FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Optimizing the Methodology to Support the Evaluation of Wind Turbine Performance

Hugo Sérgio Sousa Pinto

Master in Mechanical Engineering

Supervisor: António Ramos Silva

March 12, 2024

Optimizing the Methodology to Support the Evaluation of Wind Turbine Performance

Hugo Sérgio Sousa Pinto

Master in Mechanical Engineering

March 12, 2024

Resumo

A Europa tem demonstrado um compromisso crescente com a produção de energia a partir de fontes renováveis, visando reduzir as emissões de CO_2 para a atmosfera e ambicionando alcançar emissões líquidas nulas até 2050. Esta dissertação contribui para esse objetivo, analisando cerca de 9 anos de dados de um parque eólico em Portugal, composto por 12 turbinas. Após uma minuciosa limpeza e análise dos dados, foi possível compreender a influência de variáveis externas, como a proximidade de outras turbinas e a orografia, na eficiência das turbinas eólicas, assim como analisar as curvas de potência e o comportamento do vento. Através de uma análise PCA (Análise de Componentes Principais), evidenciou-se a correlação entre variáveis meteorológicas e a produção de energia. Recorrendo a redes neuronais com células LSTM, que são especialmente aptas para prever sequências temporais complexas, desenvolveu-se um modelo capaz de estimar a potência média produzida e a orientação da nacelle para um horizonte temporal de até 6 horas. O modelo alcançou resultados notáveis, com previsões muito próximas da realidade e erros associados à previsão extremamente baixos. A eficiência do modelo manteve-se constante para todas as turbinas, demonstrando assim uma boa capacidade de adaptação e aprendizagem para diferentes dados. Este modelo tem implicações significativas para a operação nos mercados intradiário e contínuo de eletricidade, permitindo prever com alta precisão a quantidade de energia que será injetada na rede e, consequentemente, minimizar os desvios associados à incapacidade de cumprir com a energia acordada. Além disso, o modelo pode servir de base para que, no futuro, os parques eólicos evoluam para sistemas mais inteligentes, nos quais as turbinas interajam entre si com o objetivo de maximizar a produção energética.

Palavras-chave: Energia Eólica, LSTM, Análise de Dados, Mercado de Eletricidade, Previsão de Energia, Redes Neuronais, Análise PCA (Análise de Componentes Principais), Eficiência Energética, Gestão de Parques Eólicos, Inteligência Artificial, Renováveis.

Abstract

Europe has shown a growing commitment to producing energy from renewable sources, aiming to reduce CO_2 emissions into the atmosphere and aspiring to achieve net-zero emissions by 2050. This dissertation contributes to this goal by analyzing about 9 years of data from a wind farm in Portugal, consisting of 12 turbines. After thorough cleaning and analysis of the data, it was possible to understand the influence of external variables, such as the proximity of other turbines and the orography, on the efficiency of wind turbines, as well as to analyze the power curves and wind behavior. Through a PCA (Principal Component Analysis), the correlation between meteorological variables and energy production was highlighted. Using neural networks with LSTM cells, which are especially suited for predicting complex temporal sequences, a model was developed capable of estimating the average power produced and the nacelle orientation for a time horizon of up to 6 hours. The model achieved remarkable results, with predictions very close to reality and extremely low associated prediction errors. The efficiency of the model remained constant across all turbines, thus demonstrating a good capacity for adaptation and learning to different data sets. This model has significant implications for operations in the intraday and continuous electricity markets, allowing for high-precision prediction of the amount of energy that will be injected into the grid and, consequently, minimizing deviations associated with the inability to meet the agreed energy. Furthermore, the model can serve as a basis for, in the future, wind farms evolving into smarter systems, in which turbines interact with each other with the goal of maximizing energy production.

