Faculdade de Engenharia da Universidade do Porto

Methodology to Support Optimization of Wind Power Production

Hanna V. Denysyuk

Mestrado em Engenharia de Software

Supervisor: Luís F. Teixeira

Co-Supervisor: Filipa Magalhães

August 1, 2018

Methodology to Support Optimization of Wind Power Production

Hanna V. Denysyuk

Mestrado em Engenharia de Software

August 1, 2018

Abstract

Wind energy has become one of the leading sources of renewable energy. It is regarded as a climate-friendly and cost-effective way to address the growing energy needs of the world population.

Unfortunately, faults and unscheduled shutdowns of wind turbines are costly. As the size and number of wind turbines continue to grow, monitoring for faults has become increasingly important for companies to remain competitive. Moreover, *early prediction* of status patterns may allow for predictive maintenance actions and possible avoidance of some faults. This can be achieved through identification of status patterns of wind turbine, as the health of a system or component may deteriorate gradually, rather than fail instantly.

Power curve model (or simply, power curve) of a wind turbine represents the expected relationship between output power and wind speed. The objective of a power curve is to accurately predict power output, given wind speed and other conditions of operation. Power curve model is an important characteristic of the turbine since it is used in energy assessment, warranty formulations, and performance monitoring of the turbines. Power curve models are also important for turbine health monitoring. Given an accurate power curve model, performance deviation can be detected by comparing the predicted output to the actual output, adjusted for the current conditions (typically collected in real time by a Supervisory control and data acquisition, or *SCADA*, system).

However, with the growth of the wind industry, turbines are being installed in diverse and complex terrains causing significant departure of the actual curves from the respective models. This motivates deeper research into more accurate power curve models.

The contribution of this thesis is twofold.

First, we compare the anomaly detection power of three State-of-the-Art power curve models: *i)* Data filtering by binned $mean \pm 2.57\sigma$ criterion, *ii)* Data filtering using KNN classifier method, Data filtering using INEGI internal methodologies. The power curves were reconstructed using historical wind turbine data. We analyzed the advantages and disadvantages of the three approaches. Based on this analysis, we then introduce an improved power curve model called Data filtering using *KNN&Bin* method. Our model outperforms all three previous models on the historical data used for this research. Given its practical application, our model will be suggested to INEGI.

Second, we address the problem of real-time status reporting of a wind turbine, based on status classification used at INEGI. Our approach uses data mining and several machine learning algorithms to identify and predict status patterns of turbines. For this purpose were evaluated four different machine learning algorithms for classification: *i)* Decision Tree, *ii)* Nearest Neighbors, *iii)* Multi-layer Perceptron and *iv)* Gaussian Naive Bayes. The prediction models were trained on operational and status data collected from SCADA systems of 13 wind turbines with 3 different models, and the recorded data of

the closest weather stations. We compared these algorithms with regard to their accuracy and efficiency. We showed that the Decision Tree based approach achieved the best performance in our setting. As a result, this approach was selected for further deployment in production at INEGI.

Keywords: Wind turbine, Fault detection, Data mining, Machine learning, Power curve, SCADA.

Acknowledgements

I would like to express my gratitude to everyone who contributed, in different ways, to the completion of this work.

Firstly, I would like to express a special thanks to my academic supervisor Professor Luís Teixeira at the Engineering Faculty at Porto University, and my industry supervisor Filipa Magalhães at INEGI. This thesis would not have been possible without their guidance, support, wisdom and patience.

Also, I would like to thank João Rio for his invaluable advice and support during my internship at INEGI.

Finally, I would like to express my heartfelt appreciation and love to my parents, Tetyana and Vitaliy, sister Oksana and brother Roman. I am forever grateful for their unconditional support.

Apologies to those whose names I have missed.

Hanna V. Denysyuk

"Look deep into nature,
and then you will understand everything better."

Albert Einstein

Contents

List of Figures

List of Tables

Abbreviations

LCA	Life Cycle Assessment
UK	United Kingdom
RE	Renewable Energy
RPM	Revolutions per Minute
WT	Wind Turbine
NPV	Net Present Value
IRR	Internal Rate of Return
MW	Megawatt
kW	Kilowatt
SCADA	Supervisory Control and Data Acquisition System
SPU	Signal Processing Units
HVAC	Heating, Ventilation, and Air Conditioning system
NSET	Non-linear State Estimation Technique
AI	Artificial Intelligence
PHM	Prognostic Health Management
WECs	Wind Energy Converters
CMS	Condition Monitoring System
O&M	Operation and Maintenance
DM	Data Mining
CRISP-DM	Cross Industry Standard Process for Data Mining

Chapter 1

Introduction

"Of all the forces of nature,
I should think the wind contains the greatest amount of power."

Abraham Lincoln

Energy is available in two different alternatives: non-renewable (coal, fuel, natural gas) and renewable (solar, wind, hydro, wave) sources. Especially, after the industrial revolution, in the 19th century, first coal and then fuel oil are used as primary energy sources for the needs of modern communities.

It is widely understood that fossil fuels have limited potential and, at current rates of exploitation, they are expected to deplete within the next centuries. This is one of the reasons why clean, sustainable and environmentally friendly alternative energy resources are currently being sought. The accumulation of carbon dioxide in the lower layers of the atmosphere gives way to climate change, floods, intensive rainfalls and droughts. In order to reduce these dangerous effects, it is the responsibility of each country to improve the quality of the energy resources, and if possible, to replace fossil fuels (coal and oil) with renewable alternatives wind, solar and other energy sources.

Faced with energy crises in 1973, Europe began to search for their own clean and renewable energy (RE) sources (wind, solar, biomass, etc.) which are effective but they must inevitably compete against the conventional and cheaper energy. In this competition, energy sources with huge and renewable raw materials have the advantage in the long run. Atmospheric environment is polluted due to thermoelectric power plants, and petroleum materials since the industrial revolution. The pollution crises are the catalysts for the search and development of RE sources.

The environmental impact of fossil fuels, in the form of air pollution, acid rain and greenhouse effects in addition to their limited availability gave added importance to the use of conventional and renewable alternative energy sources, such as solar, wind and solar-hydrogen energies etc. Recently, renewable and clean energy generation technological developments facilities became available on the energy market. Among the renewable alternatives, wind energy has an important potential role, and wind-power farms are becoming widely used all over the world.

1.1 Wind energy

Wind energy is a converted form of solar energy which is produced by the nuclear fusion of hydrogen (H) into helium (He) in its core. The H $\rightarrow$ He fusion process creates heat and electromagnetic radiation streams out from the sun into space in all directions. Though only a small portion of solar radiation is intercepted by the earth, it provides almost all of earth's energy needs.

Wind energy can be utilized for a variety of functions ranging from windmills to pumping water and sailing boats. With increasing significance of environmental problems, clean energy generation becomes essential in every aspect of energy consumption. Wind energy is very clean but not persistent for long periods of time. In potential wind energy generation studies, fossil fuels could be supplemented by wind energy. There are many scientific studies in wind energy domain, which have treated the problem with various approaches [11], [82]. General trends towards wind and other RE resources increased after the energy crises of the 20th century [71].

Wind energy represents a mainstream energy source of new power generation and an important player in the world's energy market. During the last decade, wind energy is developed and extended to industrial use in some European countries including Germany, Denmark and Spain. Their success in wind energy generation has encouraged other countries to consider wind energy also in their electricity generation systems. Its clean, economic, practical and renewable interaction with the environment soon draw attention from political, business circles and individuals. The concept of RE is utilized since 5000 BC, and wind energy is one of the oldest of these energy sources. Today, humanity is attempting to rediscover its lost or forgotten energy sources [65]. Opinion surveys from Europe indicate that most people support wind energy uses. Over 2000 year, water and windmills powered the world's first industries with new technology and materials.

Modern wind turbines are used to generate the clean electricity needed for lighting, heating, refrigeration and other uses. Wind energy is a rather young industry, but one which already makes good economic sense. It is a proven success and its use is increasing and the downward trend in its costs is expected to continue. Already over 20,000 turbines are producing worldwide electricity. Most are operating in "wind farms" as groups of wind turbines generating electricity on a significant scale. Single wind turbines are also being used for generating electricity, charging batteries, driving pumps and producing heat. In the search for RE power, the most important decisions are concerned with exploitation of local, clean and sustainable energy resources. The determination of wind energy potential depends very much on the meteorological measurements of the wind direction, velocity, temperature and pressure. Unfortunately, in many parts of the world, it is difficult to obtain such data.

The physics behaviour of wind shows great temporal and spatial variability. In meteorology, wind is air in motion, whose driving force is the uneven heating and cooling of the earth's surface. The horizontal movement of air parallel to the earth's surface is a measure of the wind in both direction and magnitude, which change most frequently. As a result, wind prediction is one of task needed due to random change both in wind direction and speed. This changeability adds another measure importance to wind power.

Wind power has unique characteristics for energy technology. The most significant impact on the environment is the visibility of wind turbines. Those who support the movement towards clean, sustainable energy production should take into consideration aesthetically pleasing symbols of a better future, especially when compared with the effects of acid rain, global climate change, radioactivity, land and water contamination in addition to other environmental problems associated with conventional energy sources.

History of wind energy conversion along with its present status and future prospects are discussed briefly in the following sections.

1.2 History and evolution

Despite wind being a virtually unlimited resource, at no time have humans employed more than a tiny fraction of its kinetic energy. Throughout cultures the world have paid homage to the power of wind in their legends and mythologies. Creation stories of nearly all cultures involve the power of the wind. Like the mystery of who invented the wheel, we do not know when humans first employed the force of wind to help in the work. Historians speculate that, in terms of utility, sailing seems to be one of the most likely early uses of winds, around 40 000 years ago, when humans migrated from Asia to greater Australia. Although we know little of these voyages, the far later water journeys of Polynesians in their double-hulled sailing craft are much better studied. For these early sailors, wind provided the kinetic energy for the exploration and settlement of the Pacific Ocean islands [51].

Another example of the first historic use of wind power involved its use in transportation by the Egyptians, who plied the Nile River as early as 3100BC using craft equipped with sails of linen and papyrus. In one of the major Greek poems, *The Odyssey* (attributed to Homer), Odysseus sailed the Ionian and Aegean seas, eventually making a crucial error when he angered the God of the Winds. This story reminds one of the capricious nature of the wind, that cannot be put completely under human control. For this reason, wind has been successfully employed only sparingly, with inventors and engineers preferring power sources that they can manipulate as needed. Yet, the importance of wind energy to human history is well understood. From the time of Columbus until the middle of the 19th century, nations depended on sailing vessels. Advances in navigation and sail design combined to make the great age of the sailing craft feasible [27].

Whereas historically wind power was mostly employed at the sea, today it finds most of its application on the land. The importance of sailing has faded as other means of transportation has been largely adopted, the terrestrial applications of wind have multiplied. We do not know exactly where and when people first turn to the wind to help in their works. However, artifact proof and written evidence suggests that, as early as in the 10th century A.D., windmills appeared in the Sistan region of Persia (now Iran). These were primitive windmills by modern standards, using vertical sails of reed bundles, and their owners built them to grind grain and lift water from streams to irrigate gardens. For centuries, these rudimentary machines were carried to other parts of the Wold, including India and China, where farmers employed them to pump water, grind grain and crush sugarcane.

By the 13th century, grain grinding mills were popular in most of Europe. The French adopted this technology by 1105 A.D. and the English by 1191 A.D. In contrast with the vertical axis Persian design, European mills had horizontal axis. These post mills were built with beautiful structures. The tower was circular or polygonal in cross-section and constructed in wood or brick. The rotor was manually oriented to the wind by adjusting the tail. The mill was protected against high winds by turning the rotor out of the wind or removing the canvas covering the rotor. The Dutch, with renowned designer Jan Adriaenszoon, were the pioneers in making these mills. They made many improvements in the design and invented several types of mills. Examples are the *tjasker* and *smock mills*. The rotors were made with crude airfoil profile to improve the efficiency. Apart from grain grinding, wind mills were employed to drain marshy lands in Holland. These wind mills reached America by mid-1700, through the Dutch settlers.

This is followed by the water pumping wind mill, which is still considered as one of the most successful application of wind power. An example of this is multi-bladed wind turbine that appeared in the wind energy history by the mid-1800. Relatively smaller rotors, ranging from one to several meters in diameter, were used for this application. The primary motive was to pump water from a few meters below the surface for agricultural uses. These water pumpers, with its metallic blades and better engineering design, offered good field performance. Over six million of such units were installed in the United States between 1850 and 1930.

The era of wind electric generators began close to 1900's. The first modern wind turbine, specifically designed for electricity generation, was constructed in Denmark in 1890. It supplied electricity to the rural areas. During the same period, a large wind electric generator having 17 m *picket fence* rotor was built in Cleveland, Ohio. For the first time, a speed-up gear box was introduced in the design. This system operated for 20 years generating its rated power of 12 kW. More systematic methods were adopted for the engineering design of turbines during this period. With low-solidity rotors and aerodynamically designed blades, these systems could give high field performance. By 1910, several hundreds of such machines were supplying electrical power to the villages in Denmark. By 1925, wind electric generators became commercially available in the American market. Similarly, two and three bladed propeller turbines ranging from 0.2 to 3 kW of capacity were available for charging batteries.

Intensive research on the behaviour of wind turbines occurred during 1950's. The concept of high tip speed ratio-low solidity turbines got introduced during this period. For example, light-weight constant-speed rotors were developed in Germany in 1968. They had fiber glass blades attached to simple hollow towers supported by guy ropes. The largest was of 15 m in diameter with a rated output of 100 kW [53].

The oil crisis in 1973, forced the scientists, engineers and policy makers to to explore ways to ease fossil fuel dependence. In addition to the fact the fossil fuel reserve will eventually be exhausted, it was realized that political tampering can restrict the availability and escalate the cost of fossil fuels in the meantime. Also, nuclear power was not considered a viable alternative due to safety reasons. These factors caused the revival of interest in wind energy.

1.3 Current status

Modern wind turbines tower above one of their ancestors-an old windmill used for pumping water. Humans have been harnessing the wind's energy for hundreds of years. From old Holland to farms in the United States, windmills have been used for pumping water or grinding grain. Today, the windmill's modern equivalent - a wind turbine - can use the wind's energy to generate electricity. Wind turbines, like windmills, are mounted on a tower to capture the most energy. At 30 meters or more above ground, they can take advantage of the faster and less turbulent wind. Turbines catch the wind's energy with their propeller-like blades. Usually, two or three blades are mounted on a shaft to form a rotor.

A blade acts much like an airplane wing. When the wind blows, a pocket of low-pressure air forms on the downwind side of the blade. The low-pressure air pocket then pulls the blade toward it, causing the rotor to turn. This is called lift. The force of the lift is significantly stronger than the wind's force against the front side of the blade, which is called drag. The combination of lift and drag causes the rotor to spin like a propeller, and the turning shaft spins a generator to make electricity.

Wind turbines can be used as stand-alone applications, or they can be connected to a utility power grid, or even combined with a photo-voltaic (solar cell) system. For utility-scale sources of wind energy, a large number of wind turbines are usually built close together to form a wind plant. Several electricity providers today use wind plants to supply power to their customers.

Stand-alone wind turbines are typically used for water pumping or communications. Moreover, homeowners, farmers, and ranchers in windy areas also use wind turbines as a way to reduce their electric bills. Small wind systems also have potential as distributed energy resources. Distributed energy resources refer to a variety of small, modular power-generating technologies that can be combined to improve the operation of the electricity delivery system.

1.3.1 Wind turbine design

The design and size of a turbine play a crucial role in electricity generation. Maximum wind capture and cost reductions are two primary motives of wind turbine research. Continuous research and advancement of turbine technology have taken place over the last several decades. As a result, nameplate capacity rating has risen enormously. Today, commercially available wind turbines have ratings ranging from several kilowatts to megawatts. The diameter of the turbine is an important parameter. The recent trend is toward large turbines, as longer blades sweep wind from a larger area and produce greater output energy.

Wind turbines come in two types: horizontal axis and vertical axis (Figure 1.1). Horizontal axis turbines are the more familiar *windmill* type where the blades rotate in a vertical plane about a horizontal axis and the turbine is dynamically rotated on its tower to face the wind. Vertical axis turbines do not need orientation into the wind, although the earlier versions, sometimes known as *eggbeater* turbines required a power source to start rotating because of their high torque. More recent innovations have helical blade designs that have low torque and can operate without external power. Vertical axis turbines are particularly suited to small wind power applications because they have a small environmental impact and no noise, but have not yet scaled up to the 5MW + turbine size of horizontal axis designs.

A wind turbine typically consists of blades, a rotor, a tower, a gearbox and a generator. Figure 1.2 shows all the components of a wind turbine [39], which are explained below.

Rotor. A rotor consists of large blades resembling an airplane wing. Three blades for a turbine are universally accepted, but two blade turbines are also functional. Rotor blades are very large in size. Recently, Siemens has launched 246 ft long B75 rotor blades. These will be installed into a prototype 6 MW offshore wind power system at Osterild test station in Denmark. Another component called the "pitch drive" is used to reduce the effect of lift forces in high wind speed conditions. This is necessary to guarantee that the alternator maintains a speed within a stable power system operation range of 1000–3600 RPM (revolutions per minute).

Nacelle. The nacelle is located at the top of the turbine tower. It is attached to the rotor, and contains the main technical parts, such as the rotor shaft, gearbox, and

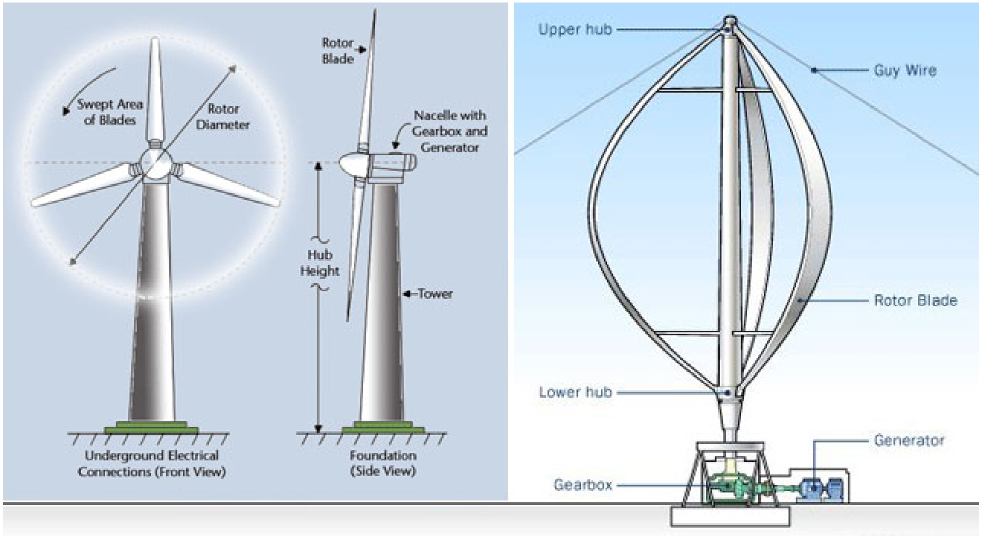


Figure 1.1: Illustration of two types wind turbine [22]
(left - horizontal axis wind turbine; right - vertical axis wind turbine).

generator. The nacelle is connected to the tower with bearings and is able to rotate with respect to the wind direction in order to harness maximum wind energy.

Gearbox. The turbine rotor typically has a speed of less than 100 RPM, but most generators need 1000–3600 RPM to generate electricity. Thus, the gear box converts low rotor speed into higher speeds in order to make the generator operational.

Generator. The generator converts the mechanical energy of the rotor into electrical energy.

Tower and foundation. Speed, quality (less turbulence and turmoil), and quantity of wind increase with the increment of height. A tower is used to place the rotor at high altitudes in order to capture more wind energy.

Additionally, a controller, anemometer, heat exchanger, and wind vane are other important components in a wind turbine. The controller is a computer operated system that controls the turbine's operation, the heat exchanger cools the generator, the anemometer measures wind speed, and the wind vane detects wind direction.

1.3.2 Wind turbine classification

Wind turbines can be classified according to the turbine generator configuration, airflow path relatively to the turbine rotor, turbine capacity, the generator-driving pattern, the power supply mode, and the location of turbine installation. Below we discuss these classifications in detail.

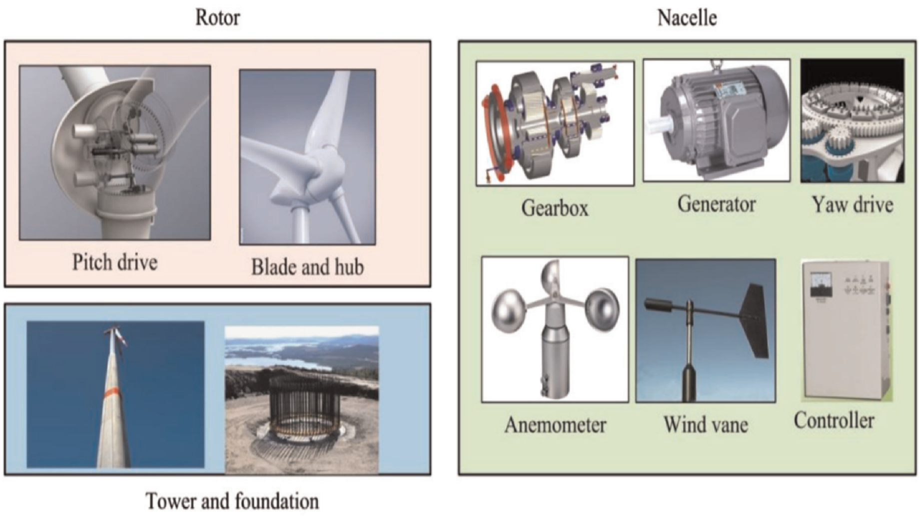


Figure 1.2: Illustration of the main components of wind turbine [39].

