitoring of wind turbine and given in Chapter 6. It can be figured out from the current literature that wind farm power forecasting is another control and monitoring approach which gives the clues on health of wind turbines. There are many kind of algorithms and approaches to forecast power production. This thesis proposes a different tool, Particle Filtering (PF), combining Artificial Neural Network (ANN) on forecasting power which allows decrease in the forecast error which is given in detailed at Chapter 7. In addition to these tools, a new ANN is developed with a new learning algorithm (Antrain ANN) to forecast wind speed, wind power and wind turbine faults in the wind energy systems which are given in Chapter 8.

This chapter only contains the general literature review on control and monitoring approaches in the wind energy systems. All other detailed literature of the contributions of this thesis are given in their chapters separately.

4.2 Literature Review

Number of wind farms increasing all over the world and so do the need for condition monitoring of these farms which minimize highly costly downtimes and maximize profit. Fault diagnosis and condition monitoring of a wind energy system can be done through many kind of methodologies and tools which are explained in following sections. Current wind turbines have their own condition monitoring systems which provide various physical data of operating wind turbines and wind farms via SCADA systems. These data are useful to evaluate the health of the turbines and the system. Unfortunately, the CM and SCADA system of a wind farm are not always able to detect and predict faults before it occurs. Early detection of faults, forecasting the production, and monitoring the production efficiency may help the managers to plan maintenance periods of any turbine before it has a fault and make an effective master production schedule for the wind turbines. In the market, there are some commercial systems (Brüel Kjær Vibro, LIOS' EN.SURE, etc.) which works together with the SCADA system to make predictions on faults, production, and the efficiency of the system. On the other hand, this thesis provides novel tools for the condition monitoring issues of wind farm using available SCADA data.

Wind turbine operation and maintenance costs are quite high because wind turbine itself is a hard to access machine (quite high towers) and they are generally erected on remote areas (on mountains or off – shore) [125]. Turbine downtimes also cause decrease in reliability of wind turbines and low power production. In the early stages of wind energy maintenance, it was commonly re-active and there were no planned maintenance schedule. A wind turbine was generally working until it had a failure. It is the fact that most of the sub-systems of wind turbines, such as generators, gearboxes, blades, rotors , etc., my failure during operation [126]. Thus, CM helps managers to avoid unwanted possible faults or performance loss.

There are many possible CM techniques which were reported in the current literature which are listed in the next section.

4.2.1 Condition monitoring methodologies in wind energy systems

Condition Monitoring techniques vary such as physical methods (acoustic emission, strain monitoring vibration analysis, etc.), thermal monitoring (thermography, thermal analysis, temperature analysis), electrical analysis, and SCADA data analysis (performance analysis, statistical analysis, and etc.). Summary definitions and current references are given below.

Acoustic monitoring: If there is a change on the structure of a metal, then acoustic monitoring can be a tool to analyze elastic waves of a sudden release of strain energy. It has similar analyzing procedure with vibration monitoring but the difference is that while vibration sensors are mounted on the component, the acoustic sensors are located in the environment and listen to the component. It can be used to detect drive train, bearings and gearboxes, blade or tower defects. There are many current studies such as [127–129]. Unfortunately, the cost of this method is still higher to be used widely. However, the tools developed in this thesis are both easy to implement and cheap. They do not need any additional physical system component except current SCADA system and a computer.

Electrical Signal monitoring: Electrical signals such as current, voltage, power, and impedance can be used to monitor for detecting unusual phenomena [130–132]. Especially, current and voltage data may provide useful information about the health of the motors, generators and accumulators. Also, electrical resistance data can be used as another predictor to detect fatigue and cracks of wind turbine components.

Oil analysis: Oil quality is valuable for wind turbines especially in the gearbox gear and bearing. It was generally used offline by taking samples for securing the oil quality by inspecting contamination by parts/moist and securing the components by inspecting characterization of parts. Today, sensor prices are at an acceptable level which allows to analyze online and are able to give information about oil temperature, contamination, and moisture [133].

Performance Monitoring: The performance of a wind turbine can be used to monitor the health of a wind turbine [134, 135]. The wind velocity, generator speed and rotor speed provide meaningful information about generated power which depends on power equation of wind energy. If there are large deviations, there must be problem causing decrease on performance that need to be investigated.

Physical condition of materials: This technique is normally offline and focused on crack detection and growth.

Radiographic inspection: In this method, the absorption level of X-ray photons passing through the inspected part gives the main clue about the health of the inspected part. Although it provides accurate and crucial information about the structural condition of the component, it is only rarely used to monitor critical structural turbine components because of the high inspection costs [136].

Strain measurement: It is a common technique for general machine parts which uses strain gauges. The possible application area of this technique for the wind energy can be prediction of life time and securing the stress level. It is useful especially for the blades[137, 138]. In the long term analysis, the strain gauges are not robust; by the way, optical fiber sensors present advantages to be more robust.

Thermocouples: These are widely used cheap and reliable parts to monitor the gearbox, nacelle, generator bearings, oil and electronic temperatures in the wind energy[139].

Thermography: It is often used to monitor health of electric and electronic components [140]. Thermography can be applied in a fast and simple way that lets the technician observe heat differences on the components which give the clues about faults of hot spots. It is an off – line technique. Infrared cameras have been used in the wind energy systems to visualize variations in blade surface temperature that is able to indicate cracks. It also has the possibility to be applicable to electronic components or generators of wind turbines.

Ultrasonic testing: It is commonly used to check the health of blades or towers of wind turbines [141–144]. Especially, this method can be used to identify the geometry of defects and to predict their approximate dimensions.

Vibration analysis: The vibration occurs nearly all rotational machines such as wind turbines. Tracking the vibration data gives meaningful information regarding the health of the rotating machines. Thus, it is the most known and applied technique for CM. Main components which can be monitored with vibration analysis are bearings, compressors, gearboxes, motors, pumps, shafts, and turbines (gas and steam). The faults of wind turbine gearbox have high costs and time consuming which can be managed using vibration analysis. In contrast, vibration analysis is not useful technique for hybrid or hydraulic-drive wind turbines. Current researches on vibration analysis can be found in the references such as [145–149]. Although many commercial wind turbines have their own vibration sensors, it is expensive to add vibration sensors on a wind turbine if it does not have it.

Visual inspection: This is the base level of monitoring of any system. Any kind of anomalies of a wind turbine can be observable by the wind energy experts just looking at the wind turbines. On the other hand, this method need more expertise to figure out faults and generally need to be used with combination of other methods [150–153]

Process parameters: The sensor systems have become more improved according to early stages of the wind energy systems so do the control and monitoring issues. The SCADA system provides more data to monitor the whole system health.

Besides traditional CM methods listed below, various signal processing methods can be implemented to the SCADA data for the same purpose as follows;

- *Fast-Fourier Transform (FFT):* This is generally used to convert a digital signals into frequency domains. FFT commonly used to figure out gearbox damages by vibration data analysis [154–157].
- Filtering: There are many time dependent SCADA data of the wind turbines. These data may be redundant and is not able to use to provide meaningful information. Filtering methods (most common used one is Kalman Filter) can be adjust varying data especially in the gearbox and blade health monitoring [158–161].
- Hidden Markov: Wind energy systems have wind speed and energy production data that needs to be classified and figured out the trends. Hidden Markov models have successfully been applied to wind speed and weather forecast to monitor and control the wind turbine [162, 163].
- Statistical methods: Processing data of a wind turbine has generally high volume and variety. Therefore, statistics is indispensable in order to understand characteristics of data. One of the most commonly used statistical approach is root mean square to diagnose of failures [164]. Many other basic statistical parameters can be listed as average, minimum, maximum, mean, mode, standard deviation, crest factor, shape factor, impulse factor, kurtosis, and etc. which are useful to identify the health of a wind turbine. Also, analysis of variances and clustering methodologies can be used as a statistical approach to identify the behavior of the SCADA data. There are many recent studies which provides a statistical approach in fault diagnosis [148, 165–167].
- Trend analysis: This methodology is similar to statistical analyses in terms of data use. The main difference is that trend analysis seeks for upward or downward trends which allow monitoring inspected data. It is commonly used on power output monitoring and control studies with many other algorithms [71, 168].
- Wavelet transforms: It is another signal converter similar to Fourier transform with an advantage of being able used for non – stationary signals.

Current literature has many examples of its short term wind speed and power prediction studies [145, 169, 170]. It may also be used for monitoring the vibration level of moving components of wind turbine [171].

- Novel techniques: Novel techniques on control and monitoring for wind energy systems are based on SCADA data analysis. Fault diagnosis and condition monitoring issues are nowadays used with intelligent algorithms to predict unwanted breakdowns. There are model based and non-model based approaches which use SCADA data. Expert systems is one of the most important novel CM methodologies to identify and predict faults with a rule-based system [172]. The acquisition and new knowledge integration are the two main capabilities of expert systems. Artificial Intelligence (AI) techniques are also more popular in novel CM methods over recent years [173, 174]. Zaher and McArthur (2007) improved a multi-agent based wind turbine condition monitoring system [175]. Qui et al. (2016) used simulation of oil temperature data to predict faults [133] A deep auto-encoder model is presented by Wang et al. (2016) to monitor wind turbine blade breakage [176]. Liu (2016) contributed an artificial immune system to predict gearbox faults [177].

The introduction of SCADA make the online CM revolutionary [178]. Empirical Mode Decomposition technique is presented as a fully intelligent condition monitoring methodology [179]. Yang and Tian (2015) studied on the monitoring power quality by the electrical signals measured from the wind turbine generator [180]. The time series modelling, Monte Carlo methods, and Markov Chains are presented also as condition monitoring approaches [126, 181, 182].

Analyzing normal behavior of SCADA data is also presented as a CM tool so that the anomalies can be identified [73, 183, 184]. Schlechtingen et al. (2013) used a machine learning approach models to learn the normal behavior of a healthy component [73]. By the way, there are many other studies which use data mining and statistical methods to analyze the jerk data obtained from SCADA system of a wind turbine and to detect abnormalities of wind turbine in time domain [70–72, 168, 185–190].

To sum up, the review of the literature presents the following conclusions;

- Higher investment rates occur due to the increasing technical capacity of a wind turbine.
- Number of installed wind turbines is rapidly increasing.
- The cost of pro – active maintenance is high, therefore, there is a common need for fault prediction and health monitoring.
- CM has a growing trend for the wind energy researchers.
- CM companies are growing in wind energy area.
- Novel techniques need to be more focused to develop smart CM and control strategies.
- SCADA data has big importance in monitoring issues of wind energy systems.
- Clustering of wind turbines is lack of literature.
- Production efficiency analysis is lack of current literature.
- Power production forecast is handled as a condition monitoring issue.

4.3 Conclusions

After a deep literature on control algorithms and methods on wind energy sector, it is seen that the common problem to control a wind farm/turbine can be handled by Condition Monitoring systems. Current literature in Chapter 4 shows that CM systems are common feature of wind turbines. Actually, new generation wind turbines have some form of condition monitoring systems which have advanced significantly over recent years. However, many of these are not designed for especially wind turbines. Therefore many commercially CM systems are expensive and they still need manual analysis to give useful information.

This PhD thesis is inspired by the perceived lack of appropriate dynamic, smart and cheaper solutions to the control and monitor of wind energy systems. In this regard, this thesis proposes different tools to condition monitoring area of the wind energy market using only current SCADA data.

Developing smart condition monitoring system approaches within the domains of wind turbine power generation in a cheaper way are the main objectives of this thesis. The thesis aims to exploit system data which is already collected as part of SCADA systems.

The all data was collected from real-world operational wind farm in Turkey. As a result, the developed solutions have demonstrated capabilities and applicability to real world condition monitoring applications. Therefore, another contribution is that how existing SCADA systems could be used to monitor large wind turbines is demonstrated.

CHAPTER 5

STATISTICAL ANALYSIS TOOL: A CASE STUDY OF AN OPERATING WIND FARM IN TURKEY

This chapter presents a statistical analyzing approach of a wind farm to discuss the similarities, differences and characteristics of each wind turbine using 10-minute SCADA data. The introduction part gives the descriptions of analyzed parameters and related literature. Then, statistical approaches are described in the methodology with numerical analyses and results and final part gives the conclusions. Even if wind turbines are in the same wind farm and have totally identical technical features, power generation performance of any turbine may differs according to working conditions and some kind of predictable or unpredictable malfunctions. Thus, the main contribution of this paper focuses on a simple and adaptable method to identify different behavioral groups of turbines. The other goal of this study is to show the importance of clustering wind turbines in a single site to have better understanding on similar or dissimilar behaviors of identical turbines. Despite the fact that the monozygous twins have generally identical genotype, many studies show the differences between monozygous twins in the life time [191–196]. Based on this fact, the identification of differences and similarities between identical wind turbines in the same wind farm has come into prominence in order to accurately analyze, monitor and control issues. An early version of this study presented in an international conference [42].

5.1 Introduction

SCADA data of a wind farm has very crucial information for early detection and prevention of potential failures in wind power plants which were previously handled in detail in Chapter 4. SCADA data has enormous size and needs to be summarized in a way. Statistical analysis is able to manage this problem by providing meaningful brief information of total data. So that, a current operational data of a wind farm in

Turkey was analyzed statistically to identify the usability of them as control and monitoring parameters.

Wind energy industry has changeable volume capacities such as more than 100 turbines in a single site (Figure 5.1) or someone may have only 1 wind turbine in the market (Figure 5.2). It is easy to manage low volume wind energy systems but what if you have a large volume one? Generally, the same brand and model of wind turbines is erected in a single wind farm to reduce investment and operation costs. On the other hand, it is a fallacy to use the same control and monitoring models for all turbines in a wind farm. Figure 5.1 shows the layout of a large wind farm in Texas. Although all wind turbines are the same model of same brand in the farm, environmental effects (terrain, wind direction, wind speed, height, etc.) and operational effects (breakdowns, faults, maintenance, etc.) may change the characteristic behavior of the wind turbine. Consequently, statistical analysis and clustering methods are able to identify characteristic changes of same – model wind turbines by SCADA data analysis.

Figure 5. 1 A large wind farm in Texas [197]

Statistical analysis of wind turbine SCADA data is not directly mentioned in the current literature to the best of researcher’s knowledge. Harb et al. (2015) used k-means clustering approach and one-way ANOVA model to underwater wireless sensor data [198]. Lam et al. (2016) presented a configuration of energy positive curtain wall [199]. As ANOVA is generally used for social survey analysis, Karatepe et al. (2012) used it to define levels of awareness about the renewable energy sources of university students in Turkey [200]. Jaber et al. (2017) presented the relationship between demographics and knowledge level of senior students about renewable energy [201].

Figure 5. 2 A private wind turbine in Freiburg/Germany

In the wind energy area, current literature on ANOVA related studies may be listed as follows;

Howard et al. (2015) provided a n experimental investigation with descriptions on the statistics of wind turbine wake meandering [202], Padron et al. (2016) presented a novel way of computing the statistics of interests of a wind farm [203], and Rai et al. (2016) studied on load statistics of wind turbines [204]. On the other side, analysis of variances (ANOVA) is widely used in wind energy studies to prove proposed methods and applied tests [205–211].

Clustering approaches are more commonly used statistical methods in wind energy researches. K – means algorithm is used to select criteria for suitable error evaluation [212]. Goh et al. (2016) compared performance of Mycielski and k – means clustering algorithms on predicting wind speed correlation [213]. An interesting study on parameters of wake effect in a wind farm is presented by Al – Shammari et al. (2016) with a comparison of different clustering methods [165]. Shukla and Singh (2016) contributed a clustering based unit commitment under wind power uncertainty [214]. Voltage dips in wind farms studied by Garcia – Sanchez et al. (2016) with clustering analysis [215].

There is a gap in the literature on the use of statistical analysis and clustering methods on SCADA data for clarifying of characteristics of operating wind turbines in a wind farm. Therefore, this study provides a new area for usage of basic statistical methods and clustering approaches as a control and monitoring tool in wind energy area. Also, proposed approaches were studied on real operating wind turbines of a farm in east Mediterranean of Turkey. There are 16 wind turbines in our wind farm having following technical features; capacity: 3 MW, rotor diameter: 90 m., hub height: 80 m. Figure 5.3 gives the actual layout of the turbines with its terrain height. The prevailing wind directions are north – west in the summer and south – east in the winter (number of arrows represents ratio of speed rates in summer and winter times with a meaning of that summer times have higher wind speeds than winter).

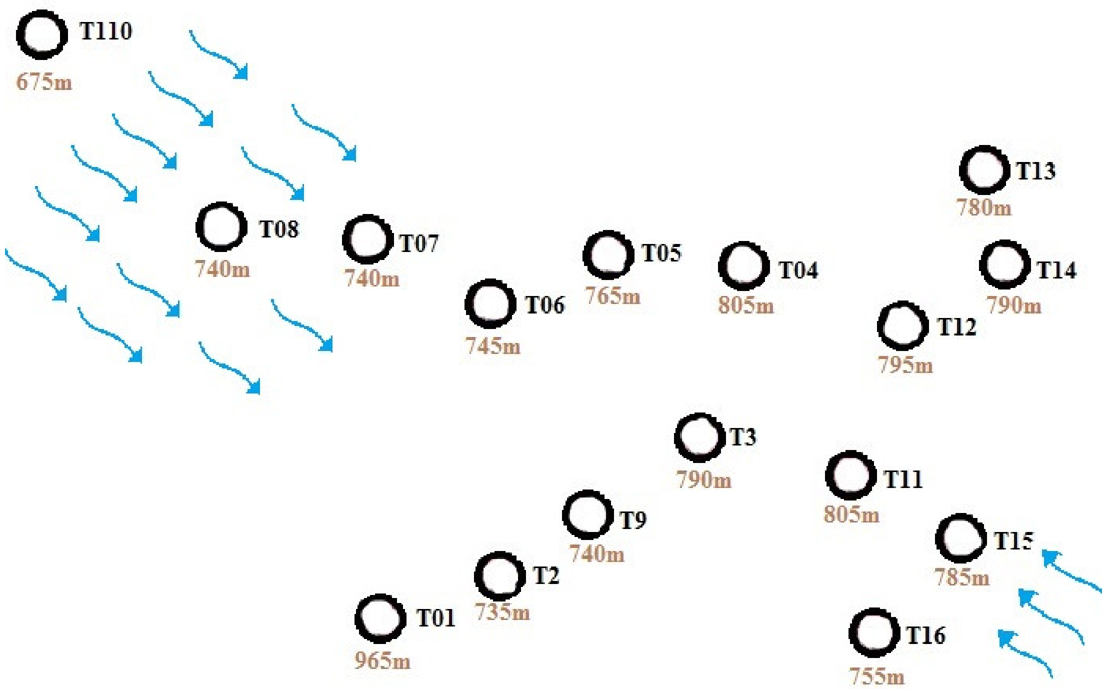

Figure 5. 3 The layout of considered wind farm in Turkey

Firstly, 10-minute interval data during one year was gathered from the SCADA system of the wind farm. Then, basic statistics such as range, minimum, maximum, mean, standard deviation, variance, skewness, and kurtosis were computed. After that, various hypotheses were tested to have better understand of the data. Finally, clustering was applied to group similar turbines and proved its usefulness on a forecasted model.

5.1.1 Descriptions of used data

The SCADA data used in this study is explained as below;

Ambient Wind Direction Absolute is the wind direction in degrees which is measured by anemometer at the top of each turbine. It is especially used for the design of wind farm layout problems to minimize the power losses [216, 217]. In spite of the fact that wind direction does not considered in power model of wind energy, designers has to take into account it to avoid higher wake effect to minimize power losses [218, 219].

Wind Speed is speed of blowing wind in meters per second (m/s) which is measured by anemometer on the nacelle of each turbine. It is the main parameter wind energy model [220]. The scatter diagram of wind speed versus power production gives the power curve of the wind turbine that may designate the anomalies of turbines. Because the wind speed data is a 10-minute interval average data, the maximum and the minimum wind speeds of each 10-minute interval are also stored. Thus, ***Wind Speed Maximum of 10-minute interval*** is defined as maximum wind speeds of 10-minute intervals, and ***Wind Speed Minimum of 10-minute interval*** is defined as minimum wind speeds of 10-minute intervals.

Active Power (Grid Production Power) is the generated power in kW for an averaged wind speed. The maximum power is 3000 kW for all turbines. Generally, it is the main objective which needs to be is maximized in wind energy investments. There are many studies take it as an objective function [221, 222] by considering layout optimization, and some others consider it in a control strategy [223, 224]. On the other hand, ***Reactive Power (Grid Production Reactive Power)*** is the amount of consumed average reactive power of each turbine in the unit VAr. It is used in control studies of wind turbines having some possibilities, limits, pros and cons which are summarized in reference [225].

Blades Pitch Angle is used to define the average of turning angle of all three blades into or out of the wind to control the power production.

Rotor Speed is the revolution per minute (RPM) of the rotor of a wind turbine.

Generator Speed is the RPM of the generator which generates the electrical power.

A nacelle is the box which contains the generator, gearbox, drive train, and brake assembly. ***Nacelle Temperature*** is the environmental temperature in ºC (degree Celsius).

Generator Bearing Temperature is the temperature value in ºC of the bearing which is a component of generator that constrains relative motion to only the desired motion.

There are many components in the wind turbine which uses hydraulic oil. Thus, ***Hydraulic Oil Temperature*** is an important data to monitor the health of the hydraulic system.

Spinner Temperature is the temperature in the hub housing where the blades are pitching on the hub.

Controller Hub Temperature gives the temperature at pitch controller.

Grid Busbar Temperature is the temperature of the busbar of a wind turbine is designed to conduct a substantial current of electricity.

5.2 Methodology

5.2.1 Descriptive statistics

Descriptive statistics is the first and easy implemented approach to have a fast look in big SCADA data. It gives base level statistical information about interested parameters. Table 5.1 gives descriptive statistics of the 10-minute SCADA data of 16 wind turbines during one year. The descriptions of the columns name are given below to understand the meanings of these basic statistical concepts.

- N: It gives the number of data. Table 5.1 shows that the number of data for each parameter is not the same exactly, there are some missing values for the same time period.
- Range: It is the differences of the maximum and minimum values of the same parameter.
- Minimum: It is the smallest value of the parameter.
- Maximum: It is the biggest value of the parameter.
- Mean: It is the average value of the parameter that is computed by dividing the sum of all data of the same parameter to the number of elements.
- Standard deviation: It is the square root of the variance of the parameter.
- Variance: It is the average of the squared differences from the mean of the parameter.
- Skewness: It gives the asymmetry of the statistical data. It is useful to identify differences of the normal distribution of the parameter by considering right or left side skew on the normal curve.
- Kurtosis: It explains the differences of the normal distribution of the parameter considering heavy – tailed or light – tailed curves relative to normal curve.

Now, one can easily read Table 5.1 using these definitions. For example, Grid Production of the wind farm for one year time period can be summarized as:

It has 832821 10-minute average production data in this file, while the minimum production is – 40,5 kW, the maximum is 3003,5 kW with a range of 3044 kW. The mean value of the production is 998,66 kW , standard deviation is 925,59 kW, and variance is 856719,71 kW. While the skewness is 0,678 which means the production parameter data is right skewed relative to normal curve, the kurtosis is – 0,804, which means the curve of production parameter is heavier (wider) than normal curve.

Table 5. 1 Descriptive Statistics for All Turbines

Parameters	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness	Kurtosis
Wind Speed Avg.	832821	31,1000	0,0000	31,1000	8,105018	3,9443545	15,558	,466	-,088
Wind Speed Max.	832821	40,5000	0,0000	40,5000	11,713147	5,3284516	28,392	,305	-,429
Wind Speed Min.	832821	19,8000	0,0000	19,8000	4,784004	2,7111144	7,350	,529	-,005
Ambient Wind Direction Absolute Avg.	832821	359,9000	0,0000	359,9000	251,891702	96,2965901	9273,033	-,702	-1,288
Grid Production Power Avg.	832821	3044,0000	-40,5000	3003,5000	998,661994	925,5915488	856719,715	,678	-,804
Grid Production Reactive Power Avg.	832821	857,2000	-87,9000	769,3000	129,167060	166,5973275	27754,670	1,251	,467
Rotor RPM Avg.	832837	16,1000	0,0000	16,1000	12,156452	4,7466980	22,531	-1,412	1,004
Generator RPM Avg.	832837	1692	0	1692	1278,52	497,938	247941,762	-1,415	1,012
Generator Bearing Temp, Avg.	832837	100	0	100	47,30	10,802	116,683	,852	,450
Hydraulic Oil Temp, Avg.	832837	55	0	55	46,40	2,319	5,376	-1,448	12,924
Nacelle Temp, Avg.	832837	53	0	53	30,62	5,153	26,552	,388	-,328
Grid Busbar Temp, Avg.	832837	53	0	53	32,78	6,181	38,203	,245	-,594
Controller Hub Temp, Avg.	832837	46	-1	45	26,76	5,219	27,242	-,035	-,702
Spinner Temp, Avg.	832837	39,0000	-1,0000	38,0000	20,620488	5,1602545	26,628	-,050	-,719
Blades Pitch Angle Avg.	832821	93,4000	-2,7000	90,7000	2,582108	12,4398975	154,751	4,781	27,521
Valid N (list wise)	832821								

Table 5. 2 Descriptive Statistics for Turbine 1

Parameters	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness	Kurtosis
Wind Speed Avg.	52476	23,9000	0,0000	23,9000	6,433290	3,4222230	11,712	,875	1,101
Wind Speed Max.	52476	31,6000	0,0000	31,6000	9,652435	5,0153813	25,154	,749	,354
Wind Speed Min.	52476	17,9000	0,0000	17,9000	3,537538	2,1042660	4,428	,892	1,201
Ambient Wind Direction Absolute Avg.	52476	359,9000	0,0000	359,9000	196,256757	95,2245261	9067,710	,460	-1,462
Grid Production Power Avg.	52476	3034,5000	-34,0000	3000,5000	615,267345	731,1646541	534601,751	1,435	1,331
Grid Production Reactive Power Avg.	52476	692,8000	-47,7000	645,1000	71,327441	127,3505479	16218,162	2,137	4,117
Rotor RPM Avg.	52476	16,1000	0,0000	16,1000	10,393182	5,1820363	26,854	-,924	-,374
Generator RPM Avg.	52476	1689	0	1689	1093,79	544,153	296102,521	-,925	-,371
Generator Bearing Temp, Avg.	52476	88	0	88	41,64	9,438	89,070	1,324	2,293
Hydraulic Oil Temp, Avg.	52476	55	0	55	45,73	2,292	5,254	-1,806	11,493
Nacelle Temp, Avg.	52476	39	0	39	25,63	3,065	9,392	1,835	2,557
Grid Busbar Temp, Avg.	52476	44	0	44	27,83	4,126	17,022	,852	,294
Controller Hub Temp, Avg.	52476	31	0	31	20,27	2,700	7,289	,166	,150
Spinner Temp, Avg.	52476	26,0000	0,0000	26,0000	14,394295	2,8735944	8,258	,239	,303
Blades Pitch Angle Avg.	52476	91,4000	-2,7000	88,7000	4,414090	14,7309689	217,001	3,842	17,775
Valid N (list wise)	52476								

Of course Table 5.1 is a general look for all data of the farm and it needs to be specialized according to turbines. Therefore Table 5.2 gives the basic descriptive statistics of the Turbine 1. It can be easily seen that the mean of the production value for the Turbine 1 is less than all over the farm. Also, other parameters have differences. This explains why deduction from general analysis of a wind farm should not be applied the same type wind turbines in a single site. Similarly, these basic statistics can be applied for all other turbines in the wind farm.

5.2.2 Graphical analysis

Graphs may also be a quick guideline for interpretation of the data other than the basic descriptive statistics. Figure 5.4 shows the scatter diagram of ambient wind speed average and power production of all farm. What you see is something like a dinosaur who eats some food overflowing of his mouth. In fact, the chin of the dinosaur is a negative shift on power production and overflowing food is loss of power efficiencies. The ideal curve of a wind speed – power production relationship must be as the red line in the Figure 5.4.

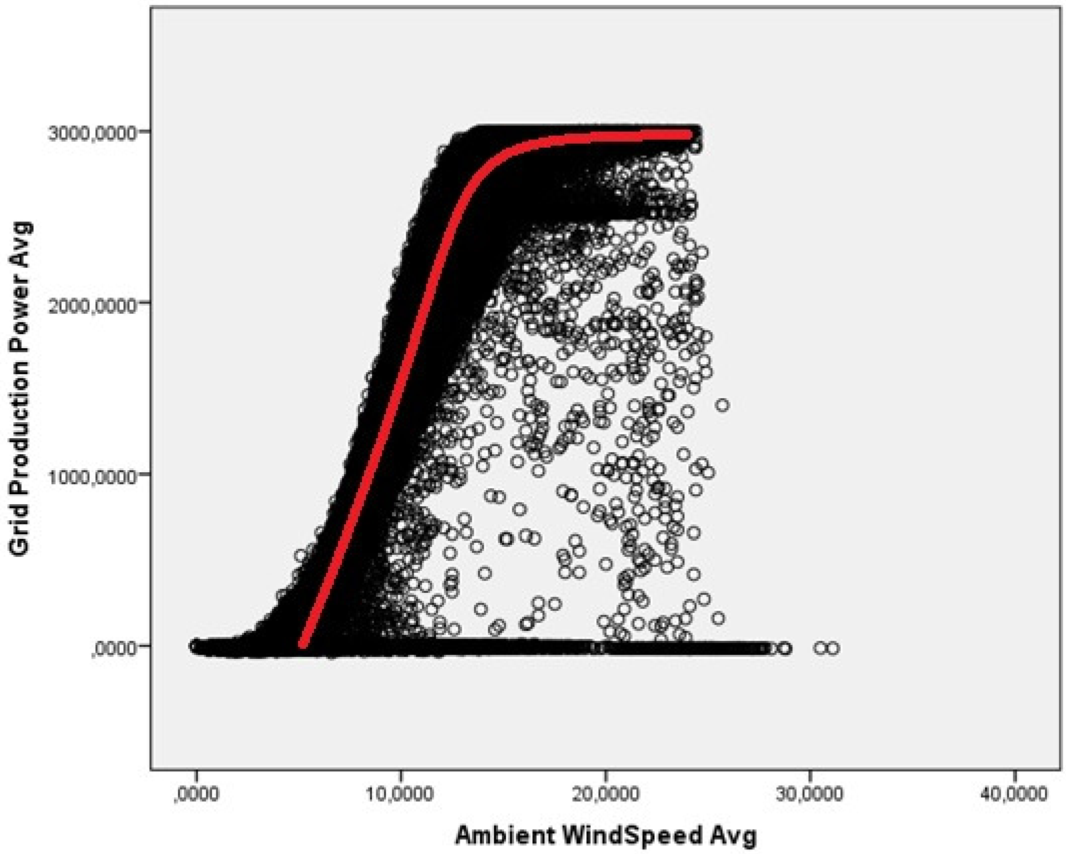

Figure 5. 4 Power production vs. wind speed characteristics of the wind farm.

Also, graphs also may help to figure out distribution clustered data. Figure 5.5 presents the wind speed categories during a one – year period of time for the wind farm. The wind speed averages are mostly in the range of cut – in speed and rated speed of the turbines. By the way, the share of less than cut – in speed is higher than rated speed’s share.

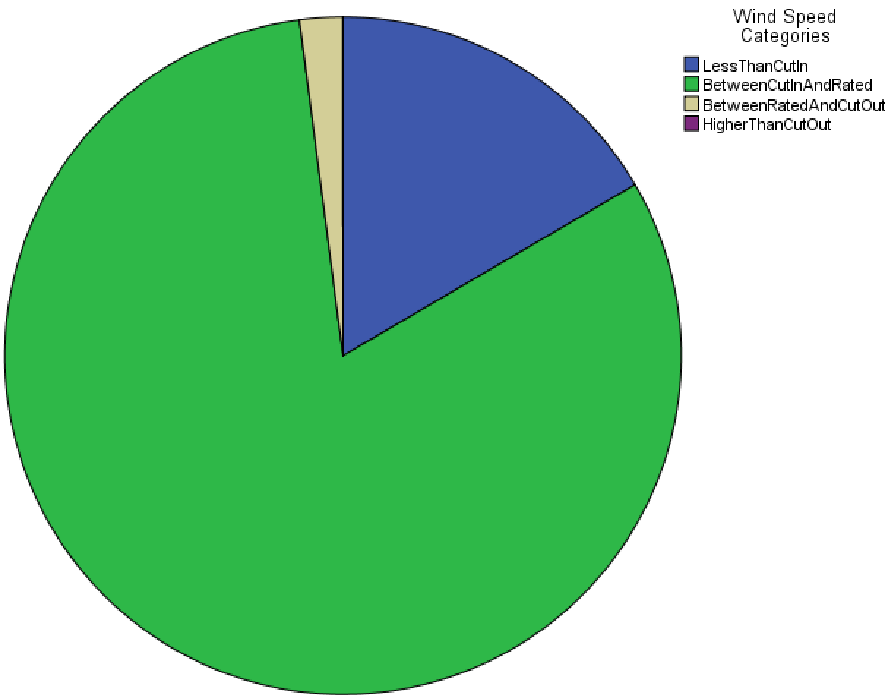

Figure 5. 5 Wind speed category shares for the wind farm

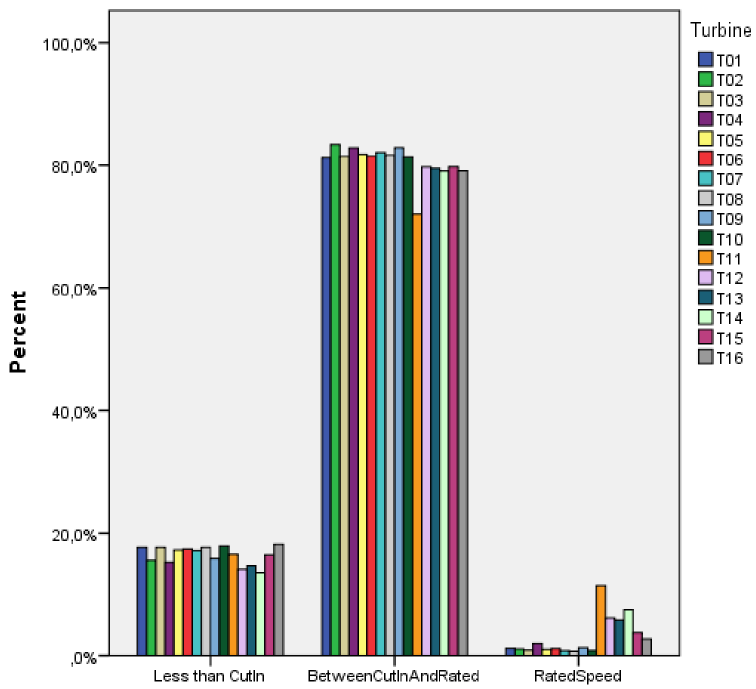

Figure 5. 6 Wind speed category shares for each turbine

Because the general tables and figure for the wind farm does not give any individual information about the turbines, Figure 5.5 and Figure 5.6 can be re-drawn for each turbine.

Figure 5.6 gives a bit more detailed information than Figure 5.4 about wind categories of turbines. Especially for the rated speed category, Turbines 11 has the biggest share followed by Turbines 12, 13, 14, 15, and 16. This figure may give a result of that the layouts of Turbines 11, 12, 13, 14, 15, and 16 is better than other turbines considering higher wind speeds.

5.2.3 Data reduction analysis

SCADA data of a wind farm contain more than 120 different parameters. Generally, data are stored in 10-minute averages for all parameters in the common commercial SCADA systems. Sometimes, 10- second time interval is also available. This situation causes to have big data in the systems to analyze. Grouping the many kinds of data for a determined period provides meaningful and summary information about which data or data group need to be focused. This procedure is generally known as data reduction technique. While Principle Component Analysis and Factor Analysis methods are used to reduction of quantitative data, Multiple Correspondence Analysis is used for categorical data. The aim of data reduction is to eliminate redundant variables from the data file.

Principle component analysis methodology is used to identify data groups. In this study, there are only 15 parameters from SCADA data. However, current SCADA systems can provide more than 120 different types of data. Moreover, this data can be stored in 10-minute intervals or even in 1-second intervals. This shows how big the size of the data obtained from SCADA systems is, and how difficult it is to analyze. Therefore, principle component analysis was used as a data reduction tool. Figure 5.7 shows the explanation ratio of components for the current data. If the number of selected components is 5, it represents a total of 91,2 % data and it is an acceptable component number.

By the way, the results of data reduction can be figured out from Table 5.3. It shows the final scores of each parameter for each component (factors). The maximum score of each parameter in the components is used to assign a parameter in a component. In this study, factors are consists of following parameters;

Factor 1: Ambient Wind Speed Avg, Ambient Wind Speed Max, Ambient Wind Speed Min, Grid Production Power Avg, Grid Production Reactive Power Avg, Generator Bearing Temp, Avg, Grid Busbar Temp, Avg.

Factor 2: Nacelle Temp, Avg, Controller Hub Temp, Avg, Spinner Temp, Avg, Blades Pitch Angle Avg.

Factor 3: Rotor RPM Avg, Generator RPM Avg.

Factor 4: Hydraulic Oil Temp, Avg

Factor 5: Ambient Wind Dir Absolute Avg

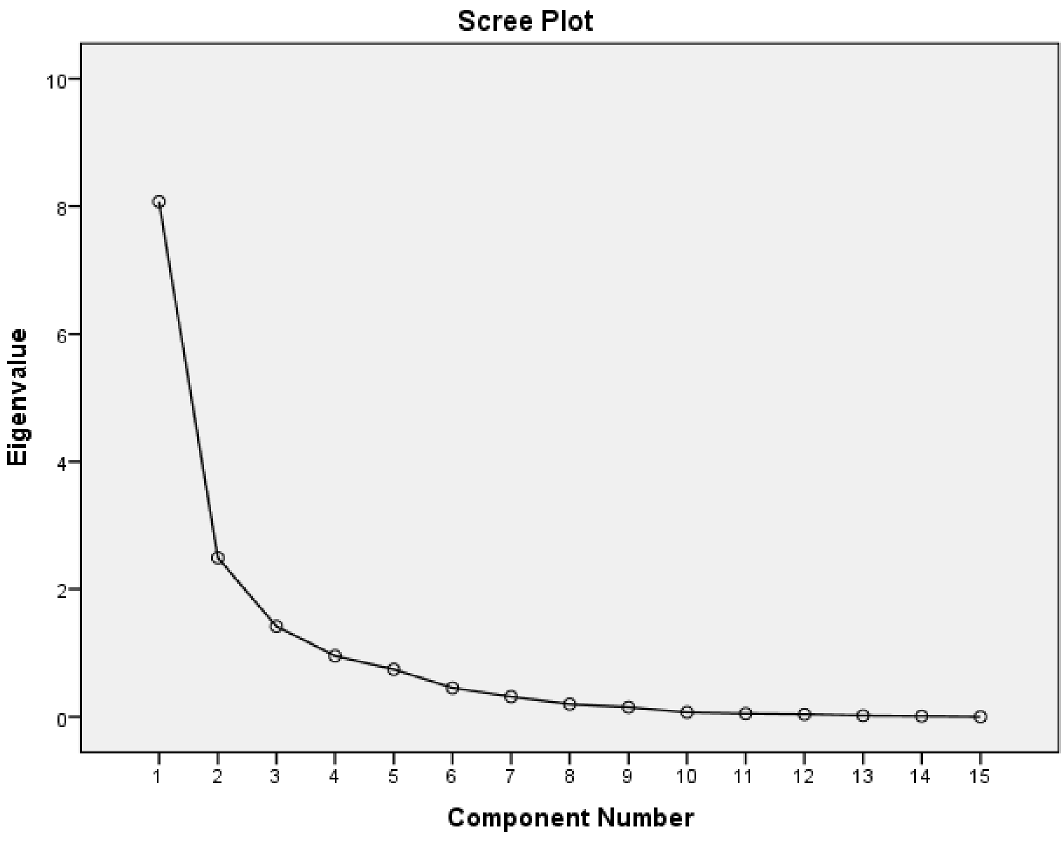

Figure 5. 7 Scree plot for principle component analysis of SCADA data

All power related parameters are grouped in the same component (Factor 1), in addition to this, bearing temperature and busbar temperatures are also in this group. If this factor is titled as "*power factors*", bearing and busbar temperatures can be handled in the effects of power.

All other environmental and non-rotational machine type heat parameters, nacelle, controller hub, spinner temperatures, are grouped in the second component. Blades pitch angle is also in the same component with least effect value (0,004) that means it has not valuable effect on this group.

Table 5. 3Principle component analysis for SCADA data

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
Ambient Wind Speed Avg.	**,961**	,111	,165	-,064	,059
Ambient Wind Speed Max.	**,944**	,110	,175	-,052	-,025
Ambient Wind Speed Min.	**,890**	,123	,179	-,090	,186
Ambient Wind Direction Absolute Avg.	,096	,255	,067	,027	**,954**
Grid Production Power Avg.	**,952**	,142	,140	,052	,047
Grid Production Reactive Power Avg.	**,899**	,135	,009	,227	,032
Rotor RPM Avg.	,659	,096	**,716**	-,078	,080
Generator RPM Avg.	,659	,096	**,716**	-,078	,079
Generator Bearing Temp, Avg.	**,676**	,270	,247	,050	-,039
Hydraulic Oil Temp, Avg.	,025	,216	,097	**,958**	,025
Nacelle Temp, Avg.	,551	**,732**	,076	-,023	,074
Grid Busbar Temp, Avg.	**,679**	,632	,151	,067	,077
Controller Hub Temp, Avg.	,082	**,943**	,020	,117	,136
Spinner Temp, Avg.	,080	**,951**	,001	,162	,133
Blades Pitch Angle Avg.	-,073	**,004**	-,952	-,169	-,022

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 6 iterations.