Keywords: Wind Energy, LSTM, Data Analysis, Electricity Earket, Energy Forecasting, Neural Networks, PCA, Energy Efficiency, Wind Farm Management, Artificial Intelligence, Renewables

Acknowledgements

This dissertation represents the end of years of great effort and dedication. These were very happy years filled with many achievements. Looking back, I recognize times of significant growth, both personal and academic. I leave here confident and prepared for the next challenges. Therefore, I must thank all the people I met during this period: colleagues, professors, and friends. It's important to specifically highlight the professors and their importance in my life. Having good teachers is a stroke of luck, and I have always been fortunate to cross paths with excellent mentors who have played a significant role in who I am today.

I am immensely grateful to my father, my mother, and my dear brother. I am thankful to my father for all the efforts to provide me with a quality life, allowing me to study and take risks without fear; to my mother, for always listening, accompanying, and helping me, support that was essential for the successful completion of the course; to Dinis, I can only say that I love you and wish your happiness above all. I am extremely lucky to have you. I thank you for always keeping me humble, for allowing me to fly as high as I wish, but with my feet on the ground. I must also thank you, Maria, for everything: for the patience, for the help, for the affection, and for the love. You play a crucial role in my life, eternally grateful for having you in it. Thank you, Sobral family.

My thanks also go to my friends Vicente and Leonardo, for all the conversations, discussions, and sharing, which I will carry with me forever. To Fábio, Gonçalo and Nuno, thank you for all the moments we spent together. A general thank you to all my friends and family friends who, in some way, contributed during this phase. A hug of longing for those who have passed away in the meantime.

Grateful to those who directly influenced this dissertation. To Professor Antonio Ramos Silva, my heartfelt thanks for agreeing to embark on this adventure with me, for all the advice and for everything you taught me. To João Ferreira, for the help and advice, to Filipa Magalhães, to José Carlos Matos, and to the entire wind energy group at INEGI.

Finally, I thank God for all the blessings and opportunities granted.

Hugo

“Do not waste your time discussing what a good man should be.
Be one”

Marcus Aurelius

Contents

List of Figures

List of Tables

Acronyms

CO_2	Carbon Dioxide
DGEG	Directorate General of Energy and Geology
EDA	Eletricidade dos Açores
EDP	Energias de Portugal
EEM	Empresa de Eletricidade da Madeira
ENSE	Nacional Entity for Energy Sector
ERSE	Energy Services Regulatory Authority
EU	European Union
GDP	Gross Domestic Product
GRUs	Gated Recurrent Units
LSTM	Long Short-Term Memory
MDD	Meteorological Data Description
MIBEL	Iberian Electricity Market
MSE	Mean Squared Error
MWD	Meteorological Wind Data
PCA	Principal Component Analysis
PES	Power Event Signals
PNEC	Portugal's National Energy and Climate Plan
REN	Redes Energéticas Nacionais
RMSE	Root Mean Squared Error
RNT	National Electricity Transmission Networ
RNN	Recurrent Neural Networks
SCADA	Supervisory Control And Data Acquisition
SWOT	Strengths, weaknesses, opportunities, and Threats
UDD	Utility Distribution Data
UCD	Utility Control Data
WSD	Wind Speed Data

Variable list

P	Theoretical power
V	Wind speed
A	Swept area
ρ	Air density
X	Data matrix (observations by variables)
μ	Means of variables in X
σ	Standard deviations of variables in X
X_{norm}	Normalized X (subtract μ, divide by σ)
Σ	Covariance of X_{norm}
v	Eigenvectors from Σ, for maximum variance
λ	Eigenvalues for v, indicating variance
v_k	Top k eigenvectors for dimension reduction
X_{PCA}	Data in principal component space
Loadings	Contribution of variables to components, using v_k and λ_k

Chapter 1

Introduction

1.1 Motivation

The world is presently faced with an imperative need for an energy transition. To mitigate the potentially irreversible consequences of climate change and safeguard life on Earth, it is imperative that we replace energy derived from fossil fuels with clean and renewable sources. Wind energy already assumes a substantial role in global energy production; nevertheless, certain constraints exist. Numerous countries presently grapple with challenges in identifying suitable locations with optimal wind characteristics for onshore wind turbine installations.