1.3.2.1 Wind turbine capacity

Based on their rated capacities, wind turbines can be divided into a number of broad categories: micro, small, medium, large, and ultra-large wind turbines.

Micro turbines. Though a restricted definition of micro wind turbines is not available, it is accepted that a turbine with the rated power less than several kilowatts can be categorized as micro wind turbine. Micro wind turbines are especially suitable in locations where the electrical grid is unavailable. They can be used on a per-structure basis, such as street lighting, water pumping, and residents at remote areas, particularly in developing countries. Because micro wind turbines need relatively low cut-in speeds at start-up and operate in moderate wind speeds, they can be extensively installed in most areas around the world for fully utilizing wind resources and greatly enhancing wind power generation availability.

Small turbines. Small wind turbines usually refer to the turbines with the output power less than 100 kW. Small wind turbines have been extensively used at residential houses, farms, and other individual remote applications such as water pumping stations, telecom sites, etc., in rural regions. Distributed small wind turbines can increase electricity supply in the regions while delaying or avoiding the need to increase the capacity of transmission lines.

Medium turbines. The most common wind turbines have medium sizes with power rating from 100 kW to 1 MW. This type of wind turbines can be used either on-grid or off-grid systems for village power, hybrid systems, distributed power, wind power plants, etc.

Large turbines. Megawatt wind turbines up to 10 MW may be classified as large wind turbines. In recent years, multi-megawatt wind turbines have become the mainstream of the international wind power market. Most wind farms presently use megawatt wind turbines, especially in offshore wind farms. Ultra-large wind turbines are referred to wind turbines with the capacity more than 10 MW. This type of wind turbine is still in the earlier stages of research and development.

1.3.2.2 Direct drive and geared drive wind turbines

According to the drive-train condition in a wind generator system, wind turbines can be classified as either direct drive or geared drive groups. To increase the generator rotor rotating speed to gain a higher power output, a regular geared drive wind turbine typically uses a multi-stage gearbox to take the rotational speed from the low-speed shaft of the blade rotor and transform it into a fast rotation on the high-speed shaft of the generator rotor. The advantages of geared generator systems include lower cost and smaller size and weight. However, utilization of a gearbox can significantly lower wind turbine reliability and increase turbine noise level and mechanical losses.

By eliminating the multi-stage gearbox from a generator system, the generator shaft is directly connected to the blade rotor. Therefore, the direct-drive concept is superior in terms of energy efficiency, reliability, and design simplicity.

1.3.2.3 On-grid and off-grid wind turbines

Wind turbines can be used for either on-grid or off-grid applications. Most medium-size and almost all large-size wind turbines are used in grid tied applications. One of the obvious advantages for on-grid wind turbine systems is that there is no energy storage problem.

As the contrast, most of small wind turbines are off-grid for residential homes, farms, telecommunications, and other applications. However, as an intermittent power source, wind power produced from off-grid wind turbines may change dramatically over a short period of time with little warning. Consequently, off-grid wind turbines are usually used in connection with batteries, diesel generators, and photovoltaic systems for improving the stability of wind power supply.

1.3.2.4 Onshore and offshore wind turbines

Onshore wind turbines, a category of turbine that is installed on land, has 50–100 m tower heights with a rotor diameter of 50–100 m. The general trend in wind turbine designs is to increase tower height and rotor blade length. A combination of high pole and long blades allows wind turbines to be installed in areas with low wind energy potential. Modern turbines work on a rotor and hub assembly speed of 12–20 RPM, much lower than those installed during the 1980s, which operated at a common speed of 60 RPM. As a result,

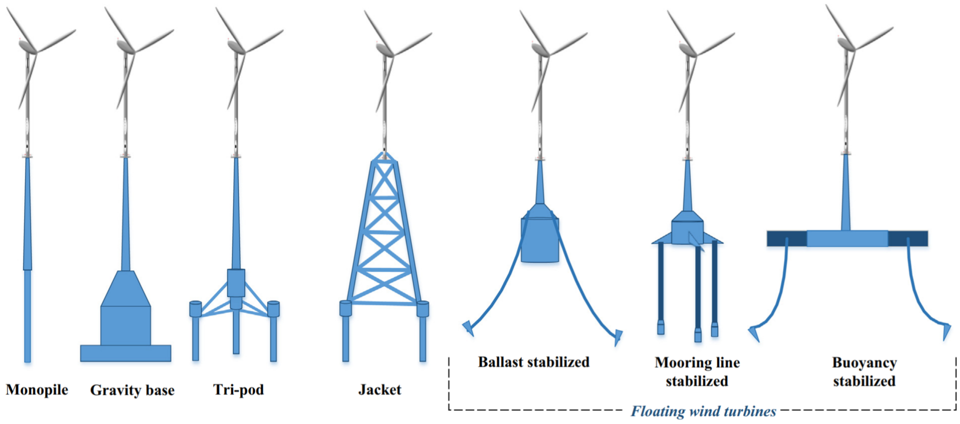


Figure 1.3: Different foundation types for offshore wind turbines [39].

modern turbines are capable of effectively generating power at much lower wind speeds. Additionally, these turbines have a significantly higher electricity generation capacity in comparison to older models. In the present day, storm controlled techniques enable wind turbines to operate even during very high wind speed conditions. Onshore wind turbines are typically grouped together into wind power plants, commonly known as wind projects or wind farms. These wind power plants are usually 5–300 MW in size, although smaller and larger plants are also in operation [39].

Wind turbines installed beyond the coast are known as offshore power systems. Development of offshore wind energy has accelerated in the past few years due to significant wind resources available over the oceans. Offshore wind flows with higher speed and more uniformity than on land. Offshore sites have the ability to establish larger power plants with larger wind turbines. Offshore wind turbines usually have more nameplate capacity ratings than onshore ones. Moreover, the world's largest cities are generally situated in coastal areas; therefore, longer distance power transmission can be avoided.

Offshore turbine technology bears striking similarity to that onshore with only minor modifications. The only significant difference is in the design of the foundations, which requires floating and/or other special foundations to account for underwater tower submergence. Major offshore wind turbine foundation types, including floating type structures, are depicted in Figure 1.3. Nowadays, many manufactures are developing floating turbines for deeper waters.

1.3.3 Wind turbine lifetime

Modern wind turbines are designed for the lifetime of 20–30 years. A critical challenge facing turbine manufacturers and wind power plants is how to achieve the lifetime goals while at the same time minimize the costs of maintenance and repair. However, improving the operational reliability and extending the lifetime of wind turbines are very difficult tasks for a number of reasons:

- Wind turbines have to be exposed to various hostile conditions such as extreme temperatures, wind speed fluctuations, humidity, dust, solar radiation, lightning, salinity and frequent onslaughts of rain, hail, snow, ice, and sandstorms.
- A modern wind turbine consists of a large number of components and systems; each of them has its own lifetime. The failure must first occur in the component or system with the shortest lifetime.
- A wind turbine is subjected to a large variety of dynamic loads due to wind fluctuations in speed and direction and numerous starts and stops of the system. Some primary parts or components have to withstand heavy fatigue loads [38].
- Advanced high-strength, fatigue-resistant materials are vital to some key components in modern large wind turbines due to the continuous increase in blade length, hub height, and turbine weight.
- A s a complex engineering system, a wind turbine must be designed at the system level rather than part/component level as a common practice in some turbine manufacturers.

1.4 Challenges to wind energy

Although wind energy is a clean and potential source of power, it has some economical, technical, social and environmental challenges. These effects may be negligible but they persist for a long time and hence directly affect the neighbouring localities; therefore, care must be taken.

1.4.1 Economic challenges

- *High capital investment:* Wind energy is a capital-intensive technology since most of the expenditures are made at time of investment. The initial expenditures can be as high as 80% of the total cost of the project. Wind turbine alone constitutes the largest cost component followed by grid connection and others. The breakup of the cost of a typical wind farm is shown in Figure 1.4. Few key parameters that govern wind power costs are capital costs (cost of turbines, foundations and grid connection), variable cost (operation and maintenance, land rent, insurance, taxes), relatively lesser plant capacity factor and economic lifespan of the investment.

  For successful implementation of wind energy, developer will need most of the funds (around 80%) during the construction phase itself. Hence, it is important to have good repayment conditions. In countries like India, a 70:30 debt equity ratio is

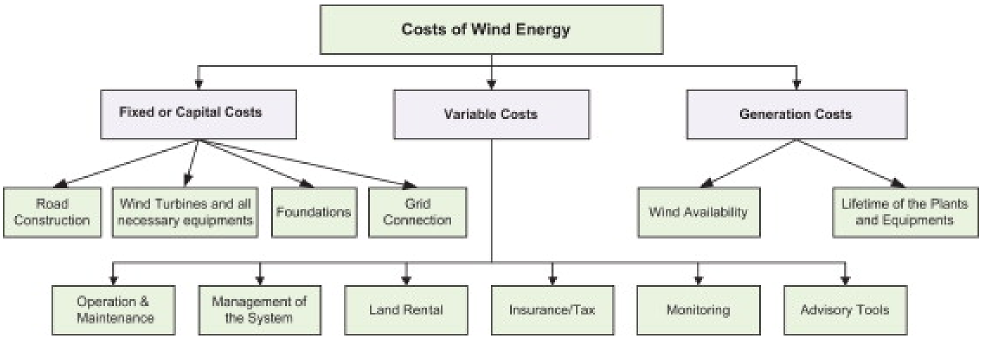


Figure 1.4: Typical breakup of overall cost of a large scale wind energy system [74].

available to many renewable energy projects. However, lesser interest rates are necessary to promote the wind energy. Due to the intermittent nature of wind (less plant capacity factor), wind power projects have higher payback periods. Hence, due to improper funding conditions during the initial phases, it may not be fruitful to harness this energy even though it may be a cheapest option. The net present value (NPV) and internal rate of return (IRR) are important parameters which are used for project investment feasibility. Petkovic [61] developed a model for economical and optimal layout of wind farm considering the interactions between turbines, various costs and wind regimes;

- *High production cost:* The production cost is determined on the basis of power produced, fixed costs (interest, land rent, insurance) and variable costs (maintenance/repair, miscellaneous). It is not always possible to implement wind energy at large scale because of the intermittent nature of wind and high investments. To reduce the production cost, power production is an important parameter which further depends on wind speed which is easily influenced by obstacles (buildings), terrain (plain, mountains) etc. Other important parameters include turbine design & selection and grid availability. The production cost is an overall parameter to assess the success or failure of the project. It can be minimized by proper planning and taking necessary actions. To improve the consistency in wind power production, power system regulators are used to make detailed schedule plans and to set reserve capacity for proper utilization of energy. In order to reduce the reserve capacity requirement, accurate forecasting methods are necessary (as stated earlier). Power output (wind farms) also varies due to the constant variations in the outputs of each wind turbine (although all are fed from same wind energy) which leads to a lesser plant capacity factor [35]. The lower plant capacity factor indicates lesser outputs thereby increasing the overall production costs. In order to decrease the production costs, it is important to increase the wind power penetration which will ultimately lead to more energy generation. This may be done by proper designing & selection, use of power regulators, flexible AC transmission systems (grid) etc. However, these will incur high costs which will further increase the production costs to some extent. Hence, it is possible to optimize the wind power plant only to a certain extent.

- *Risk:* Investors must consider whether investment in wind is profitable or not. Although the arguments for wind energy benefits are plentiful, arguments against are simple: at the present time, electricity produced by conventional energy sources (coal and gas) is more reliable and even in some places available at lower prices. However, prices of wind power are decreasing steadily. If it compared over a long period of time with all subsidies and incentives taken in to consideration, wind power would give better return on investment.

1.4.2 Technical challenges

Wind farms are usually located in rural areas due to the availability of land, higher wind speed and possibility of other works such as farming etc. There are two main problems faced in wind energy generation with respect to grid. Firstly, in many of the rural areas there are limitations of grid infrastructure and secondly, even if there is stable grid present, integration of wind energy into the grid leads to potential technical issues such as voltage fluctuations etc. due to the variations in wind energy.

Due to the limitation of grid infrastructure, most of the energy generated (if not transmitted effectively) is wasted. The availability of wind is variable due to the weather pattern, large water bodies, clouds (preferential heating), day & night cycles, storms/turbulence etc. The electricity output from wind farms/mills depends on these parameters, however, the demand does not depend on these variations. In-fact, many times it is just the opposite i.e. when the power supply is at its peak (night), the demand is very less. These losses can be minimized by using batteries/power regulators, etc., but it will be a costly supplement. In many developing countries, high borrowing cost is already creating an obstacle for wind energy sector growth. Even though project financing is available for majority of wind power projects with 70: 30 debt equity ratio but high interest rate under difficult macroeconomics conditions create problems in implementing this technology. Hence, an effective grid infrastructure is essential for wind energy.

Even if the grid is present, integration of energy generated in wind farms poses several technical issues due to the intermittent nature of wind thereby affecting the power quality. Main parameters that affect the power quality includes voltage fluctuations, power system transients and harmonics, reactive power, low power factor, electromagnetic interference, synchronization etc. [73]. Variations in voltage and grid frequency create difficulties in wind farm operations and reduce the chances for successful integration of wind energy into grid. Some of the major power quality problems encountered in wind farms due to the variations in wind energy generation are:

1. Uncontrollable reactive power and low power factor;

2. Power fluctuations and voltage distortion;

3. Voltage fluctuations & significant line losses.

To confront future wind power grid integration, we have to work on design and operation issues of the power systems with the introduction of demand side management, energy storage techniques [57], grid infrastructure issues (reinforcement and upgrade of networks), introduction of more flexible mechanisms and other issues [29]. The traditional approach of management necessitates addition of a complementary power plant to ensure supply at all times. This is done to compensate the indeterminacy of wind energy by using a more controllable source such as hydro, diesel, thermal power plants etc. The problem of grid integration can also be solved by connecting it to the energy storage systems to have an additional reserve of energy which will act as a buffer between the producer and consumer. The problem of grid integration can also be solved by using power electronics concepts. During strong wind periods there is an excess of electrical power (destabilizing of frequency), hence, it is important to limit the power produced by the wind turbines by either using electronic components or by changing the pitch angle of the blades to reduce the performance of the rotor. However, this area is under research and is being studied by many scientists/researchers worldwide. Furthermore, the reactive power produced by a wind turbine can varied through the command of inverters associated to the generator. It is possible to absorb or to supply reactive power and control the voltage level of the grid [6]. This option has been integrated in many wind turbines using doubly fed induction machines.

1.4.3 Social-environmental challenges

Human intervention of any nature has its own environmental consequences. Wind energy is not an exception. Although wind is one of the cleanest sources of energy and does not pollute the environment with harmful gases during its energy conversion process, wind energy systems pose some social and environmental problems. Visual impacts on landscape, noise pollution, radar and telecommunication interference (if within line-of-sight), and hazards to wildlife are the main socio-environmental constraints to wind turbines [66].

- *Noise pollution:* Like any other rotating mechanical systems, wind turbines also create some noise during their operation. The noise was a severe problem with the turbine designs in the 80's. During that time, environmental impacts of wind energy was not a matter of concern as it is today. Some of the turbines constructed during this period were quite noisy and could annoy people even at far away distance. In the later years, the noise emission from wind turbines has attracted more attention from the environmental groups, regulatory authorities and the wind turbine industry. A number of design modification followed in the preceding years and as a result, the modern wind turbine is a much quieter piece of machinery.

Any unwanted sound can be considered as noise. The noises generated by a wind turbine may be tonal, broad band, low frequency or impulsive. Tonal noises have discrete frequencies whereas the broad band noises have continuous frequency above a level of 100 Hz. The turbine may also produce noises with low frequency ranging from 20-100 Hz along with those due to momentary acoustic impulses. The magnitude of noise can be expressed either in terms of sound power level or sound pressure level. The sound power level indicates the acoustic power with which the noise is emitted from the source whereas the sound pressure tells us the intensity of noise experienced by the listener located at a given point. Two types of noises are generated from a wind turbine-mechanical noise and aerodynamic noise [53].

The mechanical noise is contributed by the relative motion of components of WTs (gear box, generator, yaw motors, cooling fans, hydraulic pumps and other accessories). Changes in the design of these components can significantly affect the frequency and tone of the noise. Mechanical noises are mostly of tonal type.

Flow of air around the blade causes the aerodynamic noise. When the blades of the wind turbine interacts with the air stream, a number of complex flow phenomena occurs around the blade, each contributing to the noise generated from the system. The aerodynamic noise is mostly of broad band type with some amount of low frequency or even tonal components. A more rigorous analysis of the aerodynamic noise, critically analyzing various contributing factors can be found in [63].

Noise pollution from wind turbines is not a serious issue as it is being considered. In most of the cases, the noise from wind turbines does not exceed the limits put forth by the permitting agencies. For example, a wind turbine with moderate sound power emission may be as quiet as a kitchen refrigerator at a residential area 300 m away from the source. Apart from annoyance in few cases, no health problems are reported due to the wind turbine noise. A study conducted in Europe, incorporating sixteen sites from Denmark, Germany and Netherlands showed that only 6.4 per cent of the residents felt that the noise from wind turbines as annoying [63]. With proper planning and improvements in the design, the noise emission from wind turbines will further be reduced in the coming years.

- *Visual impacts:* Another environmental concern of wind farm development is its impact on scenic beauty of the landscapes. Wind turbines are tall structures installed in open areas which make them visually prominent in the landscape. The turbines may dominate our sight up to 2 km or even more. They are often felt to be an important element in the landscape even up to a distance of 5 km. As “beauty lies in the viewer’s eyes”, some may like the sight of a wind turbines, generating energy in an environment friendly way near to their area. However, we must not forget that there may be some people who might consider the turbine as "a box on a long stick" erected to ruin the scenic beauty of the landscape. Hence, the turbines should be naturally integrated to the landscape to make them visually attractive and acceptable.

The value that one assign to the landscape and its surroundings is an important factor in moulding his opinion on the wind farm. The aesthetics value of a landscape is judged in terms of its visual, historical, ecological, socio-cultural, religious and mythological importance. Hence it is advisable to assess the sensitivity of the landscape towards these factors, before going for a wind farm project. There are several methods to establish the aesthetic sensitivity of a landscape [23]. The use of these techniques in conjunction with the public opinion surveys can give us an indication on the appropriateness of a site for wind farm installation. Many local permitting authorities have already quantified the sensitivity of their landscapes using these methods, which are available for the developers for assessing the suitability of their project.

Carefully designed and cleverly constructed wind farms can be a positive addition to the landscapes. Wind farms with layouts, naturally blending with the prevailing scenery and turbines designed with aesthetic sense can add to the scenic beauty of the landscape.

Computer models are available to assess the visual impact of wind farms on the landscape and the flora and fauna, before the farms are actually constructed. The *WindFarm* software has a special photo-montage module using which the pictures of the proposed landscape and the turbines to be installed can be superimposed to simulate the visual impact of the wind farm. The software can also animate the turbines to conceptualize the view while the turbines are actually generating [53].

The degree of visual disturbance caused by wind turbines varies from person to person based on individual perception. Put simply: some people find wind turbines attractive and non-intrusive while some feel the opposite way. However, there is some evidence that flickers and moving shadows due to rotating turbine may cause adverse effects on human health, which may include stress, annoyance, and photosensitive epilepsy seizures (a kind of neurological disorder in which seizures are triggered by visual stimuli such as moving patterns, flashing lights, and shadows) [76]. Shadow casting is not a very serious problem in wind farm management as this will affect only a small area, very close to the turbine. Interference with TV and radio signals was a problem with earlier turbines with metallic blades. However, with the introduction of composite blade materials such as fibre glass and plastics, this problem is minimized in modern turbines. However, adverse visual impacts of wind turbines can be avoided by selecting turbine sites far from residential areas.