Rotational speeds (rotor and generator) are grouped in the same factor with a same effect value 0,716.

Hydraulic oil temperature is the only group member for the 4th component which means that it needs more attention in the analysis because it is the only parameter in the fourth component. Similarly, wind direction is the 5th factor.

5.2.4 Non-parametric tests

In this section, non-parametric hypothesis test are applied to the SCADA data to figure out any distribution patterns. The Normality test is commonly used to decide whether the data fits in a normal distribution or not by using One-Sample Kolmogorov-Smirnov test. Table 5.4 gives the results of normality test for all parameters. As the test results shows, no parameters has the normal distribution with a significance level of 0,05.

Table 5. 4 Normality tests of SCADA data

Hypothesis Tests 1				
Nr.	**Null Hypothesis**	**Test**	**Significance**	**Decision**
1	The distribution of Wind Speed Average is normal with mean 8,11 and standard deviation 3,94	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
2	The distribution of Wind Speed Max is normal with mean 11,71 and standard deviation 5,33	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
3	The distribution of Wind Speed Min is normal with mean 4,78 and standard deviation 2,71	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
4	The distribution of Wind Direction is normal with mean 251,89 and standard deviation 96,30	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
5	The distribution of Production Power is normal with mean 998,66 and standard deviation 925,59	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
6	The distribution of Reactive Power is normal with mean 129,17 and standard deviation 166,6	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
7	The distribution of Rotor Speed is normal with mean 12,16 and standard deviation 4,75	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
8	The distribution of Generator Speed is normal with mean 1278,52 and standard deviation 497,94	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
9	The distribution of Generator Bearing Temperature is normal with mean 47,3 and standard deviation 10,8	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
10	The distribution of Hydraulic Oil Temperature is normal with mean 46,4 and standard deviation 2,32	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
11	The distribution of Nacelle Temperature is normal with mean 30,62 and standard deviation 5,15	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
12	The distribution of Bus - Bar Temperature is normal with mean 32,78 and standard deviation 6,18	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
13	The distribution of Controller Hub Temperature is normal with mean 26,76 and standard deviation 5,22	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
14	The distribution of Spinner Temperature is normal with mean 20,62 and standard deviation 5,16	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis
15	The distribution of Blades Pitch angle is normal with mean 2,58 and standard deviation 12,44	One - Sample Kolmogorov - Smirnov Test	0	Reject the Null Hypothesis

Table 5. 5 Similarity test of SCADA data regarding with turbine categories

Hypothesis Tests 2				
Nr.	**Null Hypothesis**	**Test**	**Significance**	**Decision**
1	The distribution of Wind Speed Average is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
2	The distribution of Wind Speed Max is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
3	The distribution of Wind Speed Minis the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
4	The distribution of Wind Direction is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
5	The distribution of Production Power is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
6	The distribution of Reactive Power is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
7	The distribution of Rotor Speed is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
8	The distribution of Generator Speed is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
9	The distribution of Generator Bearing Temperature is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
10	The distribution of Hydraulic Oil Temperature is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
11	The distribution of Nacelle Temperature is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
12	The distribution of Bus - Bar Temperature is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
13	The distribution of Controller Hub Temperature is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
14	The distribution of Spinner Temperature is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis
15	The distribution of Blades Pitch angle is the same across categories of Turbines	Independent - Sample Kruskal - Wallis Test	0	Reject the Null Hypothesis

There are many other distributions such as exponential, binomial, or random. All these distributions also tested for each parameters and results shows that no parameters are fit in these distributions. Therefore, the studied SCADA data is labeled as non-parametric data.

Homogeneity test also gives an idea of data distribution. It is also applied for all data regarding with turbine numbers and results showed that the SCADA data of each turbines are observed in non-homogeneously with a significance level of 0,05.

Now, the differences or similarities of turbines characteristic behaviors can be identified by applying non-parametric analysis methods (Chi-square tests). If the null hypothesis is that all wind turbines have the similar wind speed conditions, it can be tasted with chi-square test on wind turbines versus wind speed categories. Results showed that there is a significant differences between wind speed categories for each turbine which also observable in Figure 5.5 (significance level is 0,05 and p value is 0,000).

Analysis of variance (ANOVA) test is commonly used to identify similarities and differences of parameters regarding with any categorical groups. However, it is designed to measure similarities of variances of normal distributed data. Because studied data was failed to normality tests, Kruskal Wallis test was used to identify similarities and differences of parameters regarding with turbines. Table 5.5 shows that no parameters have similar behavior regarding with turbine categories.

In spite of the fact that all wind turbines in the considered wind farm have the same brand and model, applied hypothesis tests shows that the have different parametric properties. These results also show that wind turbines need different tools for control and monitoring. Now, it is time to apply clustering methods to identify similar-behavior turbines.

5.2.5 Clustering methods

The main purpose of statistical analysis is to predict future data as accurately as possible. Clustering tries to group meaningfully similar things together. So, rather than making one single model for a large SCADA data, it makes sense to divide the data into N similar groups and develop N simpler models. Also, clustering help in reducing the dimensionality of the problem. Following are the main characteristics of clustering methods;

- It find naturel groupings among objects/parameters/cases,
- It shows internal data structure of the whole data.

- It is used to make an easy partition on the data.
- It is useful in knowledge discovery.
- It is able to analyze different types of data (as in SCADA data).
- It is able to deal with noise and outliers of data.

There are two main categories of clustering algorithms as partitioned clustering and hierarchical clustering. In this study, a k-means partitioned clustering and hierarchical clustering algorithms are used to group SCADA data and wind turbines.

5.2.5.1 Partitioned clustering

Partitioned clustering methodology is a non-hierarchical clustering type. K-means clustering algorithm finds k centers to assign N cases regarding with neighborhood of their values to these centers. The cases are iteratively adjusted so that each of the N cases is assigned to exactly one of the k clusters.

In this study, this method is used to cluster SCADA data observations to clarify the pattern differences. By the way, Turbine 11 and 15 were eliminated for only this analysis because they had quite more missing data than others. All data are selected as inputs to clarify clusters. First of all, the number of clusters is determined by an iterative 10-fold validation test. Figure 5.8 gives the cost of cluster numbers.

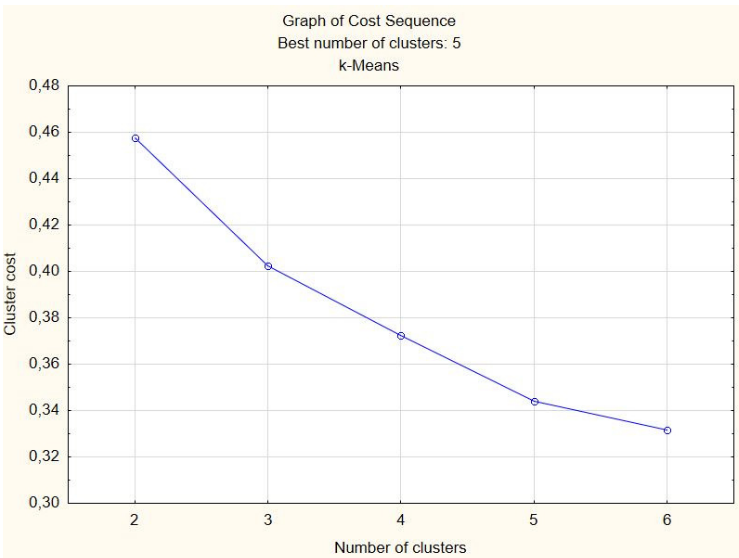

Figure 5. 8 Cost of clusters to clarify the number of clusters

The best number of cluster is 5, and the observations are clustered in 5 clusters for all SCADA data. By the way, Table 5.6 gives the centroids for each cluster. All these results are significant at 95% confidence level. In addition to this, each parameter is divided into 5 groups which are useful to make further analysis within the groups.

Table 5. 6 Centroids of each cluster for each SCADA data

Cluster ID	Wind Direction	Wind Speed (m/s)	Wind Speed Max (m/s)	Wind Speed Min (m/s)	Active Power (kW)
1	195,95	6,04	9,18	3,24	491,92
2	177,19	8,70	12,88	4,78	1059,26
3	318,81	12,23	16,86	7,68	1992,63
4	312,07	8,88	12,64	5,42	1150,97
5	262,50	5,88	8,62	3,41	447,46

Cluster ID	Reactive Power (VAr)	Pitch Angle Avg.	Rotor Speed (RPM)	Generator Speed (RPM)	Nacelle T.
1	51,29	4,88	9,84	1035,71	25,64
2	131,52	3,17	12,33	1297,13	28,23
3	288,23	0,76	15,63	1642,44	36,86
4	156,17	-0,32	13,98	1470,18	34,308
5	37,04	4,15	10,00	1052,94	30,07

Cluster ID	Generator Bearing T.	Hydraulic Oil T.	Spinner T.	Controller Hub T.	Grid Busbar T.
1	41,33	46,11	15,93	21,92	28,24
2	47,28	46,01	16,03	22,09	31,35
3	55,16	47,44	25,92	31,89	40,34
4	51,23	47,00	25,58	31,42	36,65
5	43,90	46,65	22,13	28,09	31,96

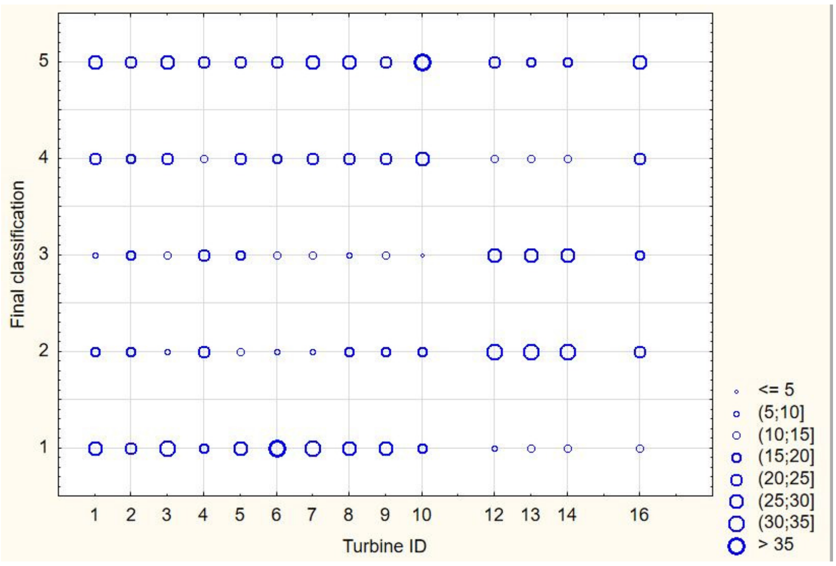

Figure 5. 9 The frequency graph of observations for Clusters versus Turbines.

Figure 5.9 gives the frequency graph of observations for the clusters and turbines. It would be better to read Table 5.6 and Figure 5.9 together to comment each turbines' behavior. For example, the observations of Turbine 10 are mostly in the cluster 5 (more than 35 observations), and generally in 4 (the observation number is between 30 and 35). Let's look at Table 5.6 to have some information about the observations of Turbine 10 (T10). The centroids (similarly center points) of wind direction of T10 are mostly 312º and 262º respectively. The wind speed centroids are generally 8,8 m/s and 5,8 m/s. The active power observations of T10 are generally in a cluster of 1150 kW and 447 kW centroid points. The rotor speed is generally around 10 and 13 rpm. The generator bearing temperature is mostly around 43,9ºC and 51,2ºC. On the other hand, the observations of turbines 12, 13, and 14 are mostly in the cluster 2 (more than 35 observations) and generally in the cluster 3 (the observation number is between 30 and 35). Thus means that turbines 12, 13, and 14 have the observed values as follows: the wind direction is mostly around 177º and generally around 318º, the average wind speed is mostly around 8,8 m/s and generally around 12,2 m/s, the active power is mostly around 1059 kW and generally around 1992 kW, the rotor speed is mostly around 12,3 rpm and generally around 15,6 rpm, and the generator bearing temperature is mostly around 47,2ºC and generally around 55,1ºC.

5.2.5.2 Hierarchical clustering

This methodology creates a hierarchical decomposition the sets of objects/Turbines. There is no need to specify the number of clusters in this method. The result of analysis is a binary tree of the data that merges similar groups of objects. It only requires a measure of similarity between grouped data. In this study, agglomerative clustering, a monotonic approach, is used with average linkages between groups. The similarity between clusters is monotone decreasing with the level of the merge. The results are given in the dendrogram plot which shows the similarity between the two merged groups.

Figure 5.10 gives the dendrogram plot of hierarchical clustering analysis of wind turbines according to their locational heights. Actually, hierarchical clustering does not give a cluster number; dendrogram helps to identify similar groups.

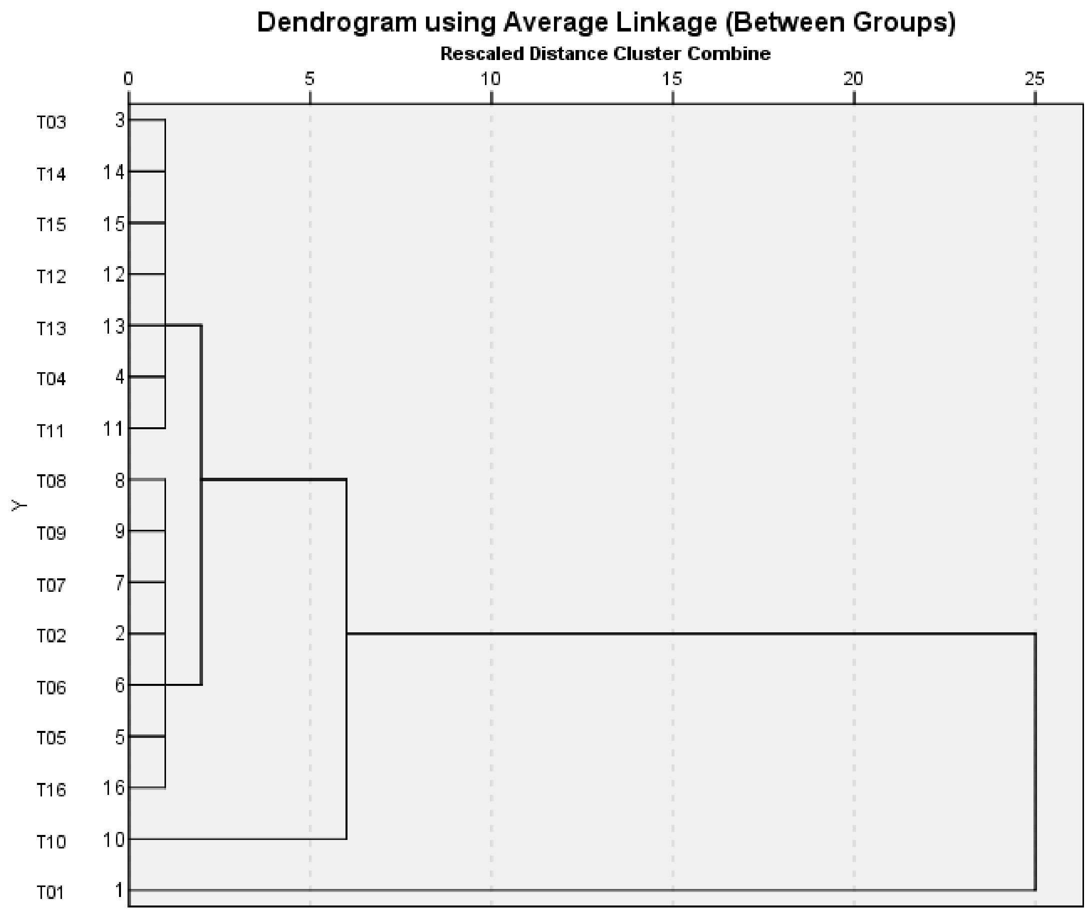

Figure 5. 10 Dendrogram for Turbine grouping using height data

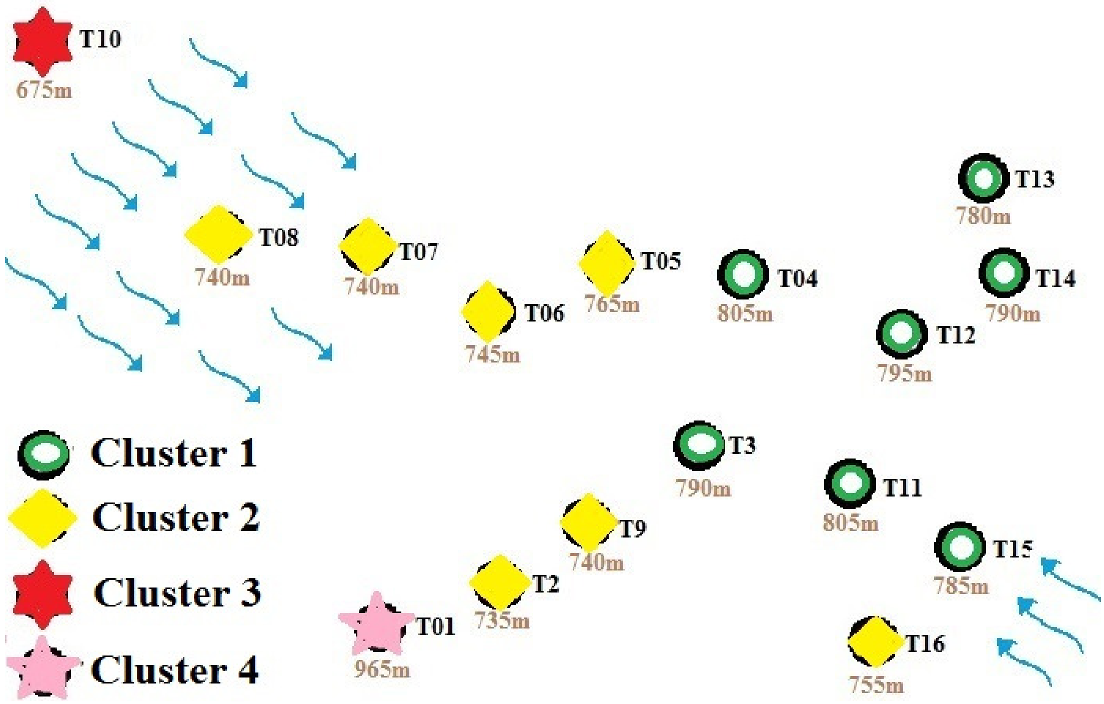

Figure 5. 11 Clusters of Turbines using height data

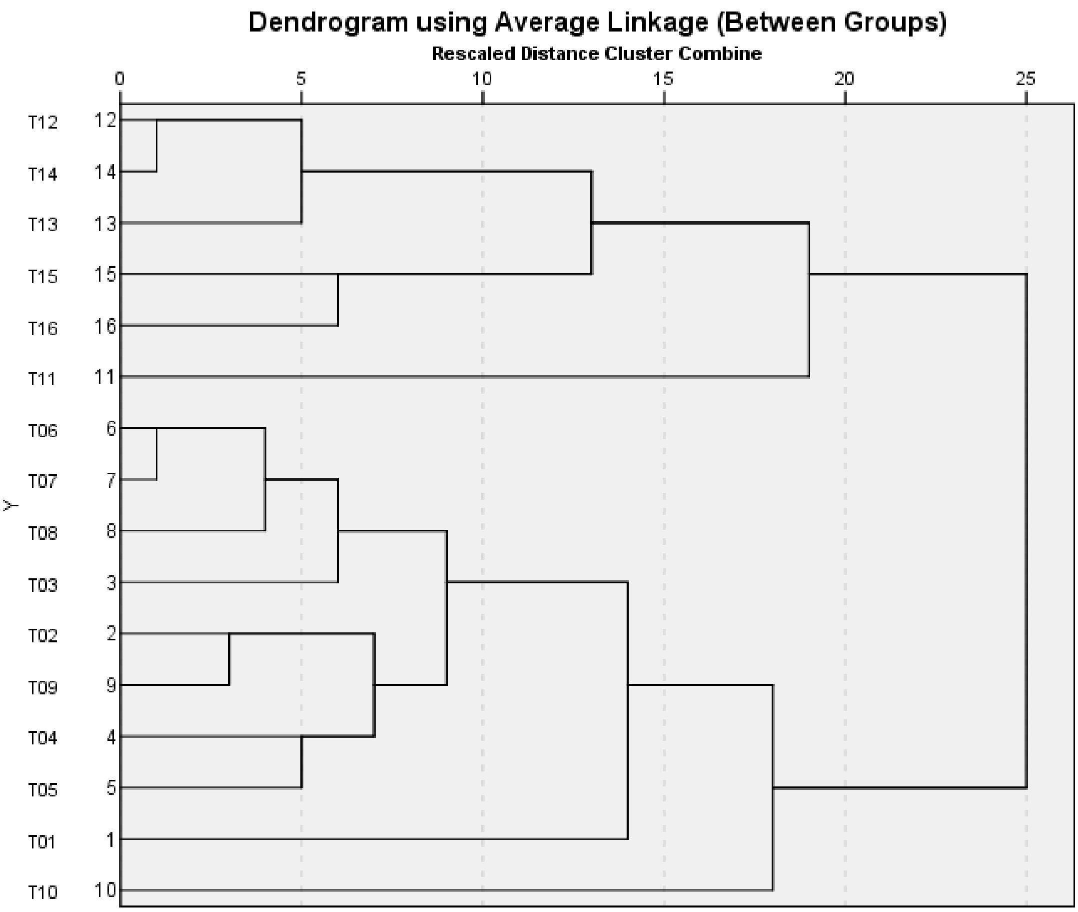

Figure 5. 12 Dendrogram for Turbine grouping using power factors data

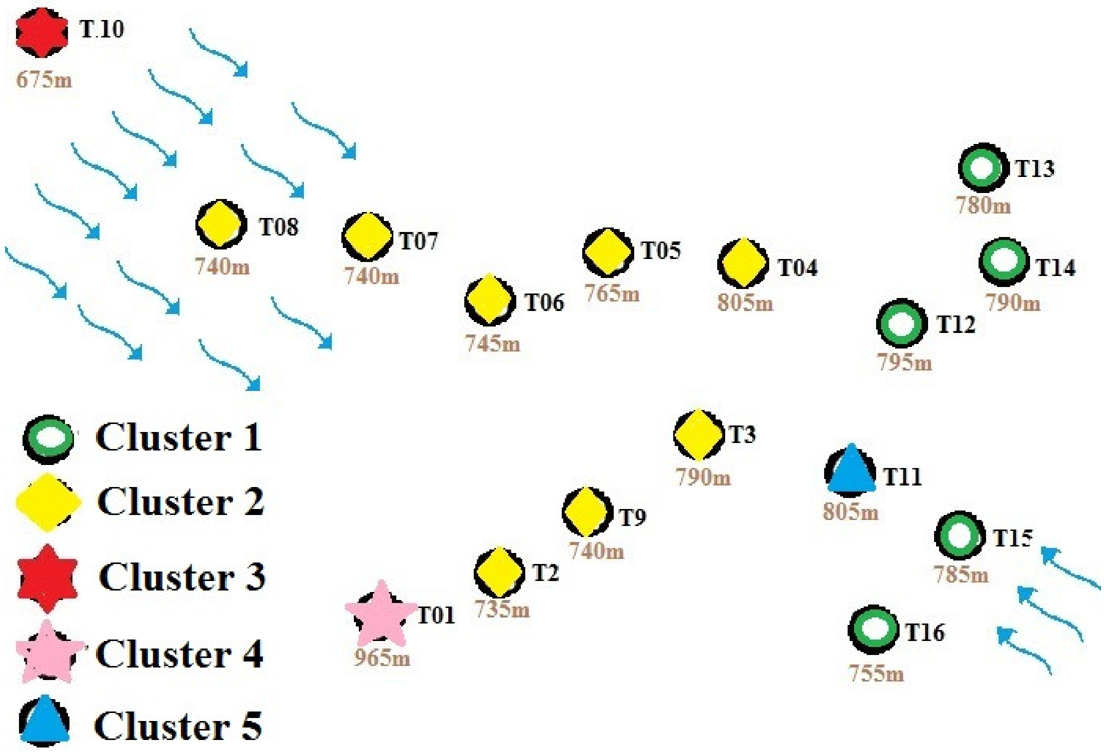

Figure 5. 13 Clusters of Turbines using power factors data

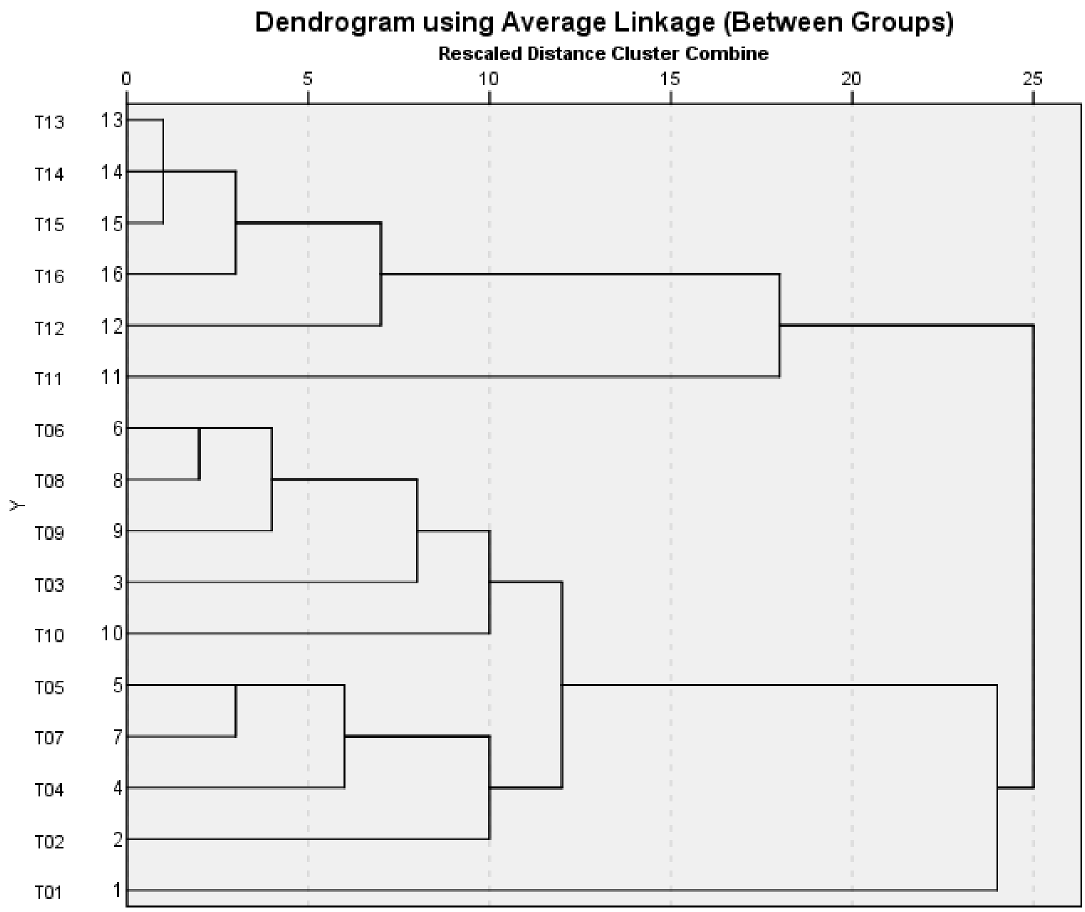

Figure 5. 14 Dendrogram for Turbine grouping using hydraulic oil temperature data

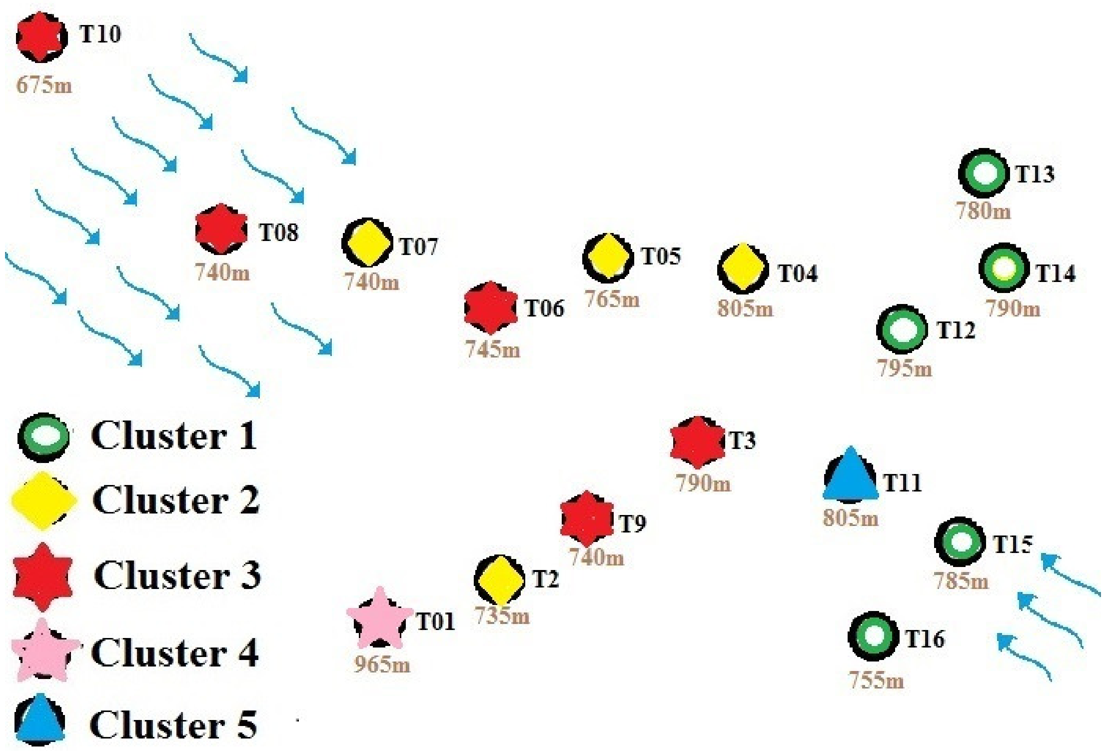

Figure 5. 15 Clusters of Turbines using hydraulic oil temperature data

It can be seen from the Figure 5.10 and Figure 5.11 that Turbine 1 is a unique group with the highest height as 965 m. Controversially, Turbine 10 is another one member group with the least height as 675 m. Turbines 3, 4, 11, 12, 13, 14, and 15 are grouped together with an average height of 790 m. The rest of turbines are grouped in other cluster

The result of data reduction analysis (principle component analysis) showed that the most important factor (the first component) consists of the following variables:

- Ambient Wind Speed Avg.,
- Ambient Wind Speed Max,
- Ambient Wind Speed Min,
- Grid Production Power Avg.,
- Grid Production Reactive Power Avg.,
- Generator Bearing Temp, Avg., and
- Grid Busbar Temp, Avg.

Thus, another clustering analysis is applied using these power related factors to group turbines. Figure 5.12 gives the dendrogram of the hierarchical clustering results and Figure 5.13 gives the clustered turbines. Turbines 1 and 10 are one member groups as in the height clusters. Turbine 11 is another one member cluster in this approach. By the way, Turbines 3 and 4 are clustered in different group considering height cluster members. However, power related clusters have similar members with height related clusters except Turbines 3, 4, and 11. This can be explained by terrain of the wind farm (layout, locations, heights) is descriptor for power related factors such as wind speed (average, maximum, and minimum) and power production (active and reactive.)

Hydraulic oil temperature itself is another important component according to data reduction analysis. This means that cluster analysis via hydraulic oil temperature may help further analysis in order to figure out faults of turbines. Figure 5.14 shows the dendrogram of the hierarchical clustering analysis regarding with hydraulic oil temperatures. Figure 5.15 gives the cluster of turbines. Turbine 1 and 11 are the only cluster members of their groups. On the other hand, these clusters have different cluster members according to previous approaches. This result shows that while Turbines 4, 5, and 7 have similar hydraulic oil temperatures, Turbine 1 and 11 are exactly different. This situation may be a clue on performance decrease or malfunctions which depend on hydraulic oil value.

5.2.6 Power forecasting modelling with Artificial Neural Network using clustered data

Power forecasting of wind energy systems can be modelled by many approaches such as regression, autocorrelation, polynomial regression, artificial neural networks (ANN), fuzzy-based models, deterministic probabilistic models, or many other heuristics [226–229]. The wind farm power prediction is generally applied to the all farm instead of predicting single or a group of turbines power. It is known that power prediction for aggregated systems have less error rate but in the case of monitoring of the systems you have to dive your systems intı groups or you should make your analysis individually.

Todays' wind energy systems generally have hundreds of wind turbines. As it is mentioned before, even if they are the same model of same brands, as it is proved previous sections, they may behave in different. This sub-section shows the usefulness of taking clustered turbines as a single in forecasting models.

It is known that wind power is a function of wind speed. An ANN model is constructed to predict power production by using only wind speed average of 10 minute before time period. Because the data reduction analysis says that Power Factor components are related with power production, the clusters of power factor is used to prove benefits of clustering in further analysis of wind systems.

First of all, 10-minute averaged wind speed data during three months is used to train, test and validation of ANN model for Turbine 4. The result of training this model gives a regression coefficient of 94%. Then, the constructed network of Turbine 4 is used to model for Turbine 5 which are in the same cluster for the power factor clustering approach. The regression coefficient is occurred as 95%. This result shows that the model for Turbine 4 can be used to predict power of Turbine 5. On the other hand, the same network is used for Turbine 10 and the result shows that the regression coefficient is highly low according to Turbine 4 and 5 as 87 %.

After all, the study is repeated for modelling Turbine 10. The network for Turbine 10 has a regression coefficient of 98%. This new network is used to predict power of Turbine 4 and 5 respectively. As it is hoped, results showed that the model for Turbine 10 gives worse regression coefficients as 84% and 85% respectively than previous network results. This example simply proves that clustering can be used as a tool to define a proper model for the turbines which are in the same cluster.

5.3 Conclusions

This section presents a statistical tool to understand SCADA data. The statistical analysis of SCADA data is a good way to figure out the whole data before doing further analysis. Firstly, the basic descriptive statistics were defined and presented to have a fast look at the studied data. Of course descriptive statistics do not give detailed information, but one can easily observe the means, minimums, maximums, standard deviations of parameters. Graphical representations of the data also help to understand the general behavior and distribution of data. Then, data reduction was applied to group parameters.

After that, one sample Kolmogorov Smirnov test was performed to see if the available SCADA data fit the normal distribution. The results showed that no parameters were fit in normal distribution. Then, Kruskal Wallis test was performed to learn categorical similarities of parameters. Consequently, it was understood that all turbines behave differently according to each parameters. After all, clustering approaches were used to group parameters (partitioned clustering) and turbines (hierarchical clustering). The similarities and differences of these clusters were explained with figures and tables. Finally, the need of clusters was proved by giving a power prediction model with ANN approach.

All these analysis can be applied by current commercial or non-profit (free) software and applications such as R, SAS, SPSS, Minitab, Statistica, or Microsoft Excel. Of course these basic statistical analyses are not adequate to discuss the performances of the turbines. Thus, the next part introduces the performance analysis of wind farm.

CHAPTER 6

PERFORMANCE ANALYSIS TOOLS FOR WIND ENERGY SYSTEMS

Recently, rapidly spreading wind power plants are considered as inevitable local sources by countries having ever-increasing energy demand. The best way to get more benefits from these energy plants, which have averagely 20 years of economic life, is to reduce operating costs or to maximize the produced energy. Performance of the turbine in a wind farm can be observed by performance analysis tools, so that, it could provide information to the plant manager about whether the current sources were used efficiently or not. In this study, performance comparisons of turbines were analyzed using SCADA data of wind farms. It is hoped that this chapter helps wind power plant managers to make decisions about maintenance planning by considering monthly performances. Data Envelopment Analysis (DEA) is used to compare wind turbines' two-year production performance in a wind farm and evaluated total factor productivities using Malmquist Index (MI). These studies were published in a journal [43]. Because the nature of DEA approach is non-parametric, Stochastic Frontier Analysis (SFA) is contributed as an alternative method to DEA to have a parametric performance analysis for the same input data. Then, a parameter based performance analysis is prepared with another study and presented in an international conference and extended version is accepted to be published in a journal [44, 45]. In the studies, operation data for the years 2013 and 2014 of a current wind farm in Turkey are used. Consequently, low-performance turbines are identified by different performance evaluation tools. It is also tried to understand the reasons for the performance losses, and development of performance-enhancing strategies were planned.

6.1 Introduction

Performance analysis is one of the good ways for condition monitoring. The basic definition of performance is that it is the whole efforts of a company to reach its all objectives [230]. Therefore, the manager can be figure out the low – utilized resources in the company by low performance levels. In this study, it is focused on the evaluation of wind turbine efficiency which is a dimension of performance. A non-parametric analytical approach, Data Envelopment Analysis (DAE), and a parametric approach, Stochastic Frontier Analysis, have been selected as a tool to evaluate the performance scores of wind turbines in a single site. Also, Malmquist Index (MI) has been used to evaluate time dependent efficiencies.

The next section gives the literature review of DEA, MI, SFA, and performance analysis of wind energy. After that, the methodology and mathematical models are described. Then, the problem statement and solution approaches are expressed. Finally, concluding remarks are given.

6.2 Literature Review

DEA is successfully applied in many fields such as health care, manufacturing, and marketing. Therefore, it is selected as one of the analytical approach to evaluate efficiency of wind turbines in a wind farm. Farrell (1957) introduced the mathematical models to analyze production techniques and efficiency in manufacturing companies [231], then, Charnes, Cooper and Rhodes extended the Farrell's approach and presented DEA which could evaluate efficiencies of decision making units in 1978 [232], lately it is called in the literature with the initials of their names as CCR model. This early model is based on constant returns to scale. Then, Banker, Charnes and Cooper (1984) relaxed the CCR model to variable returns to scale [233] which lately called as BCC model. These both approaches are the base models of DEA and both can also be modelled as input oriented or output oriented.

Despite the fact that it has been more than thirty years after the first paper published, DEA has a continuous research field and it is not seen any loss of interest [234]. This situation is caused by the following advantages of the methodology [235], [236];

- More than one inputs and outputs can be used at the same time,
- There is only a linear model to solve (which simplifies the mathematical model),
- There is no need to any other relationship between inputs and outputs except linear model,

- Any kind of measurement unit can be used for inputs and outputs (gram, meter, euro, etc.)

Of course, DEA has disadvantages as follows;

- It is deterministic, input and output data have to be selected carefully, it does not consider the random error,
- It gives relative results for each decision making unit,
- It is a non-parametric analysis which seems as an advantage by considering evaluation process but, it is impossible to make any further statistical analysis,
- It is static, it does not allow to consider time dependent analysis (this can be overcome by Malmquist Total Factor Productivity Index)
- The more input/output data, the harder to decide efficiency of units. So, it would be better as possible as less input/output data.

The basic principle process of DEA operates as in the following way; firstly, the linear model identifies the most efficient unit and set it as a limit, then, if any other unit is on that limit, the decision is that it is also efficient, else inefficient. The level of inefficiency is computed by the distance of the evaluated value of the unit to the limit.

The literature related with DEA has an increasing trend for more than thirty years, there are more than 2000 publications between the years 2010 and 2014 [237] and there were only about 4500 before the year 2010 [234]. Today, many other models can be found as alternative to base CCR and BCC models, but the purpose is still the same as to evaluate efficiencies of decision making units [234]. The main purpose of the application of DEA is to give any idea to managers about how they could improve the inefficient units [235].

Besides, DEA can analysis only in a given time, it is static, cannot give any information about efficiency changes in different time horizons. Any efficient unit at any time can lose its reference efficiency at any other time. This situation gives missing information about efficiency of any unit between changing time horizons. Malmquist Total Productivity Index can be an alternative solution to analyze different time horizons [238].

Caves, Chrisensen and Diewert contributed Malmquist Index (MI) in 1982 as a total productivity efficiency analysis which bases DEA model by considering time dimensions [239]. This approach evaluates the change on total factor efficiency of similar units having similar technologies for any two different time dimensions.

Aigner et al. [240] and Meeusen and van den Broeck [241] were independently introduced the other alternative performance analysis tool, SFA, in 1977. SFA is generally used to evaluate of different kind of models for production, revenue, profit, and cost efficiencies. In the energy area, it is also studied to define efficiency levels of energy resources/systems/market. Zhou et al. (2012) contributed a case study to measure economy – wide energy efficiency of OECD countries [242]. In addition to this, estimated total factor energy efficiency scores for OECD countries presented by Hu and Honma (2014) using SFA [243].

Although the wind energy market is becoming widespread all over the world, the publications about efficiency and productivity analysis of a wind farm are not so much widespread. DEA is not used commonly in the wind energy studies to the best of researcher's knowledge. On the other side, in the energy market, it was studied to evaluate efficiencies of energy power plants in Turkey and Spain [244, 245]. The most related work that uses DEA in the wind farms is presented by Ardente et al. (2008) with an aim of deciding whether the wind energy power plants are green or not considering investment and operating levels [246].