> "The science is clear: to avoid the worst impacts of climate change, emissions need to be reduced by almost half by 2030 and reach net-zero by 2050. To achieve this, we need to end our reliance on fossil fuels and invest in alternative sources of energy that are clean, accessible, affordable, sustainable, and reliable."
>
> United Nations [1].

It is discernible that emissions were markedly inconspicuous preceding the Industrial Revolution. The escalation of emissions maintained a relatively sluggish pace until the mid-20th century. In the year 1950, the global emission of carbon dioxide amounted to 6 billion tons. This figure nearly doubled by 1990, reaching an apex of over 22 billion tonnes. Presently, our annual carbon dioxide emissions approximate 34 billion tonnes [2].

In 2030, the Technical Carbon Dioxide Removals market is projected to average $180-220 per ton of CO_2. To meet the global net-zero target by 2050, competitive pricing with nature-based options is crucial. Accounting for additional benefits, a reasonable average price for technological Carbon Dioxide Removals is estimated at $90-100 per tonne [3].

When countries confront such challenges, several solutions become viable. Firstly, they can explore alternative clean energy sources such as solar energy. Secondly, they may opt for the deployment of offshore wind turbines. Wind energy generation is one of the most interesting alternatives since it can be installed near the consumer. The central focus of this dissertation revolves around possible efficiency enhancements of operational wind farms.

The solution will necessitate investment in all the aforementioned options, given our distance from achieving the energy objectives outlined in global policies. Increasing the efficiency of a wind farm appears to be the option demanding the least initial investment and can be applied across all wind energy production, ranging from older wind farms to more recent offshore installations. It is paramount that, as our current clean energy capacity falls short of meeting global demands, we ensure it is not squandered. To accomplish this, we must strive for energy producers operating at maximum efficiency. In this context, the spotlight is squarely on wind energy producers.

1.2 Proposed Research

The objective of this work is to examine the wind farm data and with an holistical approach identify optimization opportunities. To achieve this goal, the following subtasks were conceived:

- Analyse available data from the selected wind farm;
- Data cleaning;
- Optimization alternatives;
- Algorithm and method development;
- Improvement assessment;

1.3 Dissertation Outline

This dissertation is presented over four additional chapters. Chapter 2 addresses the state of the art. This section focuses on energy in Portugal, specifically examining how the energy sector operates. It also covers the electricity market, wind energy in Portugal, the future of wind energy in Europe, and concludes with a overview of neural networks, LSTM model and the data used to feed the model.

Chapter 3 pays special attention to the data available for the research, as these constitute the raw materials, and their quality is directly related to the final outcome of the dissertation. Initially, it identifies the available data, followed by an analysis of this data. This stage involves data processing and cleaning. Additionally, this chapter examines the wind farm and draws conclusions from the information available. It concludes with a Principal Component Analysis (PCA) to understand the relationships between the variables.

Chapter 4 develops the forecasting aspect, wherein an LSTM (Long Short-Term Memory) model is developed to predict certain variables over a 6-hour interval using data from the previous 12 hours. The variables include wind-generated power, the orientation of the nacelle, among others. This chapter not only presents and discusses the obtained results but also extrapolates how the developed work can aid real-world applications. It explores the potential implementation of the findings in actual companies, assessing the feasibility and practicality of using this LSTM model in operational environments. This section provides insight into how the research can directly contribute to the efficiency and decision-making processes in the energy sector, particularly for companies dealing with wind energy. The discussion aims to bridge the gap between academic research and practical industry applications, highlighting the model's relevance and potential impact.

The last chapter is reserved for conclusions and future work. Here, a summary of the main findings and the conclusions drawn from the research is presented. Recommendations for future research are provided, concluding with a final reflection on the dissertation.