- *Wildlife Hazards:* Bird and bat fatalities are a primary ecological consideration of wind energy systems. However, the number of bird deaths attributed to wind turbines is extremely low when compared with other more common hazards such

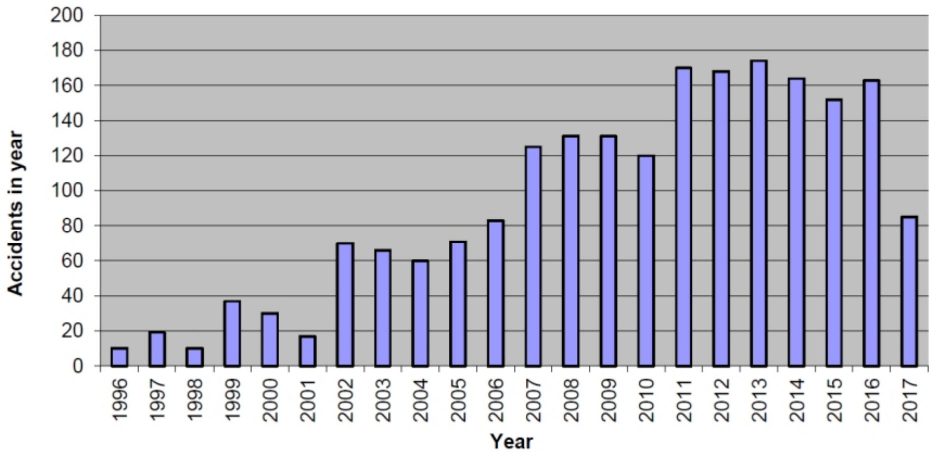


Figure 1.5: Wind turbine accidents, human fatalities, and injuries in the United Kingdom [24] (2017: To 30 September 2017 only).

as predators and buildings collisions. As estimated by J. Ryan Zimmerling et al. approximately 20,000 to 28,300 birds are killed each year due the striking of wind turbines in Canada; these mortalities are much less than the mortalities due to other causes [91]. For example, as per the Environmental Canada report on avian mortality, approximately 270 million birds died each year from human-related activities in Canada, 200 million were killed by cats, and 25 million by building collision. Besides fatalities, wind turbines may cause change in routes of migratory birds. However, the mortality rate of birds per turbine has decreased over the past few years due to special designs of rotor blades and careful monitoring. These problems can also be reduced to some extent by careful site selection [39].

- *Accidents and human casualties:* Human hazards also exist due to wind turbines. Figure 1.5 illustrates the number of accidents, human fatality, and injuries that occurred each year in the UK from the year 1997–2017.

  The trend is as expected - as more turbines are built, more accidents occur. Numbers of recorded accidents reflect this, with an average of 22 accidents per year from 1997-2001 inclusive; 70 accidents per year from 2002-2006 inclusive; 135 accidents per year from 2007-11 inclusive, and 164 accidents per year from 2012-16 inclusive [24].

The environmental impact associated with wind turbines can be evaluated by using a holistic process known as Life Cycle Assessment (LCA). LCA is a method that assesses the environmental impacts of a product over its entire life cycle [52]. Due to the inherit cleanliness of wind energy, the majority of adverse environmental emissions are incurred in the material production and manufacturing process required for creation and installation of the turbines and not during the lifespan in which the turbine is in use. Other social impacts of a wind system include land acquisition, benefit sharing, and compensation for the acquisition and use of land.

1.5 Summary

Wind energy harvesting is of prime interest today as it is the most promising renewable energy source due to its clean, environment-friendly attributes. Wind energy is not only climate-friendly and free from GHG emission, but also has cost-effective and less negative social and environmental impacts compared to other sources of energy as technology is getting more efficient and cost effective. It has the potential to reduce the energy-crisis worldwide and create employment opportunities. Wind energy is now a mature technology and there is enough evidence in favour of large-scale wind energy farms.

However, along with the positive environmental impacts, the downsides of wind energy harvesting include: social environmental impacts, such as audio and visual pollution; economic impact, such as it has high up-front costs; environmental impacts, such as death of birds, emissions during installation and future dismantling of wind farms; and technical impacts that affect the power quality of the network.

As a result, our research is motivated by the need to minimize potential negative impacts of integrating large-scale wind energy into the grid for a sustainable power system for the future. Findings of ours study are expected to be used as guidelines by the policymakers, manufacturers, industrialists and utilities for deployment of large-scale wind energy into the energy mix.

A large variety of wind turbine types and designs is available at present. Typically, efficient designs with high performance are influenced by several requirements related mainly to the application to serve and the location where the turbine is to be installed. As well, other constraints may play a role in the efficiency of the designs, e.g. size constraints, noise limitations, visual disturbances and low start-up wind speeds. Among these requirements and constrains, the performance of one particular wind turbine can better optimized compared to the others.

Chapter 2

Motivation and Objectives of Research

2.1 Motivation and Objectives

Wind energy has become one of the leading sources of renewable energy, but faults and unscheduled shutdowns of wind turbines are costly. As the size and number of wind turbines continue to rise, monitoring for faults has become increasingly important for companies to remain competitive.

Wind turbines consist of several components and sub assemblies which are likely to fail during its course of operations. Even with the advanced Supervisory control and data acquisition, or *SCADA*, systems, certain faults are difficult to characterize, or often the alarm is triggered when the fault is already occurred. Thus, fault detection is critical in identifying faults in the system in a timely manner.

Signal processing units, or *SPU*s, evaluate the process parameters and classify them into normal operations and fault situation. In this approach fault indicators are derived from process measurements via limit and trend checking of the process signals. Signal processing approaches are suitable for analyzing rotating components, i.e., turbine generators and gearboxes. Common approaches that are utilized for signal analysis in frequency domain include: Fast Fourier Transformation, spectrum analysis, envelope spectrum etc.

In system identification, the measured signal is compared against the set values. Any significant deviation from the set values indicate fault. For complex processes, where analytical models cannot be aptly applied, artificial intelligence based approaches found its scope. AI based approaches, such as machine learning approaches, can learn the complex behavior, and characterize as a fault any significant deviation in the behavior. Benefits of the AI based fault detection systems include: *i)* Avoidance of premature breakdown, *ii)*

Reduction of maintenance costs, *iii)* Remote diagnosis [13]. This insight motivates the research into utilization of machine learning based approaches in fault detection.

In the literature, turbine components, namely turbine blades, generators, and gearbox are widely researched. Description about them is provided in 4.1.

Typically, a SCADA system routinely collects wind turbine operations data which can be used for performance monitoring purposes. Even with the advanced SCADA systems, wind turbine faults are often recognized too late to perform a planned maintenance on the system. Data mining approaches are well known to extract the hidden patterns in the data. With the data-mining algorithms, faults associated with wind turbines can be identified and predicted well ahead of their occurrence. In addition, the performance of wind turbines can be continuously monitored using the operational data, such as power output, rotor speed, blade pitch angle, etc.

Over the past few years, data mining has been successfully applied in manufacturing, marketing, and medical informatics. In the energy sector, data mining based algorithms were used to forecast electricity market price [90], optimization of combustions and heating, ventilation, and air conditioning ($HVAC$) systems [40, 41, 42]. In the wind energy area, data mining based approaches are used for optimization of wind power output. At present, there is need for better solutions for wind turbine performance monitoring.

The goal of this work is to propose a performance monitoring system of wind turbines by providing accurate and robust data-mining based fault prediction models. To predict wind turbine faults, data mining and machine learning algorithms are employed to identify the relation between wind turbine performance parameters. Depending on the nature of turbine fault and available data, both classification and regressions models can be constructed. As the performance of data-mining algorithms solely depends on dataset at hand, advanced data preprocessing techniques are employed to develop robust fault prediction models.

2.2 Approach

A data mining approach with the machine learning algorithms is applied to identify and predict status patterns of wind turbines and recognize rejected data from performance curves. The identification of status patterns is an important piece of fault monitoring, as the system or component's health may gradually deteriorate, rather than failing instantly. Early prediction of status patterns may allow for predictive maintenance actions and possible avoidance of some faults. A prediction model will be trained using operational and status data collected at wind turbines.

Chapter 3

Fundamentals of Wind Energy

Energy available in wind is basically the kinetic energy of large masses of air moving over the earth's surface. Blades of the wind turbine receive this kinetic energy, which is then transformed to mechanical or electrical forms, depending on our end use. The efficiency of converting wind to other useful energy forms greatly depends on the efficiency with which the rotor interacts with the wind stream. In this chapter, let us discuss the fundamental principles involved in this wind energy conversion process.

3.1 Wind energy characteristics

Wind energy is a special form of kinetic energy in air as it flows. Wind energy can be either converted into electrical energy by power converting machines or directly used for pumping water, sailing ships, or grinding gain.

3.1.1 Wind power

Kinetic energy exists whenever an object of a given mass is in motion with a translational or rotational speed. When air is in motion, the kinetic energy in moving air can be determined as

$$E_k = \frac{1}{2}\mathrm{m}\bar{u}^2 \tag{3.1}$$

where m is the air mass and $\bar{u}$ is the mean wind speed over a suitable time period. The wind power can be obtained by differentiating the kinetic energy in wind with respect to time, i.e.:

$$P_w = \frac{\mathrm{d}E_k}{\mathrm{d}t} = \frac{1}{2}\mathrm{m}\bar{u}^2 \tag{3.2}$$

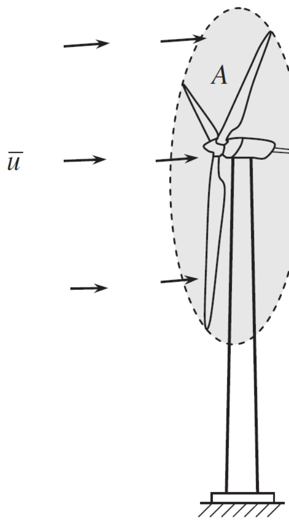


Figure 3.1: Swept area of wind turbine blades.

However, only a small portion of wind power can be converted into electrical power. When wind passes through a wind turbine and drives blades to rotate, the corresponding wind mass flow-rate is

$$m = \rho \mathrm{A}\bar{\mathrm{u}} \tag{3.3}$$

where ρ is the air density and A is the swept area of blades, as shown in Figure 3.1 . Substituting equation 3.3 into 3.2, the available power in wind P_w can be expressed as

$$P_w = \frac{1}{2}\rho A\bar{u}^3 \tag{3.4}$$

An examination of equation 3.4 reveals that in order to obtain a higher wind power, it requires a higher wind speed, a longer length of blades for gaining a larger swept area, and a higher air density. Because the wind power output is proportional to the cubic power of the mean wind speed, a small variation in wind speed can result in a large change in wind power [81].

3.1.1.1 Blade swept area

As shown in Figure 3.1, the blade swept area can be calculated from the formula:

$$A = \pi \mathrm{l}(\mathrm{l}+2\mathrm{r}) \tag{3.5}$$

Table 3.1: Classes of wind power density [81].

	10 m height		50 m height	
Wind power class	Wind power density (W/m^2)	Mean wind speed (m/s)	Wind power density (W/m^2)	Mean wind speed (m/s)
1	<100	<4.4	<200	<5.6
2	100-150	4.4-5.1	200-300	5.6-6.4
3	150-200	5.1-5.6	300-400	6.4-7.0
4	200-250	5.6-6.0	400-500	7.0-7.5
5	250-300	6.0-6.4	500-600	7.5-8.0
6	300-350	6.4-7.0	600-800	8.0-8.8
7	>400	>7.0	>800	>8.8

where l is the length of wind blades and r is the radius of the hub. Thus, by doubling the length of wind blades, the swept area can be increased by the factor up to 4. When $l >> 2r$, A $\approx \pi$l^2 .

3.1.1.2 Air density

Another important parameter that directly affects the wind power generation is the density of air, which can be calculated from the equation of state:

$$\rho \quad = \quad \frac{p}{\mathrm{R}T} \tag{3.6}$$

where p is the local air pressure, R is the gas constant (287 J/kg-K for air), and T is the local air temperature in Kelvin.

3.1.1.3 Wind power density

Wind power density is a comprehensive index in evaluating the wind resource at a particular site. It is the available wind power in airflow through a perpendicular cross-sectional unit area in a unit time period. The classes of wind power density at two standard wind measurement heights are listed in Table 3.1.

Some of wind resource assessments utilize 50 m towers with sensors installed at intermediate levels (10 m, 20m, etc.). For large-scale wind plants, class rating of 4 or higher is preferred.

3.1.2 Wind characteristics

Wind varies with the geographical locations, time of day, season, and height above the earth's surface, weather, and local landforms. The understanding of the wind characteristics will help optimize wind turbine design, develop wind measuring techniques, and select wind farm sites.

3.1.2.1 Wind speed

Wind speed is one of the most critical characteristics in wind power generation. In fact, wind speed varies in both time and space, determined by many factors such as geographic

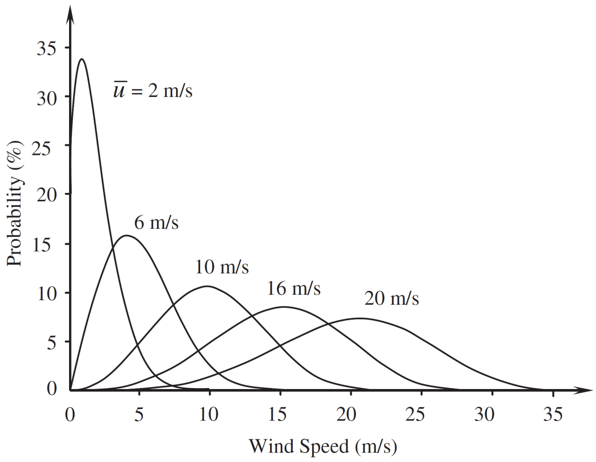


Figure 3.2: Weibull distributions for various mean wind speeds.

and weather conditions. Because wind speed is a random parameter, measured wind speed data are usually dealt with using statistical methods.

3.1.2.2 Weibull distribution

The variation in wind speed at a particular site can be best described using the Weibull distribution function [85], which illustrates the probability of different mean wind speeds occurring at the site during a period of time. The probability density function of a Weibull random variable $\bar{u}$ is:

$$f(\bar{u},\kappa,\lambda) = \begin{cases} \frac{\kappa}{\lambda}(\frac{\bar{u}}{\lambda})^{\kappa-1} & \exp\left(-(\frac{\bar{u}}{\lambda})^{\kappa}\right),\ \bar{u} \geq 0 \\ 0 & \bar{u} < 0 \end{cases}$$

where λ is the scale factor which is closely related to the mean wind speed and κ is the shape factor which is a measurement of the width of the distribution. These two parameters can be determined from the statistical analysis of measured wind speed data. As an example, the Weibull distributions for various mean wind speeds are displayed in Figure 3.2.

3.1.2.3 Wind turbulence

Wind turbulence is the fluctuation in wind speed in short time scales, especially for the horizontal velocity component. The wind speed $u(t)$ at any instant time t can be considered as having two components: the mean wind speed $\bar{u}$ and the instantaneous speed fluctuation $u'(t)$, i.e.:

$$u(t) = \bar{u} + u'(t) \tag{3.7}$$

Wind turbulence has a strong impact on the power output fluctuation of wind turbine. Heavy turbulence may generate large dynamic fatigue loads acting on the turbine and thus reduce the expected turbine lifetime or result in turbine failure.

In selection of wind farm sites, the knowledge of wind turbulence intensity is crucial for the stability of wind power production. The wind turbulence intensity l is defined as the ratio of the standard deviation σ_u to the mean wind velocity $\bar{u}$:

$$l = \frac{\sigma_u}{u} \tag{3.8}$$

where both σ_u and $\bar{u}$ are measured at the same point and averaged over the same period of time.

3.1.2.4 Wind direction

Wind direction is one of the wind characteristics. Statistical data of wind directions over a long period of time is very important in the site selection of wind farm and the layout of wind turbines in the wind farm.

The wind rose diagram is a useful tool of analyzing wind data that are related to wind directions at a particular location over a specific time period (year, season, month, week, etc.). This circular diagram displays the relative frequency of wind directions in 8 or 16 principal directions. As an example shown in Figure 3.3 , there are 16 radial lines in the wind rose diagram, with 22.5° apart from each other. The length of each line is proportional to the frequency of wind direction. The frequency of calm or near calm air is given as a number in the central circle. Some wind rose diagrams may also contain the information of wind speeds.

3.1.2.5 Wind shear

Wind shear is a meteorological phenomenon in which wind increases with the height above the ground. The effect of height on the wind speed is mainly due to roughness on the earth's surface and can be estimated using the Hellmann power equation that relates wind speeds at two different heights [19]:

$$u(Z) = u(Z_0)(\frac{Z}{Z_0})^a \tag{3.9}$$

where Z is the height above the earth's surface, Z_0 is the reference height for which wind speed $u(Z_0)$ is known, and a is the wind shear coefficient. In practice, a depends on a number of factors, including the roughness of the surrounding landscape, height, time of day, season, and locations. The wind shear coefficient is generally lower in daytime and

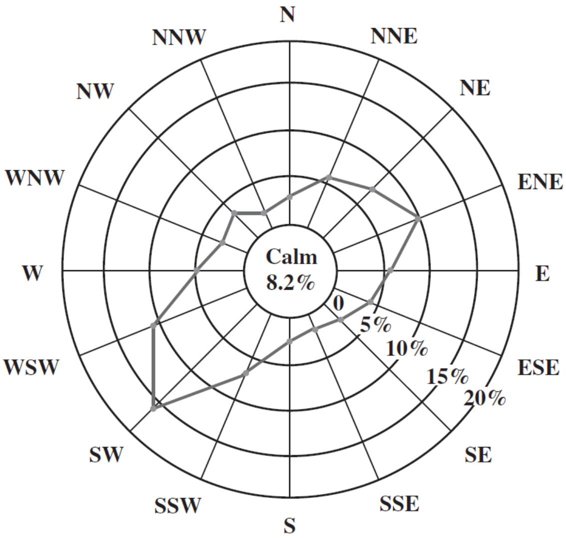


Figure 3.3: Wind rose diagram for wind directions.

higher at night. Empirical results indicate that wind shear often follows the "1/7 power law" (i.e. $a = 1/7$).

The power output of wind turbine strongly depends on the wind speed at the hub height, modern wind turbines are built at the height greater than 80 m, for capturing more wind energy and lowering cost per unit power output.

3.2 Wind power parameters

3.2.1 Power coefficient

The conversion of wind energy to electrical energy involves primarily two stages: in the first stage, kinetic energy in wind is converted into mechanical energy to drive the shaft of a wind generator. The critical converting devices in this stage are wind blades. For maximizing the capture of wind energy, wind blades need to be carefully designed.

The power coefficient C_p deals with the converting efficiency in the first stage, defined as the ratio of the actually captured mechanical power by blades to the available power in wind:

$$C_p = \frac{P_{me,out}}{p_w} = \frac{P_{me,out}}{\frac{1}{2}pA\bar{u}^3} \tag{3.10}$$

Because there are various aerodynamic losses in wind turbine systems, for instance, blade-tip, blade-root, profile, and wake rotation losses, etc., the real power coefficient C_p is much lower than its theoretical limit, usually ranging 30-45%.

3.2.2 Total power conversion coefficient and effective power output

In the second stage, mechanical energy captured by wind blades is further converted into electrical energy via wind generators. In this stage, the converting efficiency is determined by several parameters:

- *Gearbox efficiency* η_{gear} : The power losses in a gearbox can be classified as load-dependent and no-load power losses. The load-dependent losses consist of gear tooth friction and bearing losses and no-load losses consist of oil churning, windage, and shaft seal losses. The planetary gearboxes, which are widely used in wind turbines, have higher power transmission efficiencies over traditional gearboxes.

- *Generator efficiency* η_{gen}: It is related to all electrical and mechanical losses in a wind generator, such as copper, iron, load, windage, friction, and other miscellaneous losses.

- *Electric efficiency* η_{ele}: It encompasses all combined electric power losses in the converter, switches, controls, and cables.

Therefore, the total power conversion efficiency from wind to electricity η_t is the production of these parameters, i.e.:

$$\eta_t = \mathbf{C_p}\eta_{\mathbf{gear}}\eta_{\mathbf{gen}}\eta_{\mathbf{ele}} \tag{3.11}$$

The effective power output from a wind turbine to feed into a grid becomes

$$P_{eff} = \mathbf{C_p}\eta_{\mathbf{gear}}\eta_{\mathbf{gen}}\eta_{\mathbf{ele}}\mathbf{P_w} = \eta_{\mathbf{t}}\mathbf{P_w} = \frac{1}{2}(\eta_t pA\bar{u}^3) \tag{3.12}$$

3.2.3 Power curve

As can be seen from equation 3.12, the effective electrical power output from a wind turbine P_{eff} is directly proportional to the available wind power P_w and the total effective wind turbine efficiency η_t.

The power curve of a wind turbine displays the power output (either the real electrical power output or the percentage of the rated power) of the turbine as a function of the mean wind speed. Power curves are usually determined from the field measurements. As shown in Figure 3.4 , the wind turbine starts to produce usable power at a low wind speed, defined as the cut-in speed. The power output increases continuously with the increase of the wind speed until reaching a saturated point, to which the power output reaches its maximum value, defined as the rated power output. Correspondingly, the speed at this point is defined as the rated speed. At the rated speed, more increase in the wind speed will not increase the power output due to the activation of the power control. When the

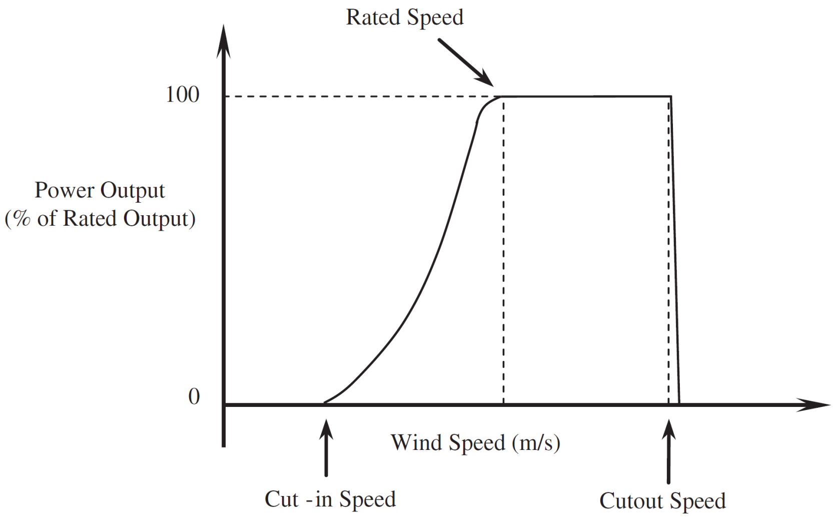


Figure 3.4: Typical wind turbine power curve.

wind speed becomes too large to potentially damage the wind turbine, the wind turbine needs to shut down immediately to avoid damaging the wind turbine. This wind speed is defined as the cut-out speed. Thus, the cut-in and cut-out speeds have defined the operating limits of the wind turbine [81].