In the current literature, DEA and SFA generally used for the same problem to discuss differences of the approaches [247–249]. DEA and SFA used to measure the productive efficiency of a group of wind farms by Iglesias et al. (2010) [247]. Their results show that SFA gives higher efficiency scores than DEA. Whereas they contribute a comparison for efficiencies of different wind farms, this study is focused on wind turbines individual efficiency levels. Acquired efficiency scores permitted to compare methodologies. Also, while the input parameters of their study are costs, labor, and other external data, this study uses only SCADA data (internal data) of wind turbines. Thus, the content and the title may have similarities with their study, but the aim, methodology, and parameters are totally different from their study.

Kusiak et al. (2012) has a magazine paper which is based on DEA by using wind farm operation data and it was the only paper that we could found in the whole literature in which DEA was directly applied on operation data of a wind farm [250]. But, they only showed that the faults on wind turbines could affect the performance and they did not present any identified performance analysis of the turbines.

Some other studies which works on productivity and efficiency of wind farms are generally focused on the layout of the farm and wake effect [251], and another study searches the power efficiency by considering wind farm topology [252].

6.3 Methodology

In this study, DEA, MI, and SFA approaches are used to analyze efficiencies of wind turbines in a wind farm, to define the inefficient wind turbines, and to help the maintenance plans of inefficient turbines by considering monthly changes on efficiencies.

6.3.1 Data Envelopment Analysis

The key idea behind DEA is the base definition of productivity that is the ratio of outputs to the inputs. Charnes et al. [232] were inspired by Farrell's work (1957) [231] and contributed the rational model as in Equations 6.1a-d;

$$Max\ h_x = \frac{\sum_{r=1}^{s} u_{rk} Y_{rk}}{\sum_{i=1}^{m} v_{ik} X_{ik}} \qquad \text{Equation 6.1a}$$

Subject to;

$$\frac{\sum_{r=1}^{s} u_{rk} Y_{rj}}{\sum_{i=1}^{m} v_{ik} X_{ij}} \le 1, \qquad j = 1, \dots, n \qquad \text{Equation 6.1b}$$

$$u_{rk} \ge 0, \qquad r = 1, \dots, s \qquad \text{Equation 6.1c}$$

$$v_{ik} \ge 0, \qquad i = 1, \dots, m \qquad \text{Equation 6.1d}$$

where m is the number of inputs, s is the number of outputs, n is the number of decision making units. X_{ij} is the amount of *ith* inputs used by *jth* decision making unit and greater than or equal to zero, and Y_{rj} is the amount of *rth* outputs produced by *jth* decision making unit and greater than or equal to zero, v_{ik} and u_{rk} are the variables which represents weights for input i and output r of kth decision making unit respectively.

Equation 6.1b stands for that the maximum efficiency of kth decision making unit is equal to 1, Equations 6.1c and 6.1d are stated for non-negativity conditions.

This model is solved for all n decision making units separately. The objective function (Equation 6.1a) tries to maximize the ratio of total weighted outputs to the total weighted inputs for kth decision making unit.

By the way, the linear form of the rational model is CCR model as in the Equations 6.2a-e;

$$Max\ h_x = \sum_{r=1}^{s} u_{rk} Y_{rk} \qquad \text{Equation 6.2a}$$

Subject to;

$$\sum_{r=1}^{s} u_{rk} Y_{rj} - \sum_{i=1}^{m} v_{ik} X_{ij} \leq 0, \qquad j = 1, \dots, n \qquad \text{Equation 6.2b}$$

$$\sum_{i=1}^{m} v_{ik} X_{ik} = 1 \qquad \text{Equation 6.2c}$$

$$u_{rk} \geq 0, \qquad r = 1, \dots, s \qquad \text{Equation 6.2d}$$

$$v_{ik} \geq 0, \qquad i = 1, \dots, m \qquad \text{Equation 6.2e}$$

where Equation 6.2d and 6.2e are stated for non-negativity conditions, the total of weighted inputs of kth decision making unit is equal to 1 Equation 6.2c, the total of weighted outputs of kth decision making unit is no more than the total of weighted inputs Equation 6.2, thus, the efficient units have the value of 1 and inefficient ones have less than 1 Equation 6.2a.

Figure 6.1 gives the flow diagram of the DEA. First of all, the data have to be collected from SCADA system for the wind energy systems. Then, pre-processing is required. In this study, one month averaged data is used, so, the collected data were averaged to the one month periods. There are many commercial or free software/tools which are able to compute DEA scores such as R, Data Envelopment Analysis Online Software, Konsi DEA Analysis, Open Source DEA, and etc. Any of those tools can be used to get DEA results. The free version of Data Envelopment Analysis Online Software was used in this study. Finally, results are concluded.

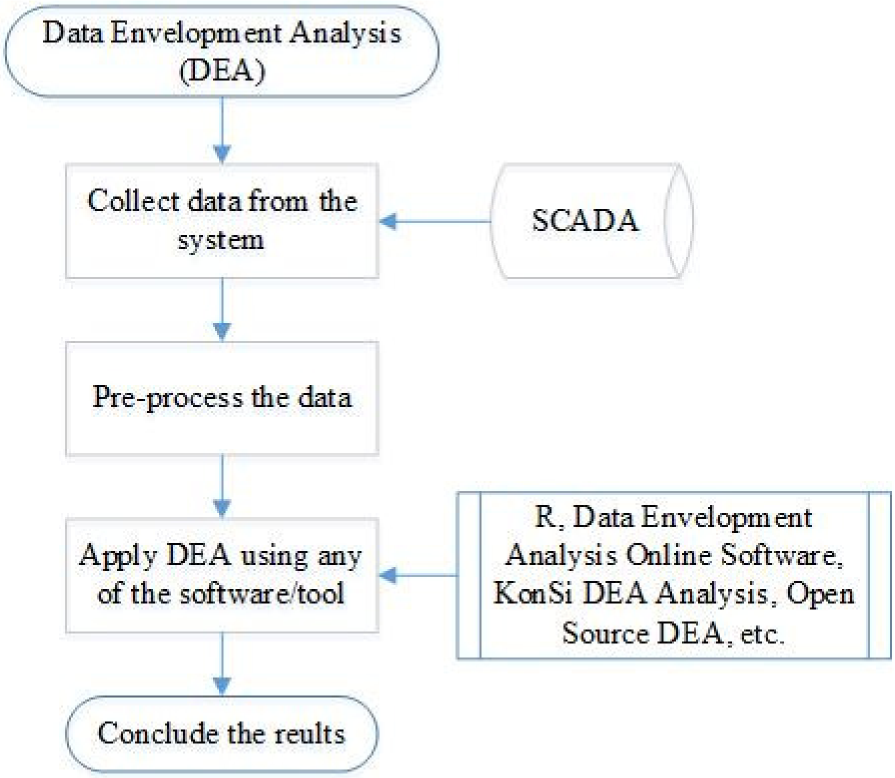

Figure 6. 1 Flow diagram of Data Envelopment Analysis

6.3.2 Malmquist Total Factor Productivity Index

MI evaluates the total factor productivity index between two data points by considering the distance ratio of each point to the joint technology. In this model, the distance function (Equation 6.3), which was firstly introduced by Malmquist in 1953 [253], was used to evaluate efficiency changes by Caves et al. (1982) [239].

$$d(x, y) = min\{\delta: (y/\delta) \in S\} \qquad \text{Equation 6.3}$$

One of the most important features of the distance function is that the production technologies having more than one input and output can be defined without giving any objectives such that minimizing costs or maximizing profit. The output oriented distance function is given in Equation 6.3. If y is on the production limit of S, $d(x,y)$ is equal to 1; if y defines an inefficient point of production limit S, less than 1; if y defines an impossible point out of the production limit S, greater than 1 [254].

The total factor productivity changes in different two data points can be defined as in Equation 6.4;

$$m(Y_S, X_S, Y_t, X_t) = \sqrt{\left[\frac{d^s(Y_t,X_t)}{d^s(Y_s,X_s)} \times \frac{d^t(Y_t,X_t)}{d^t(Y_s,X_s)}\right]} \qquad \text{Equation 6.4}$$

where; $d^s(Y_t, X_t)$ stands for distance of the time *t* observation to time *s* technology. Malmquist total factor productivity index is defined by *m(.)* and if it is greater than 1, there is an increase on MI from time *s* to time *t*; if it is less than 1, there is a decrease on MI from time *s* to time *t*. It is possible to formulate Equation 6.4 as follows (Equation 6.5);

$$m(Y_S, X_S, Y_t, X_t) = \frac{d^t(Y_t,X_t)}{d^s(Y_s,X_s)} \times \sqrt{\left[\frac{d^s(Y_t,X_t)}{d^t(Y_t,X_t)} \times \frac{d^s(Y_s,X_s)}{d^t(Y_s,X_s)}\right]} \qquad \text{Equation 6.5}$$

where, change in efficiency (Equation 6.6) and change in technology (Equation 6.7) can be evaluated separately.

$$EC = \frac{d^t(Y_t,X_t)}{d^s(Y_s,X_s)} \qquad \text{Equation 6.6}$$

Any change in output oriented efficiency (EC) between time s and time t can be evaluated by Equation 6.6.

$$TC = \sqrt{\left[\frac{d^s(Y_t,X_t)}{d^t(Y_t,X_t)} \times \frac{d^s(Y_s,X_s)}{d^t(Y_s,X_s)}\right]} \qquad \text{Equation 6.7}$$

Any change in technology (TC) between time *s* and time *t* can be evaluated by Equation 6.7.

Total factor productivity index is the multiplication of efficiency changes (EC) and technological changes (TC), thus, it easily gives hidden information in many fields to managers about the main reason for any increase or decrease in MI between any two time horizons. Thus, one can conclude about that the investments are the main reason for productivity increase or not by considering MI.

The flow diagram of the MI is given in Figure 6.2. Firstly, the data have to be collected from SCADA system for the wind energy systems. Then, pre-processing is required. In this study, one month averaged data is used, so, the collected data were averaged to the one month periods. There are many commercial or free software/tools which are able to compute MI scores such as R, Data Envelopment Analysis Online Software, Konsi Malmquist Index Software, PIM-DEA, and etc. Any of those tools can be used to get MI results. The free version of Data Envelopment Analysis Online Software was used in this study. Finally, results are concluded.

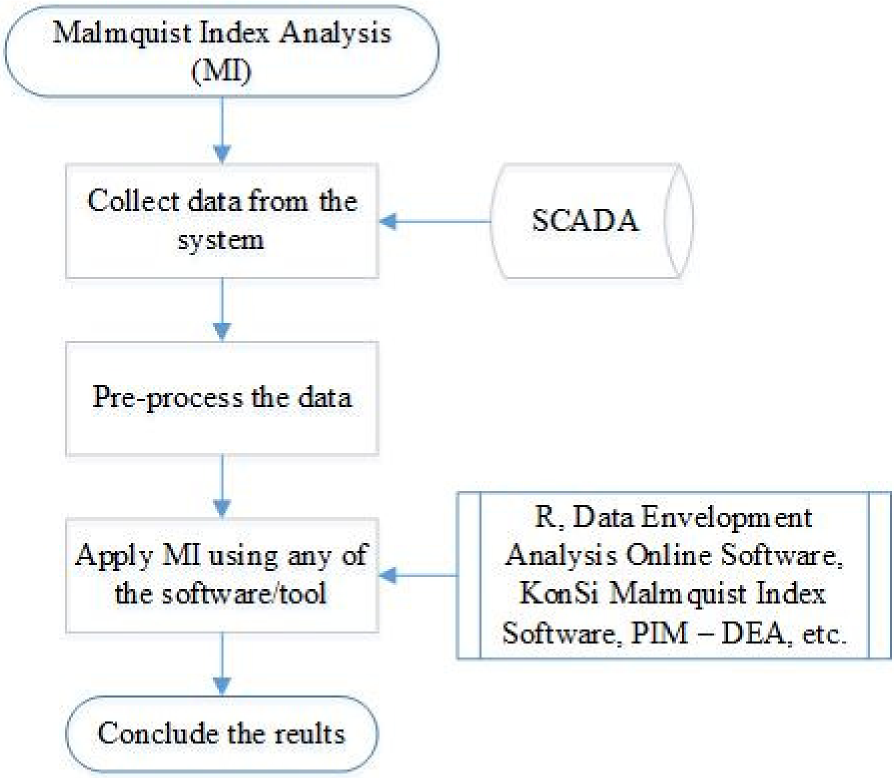

Figure 6. 2 Flow diagram of Malmquist Index Analysis

6.3.3 Stochastic Frontier Analysis

In this study, a software package program, FRONTIER 4.1, has been used to evaluate SFA efficiency scores of the wind turbines and wind farm. This SFA tool was developed by Battese and Coelli in 1995 [255]. An Econometrics Fitting Technique of Aigner et al. [240] is used to evaluate production efficiency, economic efficiency, and cost and profit frontiers. The basic SFA model describes a production function specified for cross – sectional data which had two error term components such as random symmetric statistical noise and systematic deviations from a frontier – one – side inefficiency that is given in Equation 6.8;

$y = f(X, \beta) - u + v$ Equation 6.8

where, y represents the output, X is k x 1 vector of inputs, β is a vector of unknown parameters, v is two-sided random variable, and u is non-negative random variable following a normal distribution or other distributions such as exponential, gamma, or etc. Also, u and v are assumed to be distributed independently of each other and the regressors. By the way, maximum likelihood techniques are used to estimate the frontier and the inefficiency parameter.

For the panel data, the Equation 6.8 is can be re – designed as in Equation 6.9;

$y_{it} = f(X_{it}, \beta) + \varepsilon_{it} = f(X_{it}, \beta) + (V_{it} - U_{it})$ Equation 6.9

where y_{it} is the output (generated power production in this analysis), X_{it} is the inputs vector (Wind Speed, Rotor Speed, and Generator Speed), β is a vector of unknown production parameters and ε_{it} is a random disturbance which includes statistical noise (V_{it}) and technical inefficiency ($-U_{it}$). The sub – indexes, i denotes each one of the wind turbines (i=1, 2, ..., 16, except 11 and 15) and t is the monthly averaged ten – minute – interval – data for two years (t=1, 2, …, 24).

The error parameters u and v have a relationship to explain the cause of the inefficiency levels as given in Equations 6.10 and 6.11;

$\delta^2 = \delta v^2 + \delta u^2$ Equation 6.10

$\gamma = \frac{\delta u^2}{\delta^2} \quad \epsilon \; [0, 1]$ Equation 6.11

These β (coefficients), γ (gamma) and σ (sigma) values are obtained at the end of iterative process [256]. The gamma parameter gives the clue about the cause inefficiency such that if it approaches to 0, it means cause of inefficiency is mostly by statistical noise, conversely, if it approaches to 1, it means cause of inefficiency is mostly by technical inefficiency. In the reference [257], the details of mathematical methods and program manual can be found.

Figure 6.3 gives the flow diagram of the SFA. First of all, the data have to be collected from SCADA system for the wind energy systems. Then, pre-processing is required. One month averaged data is used in this, so, the collected data were averaged to the one month periods. Also, logarithms of the collected data were evaluated to be used in the computation of SFA scores. There are many commercial or free software/tools which are able to compute SFA scores such as R, Frontier Version 4.1, LIMDEP, NLOGIT, and etc. Any of those tools can be used to get SFA results. The Frontier Version 4.1 was used in this study. Finally, results are concluded.

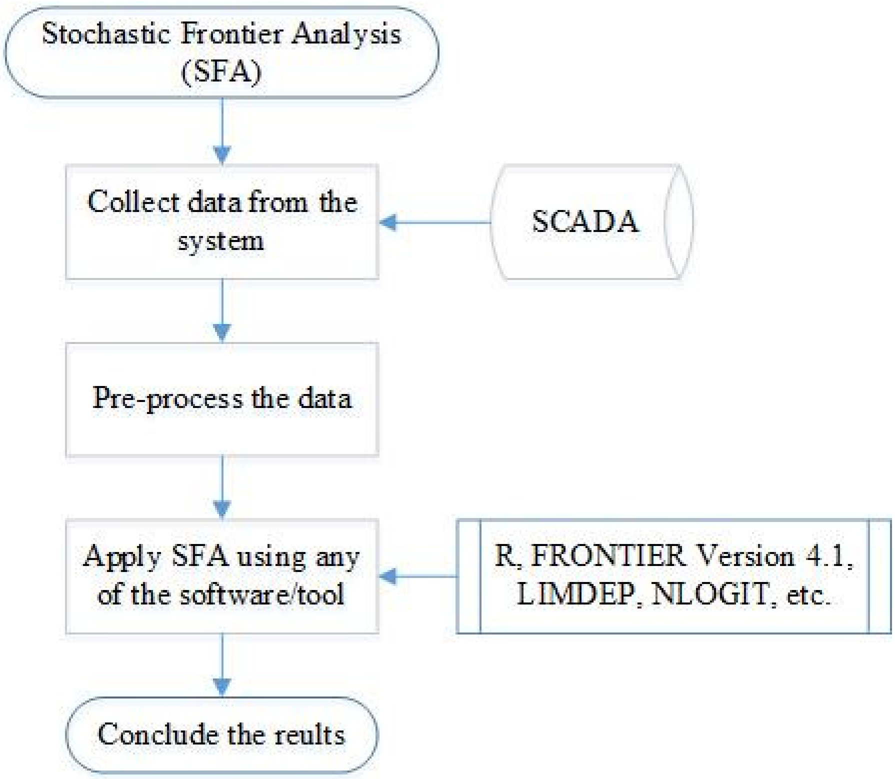

Figure 6. 3 Flow diagram of Stochastic Frontier Analysis

6.4 Problem Statement

In order to clarify the efficiencies of wind turbines in a wind farm and to identify changes on efficiencies according to time, the data were gathered from an operated wind farm in Turkey for the years 2013 and 2014. There were 16 wind turbines having following technical features; 3 MW capacity, 90 m rotor diameter, 80 m hub height. Following data are used for each turbine: generator revolution speed (Generator RPM Avg.), rotor revolution speed (Rotor RPM Avg.), wind speeds (Wind Speed Avg.), and generated power avarages (Grid Production Active Power Avg.) of 10-minute interval during two years.

As in many systems, SCADA system has some technical mufanctions which causes the loss of data. So that, actually 105210 data is gathered as 10-minute intervals during two years for each turbine. However, every turbine has some missing data as shown in Figure 6.4. Therefore, Turbine 11 and 15 were not considered in the study because of having quite high missing data than others. Collected data are avaraged as monthly periods for each of CCR DEA, MI and SFA models. The output is generated power and others are inputs for all models.

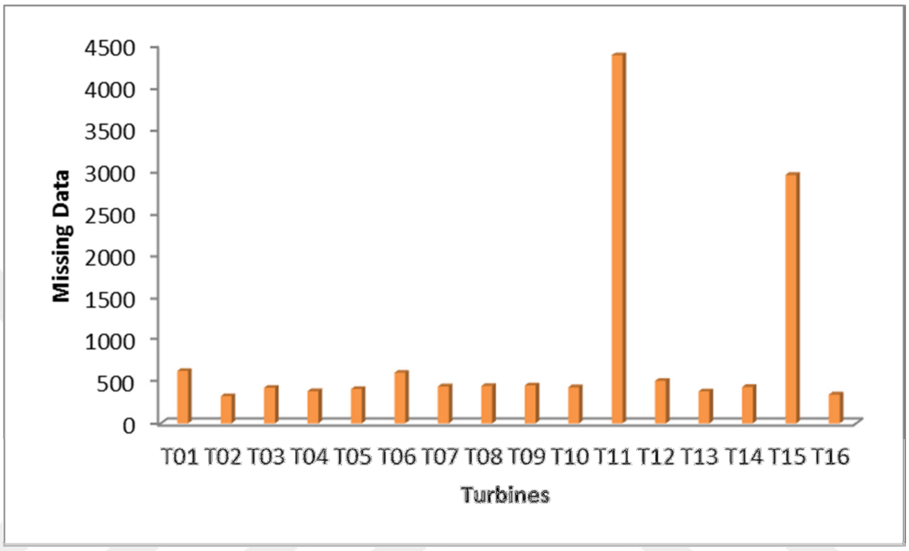

Figure 6. 4 The numbers of 10-minute missing data of turbines during two years

6.4.1 Performance analysis of wind turbines by CCR DEA model

The input oriented models are able to explain the most appropriate combinations of inputs to get a predefined output. Thus, input oriented model was selected to evaluate efficiencies. The results of input oriented CCR model of DEA for 14 turbines between January 2013 and December 2014 are listed in Table 6.1.

Because DEA is a static model, the results of the analysis have to be considered separately from each other. So that, the efficiencies for each month have to be considered independently which are given in Table 6.1. For example, the efficient Turbine is number 13 for January 2013 and all others are inefficient in certain rates by referencing Turbine 13. By the way, Turbine 6 is the most inefficient one with a performance inefficiency rate of 22%. Generally, Turbines 4, 12, 13, and 14 have greater efficiency rates than others.

The geometric average has been used to evaluate yearly average efficiencies for each turbine. Turbines 4, 12, 13, and 14 have still greater efficiency rates than all others. On the other hand, Turbine 10 has the lowest average efficiency rate for both years. It can be concluded that Turbine 10 has a problem and a maintenance plan has to be considered as soon as possible for it.

Table 6. 1 Efficiency analysis of wind turbines for each month between the years 2013 and 2014 by DEA approach

Efficiency													
Year	**Months**												
2013	1	2	3	4	5	6	7	8	9	10	11	12	**Average**
Turbine 1	0,86	0,85	0,88	0,75	0,79	0,93	0,91	0,83	0,74	0,73	0,65	0,86	**0,80**
Turbine 2	0,86	0,90	0,92	0,80	0,84	0,94	0,95	0,92	0,81	0,73	0,74	0,91	**0,85**
Turbine 3	0,81	0,66	0,83	0,73	0,78	0,97	0,98	0,89	0,79	0,71	0,66	0,90	**0,80**
Turbine 4	0,97	0,96	0,94	0,83	0,84	0,96	0,95	0,91	0,87	0,84	0,87	**1,00**	**0,91**
Turbine 5	0,82	0,87	0,83	0,72	0,74	0,88	0,91	0,85	0,74	0,70	0,69	0,88	**0,79**
Turbine 6	0,78	0,74	0,80	0,69	0,77	0,91	0,92	0,86	0,72	0,68	0,60	0,82	**0,76**
Turbine 7	0,80	0,78	0,86	0,70	0,74	0,97	0,92	0,85	0,72	0,68	0,67	0,80	**0,78**
Turbine 8	0,87	0,88	0,90	0,68	0,69	0,91	0,92	0,81	0,67	0,58	0,66	0,87	**0,77**
Turbine 9	0,85	0,87	0,87	0,75	0,80	0,92	0,93	0,90	0,76	0,71	0,67	0,84	**0,81**
Turbine 10	0,90	0,92	0,90	0,55	0,55	0,78	0,80	0,67	0,53	0,52	0,72	0,95	**0,71**
Turbine 12	0,97	**1,00**	0,99	0,98	0,97	0,99	0,99	**1,00**	**1,00**	0,96	0,90	0,98	**0,97**
Turbine 13	**1,00**	**1,00**	**1,00**	0,96	0,88	0,91	0,91	0,89	0,90	0,96	**1,00**	**1,00**	**0,94**
Turbine 14	0,94	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	0,90	0,97	**0,98**
Turbine 16	0,87	0,94	0,92	0,85	0,90	0,93	0,90	0,90	0,83	0,80	0,80	0,85	**0,87**
Average	**0,88**	**0,88**	**0,90**	**0,78**	**0,80**	**0,93**	**0,93**	**0,87**	**0,78**	**0,75**	**0,74**	**0,90**	**0,84**
2014	1	2	3	4	5	6	7	8	9	10	11	12	
Turbine 1	0,72	0,74	0,85	0,76	0,79	0,83	0,88	0,67	0,74	0,70	0,60	0,77	**0,75**
Turbine 2	0,78	0,83	0,89	0,83	0,85	0,88	0,96	0,96	0,81	0,80	0,73	0,84	**0,84**
Turbine 3	0,73	0,76	0,87	0,79	0,86	0,88	0,99	0,95	0,84	0,71	0,60	0,76	**0,80**
Turbine 4	0,92	0,92	0,96	0,89	0,90	0,93	**1,00**	0,95	0,90	0,85	0,84	0,96	**0,91**
Turbine 5	0,79	0,81	0,82	0,75	0,79	0,84	0,99	0,93	0,82	0,69	0,75	0,85	**0,81**
Turbine 6	0,68	0,67	0,79	0,74	0,79	0,85	1,00	0,93	0,80	0,64	0,60	0,71	**0,75**
Turbine 7	0,69	0,64	0,76	0,75	0,77	0,83	0,96	0,89	0,80	0,61	0,55	0,73	**0,73**
Turbine 8	0,77	0,71	0,81	0,69	0,72	0,79	0,97	0,89	0,69	0,68	0,70	0,81	**0,76**
Turbine 9	0,76	0,77	0,85	0,78	0,82	0,85	0,99	0,93	0,79	0,68	0,70	0,80	**0,80**
Turbine 10	0,80	0,70	0,79	0,60	0,59	0,63	0,85	0,75	0,58	0,59	0,71	0,84	**0,69**
Turbine 12	0,94	0,97	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	0,99	0,96	0,90	0,94	**0,97**
Turbine 13	**1,00**	**1,00**	0,97	0,95	0,90	0,91	0,96	0,94	0,94	0,96	**1,00**	**1,00**	**0,96**
Turbine 14	0,94	0,99	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	0,91	0,94	**0,98**
Turbine 16	0,82	0,94	0,92	0,90	0,93	0,93	0,98	0,99	0,83	0,90	0,83	0,87	**0,90**
Average	**0,80**	**0,81**	**0,87**	**0,81**	**0,83**	**0,86**	**0,96**	**0,91**	**0,81**	**0,76**	**0,73**	**0,84**	**0,83**

The general efficiency of the wind farm can be concluded by average efficiencies of each month which is given at the bottom of the Table 6.1. The wind farm as a whole has an efficiency rate over 90% for March, June, July and December 2013 and July - August 2014. The most inefficient month is November for both 2013 and 2014. The reason behind this situation needs to be inspected, because, there still at least three turbines which has more than or equal to 90% efficiency. Especially, inefficiency rates of some turbines (1, 3, 6, etc.) are quite high than other months and there may be a technical reason for this.

The results of a simple CCR model of DEA are able to presents meaningful information about productivity of wind turbines and wind farm as a whole. One can conclude which turbine is more efficient than others, and vice versa. Also, these efficiency information can help managers to prepare a maintenance and control plan to understand the main reason behind inefficiency. Thus, a non-parametric productivity analysis approach, DEA, may help managers to develop new strategies for maximizing operation performance of a wind energy power plant as it is generally used in many other different sectors.

6.4.2 Clarifying monthly performance changes of wind turbines by using MI

DEA has a disadvantage to analyze time dependent changes on efficiencies. So, a total factor productivity index model, Malmquist Index, was used to analyze monthly changes on efficiencies of wind turbines.

Firstly, the change between January 2013 and December 2014 was analyzed by ignoring all other months. Table 6.2 lists the efficiency change ratios as Technical Efficiency Change (TEC), Pure Efficiency Change (PEC), Scale Efficiency Change (SEC), Technological Change (TC), and Malmquist Total Factor Productivity Index (MI) respectively. If the MI is greater than 1, one can conclude that there is an increase on total productivity. On the other hand, if the MI is less than 1, a decrease on total productivity can be concluded.

Just looking efficiency change ratios of wind turbines for the months January 2013 and December 2014 in Table 6.2, it can be concluded that while 11 turbines (Turbines 2, 3, 4, 5, 8, 9, 10, 12, 13, 14, 16) have an increasing efficiency rate, efficiencies of three turbines (Turbines 1, 6, and 7) are decreasing. There is a 4% improvement on productivity of whole wind farm. Between these time period, while average technological change increases 8%, average technical efficiency change decreases 4%. The increase of technological change can be explained by increase of input values (average wind speed, average rotor revolution, average generator revolution). Despite the increase on technology, decrease in technical efficiency gives some signals about technical abnormalities on wind turbines. Also, there is a 5% decrease in scale efficiency which means that wind farm has produced 5% less energy with the same inputs. On the other hand, while the technological change increases 8%, the total productivity increases only 4% because of decrease on technical efficiency. While the decrease on total productivity is generally 1 or 2% for three turbines, the maximum increase on total productivity is occurred for Turbine 5 with 12%. The technological improvement for all turbines (between 7 and 9%) leads to total productivity improvement. But, 9 to 10 % decrease on scale efficiency of three turbines (1, 6, 7) causes the decrease on total productivity.

Table 6. 2 Malmquist Total Factor Productivity Index (MI) between the months only January 2013 and December 2014

Turbines	Technical Efficiency Change	Pure Efficiency Change	Scale Efficiency Change	Technological Change	Malmquist Index
Turbine 1	0,90	1,00	0,90	1,09	0,98
Turbine 2	0,97	1,02	0,95	1,09	1,06
Turbine 3	0,94	1,02	0,92	1,09	1,02
Turbine 4	0,99	1,00	0,99	1,09	1,07
Turbine 5	**1,03**	**1,04**	**0,99**	**1,09**	**1,12**
Turbine 6	0,91	1,00	0,91	1,09	0,99
Turbine 7	0,91	1,01	0,90	1,09	0,99
Turbine 8	0,93	1,01	0,93	1,09	1,01
Turbine 9	0,95	1,02	0,93	1,09	1,03
Turbine 10	0,93	0,99	0,94	1,09	1,01
Turbine 12	0,96	1,00	0,97	1,09	1,05
Turbine 13	1,00	1,00	1,00	1,07	1,07
Turbine 14	0,99	0,99	1,00	1,08	1,08
Turbine 16	0,99	1,00	0,99	1,09	1,08
Average	**0,96**	**1,01**	**0,95**	**1,08**	**1,04**

The technical efficiency of nearly all turbines decreases generally except Turbine 5 and 13. Turbine 1, 6, and 7 has the biggest decrease ratio with 9-10 % in technical efficiencies. The pure efficiencies generally increase or decrease in a few turbines with small change rate 1-2%. In contrast, change on scale efficiencies play a role on decrease of technical efficiencies up to 9 – 10% for some turbines (1, 3, 6, and 7).

After these general analyses for only beginning and ending time intervals of this study, the whole time horizon is analyzed by MI model for every two-month periods separately.

Table 6.3 lists the geometric average efficiency changes of each turbine during two years for every consecutive two-month time horizon. Table 6.3 shows that all turbines have increased their total productivity. The main reason is considered as that the technological change highly increases for all turbines, even though technical efficiency decreases for all turbines except Turbine 5 and 13.

Table 6. 3 Average Malmquist Total Factor Productivity Index (MI) of each turbine during the years 2013 and 2014 for every consecutive two – month time horizon

Turbines	TEC (Avg.)	PEC (Avg.)	SEC (Avg.)	TC (Avg.)	MI (Avg.)
Turbine 1	0,99	1	0,99	1,00	1,00
Turbine 2	0,99	1,00	0,99	1,01	1,01
Turbine 3	0,99	1	1,00	1,01	1,01
Turbine 4	**0,99**	**1,00**	**0,99**	**1,01**	**1,01**
Turbine 5	**1,00**	**1,00**	**0,99**	**1,01**	**1,01**
Turbine 6	0,99	1	0,99	1,01	1,01
Turbine 7	0,99	0,99	0,99	1,01	1,01
Turbine 8	0,99	1	0,99	1,01	1,00
Turbine 9	0,99	1,00	0,99	1,01	1,01
Turbine 10	0,99	0,99	0,99	1,01	1,00
Turbine 12	0,99	0,99	0,99	1,01	1,00
Turbine 13	**1**	**1**	**1**	**1,01**	**1,01**
Turbine 14	0,99	0,99	0,99	1,01	1,00
Turbine 16	0,99	1,00	0,99	1,01	1,01

Table 6.4 gives the total factor productivity indexes for the whole wind farm. The increase on total productivity of the farm is quite high for the May-June 2013 and November-December 2013 time periods with 52-53%. On the other hand, there is a big performance loss at September-October 2014 time period with 42%.

Technical efficiency is more increased at November-December 2013 time period with 21% and the biggest technical efficiency loss is at March-April 2013 time period with 16%. Whereas there is no more than ±1% change on pure efficiency for all periods, great changes on scale efficiencies are decisive on technical efficiencies. By the way, technological change differs in time periods. It has the biggest improvement rate at May-June 2013 time period with 31% and the biggest performance loss at September-October 214 time period with 37% which also caused the lowest ratio on total productivity index.

Table 6. 4 Malmquist Total Factor Productivity Index (MI) of the whole wind farm during the years 2013 and 2014 for every consecutive two-month time horizon

1st Term Year - Month	2nd Term Year - Month	TEC	PEC	SEC	TC	MI
2013 – 1	2013 – 2	1,00	1,01	0,99	0,83	0,83
2013 – 2	2013 – 3	1,03	0,99	1,03	1,03	1,06
2013 – 3	2013 – 4	0,86	1,00	0,86	0,91	0,79
2013 – 4	2013 – 5	1,03	1,00	1,03	1,17	1,20
2013 – 5	**2013 – 6**	**1,16**	**1,01**	**1,16**	**1,31**	**1,53**
2013 – 6	2013 – 7	1,00	1,00	1,00	1,11	1,10
2013 – 7	2013 – 8	0,94	0,99	0,95	1,00	0,94
2013 – 8	2013 – 9	0,90	1,00	0,89	0,78	0,70
2013 – 9	2013 – 10	0,95	1,01	0,94	0,71	0,67
2013 – 10	2013 – 11	1,00	0,99	1,01	1,19	1,18
2013 – 11	**2013 – 12**	**1,21**	**1,02**	**1,19**	**1,25**	**1,52**
2013 – 12	2014 – 1	0,89	0,99	0,90	0,88	0,79
2014 – 1	2014 – 2	1,01	1,01	1,00	0,94	0,95
2014 – 2	2014 – 3	1,08	1,00	1,08	0,99	1,07
2014 – 3	2014 – 4	0,92	1,00	0,93	1,06	0,98
2014 – 4	2014 – 5	1,03	1,00	1,03	1,18	1,22
2014 – 5	2014 – 6	1,04	0,99	1,05	1,08	1,12
2014 – 6	2014 – 7	1,12	1,01	1,10	1,22	1,37
2014 – 7	2014 – 8	0,94	0,99	0,95	0,95	0,90
2014 – 8	2014 – 9	0,90	1,00	0,89	0,84	0,75
2014 – 9	**2014 – 10**	**0,93**	**1,01**	**0,93**	**0,63**	**0,58**
2014 – 10	2014 – 11	0,97	1,00	0,97	1,27	1,23
2014 – 11	2014 – 12	1,14	1,01	1,14	1,18	1,35
Average		**0,99**	**1,00**	**0,99**	**1,00**	**1,00**

6.4.3 Performance analysis of wind turbines by SFA model

DEA and MI approaches are given in previous sections which have a characteristic of non-parametric performance analysis. DEA uses a linear program model to establish an efficient frontier of best decision making unit that envelops all other units. The model results a maximum output-input ratio for each unit with a weight that no other units overcomes the maximum reachable efficiency level 1 [258].

On the other hand, SFA is a parametric performance evaluation tool. It also allows to observe the cause of inefficiency by considering randomness or technical inefficiency [259]. The same parameters (inputs: wind speed, rotor speed, generator speed) are used as in put to figure out production efficiency (output: active power) as in the DEA and MI approaches. Equation 6.12 gives the SFA model of this study.

$$ln(Production\ Power\ Efficiency) = \beta_0 + \beta_1 * ln(Generator\ RPM) + \beta_2 * ln(Rotor\ RPM) + \beta_3 * ln(WindSpeed) + \varepsilon \qquad \text{Equation 6.12}$$

Table 6.5 summarizes the results of SFA. As in the DEA study, geometric average of each month is computed to give yearly efficiencies for each turbine. The most efficient turbines for the year 2013 are Turbine 3 and 12 and Turbines 3 and 16 for the year 2014 with an efficiency score of 0,98 of each one. On the other side, the least efficient turbines are Turbine 5 and 10 for the year 2013 and Turbines 7, 9, and 10 for the year 2014 with approximately 9% efficiency losses. Averaged monthly efficiency scores are given after the turbine rows for each year at the bottom of the Table 6.5. September, November, and December are the most inefficient months of the year 2013 for the wind farm as a whole with 0,92, 0,94, and 0,94 efficiency levels respectively. By the way, gamma is 1 for those months that means the inefficiency is caused by technically; it is not a result of data noise. The wind farm has the highest efficiency scores for the months January and March in 2013 with a rate 97%.

In the year 2014, February and April, 2014 are the most inefficient months with 93% efficiency score and the gamma value of 1,00 and 0,99 respectively which shows the cause of inefficiencies are totally technical inefficiency. By the way, May and August are the most efficient months with an efficiency level of 0,96.

There is a 5% inefficiency for the November of 2014 with a gamma value of 0,64 which means that 36% of inefficiency score of this month is caused by statistical noise and the other 64% is technical inefficiency.

As in the other similar studies of current literature of DEA and SFA, this study also presents higher efficiency scores for SFA with 99% average for all turbines than DEA that efficiencies vary between 69% and 98%. This result can be explained by the nature of the DEA model which ineffectiveness and actual random variations in production can be mixed that causes the appearance of ineffectiveness low or high than actual value [260].

Table 6. 5 Efficiency analysis of wind turbines for each month between the years 2013 and 2014 by SFA approach

EFFICIENCY													
2013	1	2	3	4	5	6	7	8	9	10	11	12	**Avg.**
Turbine 1	0,99	1,00	0,99	0,98	1,00	0,97	0,98	0,92	0,90	0,99	0,91	0,95	0,96
Turbine 2	0,93	0,94	0,95	0,88	0,90	0,92	0,94	1,00	0,90	0,96	0,91	0,95	0,93
Turbine 3	0,99	0,94	0,99	0,91	0,98	1,00	0,99	0,99	0,99	0,98	0,99	1,00	**0,98**
Turbine 4	1,00	0,99	1,00	0,95	0,94	0,96	0,97	0,95	0,93	0,97	0,99	0,97	0,97
Turbine 5	0,94	0,95	0,94	0,92	0,88	0,89	0,93	0,92	0,85	0,90	0,95	0,92	**0,91**
Turbine 6	0,95	0,92	0,94	0,98	0,97	0,98	0,95	0,96	0,92	0,94	0,87	0,92	0,94
Turbine 7	0,94	0,90	0,97	0,90	0,94	1,00	0,95	0,96	0,90	0,96	1,00	0,96	0,95
Turbine 8	0,97	0,94	0,98	0,93	0,98	0,94	1,00	0,99	0,93	0,97	0,93	0,95	0,96
Turbine 9	0,98	0,98	0,98	0,87	0,89	0,93	0,95	0,95	0,84	0,95	0,87	0,92	0,93
Turbine 10	0,97	0,99	1,00	0,97	0,89	0,85	0,94	0,86	0,86	0,89	0,91	0,96	**0,92**
Turbine 12	0,97	1,00	0,99	0,98	0,99	0,98	0,99	1,00	1,00	0,98	0,95	0,95	**0,98**
Turbine 13	0,98	0,98	0,97	0,94	0,89	0,92	0,90	0,86	0,91	0,96	1,00	0,92	0,93
Turbine 14	0,98	0,99	0,98	0,98	0,97	0,99	0,98	0,93	0,97	0,93	0,91	0,91	0,96
Turbine 16	0,98	0,96	0,95	0,99	0,95	0,93	0,91	0,92	0,93	0,96	0,95	0,95	0,95
Average	**0,97**	0,96	**0,97**	0,94	0,94	0,95	0,96	0,94	**0,92**	0,95	**0,94**	**0,94**	0,95
Gamma	1,00	1,00	1,00	0,95	1,00	1,00	1,00	1,00	1,00	**0,82**	1,00	1,00	0,98
2014	1	2	3	4	5	6	7	8	9	10	11	12	**Avg.**
Turbine 1	0,91	0,92	0,94	0,91	1,00	0,91	0,91	0,96	0,92	0,97	0,93	0,98	0,94
Turbine 2	0,90	0,94	0,92	0,91	0,93	0,93	0,95	0,98	0,92	0,94	0,97	0,96	0,94
Turbine 3	0,97	1,00	1,00	0,99	0,94	0,98	1,00	1,00	0,99	0,96	0,96	0,97	**0,98**
Turbine 4	0,99	0,98	0,99	0,95	0,99	0,96	0,97	0,97	0,92	0,96	0,97	0,99	0,97
Turbine 5	0,97	0,97	0,90	0,86	0,96	0,93	0,95	0,98	0,89	0,94	0,98	0,99	0,94
Turbine 6	0,89	0,87	0,91	0,92	0,92	0,96	0,97	0,98	0,92	0,96	0,93	0,92	0,93
Turbine 7	0,89	0,83	0,87	0,98	0,90	0,95	0,93	0,95	1,00	0,87	0,90	0,92	**0,91**
Turbine 8	1,00	0,98	0,93	0,94	0,94	0,99	0,97	0,98	0,97	0,98	0,97	1,00	0,97
Turbine 9	0,89	0,89	0,91	0,85	0,98	0,86	1,00	0,93	0,89	0,96	0,96	0,95	**0,92**
Turbine 10	0,97	0,89	0,94	0,85	0,99	0,86	0,92	0,84	0,91	0,88	0,95	0,97	**0,91**
Turbine 12	0,97	0,95	1,00	0,98	1,00	1,00	0,94	0,99	0,96	0,97	0,95	0,92	0,97
Turbine 13	0,97	0,94	1,00	0,91	0,98	0,93	0,91	0,95	0,93	0,93	0,95	0,92	0,94
Turbine 14	0,93	0,93	0,98	0,95	0,99	0,96	0,92	0,96	0,99	0,95	0,93	0,90	0,95
Turbine 16	0,93	0,98	0,98	1,00	0,97	0,99	1,00	0,99	0,96	0,99	0,98	0,98	**0,98**
Average	0,94	**0,93**	0,95	**0,93**	**0,96**	0,94	0,95	**0,96**	0,94	0,95	0,95	0,95	0,95
Gamma	1,00	1,00	1,00	0,99	1,00	1,00	1,00	1,00	1,00	**0,80**	**0,64**	1,00	0,95

6.5 Conclusions

In this study, 10-minute SCADA data for the years 2013 and 2014 of an operating wind farm is used to evaluate productivity ratio of wind turbines and wind farm by CCR DEA, MI, and SFA model for each months separately. Before starting the efficiency analysis, the gathered raw 10-minute data is re-organized and averaged for monthly periods.