Chapter 2

State of the Art

2.1 Energy in Portugal

This chapter aims to explore how the energy sector operates in Portugal, identify the responsible entities, and examine their respective policies. The energy sector in Portugal wields significant influence over the national economy. It operates as a highly complex system, involving numerous institutions and agents, and continually evolves to meet global and European challenges. The weight and importance of wind energy in Portugal and Europe, as well as future forecasts for investments in this technology, is also be addressed.

Portugal's energy sector has undergone significant advancements, particularly in the field of renewable energy. In 2023, renewable power sources contributed to 61% of Portugal's electricity, marking an increase from 49% in the previous year. This achievement set a new record, largely attributed to favorable weather conditions such as heavy rains and strong winds, which are conducive to renewable energy production [4]. Furthermore, Portugal has become a hub for energy transition initiatives. This is exemplified by the Lisbon Energy Summit & Exhibition, a leading event in Europe focusing on energy transition. Highlighting Portugal's role as a global leader in new energies and technological innovation, this event embodies the country's forward-thinking approach in energy systems.

> *"For 149 consecutive hours in November, Portugal provided a stunning example of what that could look like, as it used a mix of solar, wind, and hydropower to provide more clean energy than the entire country needed."*
>
> Darren Orf, Popular Mechanics [5].

Ambitious goals have been set in Portugal's National Energy and Climate Plan (PNEC), aiming to achieve 85% of its power generation from renewable sources by 2030. The plan includes increasing the installed wind capacity to 12.4 GW, with 2 GW coming from offshore wind farms, as part of a broader strategy to shift towards renewable energy [6].

These developments underscore Portugal's dedication to enhancing its energy sector through renewable sources, contributing to a sustainable and environmentally friendly energy landscape.

2.1.1 Operation of the Energy Sector

In the electricity sector, there are three primary entities: the regulatory body known as "ERSE" (Energy Services Regulatory Authority), the licensing authority "DGEG" (Directorate General of Energy and Geology), and the supervisory entity "ENSE" (Nacional Entity for Energy Sector). It is possible to divide the energy sector into five categories: production, transport, distribution, commercialization, and consumption, as can be observed in Figure 2.1. For more in-depth information, consult Portugal Energia website [7].

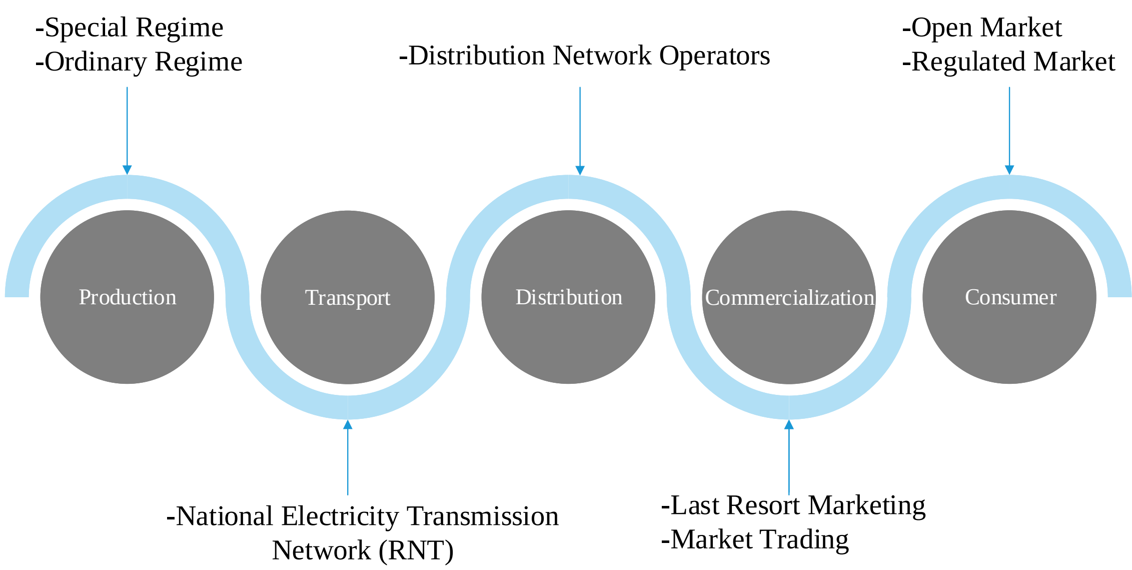


Figure 2.1: Electrical sector Portugal.