3.2.4 Wind turbine capacity factor

Due to the intermittent nature of wind, wind turbines do not make power all the time. Thus, a capacity factor of a wind turbine is used to provide a measure of the wind turbine's actual power output in a given period (e.g. a year) divided by its power output if the turbine has operated the entire time. A reasonable capacity factor would be 0.25–0.30 and a very good capacity factor would be around 0.40 [88]. In fact, wind turbine capacity factor is very sensitive to the average wind speed.

3.3 Wind turbine controls

Wind turbine control systems continue to play important roles for ensuring wind turbine reliable and safe operation and to optimize wind energy capture. The main control systems in a modern wind turbine include pitch control, stall control (passive and active), yaw control, and others.

Under high wind speed conditions, the power output from a wind turbine may exceed its rated value. Thus, power control is required to control the power output within allowable fluctuations for avoiding turbine damage and stabilizing the power output. There are two primary control strategies in the power control: pitch control and stall control. The wind turbine power control system is used to control the power output within allowable fluctuations.

3.3.1 Pitch control

The pitch control system is a vital part of the modern wind turbine. This is because the pitch control system not only continually regulates the wind turbine's blade pitch angle to enhance the efficiency of wind energy conversion and power generation stability, but also serves as the security system in case of high wind speeds or emergency situations. It requires that even in the event of grid power failure, the rotor blades can be still driven into their feathered positions by using either the power of backup batteries or capacitors or mechanical energy storage devices [25].

Early techniques of active blade pitch control applied hydraulic actuators to control all blades together. However, these collective pitch control techniques could not completely satisfy all requirements of blade pitch angle regulation, especially for MW wind turbines with the increase in blade length and hub height. This is because wind is highly turbulent fl ow and the wind speed is proportional to the height from the ground. Therefore, each blade experiences different loads at different rotation positions. As a result, more superior individual blade pitch control techniques have been developed and implemented, allowing control of asymmetric aerodynamic loads on the blades, as well as structural loads in the non-rotating frame such as tower side-side bending. In such a control system, each blade is equipped with its own pitch actuator, sensors and controller.

In today's wind power industry, there are primarily two types of blade pitch control systems: hydraulic controlled and electric controlled systems. As shown in Figure 3.5, the hydraulic pitch control system uses a hydraulic actuator to drive the blade rotating with respect to its axial centerline. The most significant advantages of hydraulic pitch control system include its large driving power, lack of a gearbox, and robust backup power. Due to these advantages, hydraulic pitch control systems historically dominate wind turbine control in Europe and North America for many years.

The electric pitch control systems have been developed alternatively with the hydraulic systems. This type of control system has a higher efficiency than that of hydraulic controlled systems (which is usually less than 55%) and avoids the risk of environmental pollution due to hydraulic fluid being split or leaked.

In an electric pitch control system as shown in Figure 3.6 , the motor connects to a gearbox to lower the motor speed to a desired control speed. A drive pinion gear engages with an internal ring gear, which is rigidly attached to the roof of the rotor blade. Alternatively, some wind turbine manufacturers use the belt-drive structure adjusting the pitch angle. The use of electric motors can raise the responsiveness rate and sensitivity of blade pitch control. To enhance operation reliability, the use of redundant pitch control systems was proposed to be equipped in large wind turbines [86].

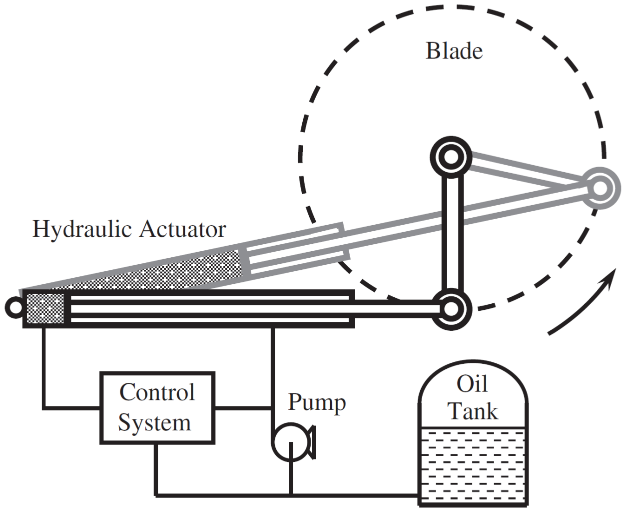


Figure 3.5: Hydraulic pitch control system.

3.3.2 Stall control

Besides pitch control, stall control is another approach for controlling and protecting wind turbines. The concept of stall control is that the power is regulated through stalling the blades after rated speed is achieved. Stall control can be further divided into passive and active control approaches. Passive stall control is basically used in wind turbines in which the blades are bolted to the hub at a fixed installing angle. In a passive stall-regulated wind turbine, the power regulation relies on the aerodynamic features of blades. In low and moderate wind speeds, the turbine operates near maximum efficiency. At high wind speeds, the turbine is automatically controlled by means of stalled blades to limit the rotational speed and power output, protecting the turbine from excessive wind speeds.

Compared with pitch control, a passive stall control system has a simple structure and avoids using a complex control system, leading to high reliability of the control system. In addition, the power fluctuations are lower for stall-regulated turbines. However, this control method has some disadvantages, such as lower efficiency, the requirement of external equipment at the turbine start, larger dynamic loads acting on the blades, nacelle, and tower, dependence on reliable brakes for the operation safety. Therefore, this control technique has been primarily used for small and medium wind turbines. Since the capacity of wind turbines has entered the multi-megawatt power range in recent years, pitch control has become dominant in the wind power market.

The active stall control technique has been developed for large wind turbines. An active stall wind turbine has stalling blades together with a blade pitch system. Since the blades at high wind speeds are turned towards stall, in the opposite direction as with pitch-control systems, this control method is also referred to as negative pitch control.

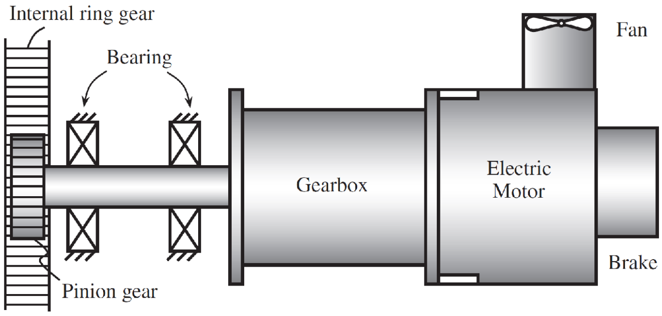


Figure 3.6: Electric pitch control system.

Compared with passive stall control, active control provides more accurate control on the power output and maintains the rated power at high wind speeds. However, with the addition of the pitch-control mechanism, the active stall control mode increases the turbine cost and decreases operation reliability. With megawatt wind turbines becoming the mainstream in the wind power industry from the late 1990s, pitch control is more favorable than stall control. It has been reported that the number of pitch-regulated turbines is four times higher than that of stall-regulated turbines and the trend is going to continue in coming decades [80].

3.3.3 Yaw control

In order to maximize the wind power output and minimize the asymmetric loads acting on the rotor blades and the tower, a horizontal-axis wind turbine must be oriented with rotor against the wind by using an active yaw control system. Like wind pitch systems, yaw systems can be driven either electrically or hydraulically. Generally, hydraulic yaw systems were used in the earlier time of the wind turbine development [30]. In modern wind turbines, yaw control is done by electric motors. The yaw control system usually consists of an electrical motor with a speed reducing gearbox, a bull gear which is fixed to the tower, a wind vane to gain the information about wind direction, a yaw deck, and a brake to lock the turbine securely in yaw when the required position is reached. For a large wind turbine with high driving loads, the yaw control system may use two or more yaw motors to work together for driving a heavy nacelle.

In practice, the yaw error signals obtained from the wind vane are used to calculate the average yaw angle in a short interval. When this average yaw angle exceeds the preset threshold, the yaw motor is activated to align the turbine with the wind direction. Thus, with heavily filtered wind direction measurements, the actions of yaw control are rather limited and slow.

3.3.4 Other control approaches

In the early time of wind turbine design, ailerons were once used to control the power output. This method involves placing move able flaps on the trailing edge of rotor blades [36]. The ailerons change the lift and drag characteristics of the blades and eventually change the rotor torque, which enable to regulate rotor speed and rotor power output. However, this method was less successful and was soon abandoned.

Another possibility is to yaw the rotor partly out of the wind to decrease power. This technique of yaw control is in practice used only for tiny wind turbines ($> 1kW$).

3.4 Summary

Wind turbines work by converting the kinetic energy in the wind first into rotational kinetic energy in the turbine and then electrical energy that can be supplied, via the national grid, for any purpose around the World. The energy available for conversion mainly depends on the wind speed and the swept area of the turbine. When planning a wind farm it is important to know the expected power and energy output of each wind turbine to be able to calculate its economic viability.

With the knowledge that it is of critical economic importance to know the power and therefore energy produced by different types of wind turbine in different conditions in this chapter was presented fundamentals mathematical models for this approach.

Chapter 4

State of the Art

Wind power is an effective form of clean, renewable energy which operate both onshore and offshore. It does not consume fuel or emit carbon emissions during its operation and is predicted to be the most cost competitive electricity source on economic by future. The primary function of a wind turbine is to harvest wind energy. It is done through converting kinetic energy of blades rotating into electrical energy.

Wind farms are unreliable sources of energy generation which require frequent maintenance. They need to continuously adapt to different wind loads as wind conditions change over time and are highly unpredictable. The conditions in which wind turbines operate often cause system failures. In [69] analyzed wind turbine failures for a wind farm in India. Was found that failure duration of yaw motor was highest, followed by failure duration of gear-oil pump, and then hydraulic unit of blade tip air brakes system.

Technology has improved significantly to minimize failure rates of components, however, failures of common parts are continually occurring. These failures are critical as it cause downtime and prevents the primary function of power generation. There have been 1868 accidents reported since 1900, of which 118 were critical, 174 structural failures and 345 blade failures [24]. Failures are highly undesirable in any system, whether it be mechanical or physical. Historical data shows a high failure rate in the gearboxes which require replacing every 5 to 7 years. Structural damage occurs frequently as a consequence of high wind loads, fire damage and wildlife impact which have high financial risks due to environmental impacts. Due to the complexity of the components, the repair cost of is very high [1]. It is the general belief that better quality of components would lead to more reliable systems, however, not much research is found in literature on predicting these failures at the design stage.

Wind turbine operation is dependent on wind speeds to generate power. Elmore and Gallagher [21] found that use of non site-specific wind velocity data, such as that available from a regional database could be a cost-effective means for predicting wind turbine performance. The concluded that Monte Carlo models could predict wind turbine performance using remote wind velocity data. Lawrence Livermore National Laboratory published the results of statistical analysis of field observations to help turbine manufacturers to refine their power curves and incorporate findings about what atmospheric processes are important in wind power forecasting [46]. However, many wind turbines' performance still declined significantly and became uneconomical to operate after 10 years of operation. The wind turbine would then be decommissioned and replaced with a newer replacement.

4.1 Wind turbine reliability and performance

4.1.1 Reliability

Wind turbine reliability depends on the reliability of some key components. For example, gearbox failures are seen as the most common and most critical failure. A gearbox failure ceases the primary function of electricity generation and faces downtime for repairs or replacement. Structural damage to the blades and tower are another common mode of failure. The blades may have experienced high wind loads or bird strikes, resulting in broken blades. In [15] analyzed the failure of a wind turbine tower caused by Typhoon "Jangmi". It was concluded that the bolts were inadequate and quality control was the cause of the damage. In other work of [45] was conducted failure analysis on wind turbine blade bolts. Laboratory testing of stress, strain, alternating loads, tensile loads, hardness and toughness were undertaken and concluded failure of bolts occurred from fatigue due to high alternating loads. Electrical system within the generator can cause an ignition, burning fuel vapors within the nacelle. Once a fire has started within the nacelle, it is highly unlikely to be extinguished due to the location and height of the fire. Since the 1980s, up to 30% of reported wind turbine accidents related to fire, with 90% of those leading to significant downtime or total loss of system [77].

4.1.2 Performance

Researchers and scientists had developed various models for the evaluation of performance of wind turbine system. A brief review of these models has been presented here. Abderrazzag had investigated the performance and energy production of a grid connected wind farm during 6 years operation and illustrated the variation in energy and wind speed on an annual and monthly basis for the whole examined period [2]. In [68] presented a probabilistic method used for the evaluation of the performance and reliability of wind-diesel energy systems. Castro and Allan built a probabilistic model of a wind frame taking into account the stochastic nature of the wind, the failure and repair processes of wind turbines, and the spatial wind-speed correlation and wake effects [70]. In [18] proposed a Monte Carlo-based method for predicting the economic performance and reliability of autonomous energy systems consisting of diesel generators and wind energy converters (WECs). Abouzahr and Ramkumar studied the performance of an autonomous WECS composed of one wind turbine feeding the load via a battery storage [3].

Billinton and Guungbai conducted studies on generating capacity adequacy associated with wind energy, using a sequential Monte-Carlo simulation procedure. The result shows that the contribution of WECs to the reliability performance of a generating system is highly dependent on the site wind condition [10]. A sequential Monte-Carlo simulation technique based on an hourly random simulation had been proposed for adequacy evaluation of a generating system including WECS. In [8] studied prediction and enhancement of performance of wind farm in India and found that there is scope for improvement in the annual plant load factor by 1–3% by improving the grid and machine availability.

A detailed parametric analysis such as available wind potential quality, examination of wind power curve, investigation of reliability for determining minimum cost is carried out concerning the optimum sizing of stand-alone wind power system by Kaldellis resulted in an appropriate decision making procedure, a significant reduction of the system dimensions may be realized leading to a remarkably diminished first installation cost [34]. In [37] studied nonlinear identification of wind turbine with a neural network, and found that variable speed wind turbine can produce 8–15% more energy output as compared to their constant speed counter parts. Wilson studied the various losses such as aerodynamic, mechanical, electrical, transmission and generator losses that reduce the power output. In that transmission and generator losses are of the order of 12% at rated power. The rotor performance is depending on the action of lift and drag forces on the blades [87].

During installation of WEGs in autonomous systems, technical constraints must be considered, since increasing wind penetration may disturb the operation of the system, leading to oscillation of voltage and frequency. Also in cases of high wind speeds the outage of WECs may damage conventional units. In order to improve the effectiveness and efficiency of research into wind energy conversion systems Noval wind turbine simulators had been developed to create a controlled test environment for drive train of wind turbine. In [12]. presented an innovative electronic system for testing the performance of wind turbines. The main goal of the system is to increase the accuracy in the measurements of torque and speed for each steady-state point of the turbine characteristic power curve. The steps to be adopted by the government agencies in order to ensure the desired growth of the wind industry in the country had been suggested by Skiha et al. and listed the suggestion to meet the technical challenges faced by the wind industry and to improve the performance of wind farms and also suggested a right choice of wind electric generator with an optimum rated wind speed will improve the wind farm performance [75].

4.2 Wind turbine modeling

Wind turbine can be modeled in different ways. Chen et al. [9] studied the functional expressions of wind turbine blade chord and twist angle span-wise distribution and developed a new common functional equation of the blade shape. A 2.3 MW wind turbine blade shape optimization model was established. In [20] presented a new design philosophy based on functional redundancies and reconfiguration that can help to increase availability of wind turbines. In the event of a fault, the new design capabilities could be used to substitute the function of a faulty component and the system's availability for operation could be analyzed. The technique Bond Graph uses power nodes to describe energy and power transfers in a dynamic system explicitly indicated by causal strokes. Bakka and Karimi [5] addressed the problem of bond graph methodology as a graphical approach for the modeling of wind turbine generating systems and validated a specific wind turbine generating system. Tapia and Medina proposed a bond graph doubly-fed wind turbine generator control. The bond graph methodology and the concept of causality were used to derive and verify the control law system of the wind turbine. Bond graph modeling methodology has the advantage of representing the energy transfers by either a flow or effort between components. Each component is represented by a bond group; source, sink, transformer, gyrator or a junction. Further to a junction (1-junction and 0-junction), passive elements such as resistance, inductance or capacitance may be assigned. These elements represent operating characteristics of the component [79].

4.3 Wind turbine maintenance

Maintenance of wind turbines are conducted twice a year at 6 month intervals. A typical maintenance would involve thorough inspection of the entire system, replacement of fluids, lubrication and servicing of mechanical parts. Repairs and replacements would be conducted if deemed necessary by the technician. These time based inspections and maintenance activities are often expensive and require undesired downtime. Prognostic Health Management (PHM), a form of conditional monitoring, was proposed by Abichou et al as the best strategy to reduce cost of operation and maintenance. Lekou et al conducted conditional health monitoring on gearboxes and bearing using vibration acoustic emissions. They utilized accelerometers and strain gauge bridges to determine the operating frequency to use as a baseline measure for conditional monitoring. Failing components were tested to measure frequencies and compared to baseline measure to determine early failure detection [47, 48, 54].

4.4 Supervisory control and data acquisition system (SCADA)

Prognostic health management (PHM) is conducted through a SCADA system. Ongoing data is collected and analyzed for abnormalities during operation. The SCADA also monitors the health of the system through various sensors which assist with optimizing maintenance scheduling and improve reliability. The sensor readings, matched with historical failure readings, are used to predict failures before they occur. In [28] developed the Nonlinear State Estimation Technique (NSET) to model turbine tower vibration to good effect, providing an understanding of the tower vibration dynamic characteristics and the main factors influencing these. SCADA data from a single wind turbine was used to validate the model. Guo et al [28] developed tower vibration model comprises two different parts: a sub-model used for the conditions that below rated wind speed and another for that above rated wind speed. With SCADA of one wind turbine, the NSET model showed good preparation for the condition monitoring for wind turbine's key components. Wang et al [84] reviewed different methods of data driven approaching for SCADA data interpretation has been reviewed and an artificial intelligence (AI) based framework for fault diagnosis and prognosis of WTs using SCADA data was proposed.

4.4.1 Wind turbine condition monitoring analysis based on SCADA data

With the continuing growth of global wind power capacity and rapid expansion of wind farms, condition monitoring and anomaly identification for WTs are of increasing importance in reduction of operation and maintenance costs. Condition monitoring is a process of monitoring the operating parameters of a physical system. From the changes in the parameters, possible failures in the system can be diagnosed. SCADA installed in most wind turbines can provide a large number of monitoring parameters that are considered valuable resources for condition monitoring and anomaly identification of WTs. The SCADA data used for condition monitoring is without additional installation of sensors and hardware devices. Therefore, WTs condition monitoring based on SCADA data has been widely used. the research on condition monitoring of WTs was mainly focused on how to identify early faults and putting forward some methods to judge the abnormal occurrence. The establishment of prediction model and the establishment of input and output are mostly based on experience or even given directly from the text and the specific reasons can not be verified. Besides, the established models are more limited which are only established for particular kinds of abnormal, like the gearbox is overheating and so on [64].

4.4.2 Wind turbine reliability analysis based on SCADA data

The turbines and associated equipment in the wind farm are connected to a SCADA system, which has installed a number of transducers for operating and controlling turbines and measuring the power generated by them. The alarm settings of SCADA system can help operator to understand the status of the turbines approximately, but SCADA data are usually thought less useful to wind turbine condition monitoring because of their low sampling rate (i.e. one sample every 10 minutes). However, in comparison with commercial CMSs, SCADA system is already there and SCADA data are ready to use. So, they are the cheapest resource for developing a cost effective CMS for wind turbines. Moreover, SCADA data are a complete record of the operating history of the turbines, through which the reliability of these machines can be fully understood. This is very helpful for the operation and maintenance (O&M) of existing turbines and the reliability design of next generation of turbines. Together with turbine service report, SCADA data record a complete O&M history of the turbines. In particular, they record the information of failures occurring in the machines (e.g., the time when the failure happens, the log number of the failure, the time when the turbine is restarted, etc). Turbine supplier regards the information as the things sensitive and confidential, because the actual quality and operation situation of the turbines could be disclosed from them. The consequence of information leak is the significant impact to market sale. However, if these data are properly used it can bring significant benefit to both turbine suppliers and owners [89].