The first part of the study consists of efficiency analysis by using input oriented CCR DEA model. The results helped to conclude whether a wind turbine was efficient or not for a constant month. Also, the inefficiency ratios could be read from the results which may help managers while planning maintenance strategies to maximize productivity. In addition to this, average efficiency ratios for each month separately may give signs for seasonal efficiency problems.

Because DEA is not able to compare time dependent changes on efficiencies, a total factor productivity analysis approach, MI model, is used in the second part of the study. First of all, for a general understanding the efficiency changes, the first time period is selected as January 2013 and the last one as December 2014 by ignoring all other months. The results showed that while the most improved total productivity is occurred on Turbine 5, generally all turbines have increased their productivities except Turbine 1, 6, and 7 at the end of the December 2014. The increase on technological changes was the main reason for increase on total productivity. Also, decrease on scale efficiencies led to decrease on technical efficiencies.

On the other hand, when all two consequent months were analyzed by MI model, the average of the total productivity indexes showed that all turbines have increased, at least a bit, their efficiencies. Because the technologic changes were in the positive direction as average, total productivity changes were also in the positive direction. The scale efficiencies were mostly decreasing which caused a bit performance loss on technical efficiencies.

When the term based averages of efficiencies were inspected, the most improvement on total productivity was occurred at May-June and November-December time periods in 2013, in contrast, the most performance loss was occurred at September-

October time periods in 2014 which was mostly caused by the highly decrease on technological changes.

Finally, SFA approach used to clarify similarities and differences between non - parametric and parametric productivity approaches. Because of the nature of the DEA, the efficiency scores are affected from the random (noisy) data. By the way, SFA itself includes the randomness in the model. Therefore, the stochastic frontier model with panel data has advantages over DEA.

The proposed approach evaluated the efficiency scores of each wind turbines to compare each other and to determine whole wind farm efficiency for each month. Results showed that the average SFA efficiency is higher than DEA.

After all these results, it may be recommended to the decision makers that they should make some other deeply analysis to find out the reasons for inefficient turbines and inefficient time horizons. This study represents an application of mathematically modelled efficiency analysis approach and it is hoped that it would help managers when preparing maintenance strategies by aiming minimize operation costs and maximize profitability.

SFA is able to figure out parametric effects on performance levels of wind turbines. Thus, a SFA methodology to evaluate productive efficiency levels of wind turbines and a whole wind farm is studied with different parameters [45]. According to the best of researcher’s knowledge it is the first time in the literature. Therefore, it makes the difference with other similar studies to measure efficiency levels in wind energy systems according to different parameters.

Consequently, if the performance analysis applied to the wind turbines in a single wind farm, the results give the following useful information about the system and turbines;

- The low production performance wind turbine can be found at a certain time using DEA,
- The low production performance time period can be found of the whole wind farm using DEA,

- The production performance changes can be found using MI for a certain time period,
- The causes of performance changes can be found using MI for a certain time period,
- The parametric production performance can be analyzed using SFA,
- The performance decrease caused by a certain parameter can be found using SFA,
- The manager of the wind farm has a powerful tool, performance analysis tool, while planning the maintenance.

CHAPTER 7

FORECASTING TOOLS FOR WIND SPEED AND GENERATED POWER OF WIND TURBINES

Every wind farm needs their power predictions to make a good production plan and trading strategy. In the current literature, there are many kinds of approaches to achieve this aim. The main purpose of this thesis is contributing monitoring and control algorithms for wind farm managers by using only current system data. Therefore, this chapter gives the simple models on wind speed and wind power modelling approach which uses only SCADA data of an operating wind turbine. In addition to this, a new approach to model wind speed and wind power, Particle Filtering (PF), is proposed. It is commonly used in video tracking problems by predicting the next frame of the current video. The algorithm basically works on the filtering of *n* predictions according to their fitness value to the actual data. This characteristics of the algorithm makes it possible to more accurate the predictions of any algorithm. In this study, the PF adaptation in artificial neural network (ANN) to predict short term wind speed and wind power forecasts. A real data set of a wind farm in Turkey is used in this study. Results are compared to traditional basic forecasting methods such as regression, moving average, weighted moving average, exponential smoothing, and simple ANN models. Early version of the study was presented in an international conference [46].

7.1 Introduction

Forecasting is one of the most important research topic in the current literature of wind energy studies [41]. Today’s energy market needs those forecast to make the energy trading in a safe way. Nevertheless, the most important problem in the wind energy forecast is the chaotic nature of the wind. There are many researches in the current literature which wants to predict wind speed or wind power for the next periods.

Pourmousavi Kani and Ardehali (2011) contributed a hybrid system of ANN and Markoc Chain model to forecast very short term wind speeds (in a few seconds) [261]. Jiang et al. (2013) studied on a time series model to predict very short term wind speed data [262]. A statistical hybrid model, which uses the most important weather forecast parameters by clustering approach to predict wind power up to 48 hours, is contributed by Özkan and Karagöz (2015) [263]. Their approach is currently used in the Wind Power Monitoring and Forecast System for Turkey. Another wind power forecasting tool, which is used in Denmark, proposed by Croonenbroeck and Ambach (2015)[264]. Their approach was based on non – parametric regression models and time series. Okumuş and Dinler (2016) proposed a valuable guide to the review of forecasting methods on wind speed and wind energy and contributed a hybrid approach of adaptive neuro –fuzzy inference system and ANN to predict 1 – hour ahead wind speeds.

The motivation for this section is that although it helps managers to plan their production ratio, every forecast has an error rate. So that, the main purpose of this study is to contribute a new approach that reduces the forecast errors for wind speeds and wind power. Therefore, a PF approach, which is commonly used to solve filtering problems in signal processing studies, is contributed to filter the forecast error of prediction methodology. First of all, classical and basic forecast approaches are used to show their prediction capabilities of the same problem, then ANN approach is presented, finally, PF is combined to ANN model to reduce forecast errors. Results showed that proposed approach is able to give better results than traditional methods by aiming short term wind power forecast.

7.2 Methodology

The wind speed average and power production average of 10-minute intervals gathered from an operating wind farm are used in this study. If the time series approaches are used, the forecasting model is generally based on Equations 7.1 and 7.2 for next periods wind speeds and generated power respectively;

$$WS_{t+1} = f(WS_t) + \varepsilon \qquad \text{Equation 7.1}$$

$$WP_{t+1} = f(WP_t) + \varepsilon \qquad \text{Equation 7.2}$$

Where WS_{t+1} is the forecast of *wind speed* for time $t+1$, $f(WS_t)$ is a *function* of *wind speed* at time t, Where WP_{t+1} is the forecast of *wind power* for time $t+1$, $f(WP_t)$ is a *function* of *wind power* at time t, and ε is the *forecast error*. On the other hand, the power output can be evaluated by using current wind speed data as in the Equation 7.3;

$$WP_t = f(WS_t) + \varepsilon \qquad \text{Equation 7.3}$$

This study presents basic time series forecasting methods such as moving averages, weighted moving averages, exponential smoothing, and regression model. Following sections gives the brief information about used forecasting methods.

7.2.1 Simple moving averages (SMA)

Moving averages model is mathematical average of the last several periods of actual observations. Equation 7.4 is the general form of the simple moving average method;

$$F_t = \frac{A_{t-n} + A_{t-n-1} + \cdots + A_{t-1}}{n} + \varepsilon \qquad \text{Equation 7.4}$$

where F_t is the forecast for time t and A is actual observation under the time indices t, and ε is the *forecast error*.

In this study, 3-time period moving averages methodology is used for the forecast of wind speeds and wind power.

7.2.2 Weighted moving averages (WMA)

In the weighted moving averages method, the past observations have weights (w) while predicting the value as seen in the Equation 7.5;

$$F_t = w_1 A_{t-1} + w_2 A_{t-2} + \cdots + w_n A_{t-n} + \varepsilon \qquad \text{Equation 7.5}$$

3-time period weighted moving averages method with $w_1 = 0{,}5$; $w_2 = 0{,}3$; and $w_3 = 0{,}2$ is used in this study.

7.2.3 Simple exponential smoothing (SES)

The simple exponential smoothing model is another method used to forecast wind speeds and power with a general formula as in the Equation 7.6;

$F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1}) + \varepsilon$ Equation 7.6

In this model, the forecast is based on previous forecast error with a smoothing constant α which taken 0,5 in this study. Also, an initial forecast is needed here (F_{t-1}). The result of 3-time period moving averages for the 4th period is used for the initial forecast of exponential smoothing model.

7.2.4 Simple linear regression (SLR)

The linear regression model is commonly used in time series studies which has a general description as in the Equation 7.7;

$Y = ax + b + \epsilon$ Equation 7.7

where Y is the predicted value, x is the time period, a is the slope of the linear regression, b is the intercept point, and ϵ is the forecast error.

7.2.5 Artificial Neural Network (ANN)

ANN is a smarter approach than aforementioned ones. Its main advantage is being able to model even nonlinear functions. It has an input vector, bias vector, weight vector for inputs, weight vector of outputs, and output vector.

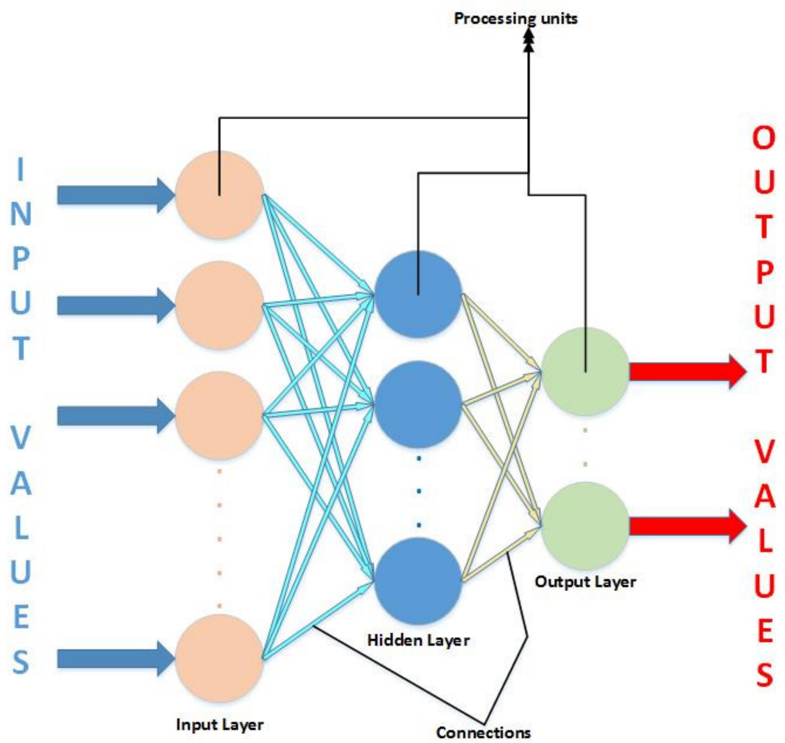

Figure 7. 1 General architecture of ANN

The basic procedure of ANN is reaching to the target output by converting inputs using a training approach which tries to finds bias vector and weights. Figure 7.1 represents the general architecture of the ANN models. Equation 7.8 gives the mathematical model of the ANN;

$$Y = f(x) + \epsilon \qquad \text{Equation 7.8}$$

where Y is the prediction, $f(x)$ is the network function, and ϵ is the forecast error.

In this study, three ANN model is constructed. ANN1 is aiming to forecast next term's wind speed (output) using currents wind speed (input), ANN2 tries to find next terms power generation (output) with the current produced power (input), ANN3 is used to predict next term's power production (output) using current wind speed (input). All those networks have same architecture (Figure 7.2) such as; 1 input, 1 output, 2 layers with 20 hidden neurons. The training function is selected as backpropagation algorithm based Levenberg-Marquardt optimization method.

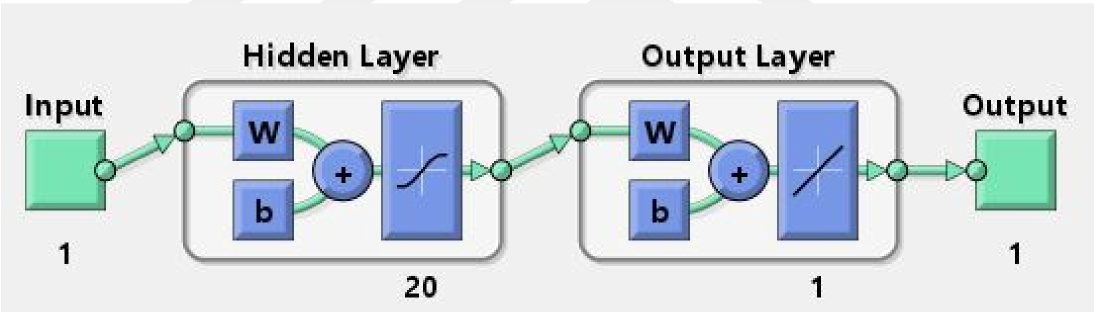

Figure 7. 2 ANN architecture of proposed approach

7.2.6 PF adapted Artificial Neural Network (ANN + PF)

In addition to traditional and basic forecasting models for wind energy systems, a PF method is combined to ANN approach to make better forecasts. The synonym for PF is Sequential Monte Carlo (SMC) methods which uses Bayesian recursion functions to estimate the posterior density of the state space.

In the early stages of the PF goes back to 1950s with a study titled as "The Poor Man's Monte Carlo" by Hammersley and Morton (1954) [265]. Then, in 1993, Gordon et al. contributed the first true implementation of the SMC methods used today [266]. The main difference of their algorithm is that it does not require any assumption about that state-space or the noise of the system. The purpose of a PF is to predict the posterior density of the state variables given the observation variables. The PF is designed for a hidden Markov Model, where the system consists of hidden and observable variables [267]. In other words, PF tries to forecast the hidden states of x using the observations y. The main model is given in Equations 7.9 and 7.10;

$$x_{t+1} = f(x_t) + w_t \qquad \text{Equation 7.9}$$

$$y_{t+1} = g(x_{t+1}) + v_t \qquad \text{Equation 7.10}$$

where x_0, x_1, … is a first order Markov process that changes according to the probability density function $p_{x_{t+1} \mid x_t}: x_{t+1} \mid x_t \sim p_{x_{t+1} \mid x_t}(x \mid x_t)$ and with an initial distribution $p(x_0)$ and y_0, y_1, … are conditionally independent provided that x_0, x_1, … are known. This conditional distribution for y_{t+1} is written as $y \mid x_{t+1} \sim p_{y \mid x}(y \mid x_{t+1})$. By the way, both w_t and v_t are mutually independent and identically distributed sequences with known probability density functions and $f(.)$ and $g(.)$ are known functions.

PF provides a set of samples that approximate the filtering distribution $p(x_{t+1} \mid y_0, \ldots, y_{t+1})$. If there are N samples from the approximate posterior distribution of x_{t+1}, where the samples are labeled with superscripts as $x_{t+1}^{(1)}, x_{t+1}^{(2)}, \ldots, x_{t+1}^{(N)}$. Then, forecasts with respect to the filtering distribution are approximated by Equation 7.11;

$$\int f(x_{t+1}) p(x_{t+1} \mid y_0, \ldots, y_{t+1}) dx_{t+1} \approx \sum_{i=1}^{N} f(x_{t+1}^{(i)}) \frac{1}{N} \qquad \text{Equation 7.11}$$

and, $f(.)$ in the usual way for Monte Carlo, can give all the moments etc. of the distribution up to some degree of approximation.

There are various applications of PF in the current literature. Yu et al. (2016) proposed a novel SMC to enhance the rotor speed regulations in high-speed wind [268]. Another paper contributed the comparison of different filtering algorithms by aiming control issues of doubly fed induction generator wind energy systems [269]. Apart from those wind energy related studies, there are many successful application of the combination with ANN of PF in different areas [270–276]. This encourages the researcher to study on combination of ANN with PF to predict wind speed and wind power.

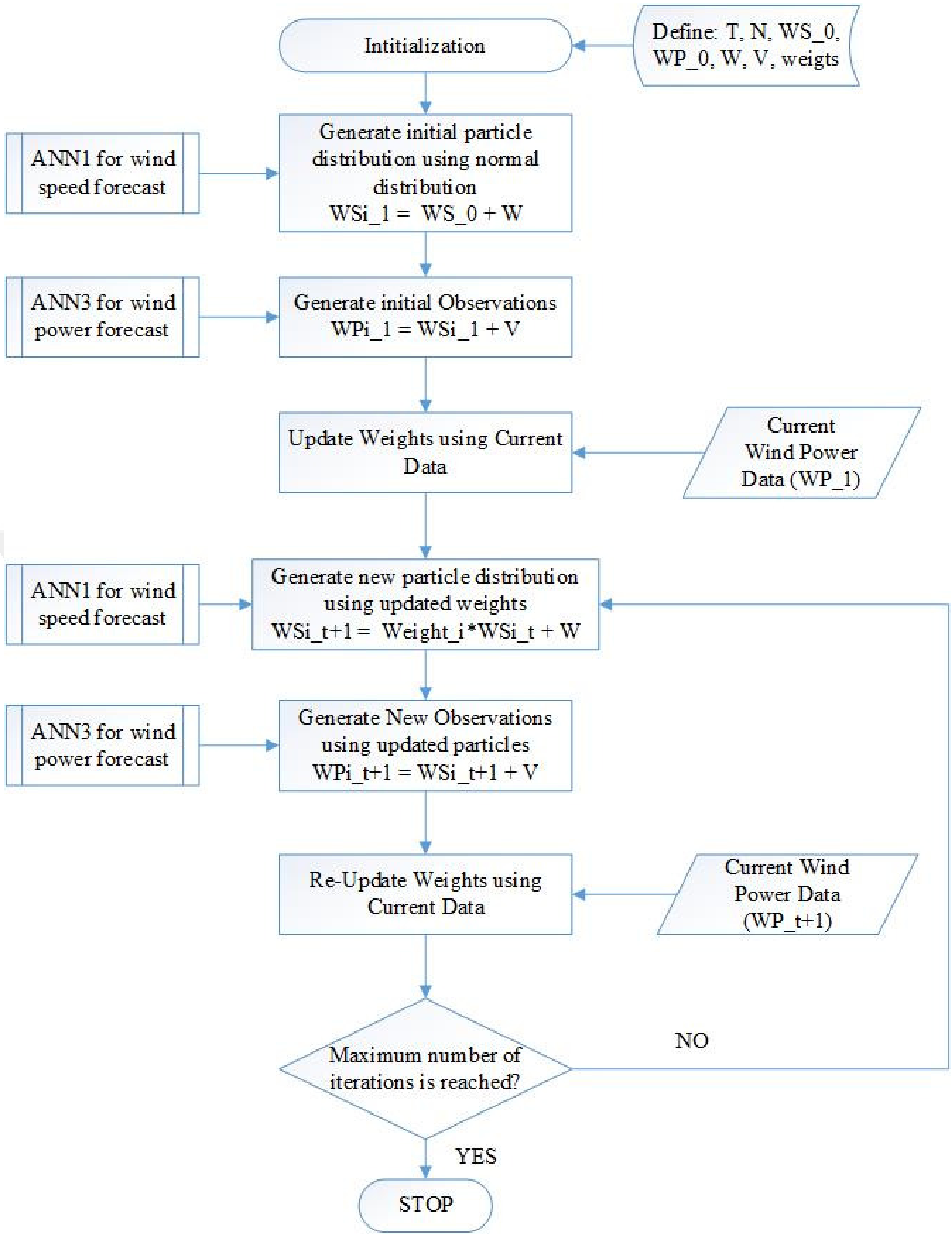

Figure 7. 3 Flow diagram of ANN + PF approach

Figure 7.3 gives the flow diagram of proposed approach. First of all, the algorithm starts with initialization of the parameters where T represents time interval (maximum iteration number), N is the number of particles, WS_0 is initial wind speed, WP_0 is initial wind power, W is the variance of wind speed prediction approach, V is the variance of wind power prediction approach, and weight are equally distributed upon each particles (1/N).

The PF is applied to reduce ANN forecast errors. Thus, ANN1 method is used to forecast next period's wind speed and ANN3 is used to predict next period's wind power. The result of ANN1 is used as the input parameter for ANN3. Then, the power forecast of each particles are compered to actual power data to re – compute the weights of particles. The Gaussian Probability Density Function (Equation 7.12) is used to re – evaluate the weights of particles.

$$\text{weight}_t^i = (1/(\sqrt{2 * v * \pi}) * e^{-\frac{(z_t - \mu_t^i)^2}{2*v}} \qquad \text{Equation 7.12}$$

where, v is the variance of power prediction, z is the actual power production, μ is the forecasted power production, I is the particle indices and t is the time period indices. This procedure continuous till the number of time periods reaches the T (maximum iteration number). The developed algorithm is given at Appendix A.1.

7.4 Results

Many different forecasting approaches such as simple moving averages (SMA), weighted moving averages (WMA), simple exponential smoothing (SES), simple linear regression (SLR), and artificial neural networks (ANN) were used in this study. Table 7.1 summarizes the methods with their parameters and labels. Six different methodologies were used to forecast wind speed and power production. It is the fact that every forecast has an error. Therefore, the performance of the forecast methods needs to be evaluated. Thus, following equations (Equations 7.13 – 7.15) are used to prove the performance of proposed algorithms;

$$Mean\ Error\ Forecast\ (MEF) = Actual\ Data - Forecast\ Data \qquad \text{Equation 7.13}$$

$$Mean\ Absolute\ Diversions\ (MAD) = |Actual\ Data - Forecast\ Data|$$

Equation 7.14

$$Mean\ Absolute\ Percentage\ Error\ (MAPE) = \frac{|Actual\ Data - Forecast\ Data|}{Actual\ Data}$$

Equation 7.15

Current literature shows that, for a few seconds to a few minutes ahead predictions, MAPE values vary in a range from 1.61% to 15.99% [277].

Table 7. 1 Forecasting methodologies and their parameters

Label	Method	Input	Output	Parameter Setting
Fa1	3 - month SMA	Wind Speed t	Wind Speed t + 1	Time Period = 500
Fa2	3 - month WMA	Wind Speed t	Wind Speed t + 1	Time Period = 500
Fa3	SES	Wind Speed t	Wind Speed t + 1	Wind Speed Forecast t = 4 comes from Fa1, Time Period = 500
Fa4	SLR	Wind Speed t	Wind Speed t + 1	Time Period = 500
Fa5	(ANN1)	Wind Speed t	Wind Speed t + 1	1 input, 1 output, 2 layer, 20 hidden neurons, Levenberg-Marquardt training algorithm, Time Period = 500
Fb1	3 - month SMA	Wind Power t	Wind Power t +1	Time Period = 500
Fb2	3 - month WMA	Wind Power t	Wind Power t +1	Time Period = 500
Fb3	SES	Wind Power t	Wind Power t +1	Wind Power Forecast t = 4 comes from Fb1, Time Period = 500
Fb4	SLR	Wind Power t	Wind Power t +1	Time Period = 500
Fb5	(ANN2	Wind Power t	Wind Power t +1	1 input, 1 output, 2 layer, 20 hidden neurons, Levenberg-Marquardt training algorithm, Time Period = 500
Fb6	ANN3	Wind Speed t	Wind Power t +1	1 input, 1 output, 2 layer, 20 hidden neurons, Levenberg-Marquardt training algorithm, Time Period = 500
Fc	ANN1 + ANN3 + PF	Wind Speed t, Wind Power t	Wind Speed t+1, Wind, Power t +1	Fb6 setting + N=100 particles, w = 0,3319 (variance of wind speed forecast), v = 17241 (variance of wind power forecast)

Forecast error performances of proposed methods are given in Table 7.2. While the most providing method is regression by considering MEF values, ANN + PF approach gives better forecasts than all others by considering MAD and MAPE scores.

By the way, power production forecast performances of the proposed methods are given in Table 7.3. Although the regression model has least mean error, ANN + PF gives better results according to MAD value than others. On the other hand, ANN2 methodology gives the best MAPE score with power input and power output strategy.

Table 7. 2 Forecast Error Performance of proposed algorithms for wind speed

Methodology	MEF	MAD	MAPE
Fa1	-0,0035	0,5003	7,7207
Fa2	-0,0028	0,4709	7,2510
Fa3	-0,0027	0,4549	7,0155
Fa4	***0,0000***	0,8884	14,0712
Fa5	-0,0042	0,4467	6,9214
Fc	-0,0051	***0,3358***	***5,3022***

Table 7. 3 Forecast Error Performance of proposed algorithms for wind power

Methodology	MEF	MAD	MAPE
Fb1	-0,5981	96,9484	22,8139
Fb2	-0,4843	90,5890	21,1031
Fb3	-0,4553	87,5584	20,5778
Fb4	**0,0000**	192,0016	59,9472
Fb5	14,7323	89,4026	20,2924
Fb6	5,6148	96,5468	22,4897
Fc	8,5115	***76,1491***	***20,4865***

According to table 7.2 and 7.2, PF approach is able to reduce the forecast error of ANN simulation on the wind speed and wind power predictions.

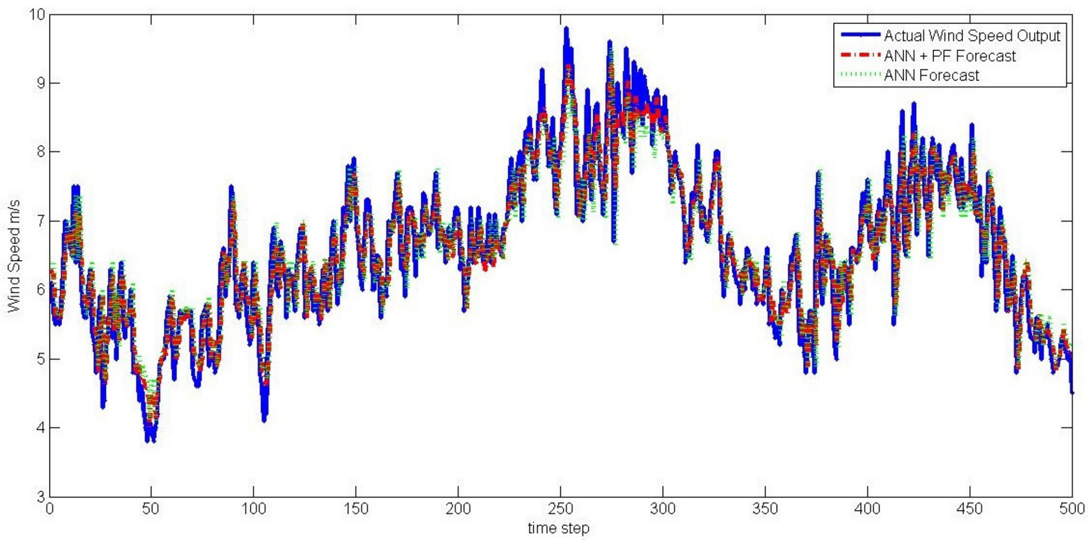

Figure 7. 4 Actual wind speed versus forecast with ANN1 and ANN + PF

Figure 7.4 shows the actual wind speed data and forecast with wind speed input – wind speed output ANN model (ANN1) and combination of ANN and PF model

(ANN1 + ANN3 + PF) with wind speed and wind power input wind speed and wind power output. Results showed that the proposed ANN + PF method has less forecast error as seen in Figure 7.5. This can be concluded as PF is able tune forecast errors of ANN model to predict short term wind speed data.

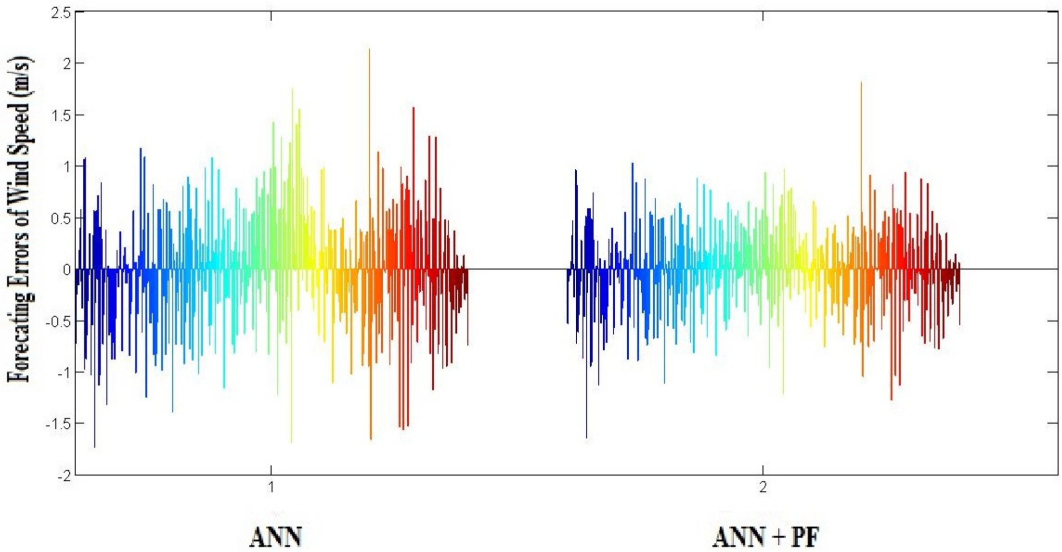

Figure 7. 5 Forecast errors for wind speed of ANN1 model and ANN + PF model

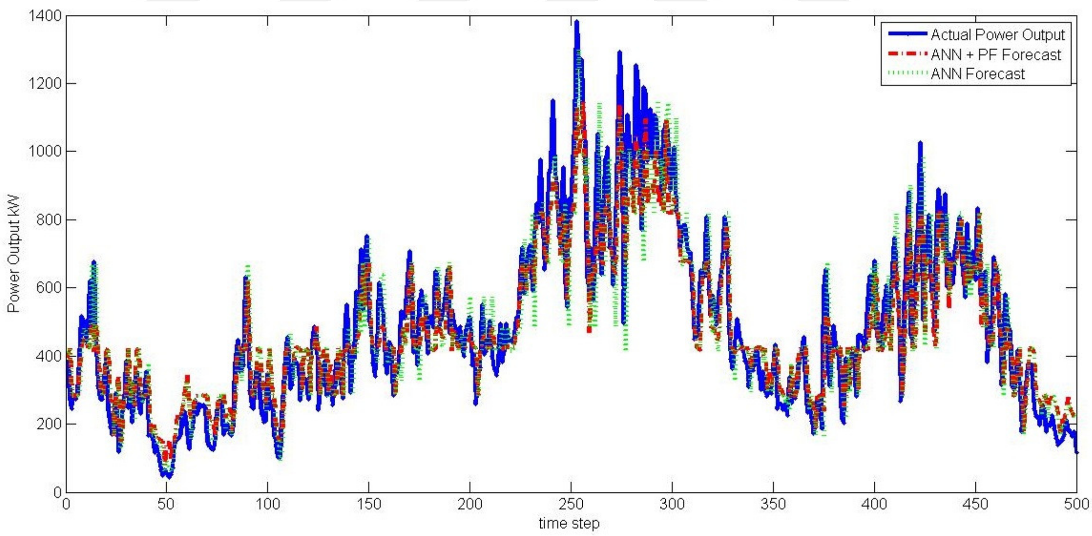

Figure 7. 6 Actual wind power versus forecast with ANN3 and ANN + PF

Similarly, Figures 7.6 and 7.7 gives the comparison of ANN models and PF combination for power production forecast. It can be seen from the Figures 7.5 and 7.7 that the power forecast error reduction is quite better than wind speed one. This can be explained by the characteristics of the PF which re – computes the particle weights according to performance on power observations.

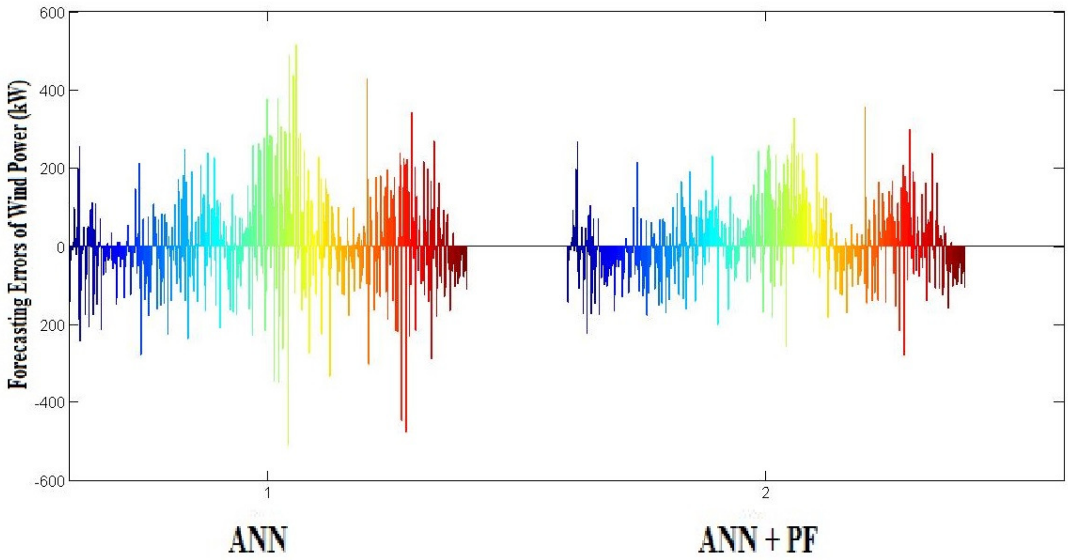

Figure 7. 7 Forecast errors for power production of ANN3 model and ANN + PF model

7.5 Conclusions

This chapter presents different approaches on wind speed and wind power forecasts. Focus of the study is on the usage of simple SCADA data to predict next terms' wind speed and wind power generation. Forecasting issue has big importance in the wind energy area. Therefore, researchers give attention to this subject. There are many different approaches aiming those issues. Simple time series methods are used to prove that forecasting in a basic way can be applicable only using current system data. In addition to those methods, ANN is used to get better results. Due to the nature of the ANN, it is able to model even nonlinear relations which provide better results than traditional time series models. Results also support these characteristics of ANN models.

On the other side, every forecast has an error which has a meaning for condition monitoring issues. This helps to managers on controlling and monitoring of power production behavior of wind turbines. The performances of proposed approaches are evaluated using various error comparison equations. MEF gives the average error forecast for all time periods. It is most commonly used when the total forecast error is important than individual periodic forecasts. Regression model is designed to minimize squared errors of whole time period; generally it gives the best results as in this study.

If the individual forecast error for each time period is important that whole period, MAD may be a good alternative performance scoring tool which evaluates mean absolute errors. In addition to this, MAPE is able to measure mean absolute percentage errors regarding with individual percentages. Therefore, those two methods may help to define which method is applicable to the problem.

ANN models generally provided suitable forecast by considering MAD and MAPE scores. The forecast precision has getting more important in a competitive energy market. So that, a PF approach was contributed to reduce forecast errors of ANN models. Current literature has successful examples of combination of ANN and PF approaches. This study presents a new methodology to forecast short term wind speed and wind power which is not studied before to the best of researcher's knowledge. This proposed approach may also help managers to monitor the wind turbines' power production characteristics.

CHAPTER 8

A NOVEL TOOL FOR NEURAL NETWORK TRAINING (ANTRAIN ANN) AND ITS APPLICATIONS TO THE WIND ENERGY SYSTEMS

Artificial neural network (ANN) models try to learn the relationship between input and output data with a training approach. There are many algorithms to train these models. On the other hand, current training algorithms, such as backpropagation, tend to be stuck in local minimums which prevent the learning process. Hence, heuristic approaches are used to train neural networks. The main purpose of the training process is to minimize the model error. The weights, which are continuous coefficients, of the network are used to evaluate the target of the inputs. The outputs of the networks have errors. Therefore any successful methodology providing global minimum (minimum of errors) can be adapted to ANN models as a training approach. This chapter presents a new neural network (ANN) training algorithm based on ant colony optimization for continuous minimization problems using a novel pheromone update strategy (ACO-NPU). First of all, a global minimization approach (ACO-NPU) contributed to search global minimum of continuous problems and presented in a journal [47]. Then, a new training algorithm, Antrain ANN, is contributed due to the ACO-NPU. After all, Antrain ANN is tested on different case problems. Results showed that the proposed approach can compete with current available algorithms.

8.1 Introduction

Traditional forecasting models may be inadequate to solve many data due to various assumptions that they require. By the way, the nature of the ANN does not require any assumptions and more data. Therefore, ANN models are used as alternative methods to traditional forecasting methods in the linear and non – linear problems. Biological neural networks can learn from the real life samples and generalize them, so do ANN models. Hence, the main advantage of ANN is learning ability from a set of input and output data. The learning process is basically defined as training that aims finding the best values of the weights which minimize the model errors. Therefore, the training can be handled as an optimization problem.

Finding the best weights in an ANN can be considered as an optimization problem. The most commonly used learning algorithms are the Levenberg–Marquardt and Backpropagation algorithms which involve derivative methodologies. By the way, there are many other methods such as particle swarm optimization, ant colony optimization, taboo search, and genetic algorithm to find best values of the neurons weights.

In the early years of the ANN model studies, there is only single-layer model which was firstly introduced by McCulloh and Pitts (1943) [278]. Those single-layer models consist of only input and output layers and a linear output function. This was the main disadvantage of the model which does not allow solving nonlinear models. Meanwhile, multi-layer neural ANN model (Multi-Layer Perceptron, MLP) introduced by Rumerhalt et al. (1986) to solve nonlinear models with the backpropagation methodology. This model includes an input layer, one or more hidden layer, and an output layer. There are many other ANN models such as generalized mean neuron [279], multiple spiking neural network model [280], and geometric mean neuron model [281]. By the way, the most common type of those models is MLP having been included much software such as Matlab, SPSS and Statistica.

This chapter presents a novel training algorithm, Antrain ANN, for MLP ANN models using ACO-NPU. The next section gives the basics of the proposed ACO-NPU methodology with benchmarking problems, and then Antrain ANN is presented. The case studies are given in the Part 8.4 with wind energy related models. Finally, results are concluded.

8.2 Ant Colony Optimization With a Novel Pheromone Updating Method

This part of the study is published in the journal of *Applied Mathematics and Computation* with a title of "*Ant colony optimization for continuous functions by using novel pheromone updating*" [47]. Thus, the proposed methodology is summarized.

A novel pheromone updating strategy is introduced in this study. The main purpose is to introduce a new global minimization heuristic approach to the literature. Therefore, proposed approach is compared with other current approaches designed to find global minimum of continuous functions in the literature as follows;

- Adaptive random search technique (ARSET) [282]
- Ant colony optimization based algorithm (ACO – BA) [283]
- Ant colony optimization with reduced search space (ACORSES) [284]
- Dynamic random search technique (DRASET) [285]
- Heuristic random optimization (HRO) [286]
- Improved genetic algorithm by random search technique (IGARSET) [287]
- Modified ant colony optimization (MACO) [288]
- Successive zooming method (SZGA) [289]

The ACO-NPU is contributed to find global minimum. The distinctive feature of the proposed algorithm is novel pheromone updating which depends on better solutions so that ants could be directed to better solutions by intensity of pheromone value. It is performed by means of solution archive and using the information provided by previous solution. Moreover, ACO-NPU differs in terms of the generated new colony with predetermined solution archive size. At the end of the each iteration, old solution archive and new archive are combined and a new extended solution archive is kept to explore better solutions. Generally, ant colony method uses a pheromone table to generate new solutions in combinatorial optimization problems. But it is not possible to obtain a pheromone table for continuous optimization problems as the number of candidate values is infinite.

Figure 8.1 summarizes the general approach of ant colony optimization methodology. The first step consists mainly in the initialization of the parameters such as number of maximum iterations (T), number of ants (m), and an initial pheromone value of each ant. The second step consists of randomly generation of ants according to their initial pheromone values as in following Equation 8.1;

$$p_{ij} = \frac{\pi_{ij0}^{\alpha}\eta_{ij}^{\beta}}{\sum_{l=1}^{N_i}\pi_{il0}^{\alpha}\eta_{il}^{\beta}} \qquad \text{Equation 8.1}$$

where m is the number of ants in population, T is the number of iterations (generations), ij are the portion of entire solution (trail), N_i is the neighborhood of location i, π_{ijt} is the amount of pheromone on trial ij at time t, η_{ij} is the heuristic regarding trail, α is the relative importance of pheromone, and β is the relative importance of heuristic.