Production can be one of two types: special regime production and ordinary regime production. In the first type, activities are licensed under specific legal regimes, with policies aimed at promoting electricity generation using indigenous renewable resources or combined heat and power technologies. Production under the special regime encompasses technologies such as wind and solar power, renewable thermal energy, and minor hydroelectric stations with an installed capacity of no more than 10 MW [8]. Electricity producers operating under the special regime, in addition to being able to establish bilateral agreements and participate in the organized market like ordinary regime producers, have the right to directly sell their electricity to the Last Resort Supplier [9]. In the second type, activities are not governed by any special legal regime.

Transport , in Portugal, is responsible for the development, operation, and maintenance of the National Electricity Transmission Network (RNT) [10]. This network ensures the coordination of production and distribution facilities, aiming to guarantee the security of the energy supply and the efficient functioning of the system. It enables the import and export of electrical energy, thereby ensuring that the country's energy needs are met. In Portugal, the operator of the Transmission Network is 'REN' (Redes Energéticas Nacionais) [11]. The compensation for the activity of electricity transportation is done through fixed tariffs for the use of the transport network, as determined by ERSE [12].

Distribution of energy flow networks enable the connection of electrical production centers and substations to the transmission network. These networks consist of overhead lines and underground cables operating at various voltage levels, including high voltage (60kV), medium voltage (30kV, 15kV, 10kV), and low voltage (400/230V) [10]. Additionally, the distribution network encompasses substations, transformation stations, public lighting installations, and connections to consumer installations[10]. There are several distribution network operators in Portugal, including EDP Distribuição, EDA (Eletricidade dos Açores), and EEM (Empresa de Eletricidade da Madeira) [12].

Commercialization is divided in two markets: last resort marketing and market trading. In the case of last resort marketing, suppliers can buy and sell electricity and have access to transmission and distribution networks upon payment of regulated tariffs [12]. In this case, the electricity price does not vary throughout the year. Consumers can freely choose their supplier. To simplify and make the supplier change more effective, there is the role of a logistics operator for changing the supplier. To protect consumers, the last resort supplier aims to guarantee the supply of electricity, particularly to the most vulnerable, under conditions of quality and service continuity [10]. Last resort suppliers are the last option when market regime suppliers do not exist. In 2025, market regime suppliers will be the only option [13].

In the case of market regime suppliers, similar to last resort suppliers, they can freely buy and sell electricity and have the right to access transmission and distribution networks upon payment of regulated fees. Consumers can freely choose their supplier[12].

The logistics operator for changing the supplier is an entity that serves as an intermediary in managing the transition from electricity suppliers to consumers in the context of liberalized energy markets, promoting market transparency [10].

Consumers have the option to choose between the open market and the regulated market [12]. The differences between the open and regulated markets include pricing. In the open market, prices are determined by companies, and discounts may be available, whereas in the regulated market, prices are fixed annually. Additionally, the regulated market typically offers fewer supplier options. Electricity prices are usually lower in the open market.

Iberian Electricity Market – MIBEL The Portuguese and Spanish governments collaborated to establish the Iberian Electricity Market (MIBEL) with the aim of integrating their electrical systems [14]. The outcomes played a pivotal role in both the European and Iberian electricity markets, serving as a crucial initial step toward creating an internal energy market.