4.5 Wind turbine power curve classification of modelling methods.

Wind energy has emerged as a promising alternative source for overcoming the energy crisis in the world. Wind power based energy is one of the most rapidly growing areas among the renewable energy sources and will continue to do so because of the growing concern about sustainability and emission reduction requirements. The uncertain nature of wind and high penetration of wind energy in power systems are a big challenge to the reliability and stability of these systems. To make wind energy a reliable source, accurate models for predicting the power output and performance monitoring of wind turbines are needed. The theoretical power captured (P) by a wind turbine is given by [44]

$$P = \frac{1}{2}\rho A_w C_p(\lambda, \beta) v^3 \tag{4.1}$$

The power production of a wind turbine (WT) thus depends upon many parameters such as wind speed, wind direction, air density (a function of temperature, pressure, and humidity) and turbine parameters. Much complexity is involved in considering the effects of all the influencing parameters properly. It is therefore difficult to evaluate the output power using the theoretical equation given in above. Power curve of a wind turbine, which gives the output power of turbine at a specific wind speed, provides a convenient way to model the performance of wind turbines. A typical power curve for a pitch regulated wind turbine is shown in Figure 4.1. In the first region when the wind speed is less than a threshold minimum, known as the cut-in speed, the power output is zero. In the second region between the cut-in and the rated speed, there is a rapid growth of power produced. In the third region, a constant output (rated) is produced until the cut-off speed is attained. Beyond this speed (region 4) the turbine is taken out of operation to protect its components from high winds; hence it produces zero power in this region. The power curve of a WT indicates its performance. Accurate models of power curves are important tools for forecasting of power and on-line monitoring of the turbines. A number of methods have been proposed in various works to model the wind turbine power curve. These methods which use data from manufacturers' specifications and actual data from the wind farms have been utilized by many researchers in various wind power applications [53, 43].

4.5.1 Need of power curve modelling

The power curve reflects the power response of a WT to various wind speeds.Accurate models of the curves are useful in a number of wind power applications. The objectives of modelling the wind turbine power curve have been discussed here.

- *Wind Power Assessment and Forecasting.* The WT power curve can be used for wind power assessment.Wind resource assessment of a region in terms of wind speed, wind power density, and wind energy potential is done to identify areas suitable for wind power development [53]. In this process, estimation of energy is done by using the available wind data and wind turbine power curve. Predicting the output power of the turbine at a candidate site is also required in sizing and cost optimization studies during the design stage of a wind energy based system. The accuracy in power prediction is important as an overestimation can result in poor reliability and an underestimation can lead to oversizing of the wind energy conversion system.Wind turbine operators who trade energy directly to the electricity market also need to forecast the power output of their turbines accurately, so that they will be able to deliver the traded amount of power [72].

  Power curves are supplied by the manufacturers in a tabular or graphical form. However, a generic equation which represents this curve accurately is required in various problems of wind power systems. Derivation of an appropriate function to describe the actual shape of the curve is a very important task. However, the manufacturer's curves are created under standard conditions therefore they may not represent the realistic conditions of the site under consideration. The turbine performance at the wind farms is also not ideal due to wear and tear and aging

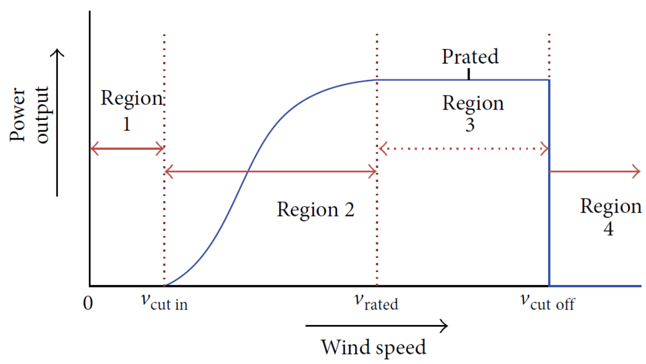


Figure 4.1: Typical power curve of a pitch regulated wind turbine.

of turbines. Another method to model the power curves is to derive them using the actual data of wind speed and power measured from the turbines [43]. The data of wind turbines collected by the SCADA can be utilized for this purpose.This method can incorporate the actual conditions at the wind farms, thus providing better accuracy in power prediction.

- *Capacity Factor Estimation.* The capacity factor of a WT is defined as the ratio of the average power output to the rated output power of the generator and is an indicator of its efficiency [33]. It is used to estimate the average energy production of a WT required for the sizing and cost optimization studies, optimum turbine-site matching, and ranking of potential sites [33, 62]. The wind turbine power curve models are used to estimate the capacity factor of a WT. A comparative analysis of four power curve modelling methods in estimation of capacity factor of wind turbine generator is presented in [14].

- *Selection of Turbines.* The power curve can be used to make generic comparison between models and can aid in the choice of turbine from the available options.The selection of the turbine characteristics which match with the wind regime of the site helps in optimizing the efficiency of wind energy system [83].

- *On-line Monitoring of Power Curves.* Power curves can be used for monitoring the performance of turbines. For this, a benchmark curve which represents the performance of a normally operating turbine is required.This reference curve can be extracted from measured power output and wind speed data of wind turbines.The actual curve of the turbine to be monitored can be compared with this benchmark curve. The deviations of the actual values from the expected output can indicate underperformance or faults [44].The wind power output of a turbine can be affected by underperformance or various faults/anomalies of the turbine such as blade faults and yaw and pitch system faults [43, 60]. Different types of faults affect the turbine system differently and will cause the power curve to depart from the expected value in a different way. Tools which can characterize and quantify these departures can aid in early identification of faults. Statistical analysis of the outlier data can give indications of the specific reason of anomaly.Wind turbine condition monitoring by use of power curve copula modelling is suggested in [78, 26] and is a topic of further research. Early recognition of the emerging faults and timely repair and maintenance of the equipment can help in improving the performance of wind turbines.

4.5.2 IEC 61400-12-1 standard

IEC 61400-12-1 is the commonly adopted international standard for power performance measurement [31]. The procedure for measuring the power performance characteristics of single wind turbines is specified in this standard. It is the most accepted standard for power curve measurement of single wind turbines.The standard describes the measurement methodology for the measured power curve which is determined by simultaneous measurement of wind speed and power output at the test site. A previous site calibration is required for certain terrain conditions. The annual energy production is calculated by applying the measured power curve to reference wind speed frequency distributions supplemented by sources of uncertainty and their effects. The standard prescribes derivation of power curve using the hub height wind speed measured with a cup anemometer in the suitable measurement sector, but if the wind speed has a large variation over the rotor swept area then there can be a significant difference between the hub height wind speed and wind speed averaged over the whole rotor swept area. The measurement methods and accuracy of measuring instruments can cause variance in measurements and can lead to large prediction errors. The impact of other measurement options such as consideration of rotor equivalent wind speed in which speed is measured at heights over the full rotor plane with the use of remote sensing technology (LIDAR and SONAR) and nacelle based anemometry is a topic of further research [55].

The IEC standard uses ten-minute averaged data grouped into wind speed intervals of 0.5 m/s (method of bins). This 10-minute averaging of data introduces systematic averaging errors and short wind fluctuations are killed off. Wind at a specific site can be affected by a number of factors such as topology of the site and obstacles and weather phenomena. Although the IEC power curve considers the wind condition of the current site it may not always be appropriate to apply to the wind conditions of other sites. Research efforts are therefore required to develop site specific power curves. These curves can incorporate the wind conditions of the particular site, thus giving better results [59, 31].

Appropriate selection of modelling method is an important requirement for during planning and operation stage of wind based system and helps in improving the performance of the system. The methods which consider only wind speed as input may not take into account the variance caused by various influencing parameters. Methods which consider the influence of these parameters on the power curve can result in more accurate models. Wind at a specific site can be affected by a number of factors such as topology of the site and obstacles and weather phenomena. It is shown in [59] that the use of developed site power curves, which used the knowledge of site and turbine parameters for modelling, resulted in more accurate energy assessment than the turbine power curve.

4.5.3 Power curve models classification

The power curve modelling methods can be classified into discrete, deterministic/probabilistic, parametric/nonparametric, and stochastic methods or they can be classified on the basis of data used for modelling.

- *Discrete Models.* In this method as described in IEC 61400-12 all the wind speeds are discretized into 0.5m/s bins [31]. The power output for each bin is then modelled.

This is a simple method as it does not require mathematical functions for describing the curve. Also it takes into account the nonlinear wind speed-power output relation. However a large number of data are required in this method to develop a reliable model.

- *Deterministic and Probabilistic Models.* A deterministic power curve model assumes a fixed relation between the output power and wind speed. But when a fleet of wind turbines are deployed on a wind farm, turbines of the same type may produce different amount of power even if the wind speed is the same. A probabilistic power curve model incorporates these power variations to characterize the relationship between wind speed and actual output powers. Most of the models available in the literature are of deterministic nature and are constructed by using the manufacturers' power curve data. A probabilistic model proposed in [32] characterizes the dynamics of output power by a normal distribution with varying mean and constant standard deviation. The method given in the paper accommodates the uncertainty of output power. The probabilistic nature of wind power output can also be modelled by deriving curves using actual data of power output and wind speed of turbines deployed in a wind farm. This method requires a large number of historical data but results in accurate models [43, 50].

- *Parametric and Nonparametric Models.* A parametric model defines the relationship between input and output by a set of mathematical equations with a finite number of parameters. In a nonparametric model, no assumption is made about the functional form of the phenomenon under observation. Parametric models of WT power curve can be built by utilizing a set of mathematical expressions having a fixed number of parameters, which are usually collected together to form a single parameter vector $\theta = (\theta_1, \theta_2, \theta_3, ..., \theta_n)$. Nonparametric models are used when it is difficult to define the underlying theory upon which the parametric model can be constructed [50].

- *Models Based on Presumed Shape, Curve Fitting, and Actual Data.* The models of power curves can be classified according to the data being used for modelling. Models of power curve based on presumed shape of curve utilize only the cut-in, cut-off, and rated speeds and the rated power of the selected turbine for calculating the parameters of expressions used in the model [16, 17, 58]. These ratings are available from the specifications of the turbines. When the manufacturer's power curve data is available, models can be developed by fitting one or more appropriate expressions to the actual curve. The parameters of the expression being fitted to the actual curve are generally calculated by using the least squares method [43].The models derived from actual data of wind farm need the actual wind speed and power output data from an operational wind farm. If the effect of the influencing parameters is also included in the model, then the data of the included parameters is also required. This data can be obtained from the wind farm's SCADA system.

- *Stochastic Models.* The stochastic method consists of characterizing the power performance of wind turbine by evaluating dynamic response against the fluctuating wind speed inputs [56]. The dynamic power output is separated into a deterministic stochastic part in this model. In [4] the Markov chain theory is used to describe the power output of WT. The resulting model is independent of turbulence intensity; however, the effect of other influencing parameters is not taken into account in this method.

4.6 Summary

A wind turbine is a complex collection of components comprising rotors, pitch system, drivetrain, gearbox, generator, electrical system, mechanical brakes, yaw system, sensors, control system and hydraulics. To understand how these intervals and what activities are required in each visit, it is important to develop mathematical models to analyse the effect of different visit intervals and maintenance strategies. This literature review examined the key modeling research relevant to wind turbine lifecycle operation analysis.

Analysis of turbine SCADA data can help people to identify the problems existing in wind farm services. For instance, the downtimes caused by turbine sub assemblies can be derived from SCADA data. Then, through comparing the calculated downtime and the actual time used for component repair and maintenance, people may easily know the major reason accounting for decreased availability of the turbine. If actual maintenance time occupies only a small percent of the downtime, maintenance is fine. The problem could happen in spare part order/delivery, limited access to turbine for weather or sea condition reasons, or the availability of the vessels for transportation and installation. However, if actual maintenance time occupies a big percent of the downtime, it is necessary to check whether turbine supplier or contracted partner has provided an instant, efficient and high quality maintenance service. This analysis is particularly useful for managing those wind farms located offshore or in other remote areas.

Accurate models of power curves are important tools for forecasting of power and online monitoring of the turbines. A number of methods have been proposed in various works to model the wind turbine power curve. The literature reviewed reveals that appropriate selection of power curve models can help in improved performance of wind energy based systems. This section presented current research of different wind turbine power curve modelling methods. We also identified the need for standard measuring and classification of models.

Chapter 5

Methodology and Results

Data mining is a creative process which requires a number of different skills and knowledge. Currently there is no standard framework in which to carry out data mining projects. This means that the success or failure of a data mining project is highly dependent on the particular person or team carrying it out and successful practice can not necessarily be repeated across the enterprise. Data mining needs a standard approach which will help translate business problems into data mining tasks, suggest appropriate data transformations and data mining techniques, and provide means for evaluating the effectiveness of the results and documenting the experience.

The Cross Industry Standard Process for Data Mining (or *CRISP-DM*) project addressed parts of these problems by defining a process model which provides a framework for carrying out data mining projects, independently of both the industry sector and the technology used. The CRISP-DM process model aims to make large data mining projects more reliable, manageable, repeatable, faster, and less costly.

We follow this methodology in the execution of the practical part of this work.

5.1 Overview CRISP-DM

The CRISP-DM reference model for data mining provides an overview of the life cycle of a data mining project. It contains the phases of a project, their respective tasks, and their outputs.

The life cycle of a data mining project is broken down in six phases which are shown in Figure 5.1. The sequence of the phases is not strict. The arrows indicate only the most important and frequent dependencies between phases, but in a particular project, it depends on the outcome of each phase which phase, or which particular task of a phase, has to be performed next.

Figure 5.1: Phases of the current CRISP-DM process model for data mining.

The outer circle in Figure 5.1 symbolizes the cyclic nature of data mining itself. Data mining is not finished once a solution is deployed. The lessons learned during the process and from the deployed solution can trigger new, often more focused business questions. Subsequent data mining processes will benefit from the experiences of previous ones [7]. The process is iterative because the results of some phases sometimes require the project cycle to go back to an earlier phase. For example, a result of the modeling stage may be that more data preparation is required new data may be needed or the existing data may have to be prepared in a different way.

Each phase of the process is described briefly below:

- *Business Understandings.* This initial phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition, and a preliminary project plan designed to achieve the objectives.

- *Data Understandings.* The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information.

  There is a close link between Business Understanding and Data Understanding. The formulation of the data mining problem and the project plan require at least some understanding of the available data.

- *Data Preparation.*The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include table, record, and attribute selection, data cleaning, construction of new attributes, and transformation of data for modeling tools.

- *Modelling.* In this phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques require specific data formats.

  There is a close link between Data Preparation and Modeling. Often, one realizes data problems while modeling or one gets ideas for constructing new data.

- *Evaluation.* At this stage in the project you have built one or more models that appear to have high quality, from a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model, and review the steps executed to construct the model, to be certain it properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

- *Deployment.* Creation of the model is generally not the end of the project. Usually, the knowledge gained will need to be organized and presented in a way that the customer can use it. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process. In many cases it will be the user, not the data analyst, who will carry out the deployment steps. In any case, it is important to understand up front what actions will need to be carried out in order to actually make use of the created models.

5.2 Selection of Tools

Many algorithm engines, tools, and platforms have been developed to implement functions and related data mining techniques. Predictive analytics today summarized the top 50 free DM software, including Orange, RapidMiner, Weka, KNIME, SpagoBI, Anaconda, Octave, and so forth. Some commercial software including Sisense, Oracle Data Mining, Microsoft SharePoint, IBM Cognos, Dundas BI, SAP Business Objects, Matlab, Statistic, SAS EM, SPSS Clementine (IBM SPSS Modeler after 2009), Tanagra, Qlik Sense, and so forth have also been widely used by researchers and practitioners [49].

In this work the main selected tool for practical environment is Anaconda. With more than 13 million downloads to date, Anaconda is blossoming into a real phenomenon in a crowded data science field. Anaconda is a completely free enterprise-ready Python distribution for large-scale data processing, predictive analytics, and scientific computing. It includes over 195 of the most popular Python packages for science, math, engineering, and data analysis. It also offers the ability to easily create custom environments by mixing and matching different versions of Python and other packages into isolated environments that individual users are free to create. Anaconda includes the ‘conda‘ package and environment manager to make managing these environments straightforward.

Anaconda comes pre-loaded with all the most popular libraries and tools for machine learning and data mining approaches. As well as Jupyter, some of the biggest Python libraries wrapped up in Anaconda include NumPy, Pandas, Matplotlib, Scikit-Learn etc. The following is a short summary of the most important tools used in deployment of this research work.

Jupyter Notebook. It is a powerful tool for interactively developing and presenting data science projects. A notebook integrates code and its output into a single document that combines visualisations, narrative text, mathematical equations, and other rich media. The workflow promotes iterative and rapid development, making notebooks an increasingly popular choice at the heart of contemporary data science, analysis, and increasingly science at large. As part of the open source Project Jupyter, this tool is also free to use.

NumPy. In the Python ecosystem, multi-dimensional arrays are provided by the NumPy library, which contains a high-performance array object and related linear algebra routines. The array interface provided by NumPy supports vector programming and is very general, underpinning many of the numerical tools widely used in the Python world. For instance, the companion SciPy package implements many useful functions for optimization, statistics, interpolation, image processing, and spatial mathematics. Many of these functions work with NumPy arrays having arbitrary size, dimensions, and data types.

Pandas. Another library which builds on the NumPy array interface is Pandas, which extends the array with labeling and an engine for transformation and analysis of structured datasets. The data structures provided by Pandas - particularly the DataFrame - greatly simplify timeseries analysis, split-apply-combine workflows, and other common research processing tasks. The Xarray package extends Pandas by providing additional data structures to handle n-dimensional labeled arrays. Most importantly, the array and labeling semantics in NumPy, Pandas, and Xarray are similar in the majority of cases. Because each package sequentially builds on the others, they can all be used within the same analysis context and data can easily be shuttled to whatever format works best for a given task.

Matplotlib. The core visualization library in the Python world is Matplotlib. Although it originated as an emulation of the graphics capabilities of MATLAB, Matplotlib has grown into the de-facto base layer for 2D graphics in Python. Matplotlib provides fine-grained control of graphics, and works natively with both base Python objects and NumPy array derived types (including Pandas Series and DataFrames and Xarray DataArrays). In fact, both Pandas and Xarray provide layers to help automate plotting numerical data through Matplotlib.

Scikit-Learn. A toolkit implement a wide variety of algorithms for supervised and unsupervised machine learning tasks, including regressions, classification, clustering, manifold learning, principal components, density estimation, and much more. It also provides many useful tools to help build pipelines for managing modeling tasks such as data processing/normalization, feature engineering, cross-validation, fitting, and prediction.

5.3 Methodology

This research project is divided into two parts: first, we employ a data mining method for recognition of data rejected sob normal operation of WT' performance curves; second, we apply machine learning algorithms for identification and prediction of status patterns of wind turbines. As mentioned above, it follows the CRISP-DM methodology. We start with the first part of this project.

5.3.1 Business understanding

Since 2000 wind technology has seen a continuous growth in Portugal, motivated by a political strategy, at European and national levels, in endogenous and renewable resources with the aim of diversifying sources, improving the security on supply, decreasing of energetic dependency and reducing the environmental impact of electro production system.

Currently, wind energy plays an invaluable role in the Portuguese electric sector: in 2017, the wind generated electricity was equivalent to almost a quarter of the total demand, in the same order of magnitude of the contribution observed in 2016.

It is worth highlighting that the wind generating capacity suffered residual growth from 2016 to 2017, just 0.6 MW. This slight variation resulted from the decommissioning of 23 wind turbines operating since 1998 and their replacement for currently available technology, more efficient and capable of generating electricity at a lower cost, materialized in only four wind turbines. Along with new generating capacity currently under construction or to license on the meantime, re-powering will certainly play an important role in the sector for the next coming years.

INEGI – Institute of Science and Innovation in Mechanical and Industrial Engineering is an interface institute between the Faculty of Engineering of the University of Porto

(FEUP) and the Industry, with the calling to have activity on the fields of Innovation and Technology Transfer.

APREN, the Portuguese Renewable Energy Association, is a non-profit association, founded in 1988 with the mission of coordinating, representing and defend the common interests of its Associates. It develops its work in close cooperation with public authorities and similar entities, both at national and international level, being an important key player in the deployment of energy policies for Portugal, enabling the promotion of endogenous renewable resources for electricity generation.

With the legal statute of non-profit private association public utility, part of the Portuguese Scientific and Technological System, and with the status of *Public Usefulness*, it aims to play an active role in the development of the Portuguese industrial sector and transformation on the competitive model of the national industry.

As we see, with the increasing number of wind farms, the renewable sources of energy have taken a key role in the Portuguese electricity mix.

Given the interest of the market players and general public to improve their knowledge about the renewable energy plants for electricity generation, including their geographic distribution and associated technologies, both APREN as INEGI seek the dissemination of the relevant information to the stakeholders and public in general.

With the willingness to coordinate efforts, both institutions came together to create and maintain a database providing relevant information about all the power plants that use renewable energy sources in Portugal.

This work is concerned with the development of methodology to support the optimization of the production of a wind turbine for planning and analyzing of the wind farms performance by INEGI.

5.3.2 Data understanding

An operational wind farm typically generates vast quantities of data which are well known SCADA data. The SCADA data contain information about every aspect of a wind farm, from power output and wind speed to any errors registered within the system. By keeping track of both wind speed and power output parameters, the overall health of the turbine can be supervised. SCADA data may be effectively used to "tune" a wind farm, providing early warning of possible failures and optimizing power output across many turbines in all conditions.