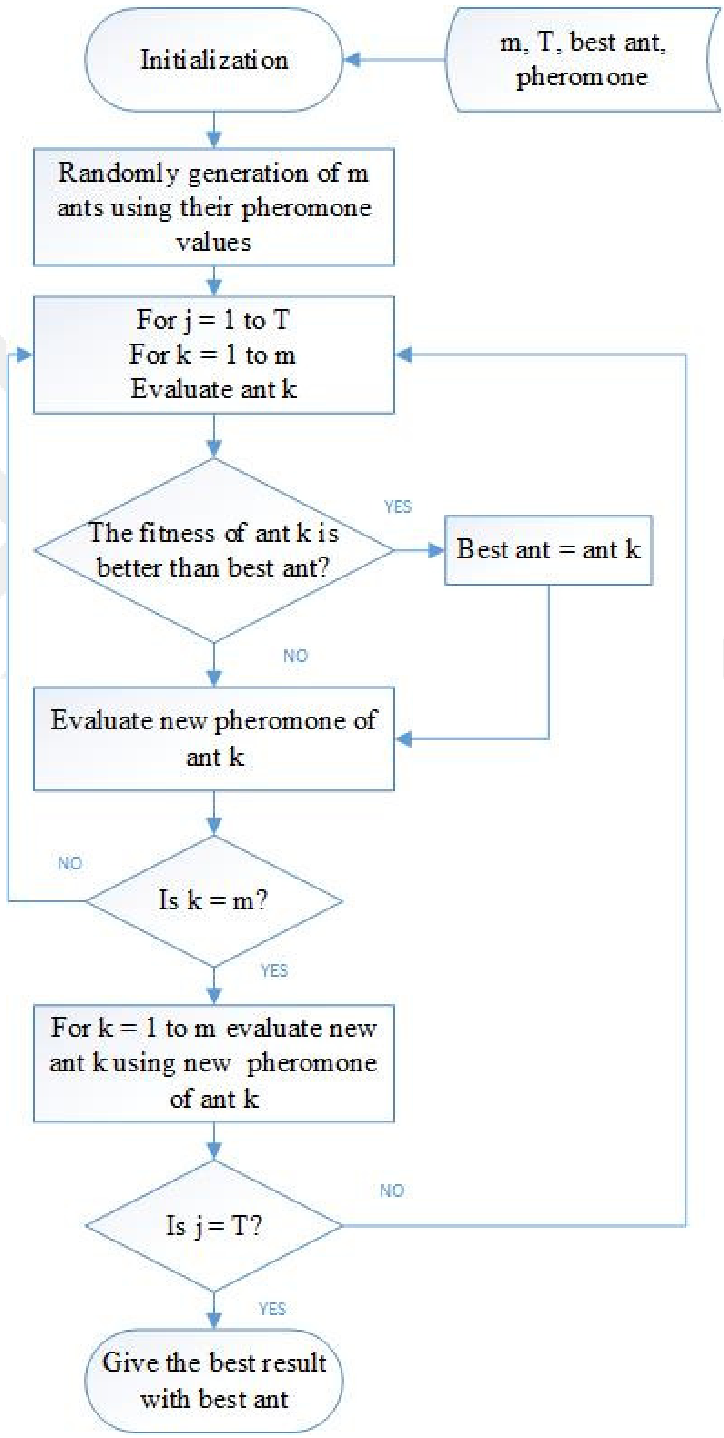

Figure 8. 1 Pseudo code of ant colony optimization

The third step updates quantity of pheromone; a global pheromone updating rule is applied in two phases. First, an evaporation phase where a fraction of the pheromone evaporates, and then a reinforcement phase where each ant deposits an amount of pheromone which is proportional to the fitness of its solution. This process is iterated until a stopping criterion.

The traditional ant ACO for discrete problems evaluates pheromone values for each candidate solution. On the other hand, in the continuous search space it is impossible. Therefore, a solution archive is used as a way of describing the pheromone distribution over the search space similar with the reference [290]. By using archive, new solutions could be independent from best solution generated by previous iterations and more than one solution vector could have a chance to generate new solutions.

Figure 8.2 illustrates the process of proposed approach. Firstly, the solution archive ($x^{k}_{initial}, k = 1,2,...,$ archive size) is initialized and $F(\, x^{k}_{initial}\,)$ is computed. These initial solutions which are to number of ants are extracted from solution archive and pheromone trails are assigned to remaining solutions. It is not necessary to assign an equal number for archive size and ant size. But, in the following step, ant size must be equal to selected solutions size which will be extracted from archive. Then new pheromone values are computed for the last selected solutions (remaining solutions). Up to this step, actually a local search is made among the candidate solutions. To approximate to the best solution on that iteration, Euclidean distance (D_i) is computed between minimum function value of F_{min} and the other selected function F_i values with Equation 8.2;

$$D_i = F_i - F_{\min} \qquad \text{Equation 8.2}$$

In order to define continuous variables, Gaussian function is selected as in Equation 8.3 to compute probabilities (ϕ_i) by using D_i.

$$\phi_i = e^{-\frac{D_i^2}{2t}} \qquad \text{Equation 8.3}$$

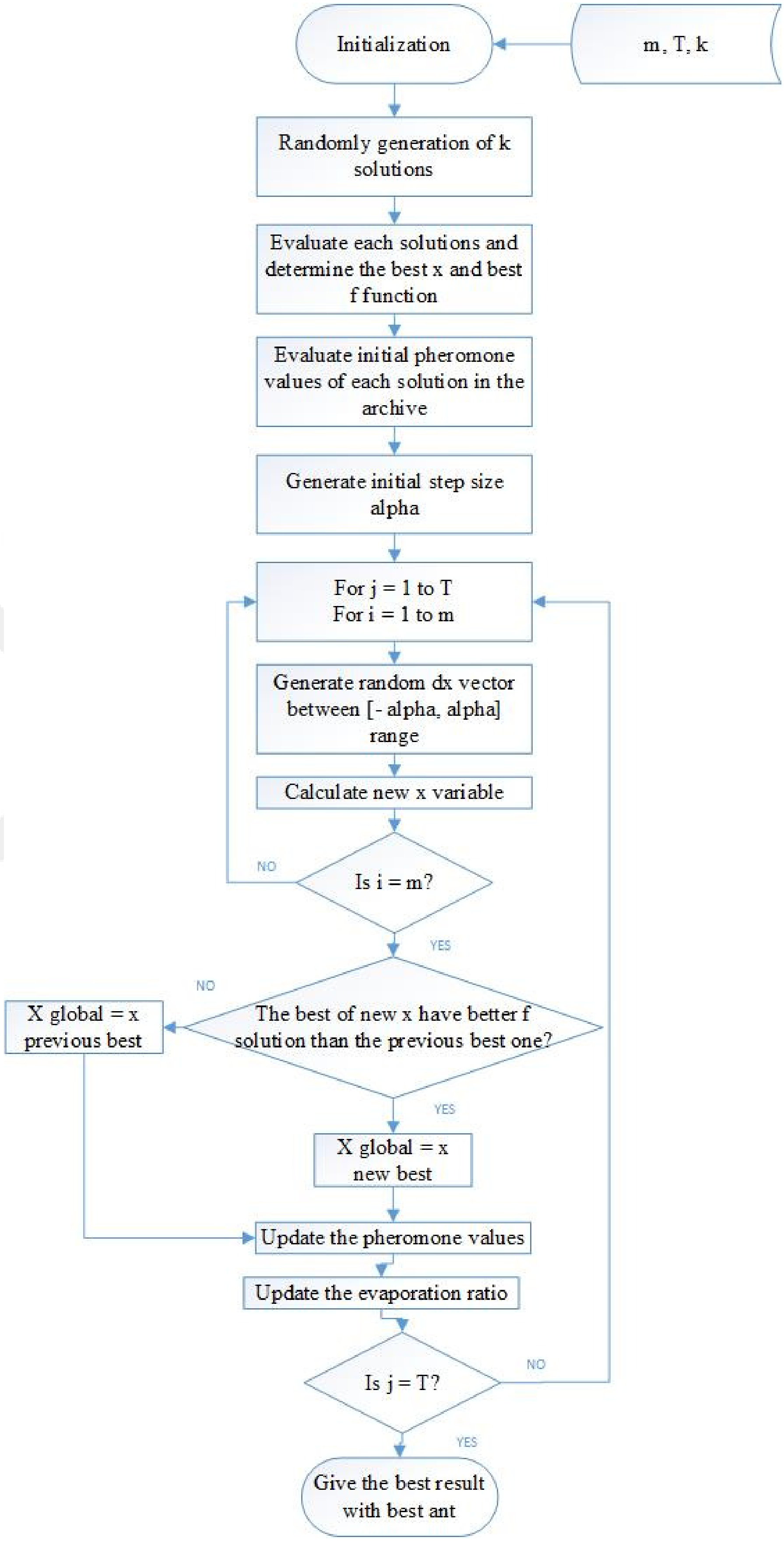

Figure 8. 2 Pseudo code of proposed ACO-NPU

Meaning of the parameter *t* in Gaussian function is a standard deviation. In order to clarify with percentiles how many ants which would go into solution, normalization is made with Equation 8.4. Normalized values (τ_i) are obtained from Gaussian function.

$$\tau_i = \frac{\phi_i}{\sum_{i=1}^{m} \phi_i} \qquad \text{Equation 8.4}$$

τ_i is pheromone value of i^{th} solution. At the end of the each iteration, quantity of pheromone is reduced to simulate the evaporation process with the following Equation 8.5;

$$\alpha_T = (0.9) \times \alpha_{T-1} \qquad \text{Equation 8.5}$$

The coefficients of evaporation equation (α) were set 0.9 at the end of the each iteration. It was realized that if the coefficient is accepted smaller or greater than 0.9, the algorithm was stopping without reaching global minimum. Also, visibility of quality solution (β) in this study is provided with exchanges in the solution archive.

The proposed ACO – NPU is tested in ten different benchmarking problems. The first two of them are one dimensional continuous minimization problems as in the following Equations 8.6 and 8.7;

Benchmarking Problem 1 (BP-1);

$$\min f(x) = \begin{Bmatrix} x^2, & \text{if } x \leq 1 \\ (x-3)^2 - 3, & \text{if } x > 1\rangle \end{Bmatrix} \qquad \text{Equation 8.6}$$

Benchmarking Problem 2 (BP-2);

$$\min f(x) = \left[x \times \sin\left(\frac{1}{x}\right) \right]^4 + \left[x \times \cos\left(\frac{1}{x}\right) \right]^4 \qquad \text{Equation 8.7}$$

While the BP-1 is an easy to find global optimum, BP-2 has too many local minimums as seen in Figure 8.3.

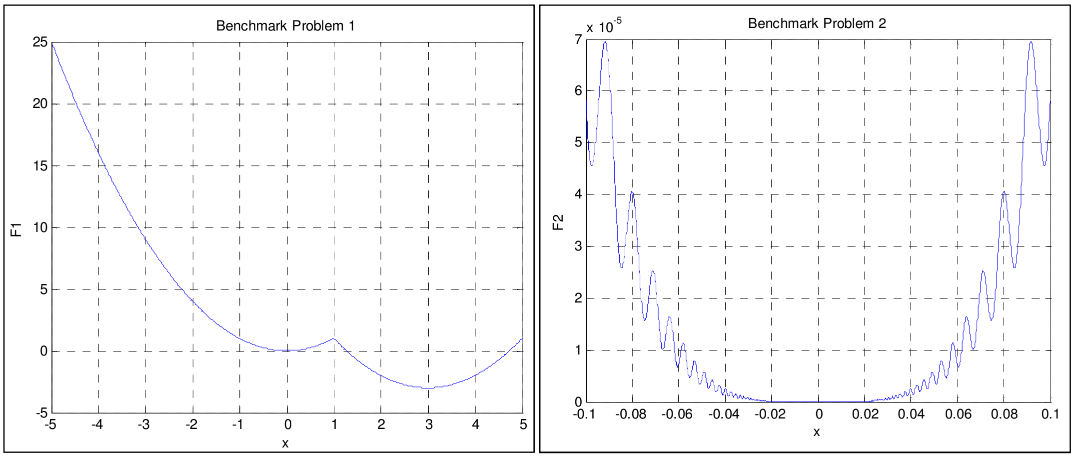

Figure 8. 3 One dimensional benchmarking problems

The comparisons with the current algorithms for the BP-1 and BP-2 are given in Table 8.1. Results showed that ACO-NPU is able to find optimum solution for the BP-1 and gives better result for the BP-2 than current algorithms. On the other hand, solution time is higher than other approaches. Of course, it can be caused by different type of computers while testing the algorithms.

Table 8. 1 Comparison of results for one dimensional benchmarking problems

		BP-1		
Algorithms	Best x	Best f(x)	Epoch number	Best solution time (s)
HRO	3,00032	-3	1000	NA
ARSET	3	-3	1000	NA
ACO-BA	3	-3	500	NA
MACO	3	-3	500	0.0090
ACORSES	3	-3	500	0.0480
IGARSET	3	-3	465	0.0420
ACO-NPU	3	-3	400	0.3959
		BP-2		
ARSET	1,90E-06	2.21E-43	50000	NA
ACO-BA	7.79E-012	1.40E-045	5000	NA
MACO	9.92E-12	5.60E -45	5000	0.032
IGARSET	NA	1.01E-74	1789	0.0669
ACORSES	NA	4.5E-83	4101	0.0550
ACO-NPU	2.64E-21	4.59E-98	1500	0.1547

The rest of the benchmarking problems are two dimensional minimization problems having following Equations 8.8 – 8.15;

Benchmarking Problem 3 (BP-3);

$$\min\ f(x,y) = \frac{(x-3)^8}{1+(x-3)^8} + \frac{(y-3)^4}{1+(y-3)^4} \qquad \text{Equation 8.8}$$

Benchmarking Problem 4 (BP-4);

$$\min f(x,y) = (100 \times (x - y^2)^2) + (1-x)^2 \qquad \text{Equation 8.9}$$

Benchmarking Problem 5 (BP-5);

$$\min f(x,y) = \frac{x}{1+|y|} \qquad \text{Equation 8.10}$$

Benchmarking Problem 6 (BP-6);

$$\min f(x,y) = x^2 + 2y^2 - 0.3\cos(3\pi x) - 0.4\cos(4\pi y) + 0.7 \qquad \text{Equation 8.11}$$

Benchmarking Problem 7 (BP-7);

$$\min f(x) = \sum_{i=1}^{n} - x_i \times \sin\sqrt{|x_i|} \qquad \text{Equation 8.12}$$

Benchmarking Problem 8 (BP-8);

$$\min f(x,y) = x^2 + y^2 - \cos(18x) - \cos(18y) \qquad \text{Equation 8.13}$$

Benchmarking Problem 9 (BP-2);

$$\min f(x,y) = \exp(\frac{1}{2}(x^2 + y^2 - 25)^2) + \sin^4(4x - 3y) + \frac{1}{2}(2x + y - 10)^2 \qquad \text{Equation 8.14}$$

Benchmarking Problem 10 (BP-10);

$$\min f(x,y) = \left[4 - 2.1x^2 + \frac{x^4}{3}\right]x^2 + xy + \left[-4 + 4y^2\right]y^2 \qquad \text{Equation 8.15}$$

The hardness of the problems can be seen in Figure 8.4 on the graphs of the mentioned benchmarked problems. By the way, results are listed in Table 8.2 for two dimensional benchmarks. The proposed method generally gives acceptable results according to other approaches. The solution time and results are providing its usefulness on the global minimization of the continuous problems. So that, ACO – NPU said to be a competitive approach to find global minimum of continuous problems.

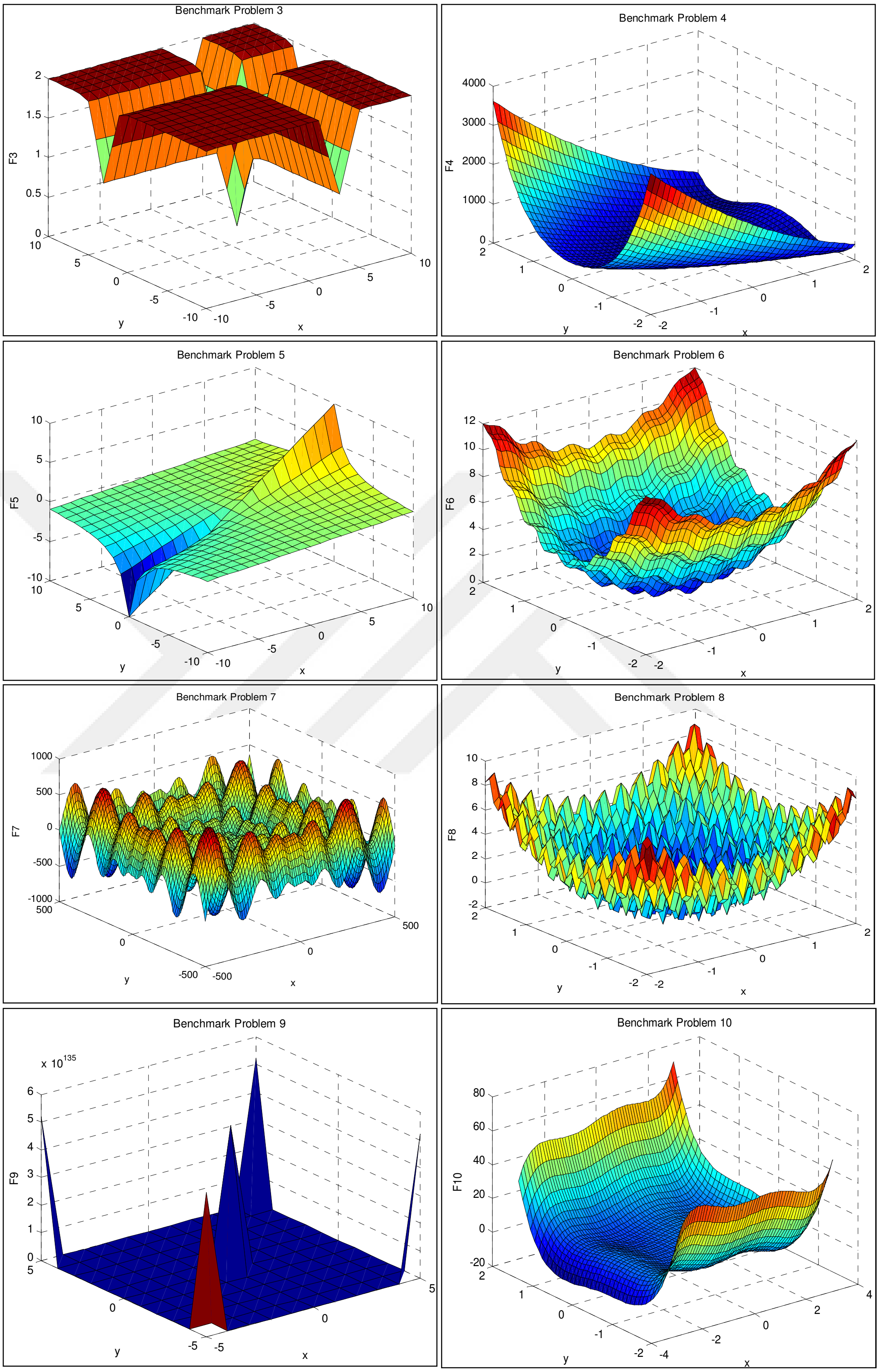

Figure 8. 4 Two dimensional benchmarking problems

Table 8. 2 Comparison of results for two dimensional benchmarking problems

Algorithms	Best x	Best y	Best f(x,y)	Epoch number	Best solution time (s)
			BP-3		
ARSET	3,0157	2,9999	3,71E-15	10000	NA
ACO-BA	3×2066E-09	3×2384E-09	2,62E-21	5000	NA
MACO	3	3	0	3750	0.0180
IGARSET	NA	NA	2.08E-27	1821	0.0666
ACORSES	NA	NA	4.06E-52	3624	0.0830
ACO-NPU	3	3	0	1500	0.0637
			BP-4		
ARSET	0,99401	0,997	3,58E-05	10000	NA
ACO-BA	1	1	0	50000	NA
MACO	1	1	0	3600	NA
IGARSET	NA	NA	0	2174	0.0568
ACORCES	NA	NA	0	3402	0.0590
ACO-NPU	1	1	0	20000	0.2256
			BP-5		
ARSET	-10	6,67E-08	-10	50000	NA
ACO-BA	-10	8,07E-11	-10	5000	NA
MACO	-10	0	-10	3750	0.0440
IGARSET	NA	NA	-10	1205	0.1043
ACORCES	NA	NA	-10	2167	0.0620
ACO-NPU	-10	-1.82E-17	-10	1125	0.0308
			BP-6		
SZGA	NA	NA	2.98E-8	4000	NA
MACO	0	0	0	3750	0.0080
IGARSET	NA	NA	0	1004	0.0485
ACORSES	NA	NA	0	1832	0.0520
ACO-NPU	-9.52E-12	-9.52E-12	0	1000	0.0556
			BP-7		
ACORSES	NA	NA	-8.379.658	1176	0.0690
ACO-NPU	420. 9687	420. 9687	-8.379.658	750	0.0289
			BP-8		
MACO	NA	NA	-2	235	0.0120
IGARSET	NA	NA	-2	2400	0.0614
ACORSES	NA	NA	-2	1610	0.0445
ACO-NPU	1.06E-05	1.06E-05	-2	75	0.0494
			BP-9		
MACO	3	4	1	36000	0.0210
DRASET	NA	NA	1	29663	14.468
IGARSET	NA	NA	1	1849	0.0537
ACORSES	NA	NA	1	1576	0.0630
ACO-NPU	3	4	1	2500	0.4252
			BP-10		
SZGA	NA	NA	-1,0316	NA	NA
ACO-NPU	-0,0898	0,7125	-1,0316	500	0.0142

Consequently, novel pheromone updating is used to compute pheromone quantity at the end of the each iteration and allows ants to generate new solutions by concentrating to better ants. The performance of the proposed algorithm was evaluated on ten benchmark problems and compared with performance of several algorithms available in literature such as ACO-BA, MACO, ARSET, SZGA, IGARSET, and ACORSES. The colony size, archive size and iteration number are given different for all benchmark problems, since these parameters requires tuning. In order to determine the parameters (archive size, ant size and iteration number) heuristically, numerous experiments were performed. Particularly, performance of ACO-NPU is dependent on ant size and iteration number as is the case of many ACO algorithms.

It is concluded that the use of a solution archive with a novel pheromone updating scheme during the steps of the ACONPU can help to find global minimum without being trapped in local minimum in selected benchmark problems within a reasonable solution time. The performance of the proposed algorithm was generally good than that of existing algorithms proposed for one or two dimensional continuous problems, so it is obvious that the algorithm by using novel pheromone updating proves useful for finding global minimum of continuous functions.

8.3 A Novel Neural Network Training Algorithm: Antrain ANN

The training process of an ANN is a crucial step in the forecasting models in order to representing the sample data. In the current literature, there are many modern approaches using heuristics in the learning process of ANN models. Chau (2006) contributed a particle swarm optimization approach to learning process of ANN model which was used to predict stage of Shing Mun River [291]. A supervised learning algorithm for multiple spiking ANN models is presented by Ghosh-Dastidar and Adeli (2009) to detect seizure in epilepsy [280]. A fuzz clustering methodology used to define training sets of ANN models to predict intrusion detection systems by Wang et al. (2010) [292]. Baş et al (2016) proposed a robust training algorithm for multiplicative ANN [293]. Their main purpose is to contribute a new ANN learning model which is able to learn with data outliers. Mohamad et al. (2016) used a genetic algorithm approach in the training of a MLP ANN model to minimize prediction errors of ripping process [294]. Lee et al. (2016) proposed a harmony search algorithm to determine the near – global optimal initial weights in the training of the ANN model which is used to predict the stability number of breakwater armor stones [295]. A dual stage multi resource data trainin with multi objective genetic algorithm is presented by Ganguly et al. (2016) and applied in alloy design problems [296]. A differential evolution algorithm is contributed to train of multiplicative neuron model for forecasting by Baş (2016) [297]. This chapter presents a heuristic approach to train MLP ANN models with a novel pheromone updating strategy of ant colony optimization called Antrain ANN. In the current literature, ACO based approaches are used in training procedures of ANN models. Lİ and Chung (2005) presented a novel backpropagation ANN model with a combination of ACO [298]. Their approach is based on the key factor of the optimal weight determination by the concentration of pheromone laid of the artificial ants moving on the connection path. Socha and Blum (2007) presented their continuous ant colony optimization technique in order to train of feed-forward neural networks for pattern classification [290]. The main characteristic of the contributed algorithm was that the ACO was able to find minimum errors of backpropagation process or standard Levenberg–Marquardt approach. Therefore, it can be said that it is a hybrid model of Backpropagation and/or Levenberg–Marquardt training algorithm. Their proposed approach was tested on a classification problem with a feed forward approach of ANN model. Beside the fact that another recent work on ACO based ANN training approach is contributed by Saghatforoush et al. (2016) [299], they only consider ACO to optimize given patterns of ANN model result.

This study proposes a new approach to train ANN models which exactly differs from another aforementioned current studies. Figure 8.5 gives the flow diagram of the proposed approach. The first step is initialization of the parameters and getting input and target data form the user. m represents the number of ants, T is the maximum iteration number, k is the archive size, sample size is defined according to sample of the input and target data, Nwi is the number of hidden layer weights according to input parameters, Now is the number of output layer weights according to number of target parameters, Bwi is the number of input bias weights, Bwo is the number of output bias weights, weight ranges are the interval of the initial weights, minError is the target minimum error defined by the user, MaxRepeat is the number of maximum repetition of the algorithm if the minError is not found.

Then, all inputs and target data are normalized between [0,1] according to Equation 8.16;

$$\text{NormalizedData} = \frac{(\text{CurrentData} - \text{MinimumOfData})}{(\text{MaximumOfData} - \text{MinimumOfData})} \qquad \text{Equation 8.16}$$

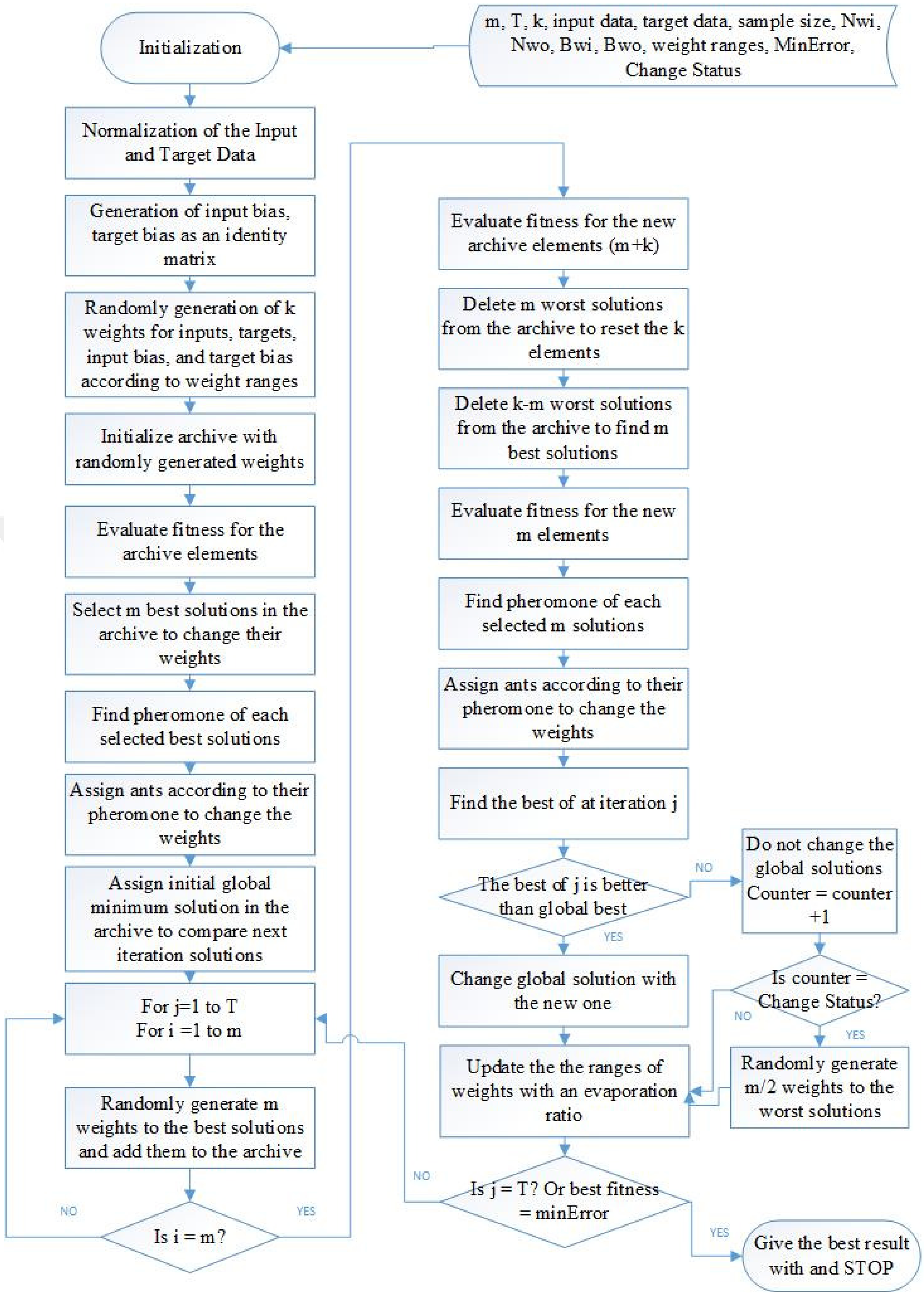

Figure 8. 5 Flow diagram of Antrain ANN training process

The input biases and target biases are identity matrix with the number of hidden layers and output layer respectively. The next step is generating an archive size of randomly weights of input layers, output layers, input biases and output biases in the range of defined intervals for each one. Now, the algorithm is ready to evaluate initial fitness values of each element in the archive.

Figure 8.6 represents the main architecture of the MLP ANN models. There is an input layer which takes the input parameters (I_1, I_2, …, I_n). The proposed approach uses the normalized input values to prevent the outliers of different kind of input parameters. The second layer is the hidden neuron layer (H_1, H_2, …, H_m). Every hidden neuron has its weights (*Wi*) which needs to be optimized. By the way, every hidden neuron has a bias value (*Bi*) with a weight of Bwi. The output of the hidden neuron is computed with the following Equation 8.17;

$$Output_{hiddeni} = \sum_{l=1}^{n} \sum_{j=1}^{m} I_l Wi_j + \sum_{j=1}^{m} Biasi_j Bwi_j \qquad \text{Equation 8.17}$$

where n is the number of inputs, m is the number of hidden neuron, Wi is the weights of hidden neurons, Bwi is the weight of hidden neuron biases. Sigmoid function is used as an activation function in this training approach. Hence, the hidden neuron outputs are activated with following Equation 8.18;

$$Net_i = 1/(1 + (e^{-Output_{hiddeni}})) \qquad \text{Equation 8.18}$$

Then, thus activated hidden neuron output is used to evaluate network output with following Equation 8.19;

$$Output_{oi} = \sum_{j=1}^{m} \sum_{p=1}^{k} Neti_j \quad Wo_p + \sum_{p=1}^{k} Biaso_p Bwo_p \qquad \text{Equation 8.19}$$

where k is the number of outputs, Wo is the output layer weights coming from input layer, Bwo is the weights of the output layer biases. Then, the result of the output layer is also activated with the sigmoid function as follows in Equation 8.20;

$$Output = 1/(1 + (e^{-Output_i})) \qquad \text{Equation 8.20}$$

Finally, the model for each sample *i* error is computed with the Equation 8.21;

$$Error_i = |Target_i - Output_i| \qquad \text{Equation 8.21}$$

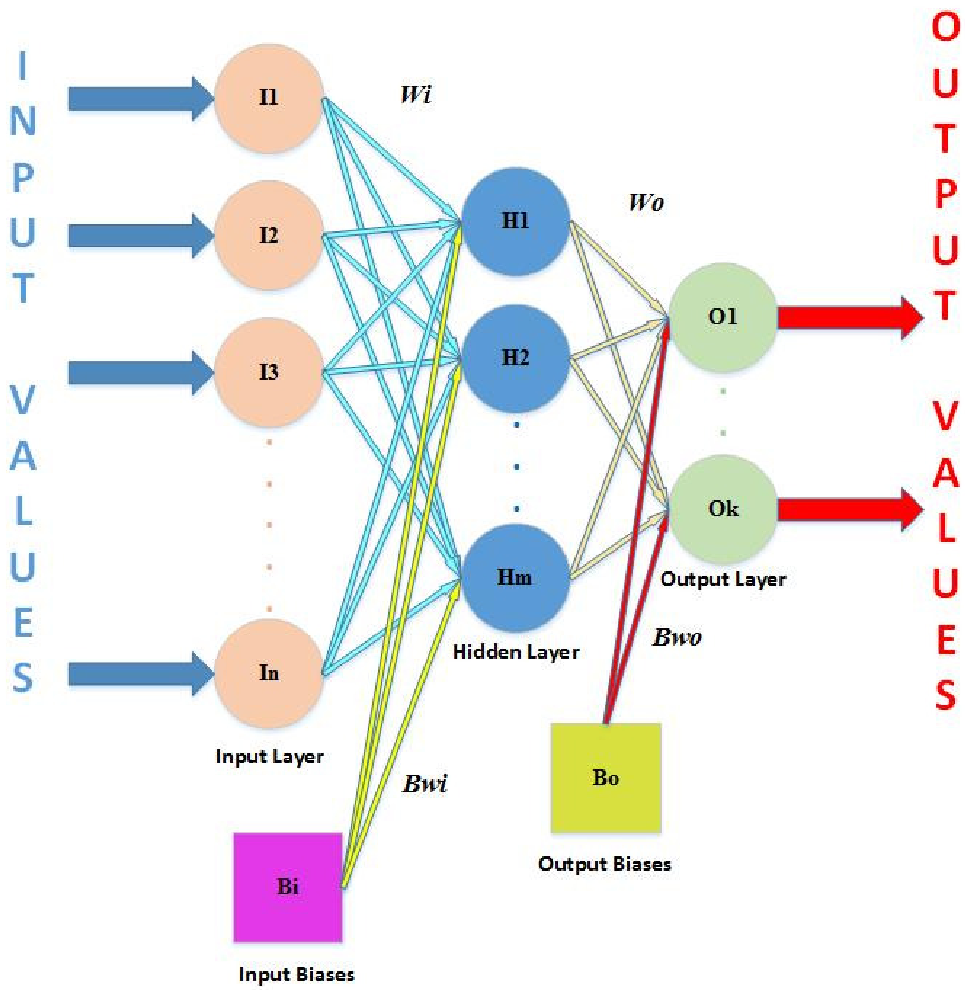

Figure 8. 6 General framework of an ANN structure

The proposed approach computes all sample evaluations as a batch process for each element in the archive. Then the sum of squared errors (SSE) is computed with following Equation 8.22;

$$SSE_k = \sqrt[2]{\sum_{i=1}^{sample} Error_i} \qquad \text{Equation 8.22}$$

After these initialization steps, ant colony procedure is prepared by selecting m best solutions according to their SSE values in the archive. The normal probabilities of those selected solutions are evaluated with the Equation 8.23;

$$Gaussian_m = e^{\frac{(SSE_m - \min(SSE))^2}{2\delta}} \qquad \text{Equation 8.23}$$

where m is the indices of the selected solution and δ is the standard deviation for a normal probability distribution. Then those probabilities are normalized to find pheromone values of each solution with Equation 8.24;

$$Pheromone_m = \frac{Gaussian_i}{\sum_{i=1}^{m} Gaussian_i} \qquad \text{Equation 8.24}$$

The minimum of those initial solutions is assigned to global minimum. Then the algorithm starts to find new solutions.

Firstly, randomly m weights are produced by using the best m solutions and add them to the archive. Then, evaluate fitness to those k + m sized solutions with Equations 8.17 – 8.22. Delete the worst m solution to gain a k sized archive. Then, define the best m solution in the archive to compute their pheromone values with Equations 8.23 – 8.24. Find the best solution of current iteration and compare it to the global best one. If the new best is less than the global one, change the best weights and solution with the new one; else continue with updating the ranges of the weights with an evaporation rate. Also, the proposed approach uses a counter to limit the worst results. If the counter reaches the Change Status value, half of the number of ants of the worst solutions in the archive changed with new weights which are randomly generated from the worst results by using global best weights. While the maximum iteration number or the minimum required error reached the algorithm stops and gives the best outputs and weights for the network. The developed algorithm is given at Appendix A.2.

8.4 Applications of Proposed Method

The new training algorithm is tested on different case problems. The parameters of the algorithm is set for nearly all cases problem as follows: m=60, k=100, T=100, Change Status=10, , Wi range is [-10, 10], Wo range is [-10, 10], Bwi range is [-10, 10], Bwo range is [-10, 10], minimum error rate is 0,005, evaporation rate for the weight ranges is 0,9, and δ=0,0005. The proposed algorithm is run for 10 times for each case problem. The results of those run trials given with the best, average, and the worst ones to compare current training algorithms such as Quick Propagation, Conjugate Gradient Descent, Quasi Newton, Limited Memory Quasi Newton, Levenberg-Marquardt, Online Backpropagation, Batch Back Propagation which are available in Alyuda NeuroIntelligence Network Software. The data of each case were partitioned to train, validation, and test sets. The validation sets were used as test sets for the Antrain ANN algorithm, because it does not require validation sets. The performance parameters are selected as mean error forecast (MEF), mean absolute deviation (MAD), Mean Absolute Percentage Error (MAPE) for continuous outputs, and Correct Classification Rate (CCR) for classification outputs. All those performance parameters were described in Chapter 7 of this thesis except CCR. The CCR is the percentage of the correct classification of proposed approach.

8.4.1 Test Problem 1: Weather Problem

The weather data set consists of fourteen samples with four input data and one output data. Inputs are the outlook of the weather, temperature in ºF, humidity percentage, and wind situation of the weather. The output is decision on playing outside or not. The sample data are given in Table 8.3 which is taken from the example data sets of a data mining tool called WEKA. First of all, the given data has non numeric values. Therefore, those non numeric data needs to be recoded with numeric ones such as for the outlook data, sunny is 1, overcast is 2, and rainy is 3; for the windy column, false is 0, true is 1; and for the output column, no is 0, yes is 1. Then, the data are partitioned as train, test, and validation sets.

Table 8. 3 Weather Data Set

No.	Status	Outlook	Temperature	Humidity	Windy	Play
1	Train	Sunny	85	85	False	No
2	Test	Sunny	80	90	True	No
3	Validation/Test	Overcast	83	86	False	Yes
4	Train	Rainy	70	96	False	Yes
5	Train	Rainy	68	80	False	Yes
6	Train	Rainy	65	70	True	No
7	Train	Overcast	64	65	True	Yes
8	Train	Sunny	72	95	False	No
9	Train	Sunny	69	70	False	Yes
10	Train	Rainy	75	80	False	Yes
11	Train	Sunny	75	70	True	Yes
12	Train	Overcast	72	90	True	Yes
13	Test	Overcast	81	75	False	Yes
14	Validation/Test	Rainy	71	91	True	No

In order to compare performance of the novel Antrain ANN algorithm with other current training algorithms, Alyuda NeuroIntelligence Software, which is available at Industrial Engineering Department of Gaziantep University, is used. The comparisons of results are given in the Table 8.4. While the worst training algorithm was Batch Back Propagation, Antrain ANN and Limited Memory Quasi Newton provided the best CCR rates with 100% correct classification. Moreover, the worst result of ten trials of Antrain ANN has given competitive result with a CCR rate of 78,5 %. On the other hand, Alyuda NeuroIntelligence Software does not provide Levenberg-Marquardt training algorithm for classification problems.

Table 8. 4 The comparison of Antrain ANN with current training algorithms for test problem 1

Algorithms	MEF	MAD	CCR %
Antrain ANN - Best	0,0000	0,0000	100,0000
Antrain ANN - Average	-0,0786	0,1642	83,5714
Antrain ANN - Worst	-0,2143	0,2143	78,5714
Quick Propagation	-0,2857	0,2857	71,4286
Conjugate Gradient Descent	0,1429	0,1429	85,7143
Quasi Newton	0,2143	0,2143	78,5714
Limited Memory Quasi Newton	0,0000	0,0000	100,0000
Levenberg-Marquardt	Not applicable	Not applicable	Not applicable
Online Backpropagation	-0,1429	0,1429	85,7143
Batch Back Propagation	-0,3571	0,3571	64,2857

8.4.2 Test Problem 2: Diabetes Problem

The diabetes problem has a total of 768 samples of diabetic patients which is taken from example data sets of WEKA. The data set consists of eight input data and one output data for the Diabetes Mellitus occurrence of women. The inputs are number of times pregnant, plasma glucose concentration at 2 hours in an oral glucose tolerance test, diastolic blood pressure in mm Hg, triceps skin fold thickness in mm, 2-hour serum insulin in μU/ml, body mass index in kg/m^2, diabetes pedigree function, and age. Those input variables have been found to be significant risk factors for diabetes [300]. The output is the occurrence of Diabetes Mellitus. The whole data set has 268 positive in Diabetes Mellitus and 500 negative ones. The first 524 data were used as train set, the next 122 data were used as validation set, the final 122 data were used as test set.