2.2 Electricity Market

In Portugal, during the restructuring of the electric sector, market agents emerged in the form of suppliers. One way in which producers and suppliers or eligible customers interact is through organized markets. This kind of market is related to a short-term horizon and can be described as the wholesale electricity market. The purpose of this market is to balance production with consumption through bids submitted by producing entities and suppliers or consumers [14, 15]. Given that electricity in te required amounts, cannot yet be stored, it is crucial to match production with demand, thus prevent wastage [14]. The MIBEL wholesale market currently comprises:

- A forward contracting market, where future commitments for the production and purchase of electricity are established.
- A spot market for on-the-spot contracting , with a component for daily contracting and an intraday adjustments component (intraday markets).
- A system services market that performs the balance adjustment of electricity production and consumption and operates in real time.
- A bilateral contracting market, where agents contract for the purchase and sale of electricity over various time horizons.

2.2.1 Forward Contracting Market

Organized forward markets include energy trades scheduled for delivery beyond the day following the contract's conclusion, offering either physical or financial settlements [14]. These markets trade energy contracts for future time frames, which can range from weekly to annual durations.

OMIP, in collaboration with OMIClear, oversees the MIBEL Derivatives Market, providing a trading platform specifically for energy-related products [14, 16]. OMIP facilitates a variety of contract types such as futures, forwards, swaps, and options, with futures being the predominant choice [17, 18].

Futures contracts are segmented into a trading phase and a delivery phase. A notable feature of these contracts is the daily computation and reconciliation of gains and losses, a process known as mark-to-market. The financial settlement is determined by the variance between the spot market reference price and the negotiated reference price on the final trading day [18].

2.2.2 Spot Market

Daily Market The unique aspect of daily markets is that they operate on the day before the corresponding buy and sell offers are executed. A day is divided into twenty-four one-hour intervals. At each interval, market participants must submit their buying and selling proposals. These include the minimum selling price, the injection node, and the availability of production for an offer. On the demand side, it is necessary to specify the maximum buying price, the amount of energy desired, and the node where the energy will be consumed. Following the negotiation period, 24 economic dispatches are determined for every hour [15].

The daily market closes at 10:00 AM the day before the supply date, with the results being announced at 11:00 AM Central European Time [14, 15]. It is at this time that the daily market session occurs across Europe, setting the prices and quantities of energy for the next twenty-four hours. In Portugal and Spain, specifically, buyers and sellers submit their proposals through the OMIE platform [15].

To ensure proper coordination in the operation of the electric system, the Market Operator and the System Operator play critical roles. The System Operator is responsible for analyzing potential technical restrictions resulting from the interaction between the daily market and bilateral contracts [14]. In contrast, the Market Operator oversees the market, gathering buy and sell proposals for the next day from participating agents, including traders, consumers ans producers. Following the receipt of these proposals, they need to be organized by constructing the supply and demand curves. In figure 2.2 (exctrated from [19]) it is possible to observe the supply and demand curve for 1/12/2023 for the first hour of the market, more information is available at OMIE. The sell offers are organized in ascending order of prices, in contrast to the purchase proposals, which are arranged in descending order.

The Market Price, representing the market's equilibrium price, is indicated by the intersection of the two curves at each hour. Buy and sell offers are not accepted if the price of the buy offers does not exceed the price of the unfulfilled sell offers [14]. Offers are accepted based on their economic merit and considering the available capacity for interconnecting different price zones. It is observed that, during a specific hour of the day, if the interconnection capacity is sufficient to allow for the flow of negotiated electricity, the electricity price remains the same across both zones. However, if the interconnection is congested at that hour, prices will differ in each zone [15]. The finite values for the transmission capacity of the lines cause a geographical dispersion