The data used in this research was generated at two wind farms with about 13 turbines and two corresponding weather stations. In the case of turbines, the data was collected by a SCADA system installed. Each SCADA system collects data on more than 120 parameters. The data is sampled at a high frequency, then averaged and stored at 10-min intervals (referred to as 10-min data). The data included one file for each turbine containing over 100 parameters of intervals of an average of 10 min from different sensors and monitoring channels. Examples of parameters include: wind speed, wind direction,

Table 5.1: Technical specifications for wind turbines E70, E82 and E92.

	Wind Turbines' Models		
Technical specifications	E 70	E 82	E 92
Rated power:	2000 kW	2000 kW	2300 kW
Rotor diameter:	71 m	82 m	92 m
Hub height in meter:	64 m	78 m	85 m
Cut-in wind speed :	2.5 m/s		
Cut-out wind speed :	28-34 m/s (with ENERCON storm control)		
Remote monitoring:	ENERCON SCADA		
Manufacturer:	Enercon		

outside temperature, and turbine control parameters; all the data samples were time stamped.

The data set was complemented with the data from nearby weather stations. A weather station is a facility, either on land or sea, with instruments and equipment for measuring atmospheric conditions to provide information for weather forecasts and to study the weather and climate. The measurements taken include temperature, atmospheric pressure, humidity, wind speed, wind direction, and precipitation amounts.

Some of the data collected was recorded in the relational database management system by INEGI. In this work, two separate sets of data were created. The description of both of data sets are described below.

5.3.2.1 Data description

The fist data set was collected during the period of one year and two months, from 2016/08/01 to 2017/09/30. This data set includes data from three wind turbines of three different models that are installed on that wind farm located in Northern Portugal. The data set is complemented with the data from the nearest weather station. The data from weather station is also averaged and stored at 10-min intervals.

The brief description of technical specifications of wind turbines selected is listed in Table 5.1. The weather station has the following specifications: anemometers and wind-vanes' height - 85 m | 63.6 m; barometer's height - 81 m; temperature/relative humidity probe's height - 81 m.

The second data set was collected during the period of six years and nine months, from 2011/01/01 to 2017/09/30. It includes data from a wind farm in its entirety and enclosing the data from weather stations. This wind farm is also located in Northern Portugal and has installed power of 20MW. The farm has 10 wind turbines in total; the model of the turbines is E 70. The technical specifications are listed in Table 5.1. The weather station selected has the following specifications: anemometers and wind-vanes' height - 65 m | 30 m; barometer's height - 63 m; temperature/relative humidity probe's height - 63 m.

It is important to emphasize that all wind turbines selected have storm control activated. ENERCON wind energy converters run with a special storm control feature. This

Table 5.2: The classification of states of wind turbine by INEGI.

Status	Code of state	Description
Normal Operation	1	The wind turbine is under normal operation.
Error SCADA	2	The error of SCADA is detected.
Unavailable	3	The wind turbine is unavailable.
No SCADA data	4	There are no SCADA records.
Betz Criterion	5	The records do not meet the Betz criterion.
Maintenance	6	The wind turbine is under maintenance state .
Ice Detection	7	The wind turbine is under Ice Detection state.
Grid Failure	8	There is a network failure in the Grid.
Park Limitation	9	There is a limitation of the power of the park.
Rejected Data	10	The wind turbine is under abnormal operation.

slows the wind turbine down so that it can continue to operate even at high wind speeds. This technique avoids numerous shutdowns, which lead to considerable losses in power output. When storm control is activated, the rated speed is linearly reduced starting at a predetermined wind speed for each turbine type. Beginning at some turbine-specific wind speed, the reduction in the turbine's rated speed also reduces active power. The turbine only shuts down at a wind speed of more than 34 m/s (10-minute average). In comparison, when storm control is deactivated, the wind turbine stops if the wind speed reaches 25 m/s in the three minute average, or 30 m/s in the 15 second average.

Both data sets are constructed by the following parameters averaged over 10 min intervals: speed from nacelle anemometer (vel_med), power produced (pot_med), rotor speed rotation (rot_med), power coefficient (cp_med), density (massa_vol), nacelle orientation (orientacao_nac), temperature (tep_med), humidity (hum_med), turbulence (int_turb), atmospheric pressure (bar), date (data) and control status state of wind turbine (idclassificacao).

The classification of states was provided by INEGI and its description is shown in Table 5.2.

5.3.2.2 Data exploration

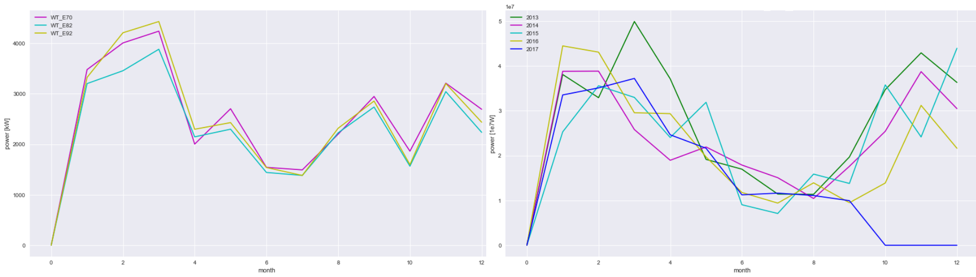
For better understanding and detecting interesting subsets in the hidden information, we start by exploring our initial data.

The distribution of all power produced per month is shown in Figure 5.2a for the first data set and in Figure 5.2b for the second data set.

We note a significant variability in the the power produced from month to month.

The 12 parameters recorded at 10-min intervals resulted in 61344 instances for the first data set and in 3546379 instances for second data set.

For further work the first data set was divided into 3 separated sub-sets that corresponded to three different wind turbines and were designated as WT_70E, WT_82E and WT_92E. The second data set was designated as WTS_FARM.



(a) Power produced from wind turbines E 70, E 82 and E 92 at [2016/08/01-2017/09/30].

(b) Power produced from wind farm at last five years [2013-2017].

Figure 5.2: Power produced by month of year.

5.3.3 Data preparation

The collected data from a wind farm is voluminous, and it is expected to contain errors caused by sensors and malfunctions of the data collection system. Such errors manifest themselves in missing values, out-of-range values, and so on.

5.3.3.1 Data cleaning

The statistics of all parameters' values and there missing data of the four data sets are summarized in Table 5.3. We found some out-of-range values in WTS_FARM data set on parameters such as: power coefficient (cp_med). The data with suspected values was deleted.

After cleaning out-of-range values, the WTS_FARM data set contains 3539485 data points with 12 parameters. The other tree data sets contain the similar volume.

5.3.3.2 Imputation of missing values

For various reasons, many real world datasets often miss values. Such values are usuallt encoded as blanks, NaNs or other placeholders. Datasets containing such values are incompatible with scikit-learn estimators which assume that all values in an array are numerical, and that all have and hold meaning. A basic strategy to use incomplete datasets is to discard entire rows and/or columns containing missing values. However, this comes at the cost of losing data which may be valuable, albeit incomplete. A better common strategy, applied in this work, is to impute the missing values, for instance, by inferring them from the known part of the data. For this purpose, we used the *Imputer* class from *sklearn.preprocessing* library.

The *Imputer* class provides basic strategies for imputing missing values, either using the mean, the median or the most frequent value of the row or column in which the missing values are located. This class also allows for different missing value encodings.

Table 5.3: Statistics description of data sets with the missing data.

	WT 70E				WT 82E				WT 92E				WTS FARM			
	missing	mean	min	max	missing	mean	min	max	missing	mean	min	max	missing	mean	min	max
vel_med	64	6.013	0.000	32.900	64	5.857	0.000	39.800	64	6.090	0.000	32.000	0.000	6.881	0.000	34.600
pot_med	64	528.453	0.000	2060.000	64	483.352	0.000	2060.000	64	522.566	0.000	2060.000	0.000	555.046	0.000	2074.000
rot_med	64	10.851	0.000	18.060	64	10.696	0.000	18.090	64	10.904	0.000	18.080	0.000	12.486	0.000	22.420
cp_med	265	0.425	0.000	1.341	272	0.437	0.000	1.071	322	0.431	0.000	1.631	30071	0.432	0.000	880.493
massa_vol	64	1.073	1.008	1.134	64	1.072	1.008	1.134	64	1.072	1.008	1.134	0.000	1.075	1.008	1.163
orientacao_nac	64	197.385	0.000	360.000	64	192.071	0.000	360.000	64	193.897	0.000	427.000	0.000	171.400	0.000	-106.000
bar	6548	87678	85000	89200	6548	87678	85000	89200	6548	87678	85000	89200	371660	87752	84000	89800
tep_med	6548	12.287	-3.200	29.800	6548	12.287	-3.200	29.800	6548	12.287	-3.200	29.800	370280	11.459	-5.400	31.300
hmd_med	6548	69.793	5.000	100.000	6548	69.793	5.000	100.000	6548	69.793	5.000	100.000	370270	72.382	1.000	100.000
int_turb	6336	0.189	0.000	1.750	6336	0.189	0.000	1.750	6336	0.189	0.000	1.750	294211	0.128	0.000	2.452

We used the median strategy for replaced our missing data. We believe this strategy is adequate because of low expected variability in adjacent samples. After cleaning and replacing manipulations, all data sets were recorded into separate files and will be used in data mining method development to recognize outliers from performance curves.

5.3.3.3 Normalization of data

Variable Normalization is one of the most important concepts of predictive modeling. The data should be normalized or standardized to bring all of the variables into proportion with one another. For example, if one variable is 100 times larger than another (on average), then the model may be better behaved if we normalize/standardize the two variables to be approximately equivalent. The normalized/standardized coefficients associated with each variable will scale appropriately to adjust for the disparity in the variable sizes. Moreover, the normalized/standardized coefficients will reflect meaningful relative activity between each variable: i.e., a positive coefficient will mean that the variable acts positively towards the objective function, and vice versa; a large coefficient versus a small coefficient will reflect the degree to which that variable influences the objective function. In contrast, the coefficients from un-normalized/un-standardized data will reflect the positive/negative contribution towards the objective function, but will be much more difficult to interpret in terms of their relative impact on the objective function.

Traditionally data normalization entails fitting the data within unity (1), so that all data values will take on a value in the interval from 0 to 1. This approach is also called Min-Max Scaling. This method is used to make equal ranges but different means and standard deviations. The formula is shown in equation 5.1.

$$\bar{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{5.1}$$

The motivation to use this scaling includes robustness to very small standard deviations of features and preserving zero entries in sparse data.

5.3.3.4 Codification of data

In several cases data and events inside a time series are seasonal. In such cases the month and the year of the event matters significantly. In such scenarios binary variables are commonly used to represent if the event is during a given month/year or not.

From our previous analysis we concluded that power produced varies from month to month (see Figures 5.2b and 5.2a). For this reason the variables from date data were binary-encoded into months. This resulted of the 12 new features in ours data sets. All data sets after data preparation process were recorded into separate files and will be used in development of prediction models.

5.3.4 Modeling

Within the field of machine learning, there are two main types of tasks: *supervised* and *unsupervised*. The main difference between the two types is that supervised learning is done using prior knowledge of what the output values for samples should be. The goal of supervised learning is to learn a function that, given a sample of data and desired outputs, best approximates the relationship between input and output observable in the data. Unsupervised learning, on the other hand, does not have labeled outputs, so its goal is to infer the implicit structure present within a set of data points.

Supervised learning is typically done in the context of classification, when we want to map input to output labels, or regression, when we want to map input to a continuous output. Common algorithms in supervised learning include logistic regression, naive bayes, support vector machines, artificial neural networks, and random forests. In both regression and classification, the goal is to find specific relationships or structure in the input data that allow us to effectively produce correct output data.

The most common tasks within unsupervised learning are clustering, representation learning, and density estimation. In all of these cases, we wish to learn the inherent structure of our data without using explicitly-provided labels. Some common algorithms include k-means clustering, principal component analysis, and autoencoders. Since no labels are provided, there is no specific way to compare model performance in most unsupervised learning methods.

It is easy to see that supervised type of task is more suitable in the context of classification, considered in this work. Recall that the objective of this work is solve two problems: detection of data rejected from performance curve of wind turbines, and development of a prediction model to identify status patterns of wind turbines.

For solving the first task we used the data sets resulting from cleaning and replacing manipulations of data, see in Section 5.3.3.2. The second task of status pattern identification will build on the results of the first one.

A wind turbine is expected to produce a certain amount of energy for a given wind speed. The relationship between the wind speed and its power output is expressed as a power curve, which has a logistic function shape. For a variety of reasons, discussed below, the power curve is not an ideal logistic function. In fact, all regions outside of the logistic curve represent power losses. This abnormality of power curves is one of the central factors in this part of our work.

The statistical results discussed point to four main sources of outliers [43]:

Wind speed. The low output power is due to the wind speed around the cut-in point (the cut-in speed is set at 2.4 m/s) or the wind speed around the cut-out point (the cut-out speed is set at 20 m/s - 32 m/s). A turbine with wind speed below the cut-in point operates abnormally because insufficient wind energy cannot power the turbine. On the other hand, wind speed above the cut-out point causes the turbine to vibrate. To avoid the negative impact of high wind speed on the turbine's lifecycle, the control system shuts down its operation.

Environmental issues. Environmental issues other than the wind speed may produce power curve outliers. Blades affected by dirt, bugs, and ice may impact power curves of individual turbines and produce outliers reflected in the wind farm power curve.

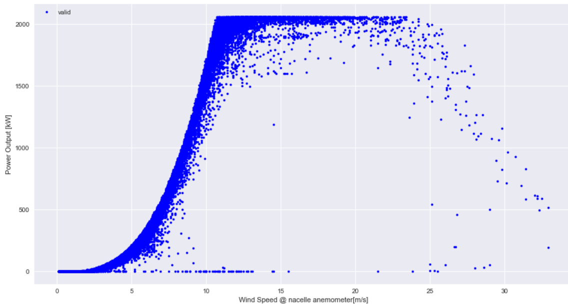
Control system issues. The conditions of the wind could be in the normal range, yet the power produced could be below the values indicated by the power curve. A possible reason is that the control parameters may be not appropriate for the wind regime. The specific reason may be attributed to the malfunction of the sensors, pitch control malfunctions, blade pitch angle errors, blade damage, control program problems, incorrect controller settings, constrained operations, and so on.

Maintenance or energy curtailment. Scheduled or unscheduled maintenance operations as well as energy curtailment due to diminished transmission capacity may lead to disrupted operations of individual turbines or the entire wind farm.

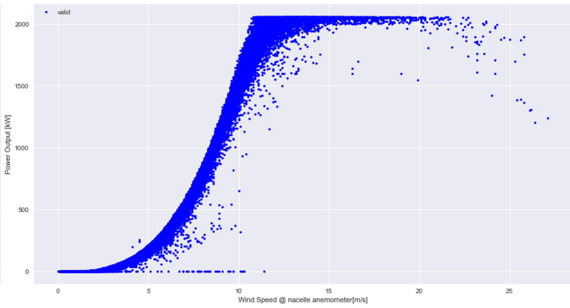
In Figure 5.3 we present a global view of the power curves for the all wind turbines under normal operation, included in our four data sets. All produced data for power output was classified as normal data, including data classified by INEGI as rejected data (see in Table 5.2).

5.3.5 Evaluation and Deployment

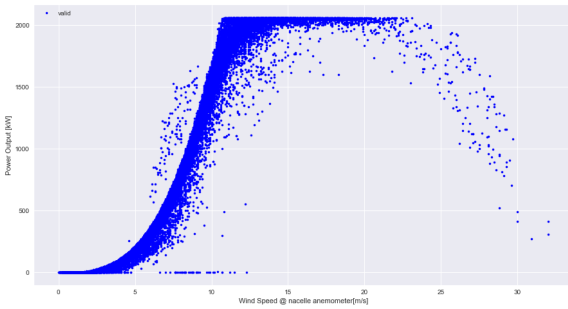
The most used approach to deal with the nonlinear characteristic of the wind turbine power curve is to consider a discrete model. For example, the method for estimating the power curve, described in the IEC 61400-12 [31] standard, takes a discretization of the wind speed range into bins of 0.5 m/s of size. For each bin the mean and standard deviation of the produced power is computed and used to characterize the power curve. The different models have been developed, applied and analyzed with the objective of detecting erroneous data.



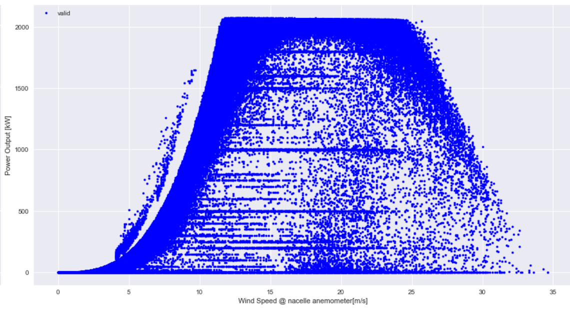
(a) Cumulative power curve of the wind turbine E 70 (data set WT_70E).



(b) Cumulative power curve of the wind turbine E 82 (data set WT_82E).



(c) Cumulative power curve of the wind turbine E 92 (data set WT_92E).



(d) Cumulative power curve of the wind farm (data set WT_FARM).

Figure 5.3: Global view of the power curves considering all data points from the four data sets.

The WT_70E, WT_82E, WT_92 and WTS_FARM data sets comprised of 52330 data points, 52343 data points, 51732 data points and 2536934 data points, respective observations were considered.

- ***Data filtering by binned mean $\pm$ 2.57σ criterion.***

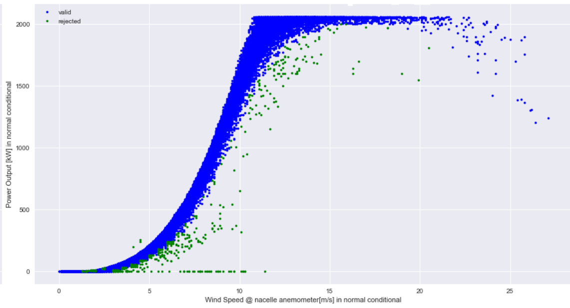
  The first method that was applied to detecting rejected data is "binned mean" [67]. It consists in computing for each bin the mean value (μ) and the standard deviation (σ) of the power, similar to the IEC standard. In this case, all the data that are bigger than $\mu + 2.57\ \sigma$ or smaller than $\mu - 2.57\sigma$ are considered as rejected data. If the noise was Gaussian, this threshold $th = 2.57$ supposes that 1% of the data eliminated is correct. The resulting power curves with the rejected data detected for the four data sets is presented in Figure 5.4.

- ***Data filtering using KNN classifier method.***

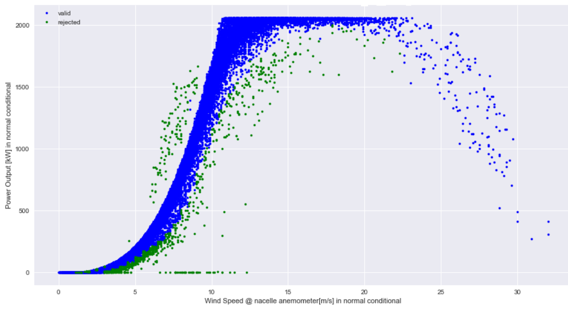
  In this method, k-NN is used to predict wind farm power based on the wind speed.

  K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure.

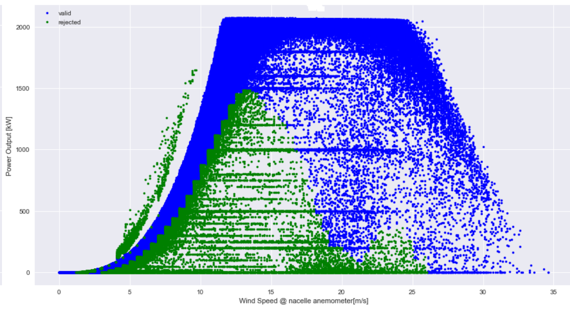
(a) Linear model for wind turbine E 70 E with binned speed partition (data set WT_70E).



(b) Linear model for wind turbine E 82 E with binned speed partition (data set WT_82E).



(c) Linear model for wind turbine E 92 E with binned speed partition (data set WT_92E).



(d) Linear model for wind farm with binned speed partition (data set WT_FARM).

Figure 5.4: Linear models for the four data sets with binned speed partition.

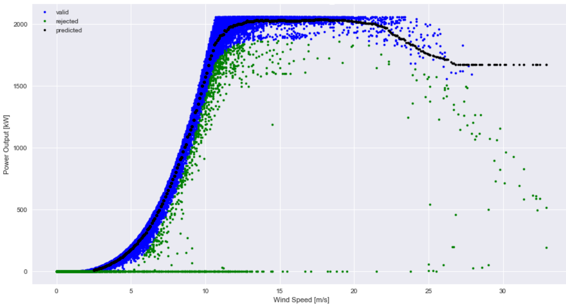
In [43] research study, the parameter k in each k-NN model was optimized for prediction accuracy, and the values of k were chosen for best accuracy. The best performance of the k-NN model was observed for $k = 250$.

The basic steps of the k-NN algorithm are as follows:

1. Represent each instance in a multi-dimensional space.
2. Divide the entire data set into training and test data sets.
3. Given a test instance, a distance metric is computed between the test instance and all training instances, then the k-nearest neighbors are selected from the training data.
4. The predicted value of the target variable for the test instance is the average value of the target variable of the selected neighbors.