Table 8.5 gives the comparison of Antrain ANN with current training algorithms for Diabetes test problem. Hence this problem is a classification problem, CCR rates provides meaningful performance measures. While the Batch Back Propagation algorithm gave the worst CCR with 65%, Quick Propagation and Limited Memory Quasi Newton algorithms provided the best CCR with around 80%. On the other hand, Antrain ANN algorithm gave an acceptable CCR rate with 74%.

Table 8. 5 The comparison of Antrain ANN with current training algorithms for test problem 2

Algorithms	MEF	MAD	CCR %
Antrain ANN - Best	0,0260	0,2578	74,2187
Antrain ANN - Average	0,0994	0,3000	70,0000
Antrain ANN - Worst	0,2786	0,3411	65,8854
Quick Propagation	0,0430	0,1992	80,0781
Conjugate Gradient Descent	0,0443	0,2318	76,8229
Quasi Newton	0,0703	0,2135	78,6458
Limited Memory Quasi Newton	0,0482	0,2044	79,5573
Levenberg-Marquardt	Not applicable	Not applicable	Not applicable
Online Backpropagation	0,1146	0,2109	78,9063
Batch Back Propagation	0,3490	0,3490	65,1042

8.4.3 Test Problem 3: Multiplication Problem

This data set is created by the researcher to test the performance of proposed algorithm on different type of nonlinear problems. There are twenty samples with two inputs and one output. The inputs are randomly generated between 1 and 10 for each input. The output is simply the multiplication of those two inputs. The data set is given in Table 8.6. The first two rows were used as test set, the next two rows were used as validation set, and the next rows were selected as training set.

The nature of this test problem is based on nonlinear fitting. Therefore, the performance measurements are used as MEF, MAD, and MAPE to compare Antrain ANN algorithm and other available training algorithms. As it can be seen from the Table 8.7, while the developed Antrain ANN algorithm has provided the best results considering with MAD and MAPE values, the Quasi Newton algorithm has given the least MEF value. If the individual forecast error is valuable to the problem, MAD and MAPE values give more meaningful information about the performance of used methods. On the other side, if the total forecast error is valuable to the problem, then it would be better to choose the least MEF value.

Table 8. 6 The multiplication data set

No.	Status	Input1	Input2	Target
1	Test	1	5	5
2	Test	3	4	12
3	Validation/Test	5	8	40
4	Validation/Test	4	6	24
5	Train	8	7	56
6	Train	9	9	81
7	Train	4	1	4
8	Train	2	4	8
9	Train	5	6	30
10	Train	7	3	21
11	Train	3	9	27
12	Train	1	4	4
13	Train	10	3	30
14	Train	8	2	16
15	Train	4	1	4
16	Train	7	9	63
17	Train	3	7	21
18	Train	5	4	20
19	Train	8	6	48
20	Train	7	2	14

Table 8. 7 The comparison of Antrain ANN with current training algorithms for test problem 3

Algorithms	MEF	MAD	MAPE %
Antrain ANN - Best	0,2873	2,3980	13,1196
Antrain ANN - Average	0,5814	3,2590	24,2697
Antrain ANN - Worst	1,1791	6,5333	63,2675
Quick Propagation	-0,6852	3,8170	34,6510
Conjugate Gradient Descent	-0,5497	3,6149	37,7120
Quasi Newton	0,0612	9,2642	68,8557
Limited Memory Quasi Newton	-0,2208	3,0324	32,1856
Levenberg-Marquardt	-5,5013	7,6491	83,7364
Online Backpropagation	-1,2503	15,0422	139,7075
Batch Back Propagation	-13,3888	22,2261	245,9412

8.4.4 Test Problem 4: Wind Speed Problem

In the current literature, ANN models are commonly used in the wind energy forecast problems [170, 227, 227, 261, 301–309]. Therefore, a case problem is created with a real wind speed data of an operating wind turbine. This problem has 499 sample of one input and one output of wind speed. The input data is the current wind speed and the output data is the next period’s wind speed. Those input and output data are 10-minute averaged time series data of an operating wind turbine of a wind farm in Turkey. Table 8.8 gives summary descriptive of the used input and output data. The all descriptive parameters were explained in the Chapter 5 of this thesis. In this test problem, the first 341 rows were selected as training set, the next 79 rows were used as validation set, and the final 79 rows were used as test set.

Table 8. 8 The descriptive statistics of Wind Speed Test Data

Descriptive	**Wind Speed$_t$ (m/s)**	**Wind Speed$_{t+1}$(m/s)**
Range	6,00	6,00
Minimum	3,80	3,80
Maximum	9,80	9,80
Mean	6,58	6,58
Std. Deviation	1,15	1,16
Variance	1,34	1,34
Skewness	,20	,20
Kurtosis	-,31	-,32

The comparison of Antrain ANN with current training algorithms of Alyuda NeuroIntelligence Software is given in Table 8.9. Results showed that Antrain ANN algorithm is able to provide better results than other algorithms considering MAD and MAPE performance values.

Table 8. 9 The comparison of Antrain ANN with current training algorithms for test problem 4

Algorithms	**MEF**	**MAD**	**MAPE %**
Antrain ANN - Best	-0,0154	0,4444	6,9026
Antrain ANN - Average	-0,0209	0,4501	7,0181
Antrain ANN - Worst	-0,0456	0,4658	7,3181
Quick Propagation	0,0274	0,5278	8,3098
Conjugate Gradient Descent	-0,0069	0,5223	8,3015
Quasi Newton	0,1538	0,4848	7,3539
Limited Memory Quasi Newton	0,0084	0,4695	7,3752
Levenberg-Marquardt	0,0130	0,4629	7,1979
Online Backpropagation	-0,0921	0,4682	7,4209
Batch Back Propagation	-0,1906	0,9606	15,6879

8.4.5 Test Problem 5: Wind Power Problem

Wind power modeling of the wind turbines is another common problem in the literature. Therefore, this case is created with 499 samples of one input and one output parameters. The input is the current wind speed and the output is the next period's wind power generation. Those input and output data are 10-minute averaged time series data of an operating wind turbine of a wind farm in Turkey. Descriptive statistics of input and output data are summarized in Table 8.10. The first 341 rows were selected as training set, the next 79 rows were used as validation set, and the final 79 rows were used as test set.

Table 8. 10 The descriptive statistics of Wind Power Test Data

Descriptive	**Wind Speed$_t$ (m/s)**	**Wind Power$_{t+1}$ (kW)**
Range	6,00	1337,00
Minimum	3,80	44,70
Maximum	9,80	1381,70
Mean	6,5892	481,2627
Std. Deviation	1,15899	257,59906
Variance	1,343	66357,275
Skewness	,200	,874
Kurtosis	-,316	,438

The trial results of comparisons are given in Table 8.11. While the Batch Back Propagation algorithm has provided the worst performance values, Antrain ANN has provided the best values considering MAD and MAPE values.

Table 8. 11 The comparison of Antrain ANN with current training algorithms for test problem 5

Algorithms	**MEF**	**MAD**	**MAPE %**
Antrain ANN - Best	2,1764	91,7354	21,8357
Antrain ANN - Average	7,3926	93,9557	23,9049
Antrain ANN - Worst	10,5266	109,2595	32,8604
Quick Propagation	10,1613	102,7276	28,5165
Conjugate Gradient Descent	89,0901	170,8074	43,6112
Quasi Newton	-2,3258	94,8691	26,0469
Limited Memory Quasi Newton	0,7669	94,8535	25,5830
Levenberg-Marquardt	-21,6717	98,4592	28,1527
Online Backpropagation	-18,4206	97,2881	27,2972
Batch Back Propagation	-151,1867	255,7342	96,9213

8.4.6 Test Problem 6: Wind Turbine Faults Problem

The final case problem is fault prediction of a wind turbine. This case is also presented with real data of a wind turbine. The data set consists of 873 samples with 12 inputs and one output. The input parameters are generator speed, hydraulic oil temperature, rotor speed, wind speed, generated power, yaw angle, current phase 1, current phase 2, current phase 3, voltage phase 1, voltage phase 2, and voltage phase 3 of the current time. Table 8.12 summarizes the descriptive statistics of inputs.

Table 8. 12 The descriptive statistics of Wind Turbines Faults Test Data

Descriptive	Generator Speed (RPM)	Hydraulic Oil Temperature (ºC)	Rotor Speed (RPM)	Wind Speed (m/s)	Generated Power (kW)	Yaw Angle (ºC)	Current Phase 1 (Ampere)	Current Phase 2 (Ampere)	Current Phase 3 (Ampere)	Voltage Phase 1 (Volt)	Voltage Phase 2 (Volt)	Voltage Phase 3 (Volt)
Range	1687,00	35,00	16,00	22,70	2899,40	357,30	1629,50	1648,10	1638,70	584,40	584,00	580,20
Minimum	0,00	17,00	0,00	0,50	-20,40	-179,20	1,00	1,00	1,00	28,80	29,00	28,70
Maximum	1687,00	52,00	16,00	23,20	2879,00	178,10	1630,50	1649,10	1639,70	613,20	613,00	608,90
Mean	531,49	45,20	5,04	7,62	335,99	0,48	198,50	201,75	203,76	569,96	570,39	567,05
Std. Deviation	638,21	4,84	6,08	4,03	620,42	37,17	344,34	350,36	352,12	57,91	57,81	57,62
Variance	407317,80	23,40	36,95	16,21	384927,16	1381,26	118567,76	122755,53	123988,13	3353,30	3342,55	3319,88
Skewness	0,60	-1,86	0,60	1,04	2,12	0,21	2,14	2,12	2,11	-7,43	-7,45	-7,42
Kurtosis	-1,39	4,58	-1,39	0,95	3,81	11,15	3,92	3,84	3,79	58,06	58,39	57,94

The output is the fault occurrence of the next period. The data set has 673 faults and 200 normal operation data. While the first 673 data are coded as fault, the rest 200 data are normal operation data. Therefore, the first 10 rows were partitioned as validation set, the next 20 rows were test set, and the next 643 rows were training set for the fault data. On the other side, the next 170 rows were training set, the next 10 rows were validation set, and the final 20 rows were test set of the normal operation data that were not fault.

Table 8.13 gives the comparison of Antrain ANN with current training algorithms. While the best algorithm was Limited Memory Quasi Newton, Antrain ANN has provided competitive CCR rate with 86%.

Table 8. 13 The comparison of Antrain ANN with current training algorithms for test problem 6

Algorithms	MEF	MAD	CCR %
Antrain ANN - Best	0,0115	0,1351	86,4833
Antrain ANN - Average	-0,0428	0,1786	82,1305
Antrain ANN - Worst	-0,2245	0,2268	77,3195
Quick Propagation	0,0149	0,1478	85,2234
Conjugate Gradient Descent	0,0389	0,1375	86,2543
Quasi Newton	0,0092	0,1329	86,7125
Limited Memory Quasi Newton	0,0641	0,1100	89,0034
Levenberg-Marquardt	Not applicable	Not applicable	Not applicable
Online Backpropagation	0,6735	0,6735	32,6460
Batch Back Propagation	-0,2291	0,2291	77,0905

8.5 Conclusions

This chapter presents a new training algorithm for MLP ANN models with an ACO approach which is called as Antrain ANN. First of all, a novel pheromone updating ACO algorithm was presented to find global minimum of continuous problems. Then it was tested on different benchmarking problems. Results of the benchmarking problems showed that the proposed ACO-NPU could compete with current algorithms. Then, this algorithm was adopted in the training process of MLP ANN models. Proposed approach was given in detailed with a flow diagram. Finally, the performance of the Antrain ANN was tested on different case problems. Because the main purpose of this thesis was contributing managerial tools for wind energy systems, wind energy related cases were also considered with the proposed learning algorithm. Because of the nature of the Antrain ANN algorithm is heuristic, there are many parameters which are able to affect the performance. By the way, parameter optimization is another crucial problem when using heuristic approaches. Thus, parameter settings of the algorithm were not considered in this study. Consequently, developed Antrain ANN algorithm is able to train much kind of MLP ANN models with different type of problems. Future works can be focus on parameter optimization of the proposed approach.

CHAPTER 9

DISCUSSIONS AND CONCLUSIONS

The main purpose of this dissertation is to present useful managerial tools for wind energy systems that are able to control and monitor the whole system using available SCADA data. The general information and the importance of the selected topic (wind energy) is summarized in Chapter 1. First of all, energy security was introduced to understand the importance of local energy sources. Then, a brief information about climate change and its effects were given to mention the need for clean energy resources. After that, many of the current energy reports were summarized to figure out the energy market of the world and Turkey separately by focusing of wind energy. The following institutions present current energy/wind energy reports regularly;

- European Wind Energy Association (EWEA),
- Global Wind Energy Council (GWEC),
- International Energy Agency (IEA),
- Turkish Wind Energy Association (TUREB),
- World Energy Council (WEC),
- World Wind Energy Association (WWEA).

The main structure of the thesis and main contributions were also mentioned in the first chapter. Because the Chapter 1 presents annually reports of energy market, it would be better to follow current versions of those reports from the aforementioned institutions.

The second chapter presents the technical background on wind energy, wind turbines, and SCADA system to non-familiar of wind energy audience of this thesis.

First of all, the wind was described with its main characteristics, then energy formula of the wind was summarized. The type of the wind turbines were also introduced according to their designs and current technological improvements on horizontal axis wind turbines were illustrated. In addition to this, advantages and disadvantages of the wind energy were mentioned. At the end of the chapter, SCADA system was introduced with its components.

Once the main study area is selected as wind energy, there is a big problem about where to begin and what to study to contribute an esteemed PhD thesis. Therefore, an intelligent way, text mining, was used to manage this problem. Defining a PhD thesis problem is generally hard. There are enormous numbers of studies in the literature and one cannot read and understand many of them. Thus, Chapter 3 was developed to define the main problem of this dissertation. The current wind energy literature of a valuable database (Web of Science) was listed in a data set with their abstracts, author information, publication year, and countries of the authors' parameters. Results showed that the most frequent words in these publications were *system*, *power*, *generation*, *control*, and *model* respectively. By the way, the most important words of this data set were *Control, offshore, solar, blade,* and *wind farm* respectively. The most important concepts were related with *System Control* studies. Therefore, this thesis is focused on health monitoring strategies and optimization methods to maximize the utilization of installed wind turbines. By the way, the wind energy literature is expending from day to day. Thus, the same approach may give different results if it is repeated at today.

The problem definition of this thesis was given in Chapter 3. The condition monitoring algorithms and methodologies were commonly used as control strategies of wind energy plants. Therefore, it was focused on developing smart control approaches by considering condition monitoring. First of all, the definition of the problem was given, and then a detailed literature review on the control and condition monitoring methodologies for the wind energy was presented. The literature study showed that the common problem to control a wind farm/turbine could be handled by Condition Monitoring and control systems. Actually, the condition monitoring technology is hidden in the system data of new generation wind turbines. Therefore, the main research area for this dissertation is aimed to fill the lack of appropriate dynamic, smart and cheaper solutions to the control and monitor of wind energy systems. Hence, the following main questions were searched;

- Can SCADA data be used to control and monitor a wind farm in order to manage the farm in a smart way?
- What kind of parameters and algorithms can be used for processing and analyzing SCADA data to achieve the mentioned problem?

This thesis provides many different tools in Chapters 5, 6, 7, and 8 which answer those questions.

The other purpose of this study is to prove the usability of proposed approaches. Thus, all tools were studied on real operational data of a wind farm in Turkey. Consequently, the contributions of this thesis demonstrated the capabilities and applicability of real world condition monitoring applications.

The fourth chapter presents statistical analyzing tools of a wind farm to identify the characteristics of each wind turbine. First of all, the descriptions of analyzed parameters were given with related literature. Then, statistical approaches were described with numerical analyses and some statistical hypothesis. The main reason of contributing this part is that the general application of wind farm monitoring and control strategies are designed for the whole wind farm. Because the wind turbines are identical in the farm, all managerial plans are generally made according to technical characteristics of identical turbines. On the other side, operational or environmental conditions may affect the characteristic behavior of any turbine in a wind farm even if they have identical technical properties. Therefore, this chapter is focused on a simple and adaptable tool to identify different behavioral groups of turbines. In addition to this, the importance of clustering wind turbines in a single site was contributed to have better understanding on similar or dissimilar behaviors of identical turbines. To sum up, identifying of identical wind turbines behavior in the same wind farm was prominence in order to accurately monitor and control issues. For this aim, statistical tools were used to understand SCADA data of each turbine which provides to figure out the whole data before doing further analysis. Initially, the basic descriptive statistics were defined and then graphical representations of the data presented. The hypothesis tests were applied such as one sample Kolmogorov Smirnov and Kruskal Wallis to learn categorical similarities of parameters. Consequently, results showed that all turbines behave differently according to each parameter. To manage this situation, clustering approaches were used to group parameters (partitioned clustering) and turbines (hierarchical clustering). Finally, the need of clusters was proved by giving a power prediction model. The further researches on this issue may conclude deep statistical analysis of turbine by turbine or parameter by parameter to figure out wind fault analysis or reliability studies.

Performance analysis of wind turbines in an econometrics approach is the lack of current literature. Therefore, traditional productivity/efficiency tools were introduced in the wind energy literature to monitor production performance of each turbine in the same wind farm. The results of performance analysis could provide information to the plant manager about whether the current sources were used efficiently or not. So that, Chapter 6 provided novel performance analysis tools in order to compare productivities of turbines using SCADA data. The main purpose of contributing this chapter was to figure out meaningful information about wind turbine production behavior which might help power plant managers in order to make decisions about maintenance planning. The first proposed tool was DEA to compare wind turbines' production performance. Then, MI was presented to analyze time depended efficiency changes on power production. Because the nature of DEA and MI approaches are non-parametric, the third proposed performance measurement tool was SFA having a parametric performance analysis for the same input data. All those studies were presented in esteemed journals. The recommendation to the decision makers of the wind farms is that they should make deep analysis according to the efficiency results in order to find out the reasons for inefficient turbines and inefficient time horizons. Consequently, Chapter 6 guides on understanding the reasons behind the performance losses which is useful in developing performance enhancing strategies. Further research of this chapter may be focused on parametric performance analysis using suitable approaches to predict faults in the system caused by a defined parameter.

One of the most common used control and monitoring tool is forecast of wind energy systems. Because the prediction of next term's energy production is crucial in the market share, there are many approaches to achieve this aim. Therefore, simple tools on wind speed and wind power modelling, which use only SCADA data of an operating wind turbine, were introduced in the Chapter 7 in addition with a novel tool, PF. The developed PF algorithm basically filters *n* predictions according to their fitness value to the actual data which allows making more accurate predictions of any algorithm. While the focus of this chapter was on the usage of simple SCADA data to predict next terms' wind speed and wind power generation, the main contribution was the adaptation of PF to ANN models while predicting short term wind speed and wind power. Every forecast has an error which may help considering condition monitoring issues. The forecast precision is an important issue in the energy market. Meanwhile, proposed PF approach provided reduced forecast errors of ANN models. In addition to this, developed PF algorithm may also help managers to monitor the wind turbines power production behavior.

The final developed tool of this thesis is a novel learning algorithm for MLP ANN models, Antrain ANN, which was presented in Chapter 8. The proposed tool was based on ant colony optimization for continuous minimization problems using a novel pheromone update strategy (ACO-NPU). The learning procedure of ANN models aims to minimize model error by defining optimum values of network weights. Therefore, the process can be handled as a continuous optimization problem. In order to achieve this aim, a global minimization approach was contributed to search global minimum of continuous problems called ACO-NPU. After proving the performance of that novel algorithm, it was remodeled as training algorithm for MLP ANN which is called Antrain ANN. The proposed new training tool was tested on different case problems. Because the main purpose of this dissertation was contributing managerial tools for wind energy systems, wind energy related cases were also considered. Because the Antrain ANN is based on heuristic, parameter tuning of the algorithm needs to be performed in order to get better results. By the way, parameter optimization is crucial problem which is not considered in the whole thesis. Future works can be focused on parameter optimization of the proposed approach.

Each part of this dissertation includes particular information with regard to the aim of the chapter. Therefore, each of them needs to be separately considered with each other. The main contributions of the thesis are given after the Chapter 4. The researcher hopes that the first three chapter help to the audience who are the unfamiliar with the wind energy systems, and the main contribution parts help the wind energy managers on how they use the current SCADA data with an aim of monitoring and control issues.

To sum up, in this thesis, it is provided that the current situation can be revealed using the system data. The current situation will show whether the system is healthy or not. It is necessary to prove that the wind farms operate effectively and under control. The fact that the investment is not wasted early depends on the correct operation of the system. This thesis,

- Checkups using basic statistical tools,
- Examines the status of the system components,
- Evaluates production efficiencies,
- Provides new forecasting tools which become a light to the planning problems,
- Provides new learning algorithms for artificial neural networks.

REFERENCES

[1] MHI Vestas Offshore Wind. (2014, October 21). V164-8.0 MW breaks world record for wind energy production. *MHI Vestas Offshore Wind*. Available at http://www.mhivestasoffshore.com/v164-8-0-mw-breaks-world-record-for-wind-energy-production/. Accessed 21.04.2015.

[2] Crabtree, C. J. (2011). *Condition Monitoring Techniques for Wind Turbines* (Doctoral Thesis). Durham University. Retrieved from http://etheses.dur.ac.uk/652/

[3] García-Gusano, D., Iribarren, D., Garraín, D. (2017). Prospective analysis of energy security: A practical life-cycle approach focused on renewable power generation and oriented towards policy-makers, *Applied Energy*, **190**, 891–901. doi:10.1016/j.apenergy.2017.01.011

[4] Wang, Q., Zhou, K. (2017). A framework for evaluating global national energy security, *Applied Energy*, **188**, 19–31. doi:10.1016/j.apenergy.2016.11.116

[5] Kester, J. (2017). Energy security and human security in a Dutch gasquake context: A case of localized performative politics, *Energy Research Social Science*, **24**, 12–20. doi:10.1016/j.erss.2016.12.019

[6] Kruyt, B., van Vuuren, D. P., de Vries, H. J. M., Groenenberg, H. (2009). Indicators for energy security, *Energy Policy*, **37**(6), 2166–2181. doi:10.1016/j.enpol.2009.02.006

[7] Winzer, C. (2012). Conceptualizing energy security, *Energy Policy*, **46**, 36–48. doi:10.1016/j.enpol.2012.02.067

[8] Ang, B. W., Choong, W. L., Ng, T. S. (2015). Energy security: Definitions, dimensions and indexes, *Renewable and Sustainable Energy Reviews*, **42**, 1077–1093. doi:10.1016/j.rser.2014.10.064

[9] Jonsson, D. K., Johansson, B., Månsson, A., Nilsson, L. J., Nilsson, M., Sonnsjö, H. (2015). Energy security matters in the EU Energy Roadmap, *Energy Strategy Reviews*, **6**, 48–56. doi:10.1016/j.esr.2015.03.002

[10] Grubb, M., Butler, L., Sinden, G. (2005). *Diversity and Security in UK Electricity Generation: The Influence of Low Carbon Objectives* (Cambridge Working Papers in Economics No. 0511). Faculty of Economics, University of Cambridge. Retrieved from https://ideas.repec.org/p/cam/camdae/0511.html

[11] Hughes, L. (2009). The four 'R's of energy security, *Energy Policy*, **37**(6), 2459–2461. doi:10.1016/j.enpol.2009.02.038

[12] Sovacool, B. K., Brown, M. A. (2010). Competing Dimensions of Energy Security: An International Perspective, *Annual Review of Environment and Resources*, **35**(1), 77–108. doi:10.1146/annurev-environ-042509-143035

[13] Umbach, F. (2010). Global energy security and the implications for the EU, *Energy Policy*, **38**(3), 1229–1240. doi:10.1016/j.enpol.2009.01.010

[14] Müftüler-Baç, M., Başkan, D. (2011). The Future of Energy Security for Europe: Turkey's Role as an Energy Corridor, *Middle Eastern Studies*, **47**(2), 361–378. doi:10.1080/00263206.2010.481176

[15] Brown, M. A., Wang, Y., Sovacool, B. K., D'Agostino, A. L. (2014). Forty years of energy security trends: A comparative assessment of 22 industrialized countries, *Energy Research Social Science*, **4**, 64–77. doi:10.1016/j.erss.2014.08.008

[16]IEA. (2015). Energy security. Available at http://www.iea.org/topics/energysecurity/. Accessed 04/22/2015.

[17] European Commission. (2014). *In-depth study of European Energy Security*. Brussels: European Commission. Retrieved from https://ec.europa.eu/energy/sites/ener/files/documents/20140528_energy_security_study.pdf

[18] Jewell, J. (2011). The IEA Model of Short-Term Energy Security (MOSES) : Primary Energy Sources and Secondary Fuels.

[19] Energy security. (2015, March 29).In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Energy_securityoldid=654038567

[20] Oreskes, N. (2004). The Scientific Consensus on Climate Change, *Science*, **306**(5702), 1686–1686. doi:10.1126/science.1103618

[21] Doran, P. T., Zimmerman, M. K. (2009). Examining the Scientific Consensus on Climate Change, *Eos, Transactions American Geophysical Union*, **90**(3), 22–23. doi:10.1029/2009EO030002

[22] Anderegg, W. R. L., Prall, J. W., Harold, J., Schneider, S. H. (2010). Expert credibility in climate change, *Proceedings of the National Academy of Sciences*, **107**(27), 12107–12109. doi:10.1073/pnas.1003187107

[23] NASA. (2015). Climate Change: Vital Signs of the Planet: Evidence, *Global Climate Change*. Available at http://climate.nasa.gov/evidence/. Accessed 04/22/2015.

[24] EPA. (2014). *Climate Change Indicators in the United States, 2014 - Third Edition* (No. EPA 430-R-14-004). U.S. Environmental Protection Agency. Retrieved from http://www.epa.gov/climatechange/pdfs/climateindicators-full-2014.pdf

[25] Core Writing Team, Pachaur, R. K., Pachaur, R. K. (2007). *Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (p. 104). Geneva, Switzerland: IPCC. Retrieved from http://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr_spm.pdf

[26] EWEA. (2015). *Wind in Power: 2014 European Statistics*. Retrieved from http://www.ewea.org/fileadmin/files/library/publications/statistics/EWEA-Annual-Statistics-2014.pdf

[27] World Energy Council. (2015). *2015 World Energy Issues Monitor - Energy price volatility: the new normal*. United Kingdom: World Energy Council. Retrieved

from http://www.worldenergy.org/wp-content/uploads/2015/01/2015-World-Energy-Issues-Monitor.pdf

[28] BP - Biritish Petrolium. (2016). BP Energy Outlook - 2016 edition. Available at https://www.bp.com/content/dam/bp/pdf/energy-economics/energy-outlook-2016/bp-energy-outlook-2016.pdf. Accessed 01/18/2017.

[29] GWEC. (2016). *Global Wind Report - Annual Market Update 2015*. Retrieved from http://www.gwec.net/wp-content/uploads/vip/GWEC-Global-Wind-2015-Report_April-2016_19_04.pdf

[30] World Energy Council. (2016). *World Energy Resources 2016*. World Energy Council. Retrieved from https://www.worldenergy.org/wp-content/uploads/2016/10/World-Energy-Resources_FullReport_2016.pdf

[31] European Commission. (2016). *EU Energy in Figures Statistical Pocketbook*. Belgium. Retrieved from https://ec.europa.eu/energy/sites/ener/files/documents/pocketbook_energy-2016_web-final_final.pdf

[32] European Commission. (2014). *Energy prices and costs report* (No. SWD(2014) 20 final/2). Brussels. Retrieved from http://ec.europa.eu/energy/sites/ener/files/documents/20140122_swd_prices.pdf

[33] World Energy Council. (2016). *2016 World Energy Issues Monitor*. United Kingdom: World Energy Council. Retrieved from https://www.worldenergy.org/wp-content/uploads/2016/03/2016-World-Energy-Issues-Monitor-Full-report.pdf

[34] Turkish Wind Energy Association. (2017). *Turkish Wind Energy Statistics Report - January 2017*. Retrieved from http://www.tureb.com.tr/files/tureb_sayfa/duyurular/2017_duyurular/subat/turkiye_ruzgar_enerjisi_istatistik_raporu_ocak_2017.pdf

[35] Deloitte Danışmanlık A.Ş., Elia Grid International. (2016). *Turkish Wind Energy Association Wind Energy and Interaction Report / (in Turkish: Türkiye Rüzgar Enerjisi Birliği Rüzgar Enerjisi ve Etkileşim Raporu)*. Retrieved from

http://www.tureb.com.tr/files/yayinlar/tureb_ruzgar_enerjisi_ve_etkilesim_raporu.pdf

[36] Turkish Wind Energy Association. (2015). *Turkish Wind Energy Statistics Report*. Retrieved from http://www.tureb.com.tr/attachments/article/420/T%C3%BCrkiye%20R%C3%BCzgar%20Enerjisi%20%C4%B0statistik%20Raporu_Ocak%202015.pdf

[37] Republic of Turkey Ministry of Foreign Affairs. (2015). From Rep. of Turkey Ministry of Foreign Affairs. *Republic of Turkey Ministry of Foreign Affairs*. Available at http://www.mfa.gov.tr/turkeys-energy-strategy.en.mfa. Accessed 04/28/2015.

[38] Energy Market Regulatory Authority. (2011). *Electricity Market Report 2010*. Retrieved from http://www.emra.org.tr/documents/electricity/publishments/ElectricityMarketReport2010.pdf

[39] Turkish Electricity Transmission Company. (2014). *Five Year Projections of Electricity Production Capacity of Turkey 2014 - 2018*. Retrieved from http://www.epdk.org.tr/documents/elektrik/rapor_yayin/UretimKapasiteProjeksiyonu_2014_2018.pdf

[40] Eroğlu, Y., Seçkiner, S. U. (2013). Text Mining for Systematic Literature Reviev on Wind Energy. In *The Second European Workshop on Renewable Energy System (EWRES)*. Presented at the The Second European Workshop on Renewable Energy Systems, Antalya, Turkey.

[41] Eroğlu, Y., Seçkiner, S. U. (2016). Trend Topic Analysis for Wind Energy Researches: A Data Mining Approach Using Text Mining, *Journal of Technology Innovations in Renewable Energy*, **5**(2), 44–58.

[42] Yunus Eroğlu, Serap U. Seçkiner. (2016). Clustering of Wind Turbines in a Single Site Using SCADA Data. In *International Conference on Advanced Technology and Science* (p. 60). Presented at the International Conference on Advanced Technology and Science, Konya/Turkey.

[43] Eroğlu, Y., Seçkiner, S. U. (2017). Performance analysis in wind farms by data envelopment analysis and Malmquist Index approaches, *Pamukkale University Journal of Engineering Sciences*, **23**(1), 45–54. doi:10.5505/pajes.2016.59023

[44] Serap U. Seçkiner, Harika Akalın, Yunus Eroğlu. (2016). Performance Analysis of Wind Turbines, In *28th European Conference on Operational Research* (p. 160). Presented at the 28th European Conference on Operational Research, Poznan / Poland.

[45] Harika Akalın, Serap U. Seçkiner, Yunus Eroğlu. (2017). Performance evaluation for wind turbines using stochastic frontier analysis, *Journal of the Faculty of Engineering and Architecture of Gazi University*, **Article in Press**.

[46] Yunus Eroğlu, Serap U. Seçkiner. (2016). Modelling of Wind Speed and Generated Power of a Wind Turbine by Using Particle Filtering to Monitor a Wind Farm. In *28th European Conference on Operational Research* (p. 160). Presented at the 28th European Conference on Operational Research, Poznan / Poland.

[47] Seçkiner, S. U., Eroğlu, Y., Emrullah, M., Dereli, T. (2013). Ant colony optimization for continuous functions by using novel pheromone updating, *Applied Mathematics and Computation*, **219**(9), 4163–4175. doi:10.1016/j.amc.2012.10.097

[48] Boyle, G. (2004). *Renewable Energy: Power for a Sustainable Future*. Oxford University Press.

[49] Munteanu, I., Bratcu, A. I., CeangĂ, E., Cutululis, N.-A. (2008). *Optimal Control of Wind Energy Systems : Towards a Global Approach*. London: Springer-Verlag London Limited.

[50] Fanchi, J. R., Fanchi, C. J. (2013). *Energy in the 21st century* (3. ed.). Singapore: World Scientific.

[51] Da Rosa, A. V. (2013). *Fundamentals of renewable energy processes* (3. ed.). Amsterdam: Elsevier.

[52] TelosNet. (2015). Early History Through 1875. Available at http://ww.telosnet.com/wind/early.html. Accessed 05/06/2015.

[53] Shepherd, D. G. (1990). *Historical Development of the Windmill* (No. 4337 DOE/NASA/5266-1). NASA. Retrieved from http://wind.nrel.gov/public/library/shepherd.pdf

[54] Matthews, J. N. A. (2009). Superconductors to boost wind power, *Physics Today*, **62**(4), 25–26. doi:10.1063/1.3120924

[55] Parfit, M. (2005). Future Power: Where will the world get its next energy fix, *National Geographic Magazine*, **208**(2), 2–31.

[56] Wikipedia. (2015, May 6). Turbine. In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Turbineoldid=661037582

[57] Lohrmann, D. (1995). Von der östlichen zur westlichen Windmühle, *Archiv für Kulturgeschichte*, **77**(1), 1–32. doi:10.7788/akg.1995.77.1.1

[58] al-Hassan, A. Y., Hill, D. R. (1992). *Islamic Technology: An Illustrated History* (1 edition.). Cambridge: Cambridge University Press.

[59] Wagner, H.-J., Mathur, J. (2009). *Introduction to Wind Energy Systems : Basics, Technology and Operation*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

[60] Redlinger, R. Y., Andersen, P. D., Morthorst, P. E. (2011). *Wind Energy in the 21st Century: Economics, Policy, Technology, and the Changing Electricity Industry*. Palgrave Macmillan. Retrieved from http://www.palgrave.com%2Fpage%2Fdetail%2FWind-Energy-in-the-21st-Century%2F%3FK%3D9780333792483

[61] Danish Wind Industry Association. (2013, July 23). A Turbine Pioneer: Charles F. Brush. Available at http://www.motiva.fi/myllarin_tuulivoima/windpower%20web/en/pictures/brush.htm . Accessed 05/06/2015.

[62] Ezell, B. C., Haimes, Y. Y., Lambert, J. H. (2001). Risks of Cyber Attack to Water Utility Supervisory Control and Data Acquisition Systems, *Military Operations Research*, **6**(2), 23–33. doi:10.5711/morj.6.2.23

[63] Cai, N., Wang, J., Yu, X. (2008). SCADA system security: Complexity, history and new developments. In *6th IEEE International Conference on Industrial Informatics, 2008. INDIN 2008* (pp. 569–574). Presented at the 6th IEEE International Conference on Industrial Informatics, 2008. INDIN 2008. doi:10.1109/INDIN.2008.4618165

[64] Bergan, C. (2011, August 1). Demystifying Satellite for the Smart Grid: Four Common Misconceptions, *Electric Light Power*, **16**(8). Retrieved from http://www.elp.com/articles/powergrid_international/print/volume-16/issue-8/features/demystifying-satellite-for-the-smart-grid-four-common-misconceptions.html

[65] Boyes, W. (2009). *Instrumentation Reference Book*. Butterworth-Heinemann.

[66] Csanyi, E. (2013, April 22). Three generations of SCADA system architectures, *Electrical Engineering Portal*. Available at http://electrical-engineering-portal.com/three-generations-of-scada-system-architectures. Accessed 05/07/2015.

[67] Bachmann. (2014). *Scada for wind 2014 The growing sophistication of data-gathering software* (Expert Report). Retrieved from http://www.bachmann.info/fileadmin/media/Unternehmen/News/Pressespiegel/2014/Fach-_und_Applikationsbeitr%C3%A4ge/SCADA_2014_Bachmann_Expert_Report.pdf

[68] Nicolae, C. (2011). SCADA - Supervisory Control And Data Acquisition, *Journal of Sustainable Energy*, **2**(4). Retrieved from http://www.energy-cie.ro/archives/2011/nr_4/v2-n4-22.pdf

[69] DEIF - Wind Power Technology. (2015). SCADA Systems for remote/ monitoring turbine control systems. Available at http://www.deifwindpower.com/wind-turbine-solutions/scada-systems. Accessed 05/07/2015.

[70] Kusiak, A., Zheng, H., Song, Z. (2010). Power optimization of wind turbines with data mining and evolutionary computation, *Renewable Energy*, **35**(3), 695–702. doi:10.1016/j.renene.2009.08.018

[71] Kusiak, A., Li, W., Song, Z. (2010). Dynamic control of wind turbines, *Renewable Energy*, **35**(2), 456–463. doi:10.1016/j.renene.2009.05.022

[72] Kusiak, A., Verma, A. (2012). A Data-Mining Approach to Monitoring Wind Turbines, *IEEE Transactions on Sustainable Energy*, **3**(1), 150–157. doi:10.1109/TSTE.2011.2163177

[73] Schlechtingen, M., Santos, I. F., Achiche, S. (2013). Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description, *Applied Soft Computing*, **13**(1), 259–270. doi:10.1016/j.asoc.2012.08.033

[74] Ananiadou, S., Rea, B., Okazaki, N., Procter, R., Thomas, J. (2009). Supporting Systematic Reviews Using Text Mining, *Social Science Computer Review*, **27**(4), 509–523. doi:10.1177/0894439309332293

[75] Chalmers, I. (2003). Trying to do more Good than Harm in Policy and Practice: The Role of Rigorous, Transparent, Up-to-Date Evaluations, *The ANNALS of the American Academy of Political and Social Science*, **589**(1), 22–40. doi:10.1177/0002716203254762

[76] Kostoff, R. N., Tshiteya, R., Pfeil, K. M., Humenik, J. A., Karypis, G. (2004). *Science and Technology Text Mining: Electric Power Sources*.

[77] Kostoff, R. N., Buchtel, H. A., Andrews, J., Pfeil, K. M. (2005). The Hidden Structure of Neuropsychology: Text Mining of the Journal Cortex: 1991-2001, *Cortex*, **41**(2), 103–115. doi:10.1016/S0010-9452(08)70885-2

[78] Miner, G., Elder, J., Hill, T., Nisbet, R., Delen, D., Fast, A. (2012). *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications* (1st ed.). Academic Press.

[79] Witten, I. H. (2004). Adaptive text mining: inferring structure from sequences, *Journal of Discrete Algorithms*, **2**(2), 137–159. doi:10.1016/S1570-8667(03)00084-4

[80] Hearst, M. A. (1999). Untangling Text Data Mining. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational*

Linguistics (pp. 3–10). Stroudsburg, PA, USA: Association for Computational Linguistics. doi:10.3115/1034678.1034679

[81] Losiewicz, P., Oard, D. W., Kostoff, R. N. (2000). Textual Data Mining to Support Science and Technology Management, *Journal of Intelligent Information Systems*, **15**(2), 99–119. doi:10.1023/A:1008777222412

[82] Zhu, D., Porter, A. L. (2002). Automated extraction and visualization of information for technological intelligence and forecasting, *Technological Forecasting and Social Change*, **69**(5), 495–506. doi:10.1016/S0040-1625(01)00157-3

[83] Drake, M. (2003). *Encyclopedia of Library and Information Science, Second Edition* -. CRC Press.

[84] Ghazinoory, S., Ameri, F., Farnoodi, S. (2013). An application of the text mining approach to select technology centers of excellence, *Technological Forecasting and Social Change*, **80**(5), 918–931. doi:10.1016/j.techfore.2012.09.001

[85] Jun, S., Park, S.-S., Jang, D.-S. (2014). Document clustering method using dimension reduction and support vector clustering to overcome sparseness, *Expert Systems with Applications*, **41**(7), 3204–3212. doi:10.1016/j.eswa.2013.11.018

[86] Fattori, M., Pedrazzi, G., Turra, R. (2003). Text mining applied to patent mapping: a practical business case, *World Patent Information*, **25**(4), 335–342. doi:10.1016/S0172-2190(03)00113-3

[87] Tseng, Y.-H., Lin, C.-J., Lin, Y.-I. (2007). Text mining techniques for patent analysis, *Information Processing Management*, **43**(5), 1216–1247. doi:10.1016/j.ipm.2006.11.011

[88] Liu, J. S., Kuan, C.-H., Cha, S.-C., Chuang, W.-L., Gau, G. J., Jeng, J.-Y. (2011). Photovoltaic technology development: A perspective from patent growth analysis, *Solar Energy Materials and Solar Cells*, **95**(11), 3130–3136. doi:10.1016/j.solmat.2011.07.002

[89] Daim, T., Iskin, I., Li, X., Zielsdorff, C., Bayraktaroglu, A. E., Dereli, T., Durmusoglu, A. (2012). Patent analysis of wind energy technology using the patent alert system, *World Patent Information*, **34**(1), 37–47. doi:10.1016/j.wpi.2011.11.001

[90] Kostoff, R. N., Eberhart, H. J., Toothman, D. R. (1997). Database tomography for information retrieval, *Journal of Information Science*, **23**(4), 301–311. doi:10.1177/016555159702300404

[91] Greengrass, E. (2000). *Information Retrieval: A Survey*. University of Maryland. Retrieved from http://www.csee.umbc.edu/csee/research/cadip/readings/IR.report.120600.book.pdf

[92] Swanson, D. R., Smalheiser, N. R. (1997). An interactive system for finding complementary literatures: a stimulus to scientific discovery, *Artificial Intelligence*, **91**(2), 183–203. doi:10.1016/S0004-3702(97)00008-8

[93] Shavinina, L. V. (2003). *The International Handbook on Innovation*. Elsevier.