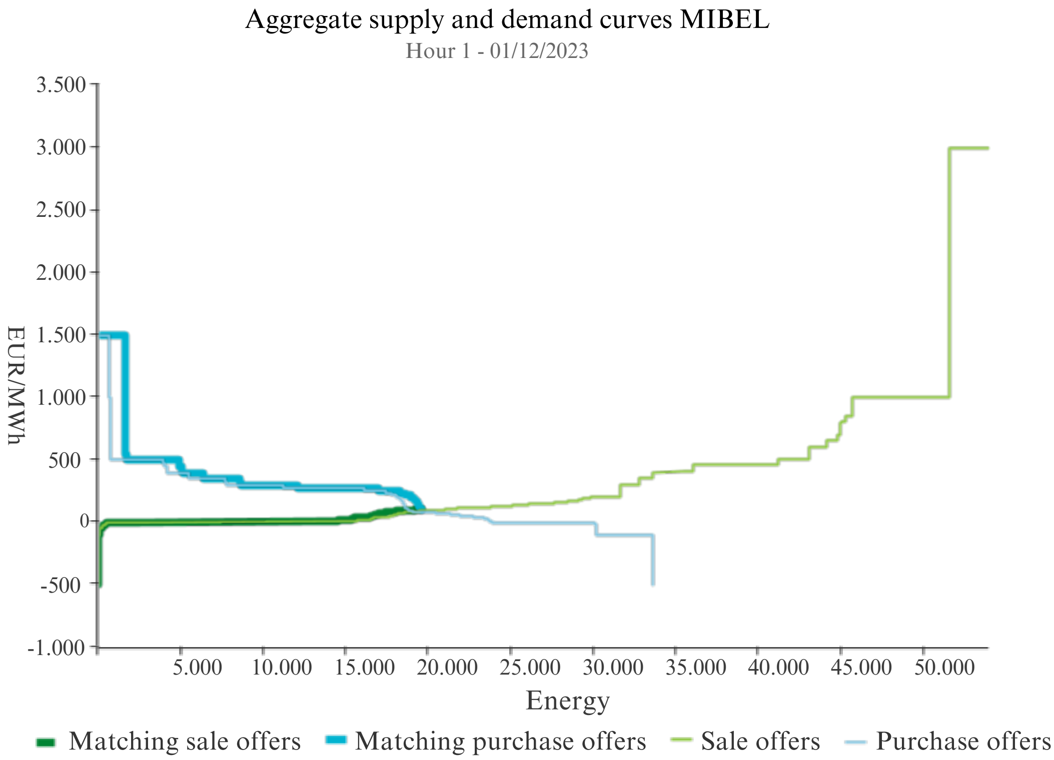


Figure 2.2: Example of a supply and demand curve.

of marginal prices, leading to an excess in costs for consumers compared to the payments received by producers. This excess is received by the organizations owning the transmission network.

Given that MIBEL is a shared electricity market between Portugal and Spain, monitoring cross-border energy transit is essential, particularly in relation to the commercially available interconnection capacities. In the event of congestion, the markets of the two countries separate and operate independently, resulting in a market split or separation [14]. Market splitting is also applied in response to insufficient interconnection capacities or specific market behaviors, affecting the structural organization of production in each area. This mechanism leads to different equilibrium prices for the Portuguese and Spanish zones, with the exporting country typically experiencing lower prices and the importing country facing higher prices [14].

Intraday Market Aimed at providing a more precise and real-time balance between supply and demand than the daily market allows, the intraday market acts as a supplement to the daily market. The intraday market corrects imbalances that may arise from the scheduling phases of the daily market [14]. This operation offers significant flexibility for the agents involved, allowing producers to purchase energy and traders to sell it [14].

The market consists of six sessions, each with its own scheduling horizon, where a marginal match between supply and demand is conducted [20]. Table 2.1 details the times for the opening, closing, and scheduling periods of MIBEL's six intraday sessions. The Auction Opening Time marks the moment when buyers and sellers are allowed to submit their bids for the purchase and sale of electric energy. The Auction Closing Time, on the other hand, indicates the point from which no further negotiation bids are accepted. The Schedule Horizon outlines the time frames within which transactions can be made, detailing the specific hours for energy delivery and reception that will be negotiated in each session. This structure ensures that all participants have a clear understanding of when they can engage in the market and plan their energy trading strategies accordingly.