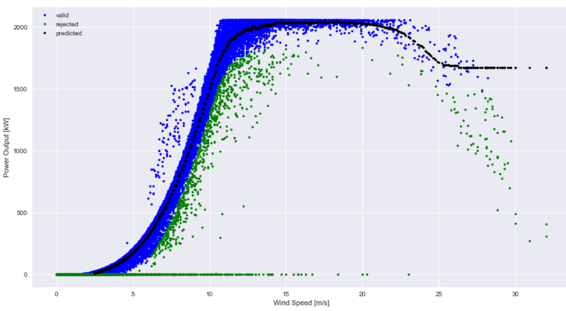
The rejected data points are those whose prediction error deviates from mean error by more than 2.57 criterion of standard deviation. The resulting power curves with the rejected data detected for the four data sets is presented in Figure 5.5.

- ***Data filtering using KNN&Bin classifier method.***

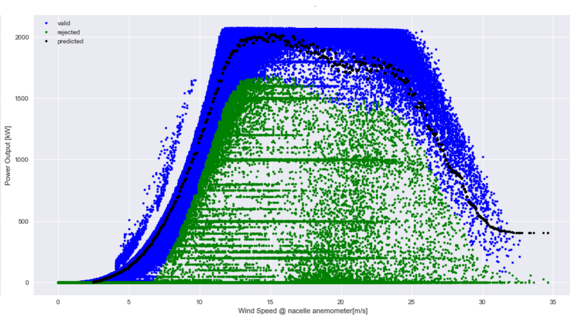


(a) Knn classification power curve for WT 92 E (data set WT_70E).

(b) Knn classification power curve for WT 82 E (data set WT_82E).



(c) Knn classification Power Curve for WT 92 E (data set WT_92E).



(d) Knn classification power curve for wind farm (data set WT_FARM).

Figure 5.5: Knn classification power curves for the four data sets.

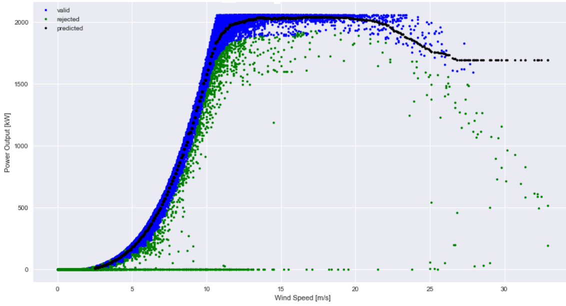
We observe that the first two methods showed poor performance. In order to significantly improve the performance curve, we propose a combination of KNN method and outliers detection based on the binned $mean \pm 2.57r$ criterion. The resulting power curves with the rejected data detected in the four data set is presented in Figure 5.6.

- ***Data filtering using INEGI internal procedures.***

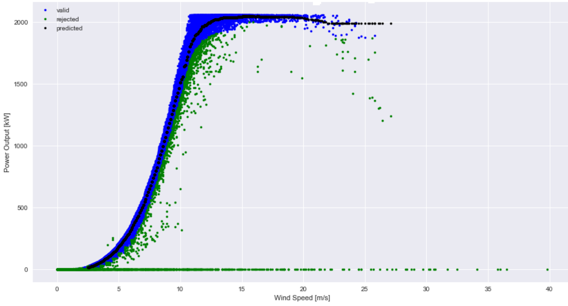
During its development, several methods were used and improved over time by INEGI company. A number of other methods have also been tested and dismissed due to their unsatisfactory performance.

Currently the rejection of data is done by calculating *Cp*. The theoretical *Cp* is given by the equation 3.10 (see in Section 3.2). The *real Cp* is derived from the SCADA values of the power and wind speed and with the values of the density provided by the meteorological station.

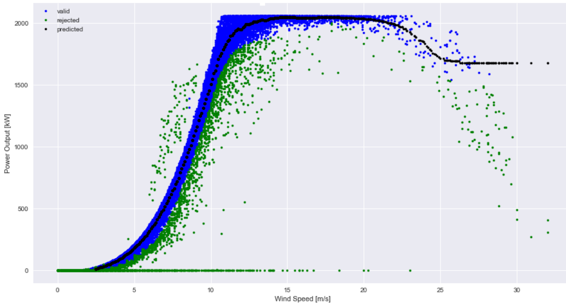
The *real Cp* and the theoretical *Cp* are calculated, and the difference between these two values is subsequently analyzed. Based on this analysis the decision is made as to whether the register should be rejected.



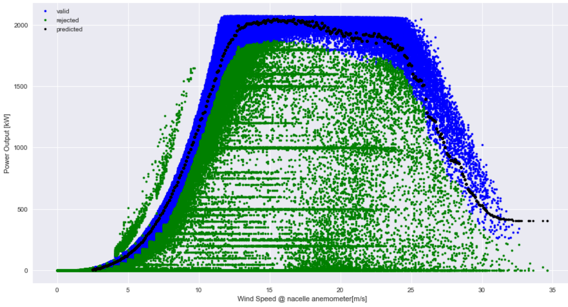
(a) Knn&Bin classification power curve for WT 92 E (data set WT_70E).



(b) Knn&Bin classification power curve for WT 82 E (data set WT_82E).



(c) Knn&Bin classification power curve for WT 92 E (data set WT_92E).



(d) Knn&Bin classification power curve for wind farm (data set WT_FARM).

Figure 5.6: Knn&Bin classification power curves for the four data sets.

We used the values of SCADA speed and density to compute *theoretical Cp*, and we interpolated from the description provided by the manufacturer the *theoretical Cp*.

By dividing the two values of Cp, we get the following equation:

$$v_{real} = v_{theor} \sqrt[3]{\frac{Cp_{real}}{Cp_{theor}}} \qquad (5.2)$$

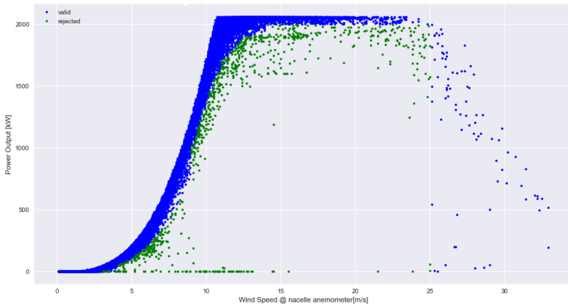
We can now calculate the theoretical speed from the above equation.

Having the two velocity values, we calculated the error between them. From here, the error limit values are set manually, and the values outside the limits are rejected. These limits are defined separately for each wind turbine and for each speed bin.

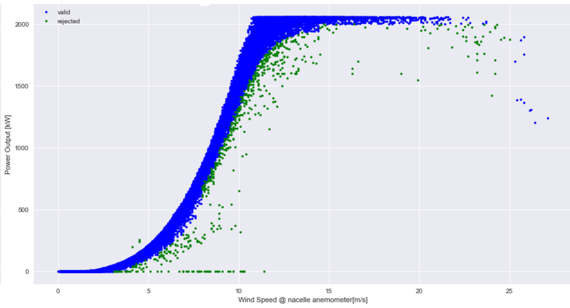
The resulting power curves with the rejected data detected in the four data sets is presented in Figure 5.7.

5.3.5.1 Evaluation of prediction models

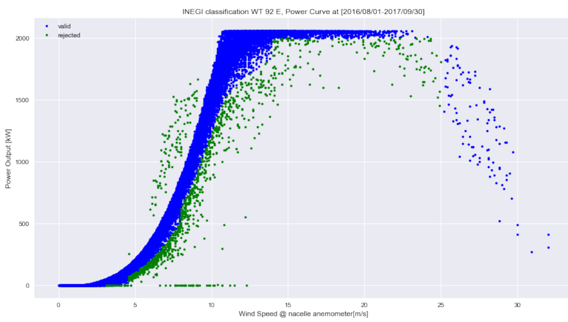
The second part of the work addresses the development of prediction models to identify and predict status patterns of wind turbines.



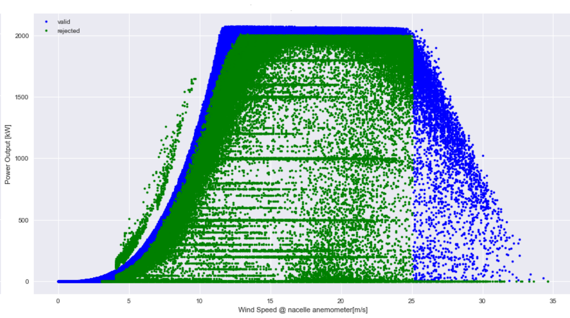
(a) INEGI method of classification power curve for WT 92 E (data set WT_70E).



(b) INEGI method of classification power curve for WT 82 E (data set WT_82E).



(c) INEGI method of classification power curve for WT 92 E (data set WT_92E).



(d) INEGI method of classification power curve for wind farm (data set WT_FARM).

Figure 5.7: INEGI method of classification for the four data sets.

For this purpose were selected four different machine learning algorithms for classification: Decision Tree, Nearest Neighbors, Multi-layer Perceptron and Gaussian Naive Bayes. The main idea of these algorithms is described briefly below:

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches. Leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

Nearest Neighbor methods look for a predefined number of training samples closest in distance to the new point, and predict the label from these. The number of samples can be a user-defined constant (k-nearest neighbor learning), or vary based on the local density of points (radius-based neighbor learning). The distance can, in general, be any metric measure: standard Euclidean distance is the most common choice

Table 5.4: Rejected data under normal operation of wind turbines for all data sets.

Techniques	WT_70E (52330 data points)	WT_82E (52343 data points)	WT_92E (51732 data points)	WTS_FARM (2536934 data points)
	% of rejected data under normal operation			
Binned mean Classifier	1.171 %	1.132 %	1.977 %	1.391 %
KNN Classifier	11.845 %	11.640 %	11.892 %	16.506 %
INEGI Classifier	1.614 %	1.203 %	2.271 %	2.448 %
KNN&Bin Classifier	12.734 %	12.911 %	13.745%	17.956 %

Multilayer perceptron classifier (MLPC) is a classifier based on the feedforward artificial neural network. MLPC consists of multiple layers of nodes. Each layer is fully connected to the next layer in the network. Nodes in the input layer represent the input data. All other nodes map inputs to outputs by a linear combination of the inputs with the node's weights and bias and applying an activation function. MLPC employs backpropagation for learning the model.

Naive Bayesian classifier is based on Bayes theorem with the independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

For development of prediction models we used all four data sets recorded after data preparation process (see Subsection 5.3.3.4), consisting of 3546334 instances with 23 features for WTS_FARM, and 61344 instances with 23 features for WT_70E, WT_82E, WT_92E.

Our target variable was status classifier of wind turbines provided by INEGI (see in Table 5.2). Note that data for rejected classifiers was replaced with classifiers that resulted from data filtering using KNN&Bin method.

We divided the available data into training (70% of data) and test (30% of data) sets 20 using different instantiations of a random number generator. In order to evaluate the predicted models were did multiple runs for each algorithm.

The machine learning algorithm with the best performance was identified and applied for training prediction model on all historical data of the four data sets.

5.4 Results

To evaluate the quality of the filtering techniques considered in this work, we compare the percentages of the rejected data under normal operation of wind turbines. The results are presented in Table 5.4.

Analyzing the on-line monitoring results of a power curve for a single turbine and wind farm, we see that the KNN&Bin model achieved the best performance. The parametric Binned mean and INEGI, and non-parametric KNN achieved average performance in detecting anomalies in the power generation.

We observe that each of the four approaches has its own advantages and disadvantages.

The k-NN algorithm is more computationally expensive than the Binned mean approach in identifying anomalies. The computational time depends on the parameter k and the size of the training data set. However, the k-NN algorithm training time is almost negligible, as it is an instanced-based scheme. In certain cases it can be more robust in detecting anomalies. Although the Binned mean method is less expensive, it showed worse result. In Figure 5.4 depicting the power curves from Binned mean filtering method, we see that data below the referencing curve (i.e., the curve close to the theoretical optimal curve) almost half of the data from abnormal operation was not detected. In contrast, the KNN model showed a much higher accuracy (please see Figure 5.5). In detecting wrong data above the referencing curve, Binned mean results were close to those of the KNN method.

The INEGI method has the disadvantage that it required manual definition of the errors limits for each wind turbine. Each time the anemometer transfer function is changed, it is necessary to change the error limits. According to INEGI, it was also verified that this method performed differently for different manufacturers. In addition to being a method that involves some initial manual manipulation, this method is not easily generalizable. In the results of our study (please see Figure 5.7), the power curves resulting from INEGI filtering method use error wind speed limit of 25 m/s. The data points outside the limit were accurately detected for rejection. However, this strategy could be applied only to wind turbines with storm control deactivated, which was not the case in all scenarios.

We see that the results obtained from our KNN&Bin method are of the highest quality in comparison to the other three State-of-the-Art power curves. This method showed high robustness rejecting wrong data, both for a single turbine with different technical descriptions and for the entire wind farm. In particular, the KNN&Bin filtering method showed comparable performance to the Binned mean and KNN methods in detecting wrong data above and below the referencing curve (please see Figure 5.6). Additionally, note that this method also takes in consideration the existence of a special storm control feature.

We now present the evaluation results of prediction models for status patterns of wind turbines. For this purpose were selected four different machine learning classification's algorithms with different behinds: *i)* Decision Tree (DTreeClas), *ii)* Nearest Neighbors (NeareNeighborsClas), *iii)* Multi-layer Perceptron (MLPClas) and *iv)* Gaussian Naive Bayes (GNaiveBayes).

To evaluate the performance of these prediction models we performed 10 different runs, selecting 1000000 random data points of data for each run, for training on WTS_FARM

Table 5.5: Performance and computational time for WT_70E, WT_82E WT_92E data sets.

Run		DTreeClas		NeareNeighborsClas		MLPClas		GNaiveBayes	
		Accuracy score	Computing time	Accuracy score	Computing time	Accuracy score	Computing time	Accuracy score	Computing time
WT_70E	1	0.988	0.214	0.885	2.109	0.974	21.767	0.158	0.034
	2	0.989	0.232	0.883	2.005	0.975	11.836	0.113	0.046
	3	0.991	0.240	0.885	1.889	0.972	13.377	0.055	0.031
	4	0.991	0.251	0.880	2.034	0.972	15.890	0.146	0.019
	5	0.989	0.240	0.882	2.066	0.975	19.005	0.149	0.046
WT_82E	1	0.987	0.179	0.881	2.441	0.971	17.098	0.060	0.046
	2	0.986	0.185	0.889	2.472	0.971	16.609	0.062	0.039
	3	0.987	0.178	0.886	2.257	0.973	17.407	0.004	0.090
	4	0.986	0.368	0.886	3.215	0.967	14.140	0.010	0.080
	5	0.985	0.271	0.886	2.371	0.974	17.520	0.017	0.049
WT_92E	1	0.986	0.380	0.878	3.286	0.973	21.966	0.189	0.031
	2	0.986	0.244	0.879	2.456	0.975	21.729	0.085	0.048
	3	0.987	0.226	0.877	3.415	0.975	19.790	0.098	0.046
	4	0.987	0.221	0.877	2.560	0.971	15.821	0.052	0.057
	5	0.987	0.275	0.881	5.914	0.971	16.666	0.025	0.032

data set, and 5 different runs selecting random data points of totality data for training on WT_70E, WT_82E, WT_92E data sets. It is well known that machine learning algorithms have stochastic nature. Repeated runs reduce the effect of randomness on the results.

In our evaluation we used the metric called Accuracy Classification Score (i.e., a ratio of correctly predicted observation to the total observations). In addition, we measured the computational time for each run. The results from the multiples runs for tree WT_70E, WT_82E, WT_92E data sets and for WTS_FARM are shown in Table 5.5 and Table 5.6 respectively.

Table 5.6: Performance and computational time on randomly selected data for WTS_FARM data set.

Run	DTreeClas		NeareNeighborsClas		MLPClas		GNaiveBayes	
	Accuracy score	Computing time	Accuracy score	Computing time	Accuracy score	Computing time	Accuracy score	Computing time
1	0.973	12.566	0.834	2063.686	0.951	554.556	0.221	1.366
2	0.974	13.750	0.834	1939.527	0.955	431.961	0.290	1.475
3	0.974	14.562	0.833	2288.701	0.945	847.745	0.307	1.311
4	0.973	12.762	0.835	2275.945	0.953	485.947	0.345	1.385
5	0.973	14.434	0.833	2049.392	0.941	589.087	0.337	1.481
6	0.973	13.981	0.834	2792.353	0.955	607.763	0.222	1.438
7	0.974	14.047	0.833	2411.119	0.954	618.685	0.297	1.475
8	0.974	13.945	0.834	2052.550	0.957	458.841	0.261	1.309
9	0.974	12.700	0.833	2035.999	0.955	494.407	0.153	1.288
10	0.974	12.329	0.833	1782.820	0.953	663.235	0.319	1.385

Table 5.7: Statistical analysis of performance and computational time on randomly selected data for all data sets.

		DTreeClas			NeareNeighborsClas			MLPClas			GNaiveBayes		
Data set	Performance	min	mean	max	min	mean	max	min	mean	max	min	mean	max
WT_70E (5 runs)	Accuracy score(%)	0.988	**0.990**	0.991	0.880	0.883	0.885	0.972	0.974	0.975	0.055	0.124	0.158
	Computing time(min)	0.214	0.235	0.251	1.889	2.021	2.109	11.836	16.375	21.767	0.019	0.035	0.046
WT_82E (5 runs)	Accuracy score(%)	0.985	**0.986**	0.987	0.881	0.886	0.889	0.967	0.971	0.974	0.004	0.030	0.062
	Computing time(min)	0.178	0.236	0.368	2.257	2.551	3.215	14.140	16.555	17.520	0.039	0.061	0.090
WT_92E (5 runs)	Accuracy score(%)	0.986	**0.986**	0.987	0.877	0.878	0.881	0.971	0.973	0.975	0.025	0.090	0.189
	Computing time(min)	0.221	0.269	0.380	2.456	3.526	5.914	15.821	19.195	21.966	0.031	0.043	0.057
WTS_FARM (10 runs)	Accuracy score(%)	0.973	**0.974**	0.974	0.833	0.834	0.835	0.941	0.952	0.957	0.153	0.275	0.345
	Computing time(min)	12.329	13.507	14.562	1782.820	2169.209	2792.353	431.961	575.223	847.745	1.288	1.391	1.481

The statistical analysis of performance and computational time for multiples runs of all machine learning algorithms and all data sets is presented in Table 5.7. We see that, out of the four models, only DtreeClass and MLPClass showed satisfactory performance.

Looking at the accuracy score from the multiple runs, we see that the DTreeClass model has the better average performance (in bold font in Table 5.7) for all four data sets. Moreover, its computational time is in the low range, when compared to other approaches. MLPClass, despite its good performance with respect to accuracy score classifier metric, is significantly more time consuming than DTreeClass. Therefore, MLPClass is more limited in its scalability.

In contrast, NeareNeighborsClass and GNaiveBayes showed poor performance. Specifically, the averages of accuracy score are several times lower than the acceptable values for prediction techniques. Since the error metric is unsatisfactory, we do not discuss the computational times of these approaches and consider NeareNeighborsClass and GNaiveBayes not suitable for our prediction task.

Finally, we note that the performance and computational time of prediction models depend on the volume of data: as the volume of data increases, the performance and computational time is expected to also increase. Therefore we take in consideration both the performance and computational time and conclude that the DtreeClass approach is the most suitable for our task of prediction of status patterns.

5.5 Summary

In this section we described in detail the methodology used for this research work. Specifically, We followed the CRISP-DM reference model. CRISP-DM reference model for data mining provides an overview of the life cycle of a data mining project. It contains the phases of a project, their respective tasks, and their outputs. We chose CRISP-DM because it has become a standard methodology for execution of data-mining projects across industries.

We then presented in detail the setting for our study and described the approaches chosen for evaluation. In particular, in the first part we evaluated three existing power curve approaches: Binned mean, KNN, and INEGI method. We presented the evaluation results and discussed the advantages and disadvantages of each approach. Based on this analysis, we proposed an improved solution that combines KNN and Binned approaches. We called this solution KNN&Bin. In our evaluation, we showed that KNN&Bin outperforms the other three existing approaches.Our findings were also presented to INEGI and it was suggested that the KNN&Bin approach is the one that best suit their needs.

In the second part, we evaluated four different techniques for prediction of status patterns of wind turbines: Decision Tree, Nearest Neighbors, Multi-layer Perceptron and Gaussian Naive Bayes. The results show that Decision Tree based approach performed best with respect to accuracy score classifier metric and computational time, on our data set. Hence, this technique was selected for further deployment in production at INEGI.

Chapter 6

Conclusions and Future Work

Wind energy is a fast growing renewable energy source, however to make wind power competitive with other sources of energy, availability, reliability and the life time of turbines will all need to be improved. As the wind energy sector grows, business economics, will demand increasingly careful management of Operation and Maintenance costs.

Unfortunately, faults and unscheduled shutdowns of wind turbines are costly. As the size and number of wind turbines continue to rise, monitoring for faults has become increasingly important for companies to remain competitive. Moreover, early prediction of status patterns may allow for predictive maintenance actions and possible avoidance of faults. This can be achieved through identification of status patterns of wind turbine, as the health of a system or component may deteriorate gradually, rather than failing instantly.