[94] Ja, G., Ww, C., Ds, P., Rm, G. (1999). Term domain distribution analysis: a data mining tool for text databases, *Methods of information in medicine*, **38**(2), 96–101.

[95] Kostoff, R. N. (2003). Bilateral asymmetry prediction, *Medical Hypotheses*, **61**(2), 265–266. doi:10.1016/S0306-9877(03)00167-1

[96] Kostoff, R. N., Green, K. A., Toothman, D. R., Humenik, J. A. (2000). Database Tomography Applied to an Aircraft Science and Technology Investment Strategy, *Journal of Aircraft*, **37**(4), 727–730. doi:10.2514/2.2659

[97] Viator, J. A., Pestorius, F. M. (2001). Investigating trends in acoustics research from 1970–1999, *The Journal of the Acoustical Society of America*, **109**(5), 1779–1783. doi:10.1121/1.1366711

[98] Kostoff, R. N., Shlesinger, M. F., Malpohl, G. (2004). Fractals text mining using bibliometrics and database tomography, *Fractals*, **12**(01), 1–16. doi:10.1142/S0218348X04002343

[99] Kostoff, R. N., Shlesinger, M. F., Tshiteya, R. (2004). Nonlinear dynamics text mining using bibliometrics and database tomography, *International Journal of Bifurcation and Chaos*, **14**(01), 61–92. doi:10.1142/S0218127404009089

[100] Huang, C.-J., Liao, J.-J., Yang, D.-X., Chang, T.-Y., Luo, Y.-C. (2010). Realization of a news dissemination agent based on weighted association rules and text mining techniques, *Expert Systems with Applications*, **37**(9), 6409–6413. doi:10.1016/j.eswa.2010.02.078

[101] Barbier, G., Liu, H. (2011). Data Mining in Social Media. In C. C. Aggarwal (Ed.), *Social Network Data Analytics* (pp. 327–352). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-8462-3_12

[102] Pang, B., Lee, L. (2008). Opinion Mining and Sentiment Analysis, *Found. Trends Inf. Retr.*, **2**(1–2), 1–135. doi:10.1561/1500000011

[103] Abdous, M., Wu He, Cherng-Jyh Yen. (2012). Using Data Mining for Predicting Relationships between Online Question Theme and Final Grade, *Journal of Educational Technology Society*, **15**(3), 77–88.

[104] He, W. (2013). Improving user experience with case-based reasoning systems using text mining and Web 2.0, *Expert Systems with Applications*, **40**(2), 500–507. doi:10.1016/j.eswa.2012.07.070

[105] He, W., Zha, S., Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry, *International Journal of Information Management*, **33**(3), 464–472. doi:10.1016/j.ijinfomgt.2013.01.001

[106] Kostoff, R. N., del Río, J. A., Humenik, J. A., García, E. O., Ramírez, A. M. (2001). Citation mining: Integrating text mining and bibliometrics for research user profiling, *Journal of the American Society for Information Science and Technology*, **52**(13), 1148–1156. doi:10.1002/asi.1181

[107] Kostoff, R. N., Tshiteya, R., Pfeil, K. M., Humenik, J. A. (2002). Electrochemical power text mining using bibliometrics and database tomography, *Journal of Power Sources*, **110**(1), 163–176. doi:10.1016/S0378-7753(02)00233-1

[108] Kongthon, A. (2004, April). *A Text Mining Framework for Discovering Technological Intelligence to Support Science and Technology Management* (Ph.D.). Georgia Institute of Technology. Retrieved from http://202.28.199.34/multim/3126708.pdf

[109] Kostoff, R. N., Tshiteya, R., Pfeil, K. M., Humenik, J. A., Karypis, G. (2005). Power source roadmaps using bibliometrics and database tomography, *Energy*, **30**(5), 709–730. doi:10.1016/j.energy.2004.04.058

[110] de Miranda Santo, M., Coelho, G. M., dos Santos, D. M., Filho, L. F. (2006). Text mining as a valuable tool in foresight exercises: A study on nanotechnology, *Technological Forecasting and Social Change*, **73**(8), 1013–1027. doi:10.1016/j.techfore.2006.05.020

[111] Kostoff, R. N., Koytcheff, R. G., Lau, C. G. Y. (2007). Global nanotechnology research literature overview, *Technological Forecasting and Social Change*, **74**(9), 1733–1747. doi:10.1016/j.techfore.2007.04.004

[112] Malheiros, V., Hohn, E., Pinho, R., Mendonca, M. (2007). A Visual Text Mining approach for Systematic Reviews. In *First International Symposium on Empirical Software Engineering and Measurement, 2007. ESEM 2007* (pp. 245–254). Presented at the First International Symposium on Empirical Software Engineering and Measurement, 2007. ESEM 2007. doi:10.1109/ESEM.2007.21

[113] Delen, D., Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining, *Expert Systems with Applications*, **34**(3), 1707–1720. doi:10.1016/j.eswa.2007.01.035

[114] Kajikawa, Y., Yoshikawa, J., Takeda, Y., Matsushima, K. (2008). Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy, *Technological Forecasting and Social Change*, **75**(6), 771–782. doi:10.1016/j.techfore.2007.05.005

[115] Kim, H., Lee, J. Y. (2008). Exploring the emerging intellectual structure of archival studies using text mining: 2001—2004.*Journal of Information Science*, **34**(3), 356–369. doi:10.1177/0165551507086260

[116] Kostoff, R. N. (2008). Literature-Related Discovery (LRD): Introduction and background, *Technological Forecasting and Social Change*, **75**(2), 165–185. doi:10.1016/j.techfore.2007.11.004

[117] Kostoff, R. N., Briggs, M. B., Solka, J. L., Rushenberg, R. L. (2008). Literature-related discovery (LRD): Methodology, *Technological Forecasting and Social Change*, **75**(2), 186–202. doi:10.1016/j.techfore.2007.11.010

[118] Thomas, J., McNaught, J., Ananiadou, S. (2011). Applications of text mining within systematic reviews, *Research Synthesis Methods*, **2**(1), 1–14. doi:10.1002/jrsm.27

[119] Kostoff, R. N. (2012). Literature-related discovery and innovation — update, *Technological Forecasting and Social Change*, **79**(4), 789–800. doi:10.1016/j.techfore.2012.02.002

[120] Tu, Y.-N., Seng, J.-L. (2012). Indices of novelty for emerging topic detection, *Information Processing Management*, **48**(2), 303–325. doi:10.1016/j.ipm.2011.07.006

[121] Küçük, D., Arslan, Y. (2014). Semi-automatic construction of a domain ontology for wind energy using Wikipedia articles, *Renewable Energy*, **62**, 484–489. doi:10.1016/j.renene.2013.08.002

[122] Yoon, J. (2012). Detecting weak signals for long-term business opportunities using text mining of Web news, *Expert Systems with Applications*, **39**(16), 12543–12550. doi:10.1016/j.eswa.2012.04.059

[123] Miller, T. W. (2005). *Data and text mining: a business applications approach*. Upper Saddle River, N.J.: Pearson Prentice Hall.

[124] Romero, C., Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005.*Expert Systems with Applications*, **33**(1), 135–146. doi:10.1016/j.eswa.2006.04.005

[125] Walford, C. A. (2006). *Wind Turbine Reliability: Understanding and Minimizing Wind Turbine Operation and Maintenance Costs* (No. SAND2006-

1100). Sandia National Laboratories. Retrieved from http://prod.sandia.gov/techlib/access-control.cgi/2006/061100.pdf

[126] McMillan, D., Ault, G. (2007). Quantification of Condition Monitoring Benefit for Offshore Wind Turbines, *Wind Engineering*, **31**(4), 267–285. doi:10.1260/030952407783123060

[127] Elías N. Montoya, Ismael P. Gómez. (2016). Guidelines for developing regulations for acoustic impact, based on the stage of operation of wind farms in Chile, *Proceedings of Meetings on Acoustics*, **26**(1), 040001. doi:10.1121/2.0000229

[128] Zajamšek, B., Hansen, K. L., Doolan, C. J., Hansen, C. H. (2016). Characterisation of wind farm infrasound and low-frequency noise, *Journal of Sound and Vibration*, **370**, 176–190. doi:10.1016/j.jsv.2016.02.001

[129] Arthur E. Newhall, Ying T. Lin, James F. Miller, Gopu R. Potty, Kathy Vigness-Raposa, Adam Frankel, … Tim Mason. (2016). Monitoring the acoustic effects of pile driving for the first offshore wind farm in the United States, *The Journal of the Acoustical Society of America*, **139**(4), 2181–2181. doi:10.1121/1.4950483

[130] Lingmei, W., Enlong, M., Jianlin, S., Dongjie, G., Shaoping, Y., Jinxiu, J., Yuehan, Y. (2017). State Monitoring and Fault Diagnosis of Wind Turbines. In G. M. Kopanos, P. Liu, M. C. Georgiadis (Eds.), *Advances in Energy Systems Engineering* (pp. 803–822). Springer International Publishing. doi:10.1007/978-3-319-42803-1_27

[131] Ma, X., Cross, P., Qian, P. (2016). A condition monitoring system for an early warning of developing faults in wind turbine electrical systems, *Insight - Non-Destructive Testing and Condition Monitoring*, **58**(12), 663–670. doi:10.1784/insi.2016.58.12.663

[132] Cablea, G., Granjon, P., Bérenguer, C., Bellemain, P. (2016). Online condition monitoring of wind turbines through three-phase electrical signature analysis, *International Journal of Condition Monitoring*, **6**(3), 66–72. doi:10.1784/204764216819708087

[133] Qiu, Y., Feng, Y., Sun, J., Zhang, W., Infield, D. (2016). Applying thermophysics for wind turbine drivetrain fault diagnosis using SCADA data, *IET Renewable Power Generation*, **10**(5), 661–668. doi:10.1049/iet-rpg.2015.0160

[134] Kulkarni, S. H., Anil, T. R., Gowdar, R. D. (2016). Wind Energy Development in India and a Methodology for Evaluating Performance of Wind Farm Clusters, *Journal of Renewable Energy*, **2016**, e6769405. doi:10.1155/2016/6769405

[135] Ravi, B., Raval, S., Rudraraju, V. R. R., Nagamani, C. (2016). Performance analysis of MVAC and MVDC offshore wind farm distribution system using direct load flow method. In *2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)* (pp. 341–346). Presented at the 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB). doi:10.1109/AEEICB.2016.7538305

[136] Yang, R., He, Y., Zhang, H. (2016). Progress and trends in nondestructive testing and evaluation for wind turbine composite blade, *Renewable and Sustainable Energy Reviews*, **60**, 1225–1250. doi:10.1016/j.rser.2016.02.026

[137] Yang, W., Peng, Z., Wei, K., Tian, W. (2016). Structural health monitoring of composite wind turbine blades: challenges, issues and potential solutions, *IET Renewable Power Generation*. doi:10.1049/iet-rpg.2016.0087

[138] Bossuyt, J., Howland, M., Meneveau, C., Meyers, J. (n.d.). Measuring power output intermittency and unsteady loading in a micro wind farm model. In *34th Wind Energy Symposium*. American Institute of Aeronautics and Astronautics. doi:10.2514/6.2016-1992

[139] Gómez Muñoz, C. Q., García Márquez, F. P., Sánchez Tomás, J. M. (2016). Ice detection using thermal infrared radiometry on wind turbine blades, *Measurement*, **93**, 157–163. doi:10.1016/j.measurement.2016.06.064

[140] Gao, B., He, Y., Woo, W. L., Tian, G. Y., Liu, J., Hu, Y. (2016). Multidimensional Tensor-Based Inductive Thermography With Multiple Physical

Fields for Offshore Wind Turbine Gear Inspection, *IEEE Transactions on Industrial Electronics*, **63**(10), 6305–6315. doi:10.1109/TIE.2016.2574987

[141] Huang, M., Lu, H., Wang, B., Lu, Y. (2016). Research on Testing the Ultrasonic Wind Sensors in Circuit Wind Tunnel. In *Proceedings of the 2015 International Conference on Communications, Signal Processing, and Systems* (pp. 859–868). Springer, Berlin, Heidelberg. doi:10.1007/978-3-662-49831-6_89

[142] Shin, D., Ko, K. (2017). Comparative analysis of degradation rates for inland and seaside wind turbines in compliance with the International Electrotechnical Commission standard, *Energy*, **118**, 1180–1186. doi:10.1016/j.energy.2016.10.140

[143] Kikuchi, Y., Ishihara, T. (2016). Assessment of weather downtime for the construction of offshore wind farm by using wind and wave simulations, *Journal of Physics: Conference Series*, **753**(9), 092016. doi:10.1088/1742-6596/753/9/092016

[144] Iliopoulos, A. N., Van Hemelrijck, D., Vlassenbroeck, J., Aggelis, D. G. (2016). Assessment of grouted samples from monopile wind turbine foundations using combined non-destructive techniques, *Construction and Building Materials*, **122**, 855–862. doi:10.1016/j.conbuildmat.2015.11.047

[145] Chen, J., Pan, J., Li, Z., Zi, Y., Chen, X. (2016). Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals, *Renewable Energy*, **89**, 80–92. doi:10.1016/j.renene.2015.12.010

[146] He, G., Ding, K., Li, W., Jiao, X. (2016). A novel order tracking method for wind turbine planetary gearbox vibration analysis based on discrete spectrum correction technique, *Renewable Energy*, **87**, **Part 1**, 364–375. doi:10.1016/j.renene.2015.10.036

[147] de da González-Carrato, R. R. H., Márquez, G., Pedro, F., Papaelias, M. (2016). Vibration-based tools for the optimisation of large-scale industrial wind turbine devices, *International Journal of Condition Monitoring*, **6**(2), 33–37.

[148] Cambron, P., Masson, C., Tahan, A., Pelletier, F. (n.d.). Control chart monitoring of wind turbine generators using the statistical inertia of a wind farm average, *Renewable Energy*. doi:10.1016/j.renene.2016.09.029

[149] Castellani, F., D'Elia, G., Astolfi, D., Mucchi, E., Giorgio, D., Terzi, L. (2016). Analyzing wind turbine flow interaction through vibration data, *Journal of Physics: Conference Series*, **753**(11), 112008. doi:10.1088/1742-6596/753/11/112008

[150]Richard, A. (2016). Engineering Foundation Characterisation of Wind Farm Site Using Integrated Data Approach - Clyde Perspective. Presented at the Near Surface Geoscience 2016 - 22nd European Meeting of Environmental and Engineering Geophysics. doi:10.3997/2214-4609.201601981

[151] Kennedy, K., Walsh, P., Mastaglio, T. W., Scully, T. (2016). Genetic optimisation for a stochastic model for opportunistic maintenance planning of offshore wind farms. In *2016 4th International Symposium on Environmental Friendly Energies and Applications (EFEA)* (pp. 1–6). Presented at the 2016 4th International Symposium on Environmental Friendly Energies and Applications (EFEA). doi:10.1109/EFEA.2016.7748820

[152]Bigdeli, N., Sadegh Lafmejani, H. (2016). Dynamic characterization and predictability analysis of wind speed and wind power time series in Spain wind farm, *Journal of AI and Data Mining*, **4**(1), 103–116. doi:10.5829/idosi.JAIDM.2016.04.01.12

[153]Schedat, M., Faber, T., Sivanesan, A. (2016). Structural health monitoring concept to predict the remaining lifetime of the wind turbine structure. In *2016 International Conference on the Domestic Use of Energy (DUE)* (pp. 1–5). Presented at the 2016 International Conference on the Domestic Use of Energy (DUE). doi:10.1109/DUE.2016.7466723

[154] Islam, M. M., Hossain, M. R., Dougal, R. A., Brice, C. W. (2016). Analysis of real-world power quality disturbances employing time-frequency distribution. In *2016 Clemson University Power Systems Conference (PSC)* (pp. 1–5). Presented at

the 2016 Clemson University Power Systems Conference (PSC). doi:10.1109/PSC.2016.7462816

[155] Stone, N. J. (2016). A Synthesis of Approaches for Wind Turbine Amplitude Modulation Data Analysis. Retrieved from http://conservancy.umn.edu/handle/11299/180873

[156] Vanfretti, L., Baudette, M., Domínguez-García, J.-L., Almas, M. S., White, A., Gjerde, J. O. (2016). A Phasor Measurement Unit Based Fast Real-time Oscillation Detection Application for Monitoring Wind-farm-to-grid Sub–synchronous Dynamics, *Electric Power Components and Systems*, **44**(2), 123–134. doi:10.1080/15325008.2015.1101727

[157] Zhang, X., Thomas, J. J., Grijalva, S. (2016). Fast evaluation of probabilistic total transfer capability considering multiple wind farms. In *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T D)* (pp. 1–5). Presented at the 2016 IEEE/PES Transmission and Distribution Conference and Exposition (T D). doi:10.1109/TDC.2016.7520074

[158] Yuzhang Che, Xindong Peng, Luca Delle Monache, Takayuki Kawaguchi, Feng Xiao. (2016). A wind power forecasting system based on the weather research and forecasting model and Kalman filtering over a wind-farm in Japan, *Journal of Renewable and Sustainable Energy*, **8**(1), 013302. doi:10.1063/1.4940208

[159] Zhong, S., Wang, X. (2016). Decentralized Model-Free Wind Farm Control Via Discrete Adaptive Filtering Methods, *IEEE Transactions on Smart Grid*, **PP**(99), 1–1. doi:10.1109/TSG.2016.2614434

[160] Doekemeijer, B. M., Wingerden, J. W. van, Boersma, S., Pao, L. Y. (2016). Enhanced Kalman Filtering for a 2D CFD NS Wind Farm Flow Model, *Journal of Physics: Conference Series*, **753**(5), 052015. doi:10.1088/1742-6596/753/5/052015

[161] Simani, S., Castaldi, P., Bonfe, M. (2016). Adaptive nonlinear filters for joint fault estimation and accommodation of a wind farm benchmark. In *2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol)* (pp. 642–647).

Presented at the 2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol). doi:10.1109/SYSTOL.2016.7739821

[162] Cadenas, E., Rivera, W., Campos-Amezcua, R., Cadenas, R. (2016). Wind speed forecasting using the NARX model, case: La Mata, Oaxaca, México, *Neural Computing and Applications*, **27**(8), 2417–2428. doi:10.1007/s00521-015-2012-y

[163] Jiang, Z., Liu, C., Akintayo, A., Henze, G., Sarkar, S. (2017). Energy Prediction using Spatiotemporal Pattern Networks.*arXiv:1702.01125 [stat]*. Retrieved from http://arxiv.org/abs/1702.01125

[164] Yan, J. (2014). Health Monitoring and Prognosis. In *Machinery Prognostics and Prognosis Oriented Maintenance Management* (pp. 101–152). John Wiley Sons Singapore Pte. Ltd. doi:10.1002/9781118638743.ch4

[165] Al-Shammari, E. T., Shamshirband, S., Petković, D., Zalnezhad, E., Yee, P. L., Taher, R. S., Ćojbašić, Ž. (2016). Comparative study of clustering methods for wake effect analysis in wind farm, *Energy*, **95**, 573–579. doi:10.1016/j.energy.2015.11.064

[166] Volker, P., Badger, Jake, Hahmann, Andrea N. (2016). Dynamical and statistical-dynamical modelling of wind farm flows with WRF. In *Ems Annual Meeting Abstracts*. Presented at the 16th EMS Annual Meeting 11th European Conference on Applied Climatology.

[167] Farajzadeh, S., Ramezani, M. H., Nielsen, P., Nadimi, E. S. (2017). Statistical modeling of the power grid from a wind farm standpoint, *Electric Power Systems Research*, **144**, 150–156. doi:10.1016/j.epsr.2016.11.020

[168] Zhang, Z., Verma, A., Kusiak, A. (2012). Fault Analysis and Condition Monitoring of the Wind Turbine Gearbox, *IEEE Transactions on Energy Conversion*, **27**(2), 526–535. doi:10.1109/TEC.2012.2189887

[169] Aghajani, A., Kazemzadeh, R., Ebrahimi, A. (2016). A novel hybrid approach for predicting wind farm power production based on wavelet transform, hybrid neural networks and imperialist competitive algorithm, *Energy Conversion and Management*, **121**, 232–240. doi:10.1016/j.enconman.2016.05.024

[170] Shao, H., Wei, H., Deng, X., Xing, S. (2016). Short-term wind speed forecasting using wavelet transformation and AdaBoosting neural networks in Yunnan wind farm, *IET Renewable Power Generation*. doi:10.1049/iet-rpg.2016.0118

[171] García Márquez, F. P., Tobias, A. M., Pinar Pérez, J. M., Papaelias, M. (2012). Condition monitoring of wind turbines: Techniques and methods, *Renewable Energy*, **46**, 169–178. doi:10.1016/j.renene.2012.03.003

[172] Tautz-Weinert, J., Watson, S. J. (2016). Using SCADA data for wind turbine condition monitoring – a review, *IET Renewable Power Generation*. doi:10.1049/iet-rpg.2016.0248

[173] Yang, W., Tavner, P. J., Crabtree, C. J., Feng, Y., Qiu, Y. (2014). Wind turbine condition monitoring: technical and commercial challenges, *Wind Energy*, **17**(5), 673–693. doi:10.1002/we.1508

[174] Kandukuri, S. T., Klausen, A., Karimi, H. R., Robbersmyr, K. G. (2016). A review of diagnostics and prognostics of low-speed machinery towards wind turbine farm-level health management, *Renewable and Sustainable Energy Reviews*, **53**, 697–708. doi:10.1016/j.rser.2015.08.061

[175] Zaher, A. S., McArthur, S. D. J. (2007). A Multi-Agent Fault Detection System for Wind Turbine Defect Recognition and Diagnosis. In *Power Tech, 2007 IEEE Lausanne* (pp. 22–27). Presented at the Power Tech, 2007 IEEE Lausanne. doi:10.1109/PCT.2007.4538286

[176] Wang, L., Zhang, Z., Xu, J., Liu, R. (2016). Wind Turbine Blade Breakage Monitoring with Deep Autoencoders, *IEEE Transactions on Smart Grid*, **PP**(99), 1–1. doi:10.1109/TSG.2016.2621135

[177] Liu, T. (2016). Fault diagnosis of gearbox by selective ensemble learning based on artificial immune algorithm. In *2016 3rd International Conference on Systems and Informatics (ICSAI)* (pp. 460–464). Presented at the 2016 3rd International Conference on Systems and Informatics (ICSAI). doi:10.1109/ICSAI.2016.7810999

[178] Sharma, S., Mahto, D. (2013). Condition Monitoring of Wind Turbines: A Review, *International Journal of Scientific Engineering Research*, **4**(8), 35–50.

[179] Yang, W., Tavner, P. J., Crabtree, C. (2009). An Intelligent Approach to the Condition Monitoring of Large Scale Wind Turbines. In *European Wind Energy Conference*. Marseille, France.

[180] Yang, W., Tian, S. W. (2015). Research on a power quality monitoring technique for individual wind turbines, *Renewable Energy*, **75**, 187–198. doi:10.1016/j.renene.2014.09.037

[181] McMillan, D., Ault, G. W. (2008). Condition monitoring benefit for onshore wind turbines: sensitivity to operational parameters, *IET Renewable Power Generation*, **2**(1), 60–72. doi:10.1049/iet-rpg:20070064

[182] Lu, B., Li, Y., Wu, X., Yang, Z. (2009). A review of recent advances in wind turbine condition monitoring and fault diagnosis. In *IEEE Power Electronics and Machines in Wind Applications, 2009. PEMWA 2009* (pp. 1–7). Presented at the IEEE Power Electronics and Machines in Wind Applications, 2009. PEMWA 2009. doi:10.1109/PEMWA.2009.5208325

[183] Schlechtingen, M., Santos, I. (2011). Condition monitoring with ordinary wind turbine SCADA data – A neuro-fuzzy approach. In *Offshore 2011*. Amsterdam, Netherlands.

[184] Schlechtingen, M., Santos, I. F. (2014). Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 2: Application examples, *Applied Soft Computing*, **14**, **Part C**, 447–460. doi:10.1016/j.asoc.2013.09.016

[185] Zhang, Z., Kusiak, A. (2012). Monitoring Wind Turbine Vibration Based on SCADA Data, *Journal of Solar Energy Engineering*, **134**(2), 021004–021004. doi:10.1115/1.4005753

[186] Kusiak, A., Verma, A. (2011). A Data-Driven Approach for Monitoring Blade Pitch Faults in Wind Turbines, *IEEE Transactions on Sustainable Energy*, **2**(1), 87–96. doi:10.1109/TSTE.2010.2066585

[187] Kusiak, A., Zheng, H., Song, Z. (2009). Models for monitoring wind farm power, *Renewable Energy*, **34**(3), 583–590. doi:10.1016/j.renene.2008.05.032

[188] Kusiak, A., Zheng, H., Song, Z. (2009). On-line monitoring of power curves, *Renewable Energy*, **34**(6), 1487–1493. doi:10.1016/j.renene.2008.10.022

[189] Kusiak, A., Zheng, H., Song, Z. (2009). Wind farm power prediction: a data-mining approach, *Wind Energy*, **12**(3), 275–293. doi:10.1002/we.295

[190] Wang, L., Singh, C., Kusiak, A. (2010). *Wind Power Systems : Applications of Computational Intelligence*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

[191] Gringras, P., Chen, W. (2001). Mechanisms for differences in monozygous twins, *Early Human Development*, **64**(2), 105–117. doi:10.1016/S0378-3782(01)00171-2

[192] Pamela Prindle Fierro. (2016, February 28). Twins That Don’t Look Alike - Why Identical Twins are Different, *Verywell*. Available at https://www.verywell.com/why-are-identical-twins-different-2447125. Accessed 05/16/2016.

[193]Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins, *Proceedings of the National Academy of Sciences of the United States of America*, **102**(30), 10604–10609. doi:10.1073/pnas.0500398102

[194] Skipper, M. (2008). Human genetics: Not-so-identical twins, *Nature Reviews Genetics*, **9**(4), 250–251. doi:10.1038/nrg2356

[195] Bruder, C. E. G., Piotrowski, A., Gijsbers, A. A. C. J., Andersson, R., Erickson, S., Diaz de Ståhl, T., … Dumanski, J. P. (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles, *American Journal of Human Genetics*, **82**(3), 763–771. doi:10.1016/j.ajhg.2007.12.011

[196] Matias, A., Silva, S., Martins, Y., Blickstein, I. (2014). Monozygotic twins: Ten reasons to be different, *Diagnóstico Prenatal*, **25**(2), 53–57. doi:10.1016/j.diapre.2013.09.003

[197] U.S. Department of Energy. (n.d.). Night-warming effect found over large wind farms in Texas, *ScienceDaily*. Available at https://www.sciencedaily.com/releases/2012/04/120430152045.htm. Accessed 03/03/2017.

[198] Harb, H., Makhoul, A., Couturier, R. (2015). An Enhanced K-Means and ANOVA-Based Clustering Approach for Similarity Aggregation in Underwater Wireless Sensor Networks, *IEEE Sensors Journal*, **15**(10), 5483–5493. doi:10.1109/JSEN.2015.2443380

[199]Lam, T. C., Ge, H., Fazio, P. (2016). Energy positive curtain wall configurations for a cold climate using the Analysis of Variance (ANOVA) approach, *Building Simulation*, **9**(3), 297–310. doi:10.1007/s12273-016-0275-6

[200] Karatepe, Y., Neşe, S. V., Keçebaş, A., Yumurtacı, M. (2012). The levels of awareness about the renewable energy sources of university students in Turkey, *Renewable Energy*, **44**, 174–179. doi:10.1016/j.renene.2012.01.099

[201] Jaber, J. O., Awad, W., Rahmeh, T. A., Alawin, A. A., Al-Lubani, S., Dalu, S. A., Al-Bashir, A. (2017). Renewable energy education in faculties of engineering in Jordan: Relationship between demographics and level of knowledge of senior students', *Renewable and Sustainable Energy Reviews*, **73**, 452–459. doi:10.1016/j.rser.2017.01.141

[202] Kevin B. Howard, Arvind Singh, Fotis Sotiropoulos, Michele Guala. (2015). On the statistics of wind turbine wake meandering: An experimental investigation, *Physics of Fluids*, **27**(7), 075103. doi:10.1063/1.4923334

[203] Padrón, A. S., Stanley, A. P. J., Thomas, J. J., Alonso, J. J., Ning, A. (2016). Polynomial chaos for the computation of annual energy production in wind farm layout optimization, *Journal of Physics: Conference Series*, **753**(3), 032021. doi:10.1088/1742-6596/753/3/032021

[204] Rai, R. K., Gopalan, H., Naughton, J. W. (2016). Effects of spatial and temporal resolution of the turbulent inflow on wind turbine performance estimation, *Wind Energy*, **19**(7), 1341–1354. doi:10.1002/we.1888

[205] Gibanica, M., Johansson, A. T., Rahrovani, S., Khorsand, M., Abrahamsson, T. (2014). Spread in Modal Data Obtained from Wind Turbine Blade Testing. In *Topics in Experimental Dynamic Substructuring, Volume 2* (pp. 207–215). Springer, New York, NY. doi:10.1007/978-1-4614-6540-9_17

[206] England, M., Sashi, N. (2015). The Effect Of Placing A Wind Turbine On Various Locations On The Backs Of A Honda Accord Hybrid In Order To Capture Wind Energy, *South Carolina Junior Academy of Science*. Retrieved from http://scholarexchange.furman.edu/scjas/2015/all/244

[207] A. M. Costa, F. Fraguela, J. A. Orosa, J. A. Pérez. (2015). A new procedure for wind energy systems maintenance design, *Journal of Renewable and Sustainable Energy*, **7**(4), 043129. doi:10.1063/1.4928872

[208] Yoon, J., He, D., Hecke, B. V. (2015). On the Use of a Single Piezoelectric Strain Sensor for Wind Turbine Planetary Gearbox Fault Diagnosis, *IEEE Transactions on Industrial Electronics*, **62**(10), 6585–6593. doi:10.1109/TIE.2015.2442216

[209] Tabatabaeikia, S., Ghazali, N. N. B. N., Chong, W. T., Shahizare, B., Izadyar, N., Esmaeilzadeh, A., Fazlizan, A. (2016). Computational and experimental optimization of the exhaust air energy recovery wind turbine generator, *Energy Conversion and Management*, **126**, 862–874. doi:10.1016/j.enconman.2016.08.039

[210] Abbasi, M., Monazzam, M. R., Ebrahimi, M. H., Zakerian, S. A., Dehghan, S. F., Akbarzadeh, A. (2016). Assessment of noise effects of wind turbine on the general health of staff at wind farm of Manjil, Iran, *Journal of Low Frequency Noise, Vibration and Active Control*, **35**(1), 91–98. doi:10.1177/0263092316628714

[211] Lo, S.-F., Wu, C.-Y. (2015). Evaluating the performance of wind farms in China: An empirical review, *International Journal of Electrical Power Energy Systems*, **69**, 58–66. doi:10.1016/j.ijepes.2014.12.034

[212] Wu, J., Wang, J., Qin, S., Lu, H. (2016). Suitable error evaluation criteria selection in the wind energy assessment via the K-means clustering algorithm, *International Journal of Green Energy*, **13**(11), 1145–1162. doi:10.1080/15435075.2016.1175351

[213] Goh, H. H., Lee, S. W., Chua, Q. S., Goh, K. C., Teo, K. T. K. (2016). Wind energy assessment considering wind speed correlation in Malaysia, *Renewable and Sustainable Energy Reviews*, **54**, 1389–1400. doi:10.1016/j.rser.2015.10.076

[214] Shukla, A., Singh, S. N. (2016). Clustering based unit commitment with wind power uncertainty, *Energy Conversion and Management*, **111**, 89–102. doi:10.1016/j.enconman.2015.12.040

[215]García-Sánchez, T., Gómez-Lázaro, E., Muljadi, E., Kessler, M., Molina-García, A. (2016). Statistical and Clustering Analysis for Disturbances: A Case Study of Voltage Dips in Wind Farms, *IEEE Transactions on Power Delivery*, **31**(6), 2530–2537. doi:10.1109/TPWRD.2016.2522946

[216] Eroğlu, Y., Seçkiner, S. U. (2013). Wind farm layout optimization using particle filtering approach, *Renewable Energy*, **58**, 95–107. doi:10.1016/j.renene.2013.02.019

[217] Eroğlu, Y., Seçkiner, S. U. (2012). Design of wind farm layout using ant colony algorithm, *Renewable Energy*, **44**, 53–62. doi:10.1016/j.renene.2011.12.013

[218] Abkar, M., Porté-Agel, F. (2015). A new wind-farm parameterization for large-scale atmospheric models, *Journal of Renewable and Sustainable Energy*, **7**(1), 013121. doi:10.1063/1.4907600

[219] Bulaevskaya, V., Wharton, S., Clifton, A., Qualley, G., Miller, W. O. (2015). Wind power curve modeling in complex terrain using statistical models, *Journal of Renewable and Sustainable Energy*, **7**(1), 013103. doi:10.1063/1.4904430

[220] Lydia, M., Suresh Kumar, S., Immanuel Selvakumar, A., Edwin Prem Kumar, G. (2015). Wind resource estimation using wind speed and power curve models, *Renewable Energy*, **83**, 425–434. doi:10.1016/j.renene.2015.04.045

[221] Song, M. X., Chen, K., Zhang, X., Wang, J. (2015). The lazy greedy algorithm for power optimization of wind turbine positioning on complex terrain, *Energy*, **80**, 567–574. doi:10.1016/j.energy.2014.12.012

[222] Herp, J., Poulsen, U. V., Greiner, M. (2015). Wind farm power optimization including flow variability, *Renewable Energy*, **81**, 173–181. doi:10.1016/j.renene.2015.03.034

[223] Bin, Z., Hui, L., Mingyu, W., Yaojun, C., Shengquan, L., Dong, Y., … Zhe, C. (2015). An active power control strategy for a DFIG-based wind farm to depress the subsynchronous resonance of a power system, *International Journal of Electrical Power Energy Systems*, **69**, 327–334. doi:10.1016/j.ijepes.2015.01.002

[224] Zhang, K., Mao, C., Lu, J., Wang, D., Zhang, J., Chen, X. (2015). Power Control of Directly Driven Wind Generation System Based on Superconducting Magnetic Energy Storage, *Electric Power Components and Systems*, **43**(7), 747–758. doi:10.1080/15325008.2014.1002586

[225] Hunyar, M., Veszpremi, K. (2014). Reactive power control of wind turbines. In *Power Electronics and Motion Control Conference and Exposition (PEMC), 2014 16th International* (pp. 348–352). Presented at the Power Electronics and Motion Control Conference and Exposition (PEMC), 2014 16th International. doi:10.1109/EPEPEMC.2014.6980517

[226] Kavousi-Fard, A., Khosravi, A., Nahavandi, S. (2016). A New Fuzzy-Based Combined Prediction Interval for Wind Power Forecasting, *IEEE Transactions on Power Systems*, **31**(1), 18–26. doi:10.1109/TPWRS.2015.2393880

[227] Kaur, T., Kumar, S., Segal, R. (2016). Application of artificial neural network for short term wind speed forecasting. In *2016 Biennial International Conference on Power and Energy Systems: Towards Sustainable Energy (PESTSE)* (pp. 1–5). Presented at the 2016 Biennial International Conference on Power and Energy Systems: Towards Sustainable Energy (PESTSE). doi:10.1109/PESTSE.2016.7516458

[228] Xu, M., Pinson, P., Lu, Z., Qiao, Y., Min, Y. (2016). Adaptive robust polynomial regression for power curve modeling with application to wind power forecasting, *Wind Energy*, **19**(12), 2321–2336. doi:10.1002/we.1985

[229] Yan, J., Li, K., Bai, E. W., Deng, J., Foley, A. M. (2016). Hybrid Probabilistic Wind Power Forecasting Using Temporally Local Gaussian Process, *IEEE Transactions on Sustainable Energy*, **7**(1), 87–95. doi:10.1109/TSTE.2015.2472963

[230]Akal, Z. (1992). *İşletmelerde performans ölçüm ve denetimi: çok yönlü performans göstergeleri*. Milli Prodüktivite Merkezi.

[231] Farrel, M. J. (1957). The measurement of productive efficiency, *Journal of the Royal Statistical Society. Series A (General)*, **120**(3), 253–290.

[232] Charnes, A., Cooper, W. W., Rhodes, E. (1978). Measuring the efficiency of decision making units, *European Journal of Operational Research*, **2**(6), 429–444. doi:10.1016/0377-2217(78)90138-8

[233] Banker, R. D., Charnes, A., Cooper, W. W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science*, **30**(9), 1078–1092. doi:10.1287/mnsc.30.9.1078

[234] Liu, J. S., Lu, L. Y. Y., Lu, W.-M., Lin, B. J. Y. (2013). Data envelopment analysis 1978–2010: A citation-based literature survey, *Omega*, **41**(1), 3–15. doi:10.1016/j.omega.2010.12.006

[235] Özden, Ü. H. (2008). Veri zarflama analizi (VZA) ile Türkiye'deki vakıf üniversitelerinin etkinliğinin ölçülmesi, *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, **37**(2), 167–185.

[236] Bircan, H. (2011). Veri zarflama analizi ile sivas ili merkez sağlık ocaklarının etkinliğinin ölçülmesi, *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, **12**(1), 331–347.

[237] Liu, J. S., Lu, L. Y. Y., Lu, W.-M. (2016). Research fronts in data envelopment analysis, *Omega*, **58**, 33–45. doi:10.1016/j.omega.2015.04.004

[238] Dinçer, S. E. (2008). Veri zarflama analizinde malmquist endeksiyle toplam faktör verimliliği değişiminin incelenmesi ve İMKB üzerine bir uygulama, *Marmara Üniversitesi İİBF Dergisi*, **25**(2), 825–846.

[239] Caves, D. W., Christensen, L. R., Diewert, W. E. (1982). The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity, *Econometrica*, **50**(6), 1393–1414. doi:10.2307/1913388

[240] Aigner, D., Lovell, C. A. K., Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics*, **6**(1), 21–37. doi:10.1016/0304-4076(77)90052-5

[241] Meeusen, W., van Den Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error, *International Economic Review*, **18**(2), 435–444. doi:10.2307/2525757

[242]Zhou, P., Ang, B. W., Zhou, D. Q. (2012). Measuring economy-wide energy efficiency performance: A parametric frontier approach, *Applied Energy*, **90**(1), 196–200. doi:10.1016/j.apenergy.2011.02.025

[243] Hu, J.-L., Honma, S. (2014). A Comparative Study of Energy Efficiency of OECD Countries: An Application of the Stochastic Frontier Analysis, *Energy Procedia*, **61**, 2280–2283. doi:10.1016/j.egypro.2014.12.447

[244] Sarıca, K., Or, I. (2007). Efficiency assessment of Turkish power plants using data envelopment analysis, *Energy*, **32**(8), 1484–1499. doi:10.1016/j.energy.2006.10.016

[245] San Cristóbal, J. R. (2011). A multi criteria data envelopment analysis model to evaluate the efficiency of the Renewable Energy technologies, *Renewable Energy*, **36**(10), 2742–2746. doi:10.1016/j.renene.2011.03.008

[246] Ardente, F., Beccali, M., Cellura, M., Lo Brano, V. (2008). Energy performances and life cycle assessment of an Italian wind farm, *Renewable and Sustainable Energy Reviews*, **12**(1), 200–217. doi:10.1016/j.rser.2006.05.013

[247] Iglesias, G., Castellanos, P., Seijas, A. (2010). Measurement of productive efficiency with frontier methods: A case study for wind farms, *Energy Economics*, **32**(5), 1199–1208. doi:10.1016/j.eneco.2010.03.004

[248] Akan, Y., Çalmaşur, G. (2011). Comparison Of Data Envelopment Analysis And Stochastic Frontier Approach Method in Calculation Of Efficiency (An Application On TRA1 Subregion), *Atatürk Ü. İİBF Dergisi*, (10). Retrieved from http://e-dergi.atauni.edu.tr/atauniiibd/article/download/1025007766/1025007136

[249] Atılgan Emre. (2012). *Evaluation of Hospital Efficiency with Stochastic Frontier Analysis: An Application for Turkish Ministry of Health Hospitals, P.hd. Dissertation* (Doctoral Thesis). Hacettepe University, Ankara. Retrieved from http://emreatilgan.com/wp-content/uploads/2014/02/E.ATILGAN-TEZ-SON.pdf

[250] Kusiak, A., Verma, A., Wei, X. (2012). Wind turbine capacity frontier from SCADA, *Wind System Magazine*, **3**(9), 36–39.