In our research we considered power curve models of wind turbines. Power curve is an important characteristic since it is used in energy assessment, warranty formulations, and performance monitoring of the turbines. Power curve models are also important for turbine health monitoring. Given an accurate power curve model, performance deviation can be detected by comparing the predicted output to the actual output, adjusted for the current conditions (typically collected in real time by a SCADA system).

Accurate characterization of the wind turbine power curve is necessary to optimize the operation and maintenance of a wind farm. The accuracy of such characterizations depends on the amount and quality of measurement data collected from real turbines in operation. It is well known that such measurements are prone to outliers, so the application of automatic filtering techniques is essential to deal with this problem.

We used a robust statistical technique to filter the raw data taken from a real wind farm. To study the nonlinear nature of the power curve in function of the wind speed and power output, we tested and compared different models.

More specifically, in the first part of this work, were compared the anomaly detection power of three existing power curve models using the following techniques: *i)* Data filtering by binned $mean \pm 2.57\sigma$ criterion, *iii)* Data filtering using KNN classifier method, *iii)*

Data filtering using INEGI method. The power curves were developed using historical wind turbine data.

We showed that the above techniques reduce the need for various filtering steps usually done manually to reject outliers, reducing the time and costs required for the process in a great factor. Moreover, this techniques have shown potential to be employed in automatic systems that evaluate the efficiency deviations of the installed wind power turbines. However, as we discussed in the previous chapter, each approach showed some technical limitations. In particular, Data filtering by binned $mean \pm 2.57\sigma$ criterion showed high level of inaccuracy detecting abnormal data points. Data filtering using KNN classifier method showed better accuracy, but it is computationally more expensive. Finally, the INEGI method reveal the required manual steps for setting error limits, which greatly hurts its generality and scalability.

Based on the above analysis, we introduced an improved power curve model called Data filtering using KNN&Bin method. We showed that this model outperforms all three existing models on the historical data used for this research. The KNN&Bin method demonstrated high accuracy at no additional computational cost, compared to other solutions. As a result of its demonstrated accuracy and practicality, our approach will be suggested to INEGI.

The second part of the research addressed the problem of real-time status reporting of a turbine, based on status classification used at INEGI.

Automated monitoring of the performance of wind turbines and early fault prediction is an effective way to reduce operational costs. However, traditional maintenance strategies, such as reactive maintenance or periodic maintenance, are more prevalent in the wind industry. Fortunately, over the last couple of years, the research pertaining to wind turbines started to also address the condition monitoring and maintenance. The goal of condition monitoring is to provide continuous monitoring of the wind turbines and identify fault signatures in the event of faults. Most of the studies reported in literature are preliminary: they are based on the simulated dataset or constrained experiments. In reality, the external environment plays an important role in governing the turbine operations. Note that condition monitoring bares additional cost of installations of specific sensors and other equipment. Therefore, optimization of the process is highly relevant.

Our approach uses data mining and several machine learning algorithms to identify and predict status patterns of wind turbines. Monitoring the performance of overall wind farm provides the current status of all wind turbines installed in a wind farm. The goal of prediction is to facilitate planned future maintenance. As a result, the performance of overall wind farm can be assessed on a daily, weekly, or monthly basis, depending on the requirements.

For our study we selected four different machine learning algorithms for classification: Decision Tree, Nearest Neighbors, Multi-layer Perceptron and Gaussian Naive Bayes. The prediction models were trained on operational and status data collected from SCADA systems of 13 wind turbines of 3 different models, and the recorded data of the closest weather stations. We compared the four approaches with regard to accuracy classification score metric and computational time. We showed that he best performance was displayed by the Decision Tree based approach, which we selected for further deployment in production.

In future research, it will be interesting to focus on development of more comprehensive power curves with more parameters such as wind direction, rotor speed, etc. We believe that such models will be able to further minimize the prediction errors and will be more suitable for on-line monitoring of turbines.

Other promising techniques of the machine learning approach, such as Artificial Neural Networks, can be applied to predict fault status patterns of wind turbines. The quality of such networks is improved as more data is collected and ingested into the model. Unfortunately, handling large volumes of data incurs high computational cost and is often time consuming. This can be addressed by parallel computation techniques, such as map-reduce.

References

[1] Irena greeted as milestone for renewables. *Renewable Energy Focus*, 9(7):19, 2009.

[2] M.h. Abderrazzaq. Energy production assessment of small wind farms. *Renewable Energy*, 29(15):2261–2272, 2004.

[3] I. Abouzahr and R. Ramakumar. Loss of power supply probability of stand-alone wind electric conversion systems: a closed form solution approach. *IEEE Transactions on Energy Conversion*, 5(3):445–452, 1990.

[4] E. Anahua, St Barth, and J. Peinke. Markovian power curves for wind turbines. *Wind Energy*, 11(3):219–232, 2008.

[5] Tore Bakka and Hamid Reza Karimi. Bond graph modeling and simulation of wind turbine systems. *Journal of Mechanical Science and Technology*, 27(6):1843–1852, 2013.

[6] John P. Barton and David G. Infield. Energy storage and its use with intermittent renewable energy. *IEEE Transactions on Energy Conversion*, 19(2):441–448, 2004.

[7] Gordon S. Berry, Michael J. A. ; Linoff. *Data mining techniques*, volume 25. Wiley, Inc., second edition edition, 1996.

[8] M. Siddhartha Bhatt and S. Jothibasu. Performance enhancement in coal fired thermal power plants. part i: boilers. *International Journal of Energy Research*, 23(6):465–487, 1999.

[9] Qingzhen Bi, Hua Chen, Ding Tong, and Yaoan Lu. A design method for globally curvature-smooth centrifugal compressor blades. *Volume 2C: Turbomachinery*, 2016.

[10] R. Billinton and G. Bai. Generating capacity adequacy associated with wind energy. *IEEE Transactions on Energy Conversion*, 19(3):641–646, 2004.

[11] Shailendra Kumar Bohidar, Ritesh Sharma, and Prabhat Ranjan Mishra. Wind Energy – A Step towards Non-Conventional Energy Resources. *International Journal of Research*, 1(6):229–234, 2014.

[12] S. M. Camporeale, B. Fortunato, and G. Marilli. Automatic system for wind turbine testing. *Journal of Solar Energy Engineering*, 123(4):333, 2001.

[13] Peter Caselitz and Jochen Giebhardt. Rotor Condition Monitoring for Improved Operational Safety of Offshore Wind Energy Converters. *Journal of Solar Energy Engineering*, 127(2):253, 2005.

[14] Tian Pau Chang, Feng Jiao Liu, Hong Hsi Ko, Shih Ping Cheng, Li Chung Sun, and Shye Chorng Kuo. Comparative analysis on power curve models of wind turbine generator in estimating capacity factor. *Energy*, 73:88–95, 2014.

[15] Jui-Sheng Chou and Wan-Ting Tu. Failure analysis and risk management of a collapsed large wind turbine tower. *Engineering Failure Analysis*, 18(1):295–313, 2011.

[16] S. Diaf, M. Belhamel, M. Haddadi, and A. Louche. Technical and economic assessment of hybrid photovoltaic/wind system with battery storage in Corsica island. *Energy Policy*, 36(2):743–754, 2008.

[17] S. Diaf, G. Notton, M. Belhamel, M. Haddadi, and A. Louche. Design and techno-economical optimization for hybrid PV/wind system under various meteorological conditions. *Applied Energy*, 85(10):968–987, 2008.

[18] P.s. Dokopoulos, A.c. Saramourtsis, and A.g. Bakirtzis. Prediction and evaluation of the performance of wind-diesel energy systems. *IEEE Transactions on Energy Conversion*, 11(2):385–393, 1996.

[19] Caye M Drapcho, Nghiem Phu Nhuan, Terry H Walker, Wim Soetaert, Erick J Vandamme, John Wiley, V. Roman, D.; Villaverde, Jack Casazza, Frank Delea, V. Roman, D.; Villaverde, and Louis D Francis M. Vanek and Albright. *Energy Systems Enginneering*. Number 1. 2014.

[20] E. Echavarría, G. J. W. Bussel, and T. Tomiyama. Finding functional redundancies in offshore wind turbine design. *Wind Energy*, 15(4):609–626, 2011.

[21] Andrew Curtis Elmore and Ron Gallagher. Using regional climate center data to predict small wind turbine performance. *Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management*, 13(1):14–19, 2009.

[22] English Eco Energy. Wind turbines. Available in `http://www.englishecoenergy.com/wind_turbines_blackpool.html`, accessed in December 20, 2017.

[23] Ironside Farrar. Strategic Landscape Capacity Assessment for Wind Energy in Angus Final Report. (March), 2014.

[24] Caithness Windfarm Information Forum. Summary of wind turbine accident data to 30 september 2017. Available in `http://www.caithnesswindfarms.co.uk/AccidentStatistics.htm`, accessed in December 20, 2017.

[25] Richie Gao and Zhiwei Gao. Pitch control for wind turbine systems using optimization, estimation and compensation. *Renewable Energy*, 91:501–515, 2016.

[26] Simon Gill, Bruce Stephen, and Stuart Galloway. Wind turbine condition assessment through power curve copula modeling. *IEEE Transactions on Sustainable Energy*, 3(1):94–101, 2012.

[27] Paul Gipe. *Wind Energy Comes of Age*. 1995.

[28] Peng Guo and David Infield. Wind turbine tower vibration modeling and monitoring by the nonlinear state estimation technique (nset). *Energies*, 5(12):5279–5293, 2012.

[29] T. J. Hammons. Integrating renewable energy sources into European grids. *International Journal of Electrical Power and Energy Systems*, 30(8):462–475, 2008.

[30] S. Hildingsson and T. Westin. Yawing system for adjusting a wind turbine into a required wind direction by turbine the turbine about a yawing axle. 1999.

[31] IEC 61400-12-1, Wind turbines- Power performance measurement of electricity producing wind turbines. International standard, 2005.

[32] Tongdan Jin and Zhigang Tian. Uncertainty analysis for wind energy production with dynamic power curves. *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems*, pages 745–750, 2010.

[33] Fawzi A.L. Jowder. Wind power analysis and site matching of wind turbine generators in Kingdom of Bahrain. *Applied Energy*, 86(4):538–545, 2009.

[34] J.k. Kaldellis. Parametric investigation concerning dimensions of a stand-alone wind-power system. *Applied Energy*, 77(1):35–50, 2004.

[35] Rajesh Karki and Roy Billinton. Cost-effective wind energy utilization for reliable power supply. *IEEE Transactions on Energy Conversion*, 19(2):435–440, 2004.

[36] Krishnan R. Kazmiekowski, M.P. and F. Blaabjerg. *Control in PowermElectronics*. Academic Press, 2002.

[37] S. Kelouwani and K. Agbossou. Nonlinear model identification of wind turbine with a neural network. *IEEE Transactions on Energy Conversion*, 19(3):607–612, 2004.

[38] Christoph W. Kensche. Fatigue of composites for wind turbines. *International Journal of Fatigue*, 28(10 SPEC. ISS.):1363–1374, 2006.

[39] Yogesh Kumar, Jordan Ringenberg, Soma Shekara Depuru, Vijay K. Devabhaktuni, Jin Woo Lee, Efstratios Nikolaidis, Brett Andersen, and Abdollah Afjeh. Wind energy: Trends and enabling technologies. *Renewable and Sustainable Energy Reviews*, 53:209–224, 2016.

[40] Andrew. Kusiak. Short-Term Prediction of Wind Farm Power: A Data Mining Approach. *IEEE Transactions on Energy Conversion*, 24(1):125–136, 2009.

[41] Andrew Kusiak and Wenyan Li. Virtual models for prediction of wind turbine parameters. *IEEE Transactions on Energy Conversion*, 25(1):245–252, 2010.

[42] Andrew Kusiak and Anoop Verma. The Future of Wind Turbine diagnostics. *Wind System Magazine*, 2(4):66–71, 2010.

[43] Andrew Kusiak, Haiyang Zheng, and Zhe Song. Models for monitoring wind farm power. *Renewable Energy*, 34(3):583–590, 2009.

[44] Andrew Kusiak, Haiyang Zheng, and Zhe Song. On-line monitoring of power curves. *Renewable Energy*, 34(6):1487–1493, 2009.

[45] Donghuan Liu and Xinchun Shang. Failure investigation of the wind turbine blade root bolt. *Journal of Failure Analysis and Prevention*, 13(3):333–339, 2013.

[46] Fuzzy Logic, Controller For, Predicting Wind, and Turbine Power. Predicting Wind Turbine Power. (May):4–12, 2014.

[47] Bin Lu, Yaoyu Li, Xin Wu, and Zhongzhou Yang. A review of recent advances in wind turbine condition monitoring and fault diagnosis. *2009 IEEE Power Electronics and Machines in Wind Applications*, 2009.

[48] Wenxiu Lu and Fulei Chu. Condition monitoring and fault diagnostics of wind turbines. *2010 Prognostics and System Health Management Conference*, 2010.

[49] Shengping Lv, Hoyeol Kim, Binbin Zheng, and Hong Jin. A Review of Data Mining with Big Data towards Its Applications in the Electronics Industry. *Applied Sciences*, 8(4):582, 2018.

[50] M. Lydia, A. I. Selvakumar, S. S. Kumar, and G. E. P. Kumar. Advanced Algorithms for Wind Turbine Power Curve Modeling. *IEEE Transactions on Sustainable Energy*, 4(3):827–835, 2013.

[51] Martin J. Pasqualetti. History of wind energy. In Cutler Cleveland, editor, *Encyclopedia of Energy*, number January 2004, pages 419–433. Academic Press.

[52] E. Martínez, F. Sanz, S. Pellegrini, E. Jiménez, and J. Blanco. Life cycle assessment of a multi-megawatt wind turbine. *Renewable Energy*, 34(3):667–673, 2009.

[53] Sathyajith Mathew. *Wind Energy Fundamentals, Resource Analysis and Economics.* Springer, India, 2010.

[54] D. Mba. Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines: Bearings, pumps, gearboxes, engines, and rotating structures. *The Shock and Vibration Digest*, 38(1):3–16, Jan 2006.

[55] Harald Mellinghoff. Development of Power Curve Measurement Standards. *DEWI Magazin*, 2013.

[56] Patrick Milan, Matthias Wächter, and Joachim Peinke. Stochastic modeling and performance monitoring of wind farm power production. 033119(2014), 2015.

[57] Pedro S. Moura and Aníbal T. de Almeida. The role of demand-side management in the grid integration of wind power. *Applied Energy*, 87(8):2581–2588, 2010.

[58] L. Nanda Kishore and E. Fernandez. Reliability well-being assessment of PV-wind hybrid system using Monte Carlo simulation. *2011 International Conference on Emerging Trends in Electrical and Computer Technology, ICETECT 2011*, pages 63–68, 2011.

[59] Zaccheus O. Olaofe and Komla A. Folly. Wind energy analysis based on turbine and developed site power curves: A case-study of Darling City. *Renewable Energy*, 53:306–318, 2013.

[60] Joon-Young Park, Jae-Kyung Lee, Ki-Yong Oh, and Jun-Shin Lee. Development of a Novel Power Curve Monitoring Method for Wind Turbines and Its Field Tests. *IEEE Transactions on Energy Conversion*, 29(1):119–128, 2014.

[61] Dalibor Petković. Adaptive Neuro-Fuzzy Optimization of the Net Present Value and Internal Rate of Return of a Wind Farm Project under Wake Effect. 8(1):11–28, 2015.

[62] W. Richard Powell. An analytical expression for the average output power of a wind machine. *Solar Energy*, 26(1):77–80, 1981.

[63] Dipl.-Ing. Gianfranco Guidati (auth.) Prof. Dr.-Ing. Siegfried Wagner, Dipl.-Ing. Rainer Bareiß. *Wind Turbine Noise*. Springer-Verlag Berlin Heidelberg, 1 edition, 1996.

[64] Yan Ren, Fuming Qu, Jinhai Liu, Jian Feng, and Xiaodong Li. A universal modeling approach for wind turbine condition monitoring based on SCADA data. *Proceedings of 2017 IEEE 6th Data Driven Control and Learning Systems Conference, DDCLS 2017*, 2:265–269, 2017.

[65] Ahmet Duran Sahin. Progress and recent trends in wind energy. *Progress in Energy and Combustion Science*, 30(5):501–543, 2004.

[66] R. Saidur, N. A. Rahim, M. R. Islam, and K. H. Solangi. Environmental impact of wind energy. *Renewable and Sustainable Energy Reviews*, 15(5):2423–2430, 2011.

[67] E. Sainz, A. Llombart, and J. J. Guerrero. Robust filtering for the characterization of wind turbines: Improving its operation and maintenance. *Energy Conversion and Management*, 50(9):2136–2147, 2009.

[68] A.c. Saramourtsis, A.g. Bakirtzis, P.s. Dokopoulos, and E.s. Gavanidou. Probabilistic evaluation of the performance of wind-diesel energy systems. *IEEE Transactions on Energy Conversion*, 9(4):743–752, 1994.

[69] Bharat Kumar Saxena and K. V. S. Rao. Wind turbine failure analysis for wind farm at devgarh in rajasthan. *2013 International Conference on Renewable Energy and Sustainable Energy (ICRESE)*, 2013.

[70] F. Castro Sayas and R.n. Allan. Generation availability assessment of wind farms. *IEE Proceedings - Generation, Transmission and Distribution*, 143(5):507, 1996.

[71] Ali Sayigh. Renewable energy Ð the way forward. 64:15–30, 1999.

[72] Meik Schlechtingen, Ilmar Ferreira Santos, and Sofiane Achiche. Using data-mining approaches for wind turbine power curve monitoring: A comparative study. *IEEE Transactions on Sustainable Energy*, 4(3):671–679, 2013.

[73] G M Shafiullah, Amanullah M. T. Oo, Dennis Jarvis, A B M Shawkat Ali, and Peter Wolfs. Potential challenges: Integrating renewable energy with the smart grid. *2010 20th Australasian Universities Power Engineering Conference*, pages 1–6, 2010.

[74] G.M. Shafiullah, Amanullah M.T. Oo, A.B.M. Shawkat Ali, and Peter Wolfs. Potential challenges of integrating large-scale wind energy into the power grid–A review. *Renewable and Sustainable Energy Reviews*, 20:306–321, 2013.

[75] Shikha, T. S. Bhatti, and D. P. Kothari. Wind energy in india: Shifting paradigms and challenges ahead. *Journal of Energy Engineering*, 130(3):67–80, 2004.

[76] Andrew R.D. Smedley, Ann R. Webb, and Arnold J. Wilkins. Potential of wind turbines to elicit seizures under various meteorological conditions. *Epilepsia*, 51(7):1146–1151, 2010.

[77] Colin Smith. Fires are major cause of wind farm failure, according to new research, Jul 2014.

[78] B Stephen, S J Galloway, D McMillan, D C Hill, and D G Infield. A Copula Model of Wind Turbine Performance. *Power Systems, IEEE Transactions on*, 26(2):965–966, 2011.

[79] Roberto Tapia and A. Medina. Doubly-fed wind turbine generator control: A bond graph approach. *Simulation Modelling Practice and Theory*, 53:149–166, 2015.

[80] Wind Energy the Facts. Wind turbines technology. Available in https://www.wind-energy-the-facts.org/the-technical-challenge-of-a-unique-technology.html, accessed in December 29, 2017.

[81] Wei Tong. Fundamentals of wind energy, 2010.

[82] Ib Troen and Erik Lundtang Petersen. *European wind atlas*. 1989.

[83] Taner Üstüntaş and Ahmet Duran Şahin. Wind turbine power curve estimation based on cluster center fuzzy logic modeling. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(5):611–620, 2008.

[84] Ke-Sheng Wang, Vishal S. Sharma, and Zhen-You Zhang. SCADA data based condition monitoring of wind turbines. *Advances in Manufacturing*, 2(1):61–69, 2014.

[85] Waloddi Weibull. Wide applicability. *Journal of applied mechanics*, 103:293–297, 1951.

[86] Lütze H. Riesberg Riesberg-A. Weitkamp, R. and J. Anemüller. Means for adjusting the rotor blade of a wind power plant rotor. 2004.

[87] R.e. Wilson. Wind-turbine aerodynamics. *Journal of Wind Engineering and Industrial Aerodynamics*, 5(3-4):357–372, 1980.

[88] University of Massachusetts Amherst Wind Energy Center. Capacity Factor , Intermittency , Wind Power : Wind Power on the Community Scale Introduction : the variability of wind. pages 0–3, 2009.

[89] Wenxian Yang and Jiesheng Jiang. Wind turbine condition monitoring and reliability analysis by SCADA information. *2011 2nd International Conference on Mechanic Automation and Control Engineering, MACE 2011 - Proceedings*, pages 1872–1875, 2011.

[90] Jun Hua Zhao, Zhao Yang Dong, Xue Li, and Kit Po Wong. A framework for electricity price spike analysis with advanced data mining methods. *IEEE Transactions on Power Systems*, 22(1):376–385, 2007.

[91] J Ryan Zimmerling, Andrea C Pomeroy, Marc V D'Entremont, and Charles M Francis. Canadian estimate of bird mortality due to collisions and direct habitat loss associated with wind turbine developments. *Avian Conservation and Ecology*, 8(2):10, 2013.