[251] Barthelmie, R. J., Jensen, L. E. (2010). Evaluation of wind farm efficiency and wind turbine wakes at the Nysted offshore wind farm, *Wind Energy*, **13**(6), 573–586. doi:10.1002/we.408

[252] de Prada Gil, M., Gomis-Bellmunt, O., Sumper, A., Bergas-Jané, J. (2012). Power generation efficiency analysis of offshore wind farms connected to a SLPC (single large power converter) operated with variable frequencies considering wake effects, *Energy*, **37**(1), 455–468. doi:10.1016/j.energy.2011.11.010

[253] Malmquist, S. (1953). Index Numbers and Difference Surfaces, *Trabajos De Estadistica*, **4**(2), 209–242. doi:10.1007/BF03006863

[254] Cingi, S., Tarım, Ş. A. (2000). *Türk Banka Sisteminde Performans Ölçümü DEA Malmquist TFP Endeksi Uygulaması* (No. Araştırma Tebliğleri Serisi; Sayı:2000-1). Türkiye Bankalar Birliği.

[255]Battese, G. E., Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data, *Empirical Economics*, **20**(2), 325–332. doi:10.1007/BF01205442

[256] Battese, G. E., Corra, G. S. (1977). Estimation of a Production Frontier Model: With Application to the Pastoral Zone of Eastern Australia, *Australian Journal of Agricultural Economics*, **21**(3), 169–179. doi:10.1111/j.1467-8489.1977.tb00204.x

[257] Coelli, T. J. (1996). A guide to FRONTIER version 4.1: a computer program for stochastic frontier production and cost function estimation, *CEPA Working papers*, **7**, 1–96.

[258]Iglesias, G., Castellanos, P., Seijas, A. (2010). Measurement of productive efficiency with frontier methods: A case study for wind farms, *Energy Economics*, **32**(5), 1199–1208. doi:10.1016/j.eneco.2010.03.004

[259] Bryce, C. L., Engberg, J. B., Wholey, D. R. (2000). Comparing the agreement among alternative models in evaluating HMO efficiency, *Health Services Research*, **35**(2), 509–528.

[260] Newhouse, J. P. (1994). Frontier estimation: How useful a tool for health economics?, *Journal of Health Economics*, **13**(3), 317–322. doi:10.1016/0167-6296(94)90030-2

[261] Pourmousavi Kani, S. A., Ardehali, M. M. (2011). Very short-term wind speed prediction: A new artificial neural network–Markov chain model, *Energy Conversion and Management*, **52**(1), 738–745. doi:10.1016/j.enconman.2010.07.053

[262] Jiang, Y., Song, Z., Kusiak, A. (2013). Very short-term wind speed forecasting with Bayesian structural break model, *Renewable Energy*, **50**, 637–647. doi:10.1016/j.renene.2012.07.041

[263] Ozkan, M. B., Karagoz, P. (2015). A Novel Wind Power Forecast Model: Statistical Hybrid Wind Power Forecast Technique (SHWIP), *IEEE Transactions on Industrial Informatics*, **11**(2), 375–387. doi:10.1109/TII.2015.2396011

[264] Croonenbroeck, C., Ambach, D. (2015). A selection of time series models for short- to medium-term wind power forecasting, *Journal of Wind Engineering and Industrial Aerodynamics*, **136**, 201–210. doi:10.1016/j.jweia.2014.11.014

[265] Hammersley, J. M., Morton, K. W. (1954). Poor Man's Monte Carlo, *Journal of the Royal Statistical Society. Series B (Methodological)*, **16**(1), 23–38.

[266] Gordon, N. J., Salmond, D. J., Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *Radar and Signal Processing, IEE Proceedings F*, **140**(2), 107–113.

[267] World Heritage Encyclopedia. (n.d.). Sequential Monte Carlo methods. World Heritage Encyclopedia. Retrieved from http://self.gutenberg.org/articles/eng/Sequential_Monte_Carlo_methods

[268] Yu, S., Fernando, T., Savkin, A. V., Iu, H. H. C. (2016). A dynamic state estimation based sliding mode controller for wind energy generation system connected to multimachine grids. In *2016 IEEE Conference on Control Applications (CCA)* (pp. 1556–1559). Presented at the 2016 IEEE Conference on Control Applications (CCA). doi:10.1109/CCA.2016.7588022

[269] Yu, S., Fernando, T., Iu, H. H. C., Emami, K. (2016). Realization of State-Estimation-Based DFIG Wind Turbine Control Design in Hybrid Power Systems Using Stochastic Filtering Approaches, *IEEE Transactions on Industrial Informatics*, **12**(3), 1084–1092. doi:10.1109/TII.2016.2549940

[270] Outemzabet, S., Nerguizian, C. (2008). Accuracy Enhancement of an Indoor ANN-based Fingerprinting Location System Using Particle Filtering and a Low-Cost Sensor. In *VTC Spring 2008 - IEEE Vehicular Technology Conference* (pp. 2750–2754). Presented at the VTC Spring 2008 - IEEE Vehicular Technology Conference. doi:10.1109/VETECS.2008.601

[271] Peter Torma, Csaba Szepesvári. (2003). Combining Local Search, Neural Networks and Particle Filters to Achieve Fast and Reliable Contour Tracking. In *Proceedings of the IEEE International Symposium on Intelligent Signal Processing*. Retrieved from http://www.peerevaluation.org/read/libraryID:1655

[272] Wan, Eric A., Rudolph Van Der Merwe, Alex T. Nelson. (1999). Dual Estimation and the Unscented Transformation - 1681-dual-estimation-and-the-unscented-transformation.pdf.*NIPS*, **12**, 666–672.

[273] Baraldi, P., Compare, M., Sauco, S., Zio, E. (2013). Ensemble neural network-based particle filtering for prognostics, *Mechanical Systems and Signal Processing*, **41**(1–2), 288–300. doi:10.1016/j.ymssp.2013.07.010

[274] Chen, C., Zhang, B., Vachtsevanos, G., Orchard, M. (2011). Machine Condition Prediction Based on Adaptive Neuro Fuzzy and High-Order Particle Filtering, *IEEE Transactions on Industrial Electronics*, **58**(9), 4353–4364. doi:10.1109/TIE.2010.2098369

[275] Zhan, R., Wan, J. (2006). Neural network-aided adaptive unscented Kalman filter for nonlinear state estimation, *IEEE Signal Processing Letters*, **13**(7), 445–448. doi:10.1109/LSP.2006.871854

[276] Shoorehdeli, M. A., Teshnehlab, M., Sedigh, A. K. (2009). Training ANFIS as an identifier with intelligent hybrid stable learning algorithm based on particle swarm optimization and extended Kalman filter, *Fuzzy Sets and Systems*, **160**(7), 922–948. doi:10.1016/j.fss.2008.09.011

[277] Okumus, I., Dinler, A. (2016). Current status of wind energy forecasting and a hybrid method for hourly predictions, *Energy Conversion and Management*, **123**, 362–371. doi:10.1016/j.enconman.2016.06.053

[278] McCulloch, W. S., Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity, *The Bulletin of Mathematical Biophysics*, **5**(4), 115–133. doi:10.1007/BF02478259

[279] Yadav, R. N., Kumar, N., Kalra, P. K., John, J. (2006). Learning with generalized-mean neuron model, *Neurocomputing*, **69**(16–18), 2026–2032. doi:10.1016/j.neucom.2005.10.006

[280] Ghosh-Dastidar, S., Adeli, H. (2009). A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection, *Neural Networks*, **22**(10), 1419–1431. doi:10.1016/j.neunet.2009.04.003

[281]Shiblee, M., Chandra, B., Kalra, P. K. (2010). Learning of geometric mean neuron model using resilient propagation algorithm, *Expert Systems with Applications*, **37**(12), 7449–7455. doi:10.1016/j.eswa.2010.04.018

[282] Hamzaçebi, C., Kutay, F. (2006). A heuristic approach for finding the global minimum: Adaptive random search technique, *Applied Mathematics and Computation*, **173**(2), 1323–1333. doi:10.1016/j.amc.2005.05.002

[283]Toksari, M. D. (2006). Ant colony optimization for finding the global minimum, *Applied Mathematics and Computation*, **176**(1), 308–316. doi:10.1016/j.amc.2005.09.043

[284] Baskan, O., Haldenbilen, S., Ceylan, H., Ceylan, H. (2009). A new solution algorithm for improving performance of ant colony optimization, *Applied Mathematics and Computation*, **211**(1), 75–84. doi:10.1016/j.amc.2009.01.025

[285] Hamzaçebi, C., Kutay, F. (2007). Continuous functions minimization by dynamic random search technique, *Applied Mathematical Modelling*, **31**(10), 2189–2198. doi:10.1016/j.apm.2006.08.015

[286] Li, J., Russell Rhinehart, R. (1998). Heuristic random optimization, *Computers Chemical Engineering*, **22**(3), 427–444. doi:10.1016/S0098-1354(97)00005-7

[287] Hamzaçebi, C. (2008). Improving genetic algorithms' performance by local search for continuous function optimization, *Applied Mathematics and Computation*, **196**(1), 309–317. doi:10.1016/j.amc.2007.05.068

[288] Duran Toksarı, M. (2007). A heuristic approach to find the global optimum of function, *Journal of Computational and Applied Mathematics*, **209**(2), 160–166. doi:10.1016/j.cam.2006.10.074

[289]Kwon, Y.-D., Kwon, S.-B., Jin, S.-B., Kim, J.-Y. (2003). Convergence enhanced genetic algorithm with successive zooming method for solving continuous optimization problems, *Computers Structures*, **81**(17), 1715–1725. doi:10.1016/S0045-7949(03)00183-4

[290]Socha, K., Blum, C. (2007). An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training, *Neural Computing and Applications*, **16**(3), 235–247. doi:10.1007/s00521-007-0084-z

[291] Chau, K. W. (2006). Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River, *Journal of Hydrology*, **329**(3–4), 363–367. doi:10.1016/j.jhydrol.2006.02.025

[292] Wang, G., Hao, J., Ma, J., Huang, L. (2010). A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering, *Expert Systems with Applications*, **37**(9), 6225–6232. doi:10.1016/j.eswa.2010.02.102

[293] Bas, E., Uslu, V. R., Egrioglu, E. (2016). Robust learning algorithm for multiplicative neuron model artificial neural network, *Expert Systems with Applications*, **56**, 80–88. doi:10.1016/j.eswa.2016.02.051

[294] Mohamad, E. T., Faradonbeh, R. S., Armaghani, D. J., Monjezi, M., Majid, M. Z. A. (2016). An optimized ANN model based on genetic algorithm for predicting ripping production, *Neural Computing and Applications*, 1–14. doi:10.1007/s00521-016-2359-8

[295] Lee, A., Geem, Z. W., Suh, K.-D. (2016). Determination of Optimal Initial Weights of an Artificial Neural Network by Using the Harmony Search Algorithm: Application to Breakwater Armor Stones, *Applied Sciences*, **6**(6), 164. doi:10.3390/app6060164

[296] Ganguly, S., Patra, A., Chattopadhyay, P. P., Datta, S. (2016). New training strategies for neural networks with application to quaternary Al–Mg–Sc–Cr alloy design problems, *Applied Soft Computing*, **46**, 260–266. doi:10.1016/j.asoc.2016.05.017

[297] Bas, E. (2016). The Training Of Multiplicative Neuron Model Based Artificial Neural Networks With Differential Evolution Algorithm For Forecasting, *Journal of Artificial Intelligence and Soft Computing Research*, **6**(1), 5–11. doi:10.1515/jaiscr-2016-0001

[298] Li, J.-B., Chung, Y.-K. (2005). A Novel Back-propagation Neural Network Training Algorithm Designed by an Ant Colony Optimization. In *2005 IEEE/PES Transmission Distribution Conference Exposition: Asia and Pacific* (pp. 1–5). Presented at the 2005 IEEE/PES Transmission Distribution Conference Exposition: Asia and Pacific. doi:10.1109/TDC.2005.1546990

[299] Saghatforoush, A., Monjezi, M., Faradonbeh, R. S., Armaghani, D. J. (2016). Combination of neural network and ant colony optimization algorithms for prediction and optimization of flyrock and back-break induced by blasting, *Engineering with Computers*, **32**(2), 255–266. doi:10.1007/s00366-015-0415-0

[300] Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., Johannes, R. S. (1988). Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus, *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 261–265.

[301] Sfetsos, A. (2000). A comparison of various forecasting techniques applied to mean hourly wind speed time series, *Renewable Energy*, **21**(1), 23–35. doi:10.1016/S0960-1481(99)00125-1

[302] Sfetsos, A. (2002). A novel approach for the forecasting of mean hourly wind speed time series, *Renewable Energy*, **27**(2), 163–174. doi:10.1016/S0960-1481(01)00193-8

[303]Barbounis, T. G., Theocharis, J. B. (2007). Locally recurrent neural networks for wind speed prediction using spatial correlation, *Information Sciences*, **177**(24), 5775–5797. doi:10.1016/j.ins.2007.05.024

[304] Cadenas, E., Rivera, W. (2007). Wind speed forecasting in the South Coast of Oaxaca, México, *Renewable Energy*, **32**(12), 2116–2128. doi:10.1016/j.renene.2006.10.005

[305] Monfared, M., Rastegar, H., Kojabadi, H. M. (2009). A new strategy for wind speed forecasting using artificial intelligent methods, *Renewable Energy*, **34**(3), 845–848. doi:10.1016/j.renene.2008.04.017

[306] Palomares-Salas, J. C., Rosa, J. J. G. de la, Ramiro, J. G., Melgar, J., Aguera, A., Moreno, A. (2009). ARIMA vs. Neural networks for wind speed forecasting.In *2009 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications* (pp. 129–133).

[307] Barszcz, T., Bielecki, A., Wójcik, M. (2010). ART-Type Artificial Neural Networks Applications for Classification of Operational States in Wind Turbines.In L. Rutkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, J. M. Zurada (Eds.), *Artifical Intelligence and Soft Computing* (pp. 11–18). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-13232-2_2

[308]Li, S., Wunsch, D. C., O'Hair, E. A., Giesselmann, M. G. (2001). Using neural networks to estimate wind turbine power generation, *IEEE Transactions on Energy Conversion*, **16**(3), 276–282. doi:10.1109/60.937208

[309]Brandao, R. F. M., Carvalho, J. A. B., Barbosa, F. P. M. (2010). Neural networks for condition monitoring of wind turbines. In *Modern Electric Power Systems (MEPS), 2010 Proceedings of the International Symposium* (pp. 1–4). Presented at the Modern Electric Power Systems (MEPS), 2010 Proceedings of the International Symposium.

APPENDIX A

MATLAB CODES OF DEVELOPED APPROACHES

A.1 Matlab Codes of Particle Filtering for ANN Models

```
%Particle Filtering approach to wind power estimation
%In this program, wind power estimation error is tuned by PF. Wind
power estimation is very important issue in wind energy market.
There are many methods to estimate power production. Nevertheless,
it is the fact that every forecast has an error.
%% initialize the variables
x=i1(1); %initial wind speed at time = 1
x_N = 0.3319; % Noise covariance in the system (i.e. process noise
in the state update, here, we'll use a Gaussian.)
x_R = 17241; % Noise covariance in the measurement
T = 500; % duration of the forecast (i.e. number of iterations).
N = 100; % The number of particles the system generates. The larger
this is, the better your approximation, but the more computation you
need.
nn_outs = output(1);
nn_outp = output3(1);
%initialize our initial, prior particle distribution as a Gaussian
around the true initial value
V = 1; %define the variance of the initial estimate
% make the randomly generated particles from the initial prior
Gaussian distribution
for i = 1:N
    x_P(i) = sim(net, x) + sqrt(x_N)*randn;
    z_P(i)=sim(net3,x_P(i))+ sqrt(x_R)*randn;
    P_w(i)= normpdf(t2(1),z_P(i),sqrt(x_R));
end
%generate the observations from the randomly selected particles,
based upon the given function
sumPw=sum(P_w);
% Normalize to form a probability distribution (i.e. sum to 1).
for i=1:N
    P_w(i) = P_w(i)./sumPw;
end
z_out = t2(1);  %the actual output vector for measurement values.

x_out = [x];  %the actual output vector for measurement values.
x_est = mean(x_P); % time by time output of the particle filters
estimate
x_est_out = [x_est]; % the vector of particle filter estimates.
z_est_out=mean(z_P);
for t = 2:T
x = i1(t);
z = t2(t);
%Here, we do the particle filter
for i = 1:N
%given the prior set of particle
```

```
x_P_update(i) = sim(net, x)  + sqrt(x_N)*randn;
%with these new updated particle locations, update the observations
for each of these particles.
z_update(i) = sim(net3,x_P_update(i))+ sqrt(x_R)*randn;
%Generate the weights for each of these particles. The weights are
based upon the probability of the given observation for a particle,
GIVEN the actual observation. That is, if we observe a location z,
and we know our observation error is Gaussian with variance x_R,
then the probability of seeing a given z centered at that actual
measurement is (from the equation of a Gaussian). Probability
Density Function of each particle's forecast
P_w(i)= normpdf(z,z_update(i),sqrt(x_R));
end
sumPw=sum(P_w);
% Normalize to form a probability distribution (i.e. sum to 1).
for i=1:N
    P_w(i) = P_w(i)./sumPw;
    x_P(i) = x_P_update(i)*P_w(i);
end
    t
%% Resampling: From this new distribution, now we randomly sample
from it to generate our new estimate particles what this code
specifically does is randomly, uniformly, sample from the cumulative
distribution of the probability distribution generated by the
weighted vector P_w.  If you sample randomly over this distribution,
you will select values based upon there statistical probability, and
thus, on average, pick values with the higher weights. Store this
new value to the new estimate which will go back into the next
iteration.
    x_est = sum(x_P);
    z_est = sim(net3,x_est);
    x_out = [x_out t1(t)];
    z_out = [z_out t2(t)];
    nn_outs = [nn_outs output(t)];
    nn_outp = [nn_outp output3(t)];
    x_est_out = [x_est_out x_est];
    z_est_out= [z_est_out z_est];

end
t = 1:T;
figure(3);
clf
plot(t, x_out, '.-b', t, x_est_out, '-.r',t,nn_outs,
'g:','linewidth',3);
set(gca,'FontSize',12); set(gcf,'Color','White');
xlabel('time step'); ylabel('Wind Speed m/s');
legend('Actual Wind Speed Output', 'ANN + PF Forecast','ANN
Forecast');
t = 1:T;
figure(4);
clf
plot(t, z_out, '.-b', t, z_est_out, '-.r',t,nn_outp,
'g:','linewidth',3);
set(gca,'FontSize',12); set(gcf,'Color','White');
xlabel('time step'); ylabel('Power Output kW');
legend('Actual Power Output', 'ANN + PF Forecast','ANN Forecast');
```

A.2 Matlab Codes of Antrain ANN

```
%Antrain ANN
reset(RandStream.getDefaultStream,sum(100*clock))
%to reset random sets
k=100; %archive size
m=60; %number of antsant
it=100; %number of iterations
[Nxx Nx]=size(i1); %number of inputs Nx
[Nyy Ny]=size(t1); %number of outputs Ny
for i=1:Nxx
   X(i,:)=(i1(i,:)-min(i1))./(max(i1)-min(i1)); % inputs
end
for i=1:Nyy
    Y(i,:)=(t1(i,:)-min(t1))./(max(t1)-min(t1)); %targets
end
 [Nxx Nx]=size(X); %number of inputs Nx
 [Nyy Ny]=size(Y); %number of outputs Ny
 Biasi=ones(1, Nx);
 Biaso=ones(1, Ny);
 sample=Nxx;
 Nwi=length(Biasi);  %number  of  hidden  neurons'  weights   for  the
first layer
 Nwo=Ny; %number of output layer weights
 E=zeros(sample,k);
 O=zeros(sample,k);
 counter=0;
 ChangeStatus=10;
 wiAlphaS=10;
 woAlphaS=10;
 bwiAlphaS=10;
 bwoAlphaS=10;
 MinErrorRate=0.005;
 MaxRepeatition=1;
%randomization of weights in the archive
for i=1:k
     AWi(i,:,:)=(-1+2*rand(Nx))*wiAlphaS;  %randomization  of  weights
in the archive for inputs
     AWo(i,:,:)=(-1+2*rand(1,Nwi))*woAlphaS;    %randomization    of
weights in the archive for outputs
     ABwi(i,:,:)=(-1+2*rand(1,Nwi))*bwiAlphaS;   %randomization   of
bias weights in the archive for hidden neurons
     ABwo(i,:,:)=(-1+2*rand(1,Nwo))*bwoAlphaS;   %randomization   of
bias weights in the archive for output
end
%initial archive assignments
arsivWi=AWi;
arsivWo=AWo;
arsivBwi=ABwi;
arsivBwo=ABwo;
Wi=AWi;
Wo=AWo;
Bwi=ABwi;
Bwo=ABwo;
s=1; %sample number
for h=1:sample
    [O(h,:)
E(h,:)]=evaluation10(h,k,Nwi,X,Y,Wi,Biasi,Bwi,Nwo,Wo,Biaso,Bwo);
end
```

```
for i=1:k
    SSE(i)=sqrt(sumsqr(E(:,i)));
end
 cv=[]; %cv is a check operator to eliminate the excess value of
archive elements. If the number of archive size greater than number
of ants, the excess elements need to be eliminate
%The codes below eliminates the worst results until the number of
results and the number of ants are the same
for h=1:(k-m) %the excess value (number of archive - number of ants)
     [maxf1,maxf1ind]=max(SSE); %the indices of maximum values of
outputs - the worst results
    cv(h)=maxf1ind; %the maximum error indices are assigned to cv
    SSE(maxf1ind)=-inf; %change the maximum values with -inf
end
%delete excess values
O(:,cv)=[];
E(:,cv)=[];
SSE(cv)=[];
Wi(cv,:,:)=[];
Wo(cv,:,:)=[];
Bwi(cv,:,:)=[];
Bwo(cv,:,:)=[];
cv=[]; %refresh cv
minf1=min(SSE); %finding the minimum value of errors
%compute the distance to the best solution - best ant
diff=SSE-minf1; %Di
gauss=exp(-((diff.^2)/(2*0.0005))); %compute the gauss distribution
of the distances
feromen=[(gauss./(sum(gauss)))]; %normalize the distribution and
assigned the values as pheromone
ants=round(feromen*m); %assign the ants to the solution vectors
according to pheromone amounts of the solutions
v=sum(ants); %sum the assigned ants to check the total number of
ants
[maxant,maxantind]=max(ants); %define the indices of maximum number
of ants

while v~=m %if the total assigned ants differ from number of total
ants, balance the assigned ants
if v>m
   % p=m-v;
   [maxant,maxantind]=max(ants);
    maxant=maxant-1;
    ants(maxantind)=maxant;
    v=sum(ants) ;
elseif v<m
    t=round(rand*m);
    while t==0
        t=round(rand*m);
    end
    r=ants(t) ;
    r=r+1 ;
    ants(t)=r ;
    v=sum(ants) ;
end
end
[fbest(1),fbestind]=min(SSE);
Ebest(1)=E(s,fbestind);
EbestG=Ebest(1);
fGlobalMin=fbest(1);
```

```
founditer=1;
WiBest(1,:,:)=Wi(fbestind,:,:);
WoBest(1,:,:)=Wo(fbestind,:,:);
BwiBest(1,:,:)=Bwi(fbestind,:,:);
BwoBest(1,:,:)=Bwo(fbestind,:,:);
WiGlobal(1,:,:)=WiBest(1,:,:);
WoGlobal(1,:,:)=WoBest(1,:,:);
BwiGlobal(1,:,:)=BwiBest(1,:,:);
BwoGlobal(1,:,:)=BwoBest(1,:,:);
Gcounter=0;
O=[O zeros(sample,(k-m))];
E=[E zeros(sample,(k-m))];
tic ;% timer
while fGlobalMin>MinErrorRate  Gcounter<MaxRepeatition;
    wiAlpha=wiAlphaS;
    woAlpha=woAlphaS;
    bwiAlpha=bwiAlphaS;
    bwoAlpha=bwoAlphaS;
    for j=1:it
        Ar2Wi=arsivWi;
        Ar2Wo=arsivWo;
        Ar2Bwi=arsivBwi;
        Ar2Bwo=arsivBwo;
        if s>sample
            s=1;
        end

    for i=1:m
        if ants(i)>0
            for t=1:ants(i)
                CWi=Wi(i,:,:)+(-1+2*rand(1,Nx,Nx))*wiAlpha;
                CWo=Wo(i,:,:)+(-1+2*rand(1,Nwo,Nwi))*woAlpha;
                CBwi=Bwi(i,:,:)+(-1+2*rand(1,1,Nwi))*bwiAlpha;
                CBwo=Bwo(i,:,:)+(-1+2*rand(1,Nwo,Ny))*bwoAlpha;
                Ar2Wi=[Ar2Wi;CWi];
                Ar2Wo=[Ar2Wo;CWo];
                Ar2Bwi=[Ar2Bwi;CBwi];
                Ar2Bwo=[Ar2Bwo;CBwo];
            end
        end
    end
    O=[O zeros(sample,m)];
    E=[E zeros(sample,m)];
    for h=1:sample
    [O(h,:)
E(h,:)]=evaluation10(h,length(Ar2Wi),Nwi,X,Y,Ar2Wi,Biasi,Ar2Bwi,Nwo,
Ar2Wo,Biaso,Ar2Bwo);
    end
    for i=1:length(Ar2Wi)
        SSE(i)=sqrt(sumsqr(E(:,i)));
    end
    for h=1:m
        [maxfab,maxfabind]=max(SSE);
        cv(h)=maxfabind;
        SSE(maxfabind)=-inf;
    end
    O(:,cv)=[];
    E(:,cv)=[];
    SSE(cv)=[];
    Ar2Wi(cv,:,:)=[];
```

```
    Ar2Wo(cv,:,:)=[];
    Ar2Bwi(cv,:,:)=[];
    Ar2Bwo(cv,:,:)=[];
    cv=[]; %refresh cv
    arsiv1Wi=Ar2Wi;
    arsiv1Wo=Ar2Wo;
    arsiv1Bwi=Ar2Bwi;
    arsiv1Bwo=Ar2Bwo;
    arsivWi=Ar2Wi;
    arsivWo=Ar2Wo;
    arsivBwi=Ar2Bwi;
    arsivBwo=Ar2Bwo;
    cvb=[]
    for h=1:(k-m)
        [maxfab,maxfabind]=max(SSE);
        cvb(h)=maxfabind;
        SSE(maxfabind)=-inf;
    end
    O(:,cvb)=[];
    E(:,cvb)=[];
    SSE(cvb)=[];
    arsiv1Wi(cvb,:,:)=[];
    arsiv1Wo(cvb,:,:)=[];
    arsiv1Bwi(cvb,:,:)=[];
    arsiv1Bwo(cvb,:,:)=[];
    Wi=arsiv1Wi;
    Wo=arsiv1Wo;
    Bwi=arsiv1Bwi;
    Bwo=arsiv1Bwo;
    cvb=[];
    for h=1:sample
    [O(h,:)
E(h,:)]=evaluation10(h,m,Nwi,X,Y,Wi,Biasi,Bwi,Nwo,Wo,Biaso,Bwo);
    end
    for i=1:m
        SSE(i)=sqrt(sumsqr(E(:,i)));
    end
    minf=min(SSE);
    diff=SSE-minf; %Di
    gauss=exp(-((diff.^2)/(2*0.0005)));
    feromen=[(gauss./(sum(gauss)))];
    ants=round(feromen*m);
    v=sum(ants);
    [maxant,maxantind]=max(ants);
while v~=m %if the total assigned ants differ from number of total
ants, balance the assigned ants
if v>m
   [maxant,maxantind]=max(ants);
    maxant=maxant-1;
    ants(maxantind)=maxant;
    v=sum(ants) ;
elseif v<m
    t=round(rand*m);
    while t==0
        t=round(rand*m);
    end
    r=ants(t) ;
    r=r+1 ;
    ants(t)=r ;
    v=sum(ants) ;
```

```
end
end
    [fbest(j+1),fbest1ind]=min(SSE);
    Ebest(j+1)=E(s,fbest1ind);
    WiBest(j+1,:,:)=Wi(fbest1ind,:,:);
    WoBest(j+1,:,:)=Wo(fbest1ind,:,:);
    BwiBest(j+1,:,:)=Bwi(fbest1ind,:,:);
    BwoBest(j+1,:,:)=Bwo(fbest1ind,:,:);
    O=[O zeros(sample,(k-m))];
    E=[E zeros(sample,(k-m))];
    for h=1:sample
    [O(h,:)
E(h,:)]=evaluation10(h,k,Nwi,X,Y,arsivWi,Biasi,arsivBwi,Nwo,arsivWo,
Biaso,arsivBwo);
    end
    for i=1:k
        SSE(i)=sqrt(sumsqr(E(:,i)));
    end
    if fbest(j+1)<fGlobalMin
        WiGlobal(1,:,:)=WiBest(j+1,:,:);
        WoGlobal(1,:,:)=WoBest(j+1,:,:);
        BwiGlobal(1,:,:)=BwiBest(j+1,:,:);
        BwoGlobal(1,:,:)=BwoBest(j+1,:,:);
        fGlobalMin=fbest(j+1)
        EbestG=Ebest(j+1);
        founditer=j+1
    else
        counter=counter+1;
        if counter==ChangeStatus
            counter=0;
            for  h=1:(round(k/2))  %the  excess  value  (number  of
archive - number of ants)
                [maxf1,maxf1ind]=max(SSE);  %the  indices  of  maximum
values of outputs - the worst results
                SSE(maxf1ind)=-inf; %change the maximum values with
-inf
arsivWi(maxf1ind,:,:)=arsivWi(maxf1ind,:,:)+randn*WiGlobal(1,:,:);
arsivWo(maxf1ind,:,:)=arsivWo(maxf1ind,:,:)+randn*WoGlobal(1,:,:);
arsivBwi(maxf1ind,:,:)=arsivBwi(maxf1ind,:,:)+randn*BwiGlobal(1,:,:)
;
arsivBwo(maxf1ind,:,:)=arsivBwo(maxf1ind,:,:)+randn*BwoGlobal(1,:,:)
;
            end
        end
    end
    wiAlpha=0.9*wiAlpha;
    woAlpha=0.9*woAlpha;
    bwiAlpha=0.9*bwiAlpha;
    bwoAlpha=0.9*bwoAlpha;
    s=s+1;
    j
end
Gcounter=Gcounter+1
end
ElapsedTimeInSec=toc %timer
Gcounter
O=[];
E=[];
for s=1:sample
```

```
[O(s,:)
E(s,:)]=evaluation10(s,1,Nwi,X,Y,WiGlobal,Biasi,BwiGlobal,Nwo,WoGlob
al,Biaso,BwoGlobal);
end
SSE=[];
    for i=1:1
        SSE(i)=sqrt(sumsqr(E(:,i)));
    end
    yyy=(max(t1)-min(t1)).*O+min(t1);
    founditer
[Nxx Nx]=size(Vi1); %number of inputs Nx

[Nyy Ny]=size(Vt1); %number of outputs Ny

for i=1:Nxx
   X(i,:)=(Vi1(i,:)-min(Vi1))./(max(Vi1)-min(Vi1)); % inputs
end
for i=1:Nyy
    Y(i,:)=(Vt1(i,:)-min(Vt1))./(max(Vt1)-min(Vt1)); %targets
end
Biasi=ones(1, Nx);
Biaso=ones(1, Ny);
sample=Nxx; %number of samples for training Nxx and Nyy are the same
Nwi=length(Biasi); %number of hidden neurons' weights  for the first
layer
Nwo=Ny; %number of output layer weights
V_E=zeros(sample,1);
V_O=zeros(sample,1);
for s=1:sample
[V_O(s,:)
V_E(s,:)]=evaluation10(s,1,Nwi,X,Y,WiGlobal,Biasi,BwiGlobal,Nwo,WoGl
obal,Biaso,BwoGlobal);
end
V_SSE=[];
    for i=1:1
        V_SSE(i)=sqrt(sumsqr(V_E(:,i)));
    end
    Vyyy=(max(Vt1)-min(Vt1)).*V_O+min(Vt1);

function[O              E]             =             evaluation10
(s,k,Nwi,X,Y,Wi,Biasi,Bwi,Nwo,Wo,Biaso,Bwo)
for i=1:k
    for l=1:Nwi
        for ii=1:Nwi
            hin(ii)=X(s,ii)*Wi(i,ii,l);
        end
        Neti(i,l)=sum(hin)+Biasi(l)*Bwi(i,l);  %inputs  and  bias  are
added by multiplying their weights in the archive for hidden neurons
    end
end
Netii=1./(1+exp(-Neti)); %sigmoid function for the hidden neuron net
outputs in the archive
for i=1:k
    for l=1:Nwo
        for ii=1:Nwi
            hon(ii)=Netii(i,ii)*Wo(i,l);
        end
```

```
        Neto(i,l)=sum(hon)+Biaso*Bwo(i);    %first   hidden   neuron
outputs and bias are added by multiplying their weights in the
archive for output
    end
end

for i=1:k
    O(i)=1./(1+exp(-Neto(i)));
end
for i=1:k
    E(i)=abs(Y(s)-O(i));
end
```

APPENDIX B

CURRICULUM VITAE

PERSONAL INFORMATION

Name and Surname: Yunus EROĞLU

Nationality: Turkish (TC)

Birth place and date: Adana / 24.06.1984

Marital status: Married

Phone number: +90 544 643 9063

Fax:

Email:eroglu@gantep.edu.tr, erogluyunus@hotmail.com, erogluyunus@gmail.com

EDUCATION

	Graduate school	**Year**
Master:	Industrial Engineering Department, Gaziantep University	2011
Bachelor:	Industrial Engineering Department, Gaziantep University	2007
High School:	Adana Erkek High School,	2002

Work experience

	Place	Enrollment
2009-Present	Industrial Engineering Department, Gaziantep University	Research Assistant
2008-2009	Mert Teknik A.Ş./ Adana Office/	Project/Sales Engineer

PUBLICATIONS

International Refereed Journals

EROĞLU, Y., SEÇKİNER S.U.,"Trend Topic Analysis for Wind Energy Researches: A Data Mining Approach Using Text Mining",Journal of Technology Innovations in Renewable Energy, Volume 5, No. 2, 2016

SEÇKİNER S.U., **EROĞLU, Y.,** "Evaluation of a Harmony Search Algorithm for Manual Lifting Tasks Optimization", International Journal of Computer Applications, Volume 131 - Number 6, December 2015, Pages 10-17, Published by Foundation of Computer Science (FCS), NY, USA. DOI:10.5120/ijca2015907442

EROĞLU, Y., SEÇKİNER S.U., "Wind Farm Layout Optimization using Particle Filtering Approach", Renewable Energy, Volume 58, October 2013, Pages 95–107.

SEÇKİNER S.U., **EROĞLU, Y.,** EMRULLAH, M., DERELİ, T., "Ant colony optimization for continuous functions by using novel pheromone updating", Applied Mathematics and Computation, Volume 219, Issue 9, 1 January 2013, Pages 4163–4175

EROĞLU, Y., SEÇKİNER S.U., "Design of wind farm layout using ant colony algorithm", Renewable Energy, Volume 44, August 2012, Pages 53–62

National Refereed Journals

EROĞLU, Y., SEÇKINER, S. U. “Performance analysis in wind farms by data envelopment analysis and Malmquist Index approaches,” *Pamukkale Univ. J. Eng. Sci.*, vol. 23, no. 1, pp. 45–54, 2017.

AKALIN, H., SEÇKINER, S. U., **EROĞLU, Y.,** “Performance evaluation for wind turbines using stochastic frontier analysis,” J. Fac. Eng. Archit. Gazi Univ., vol. Article in Press, 2017.

International Congress and Conferences

EROĞLU, Y., SEÇKINER, S. U. “Clustering of Wind Turbines in a Single Site Using SCADA Data,” in International Conference on Advanced Technology and Science, Konya/Turkey, 2016, p. 60.

SEÇKINER, S. U., AKALIN, H., **EROĞLU, Y.,**, “Performance Analysis of Wind Turbines,” in 28th European Conference on Operational Research, Poznan / Poland, 2016, p. 160.

EROĞLU, Y., SEÇKINER, S. U., “Modelling of Wind Speed and Generated Power of a Wind Turbine by Using Particle Filtering to Monitor a Wind Farm,” in 28th European Conference on Operational Research, Poznan / Poland, 2016, p. 160.

EROĞLU, Y., SEÇKİNER, S.U., "Optimum Design of Lifting Tasks Considering Revised NIOSH Equation by Using Harmony Search Algorithm", Joint Symposium on IMSS14 and CIE44, 14 – 16 Ekim 2014, İstanbul, Türkiye

EROĞLU, Y., SEÇKİNER, S.U., "Text Mining For Systematic Literature Review on Wind Energy", The Second European Workshop on Renewable Energy Systems - EWRES, 20 – 22 Eylül, 2013, Antalya, Türkiye

EROĞLU, Y., SEÇKİNER, S.U. , “Power Optimization of Wind Farm with Ant Colony Optimization”, 2nd International Conference on Nuclear Renewable Energy Resources, 4 – 7 Temmuz 2010, Ankara, Turkey

EROĞLU, Y., SEÇKİNER, S.U. , "Wind Farm Layout Design Optimization: Continuous versus Discrete Search Spaces", 25th Conference of European Chapter on Combinatorial Optimization, 26 – 28 Nisan, 2012, Antalya, Turkey

National Congress and Conferences

EROĞLU, Y., SEÇKİNER, S.U., MITIŞ, M., "Mobil uygulamaların veri madenciliği yöntemi ile analizi", Yöneylem Araştırması ve Endüstri Mühendisliği 34. Ulusal Kongresi, 25 – 27 Haziran 2014, Bursa, Türkiye

EROĞLU, Y., SEÇKİNER, S.U., MITIŞ, M., "Metin madenciliği yöntemi ile mobil uygulamalara genel bir bakış", Yöneylem Araştırması ve Endüstri Mühendisliği 34. Ulusal Kongresi, 25 – 27 Haziran 2014, Bursa, Türkiye

EROĞLU, Y., SEÇKİNER, S.U., ÜLKER, E., "Güneş enerjisi alanında yapılan çalışmaların metin madenciliği kullanılarak analizi", Yöneylem Araştırması ve Endüstri Mühendisliği 34. Ulusal Kongresi, 25 – 27 Haziran 2014, Bursa, Türkiye

SEÇKİNER, S.U., **EROĞLU, Y.,** "Gürültü Maruziyetini Sınırlandıran İş Çizelgeleri", 19. Ulusal Ergonomi Kongresi, 27 – 29 Eylül 2013, Balıkesir, Türkiye

EROĞLU, Y., SEÇKİNER, S.U., ŞİRİN, Y., GEYİK, F., "Nano Teknoloji Uygulamalarının Sağlık ve Güvenlik Açısından Metin Madenciliği ile Değerlendirilmesi", 18. Ulusal Ergonomi Kongresi, 16 – 18 Kasım, 2012 Gaziantep, Türkiye

EROĞLU, Y., ŞİRİN, Y., GEYİK, F., SEÇKİNER, S.U., "Güneydoğu Anadolu Bölgesi'ndeki Organize Sanayi Bölgelerinin Enerji Tüketimi Etkinlik Analizi", Yöneylem Araştırması ve Endüstri Mühendisliği 32. Ulusal Kongresi, 20 – 22 Haziran 2012, İstanbul, Türkiye

SEÇKİNER, S. U., **EROĞLU, Y.,** TORAMAN, N. , "Farklı Fakültelerdeki Öğrencilerin Görme Yeteneklerinin Ölçülmesi ve Sonuçların Veri Madenciliği ile Değerlendirilmesi", 17.Ulusal Ergonomi Kongresi, 14-16 Ekim 2011 Eskişehir, Türkiye

EROĞLU, Y., SEÇKİNER, S.U. , "Jeotermal Enerji Problemleri ve Çözümlerine Genel Bir Bakış", Yöneylem Araştırması ve Endüstri Mühendisliği 31. Ulusal Kongresi, 5 – 7 Temmuz 2011, Sakarya, Türkiye

WORKSHOPS

Public Workshop on Big Data Analytics and Security, 11 February 2016 / Ankara / Turkey

CO@Work 2015 / Combinatorial Optimization at Work, 28 September - 10 October 2015 / Berlin / Germany

FRICO 2015 / 19th Future Research in Combinatorial Optimization, 11 - 14 August 2015 / Cologne / Germany

EWEA's Technology Workshop on Analysis of Operating Wind Farms, 9 - 10 December 2014, Malmö / Sweden, sponsored by TUBİTAK (Program Code: 2224 - A)

The Second European Workshop on Renewable Energy Systems (EWRES 2013), 7.5 ECTS, 20 - 29 September 2013, Antalya, Turkey, Partially sponsored by ERASMUS (ERASMUS IP: 2013 - 1 - TR1 - ERA10 - 48722)

SCHOLARS AND AWARDS

01 April - 30 September 2015 Research on "Control algorithms for wind energy systems" at Department of Electrical and Power Systems of Duisburg Essen University / 2214 - A TUBITAK International Doctoral Research Fellowship Programme / Germany (6 months)

University of Gaziantep / Graduate School of Natural and Applied Science/ M.Sc.Thesis Honor Award of the Years 2011-2012

FOREIGN LANGUAGE

ENGLISH (YDS 2016 Autumn Exam Result is 80)

HOBBIES

Reading science fiction, exploring new research areas, writing poems, painting, biking, jogging