Table 2.1: Structure of Intraday Auction Market

	Session 1	Session 2	Session 3	Session 4	Session 5	Session 6
Auction Opening Time	14:00	17:00	21:00	1:00	4:00	9:00
Auction closing Time	15:00	17:50	21:50	1:50	4:50	9:50
Schedule Horizon (Horizon Periods)	24 hours (1-24 D+1)	28 hours (21-24 y 1-24 D+1)	24 hours (1-24 D+1)	20 hours (5-24)	17 hours (8-24)	12 hours (13-24)

Within intraday trading, the continuous intraday market, also referred to as the single intraday coupling, comes into play. Similar to the standard intraday market, it enables market participants to rectify any potential energy imbalances [15]. The primary distinctions between these two types of markets are:

- The continuous market not only offers access to local market liquidity but also to the liquidity present in markets across different European regions, assuming there is sufficient cross-border transportation capacity available [21];
- Adjustments in the continuous market can be executed up until one hour prior to the energy being delivered [21].

The purpose of the continuous market is to facilitate more straightforward energy transactions between different regions in Europe, thus increasing the overall transactional efficiency within intraday markets [21]. It is possible to consult the contracts in negotiation by day, period, and time in OMIE.

2.2.3 Bilateral Contracts

The fundamental aim of bilateral contracts in the electricity market is primarily to reduce the risks associated with the fluctuating nature of short-term market operations [22]. Moreover, these agreements offer consumers the autonomy to select their energy supplier, establishing a more personalized relationship with their chosen provider [22].

In a fixed-price bilateral contract for energy, the deal acts as a safeguard against unexpected shifts in market prices [22]. While these contracts offer protection from price changes, they can sometimes lead to an imbalance, where one party may either gain or lose more than the other [22]. Specifically, if market prices fall below the contract's fixed price, the buyer (retailer) will pay more for energy than the current market rate. On the other hand, if market prices rise above the contract price, the seller (producer) ends up selling energy at a lower price than the market offers.

The System Operator plays a crucial role in overseeing electricity contracts, ensuring they adhere to all technical safety standards for energy supply and network congestion [22, 23]. If the grid lacks the capacity to handle the requested energy, the System Operator has the authority to either reduce or cancel the energy volume specified in the contract [22, 23].

2.2.4 System Services Market

The System Services Market pertains to the procurement of products beyond just active energy, that is, it is tasked with the contracting of secondary regulation and reserve regulation or tertiary reserve [24]. Given that alternating current electrical systems operate with constant frequency and voltage, regulation is a necessary function to ensure the stability of these quantities. System services are supplemental services responsible for the quality, reliability, stability, continuity, and security of the electrical system's operation [23]. These requirements are upheld through the control of frequency and active power, as well as voltage and reactive power control. The management of the system services market falls under the purview of REN, which secures contracting and clearing processes to ensure a balance between energy generation and usage [25].

2.3 Wind Energy in Portugal

In recent years, investments in onshore wind energy in Portugal have stabilized, largely due to the maturity of the market and the occupation of most suitable locations for harnessing wind resources. Since 2017, the installed capacity has remained relatively stable, totaling 5,819 MW as of 2022 [26]. To further boost renewable electrical energy production, Portugal faces the challenge of either fully or partially replacing existing wind turbines with more modern and efficient units, possibly integrated with storage solutions. The modernization of wind energy infrastructure not only promises higher efficiency but also aligns with technological advancements and economic considerations.

Currently, Portugal's offshore wind energy capacity stands at a modest 25 MW [26]. In comparison, Europe has an installed wind power capacity of 255 GW, with 88% (225 GW) deployed onshore and 12% (30 GW) offshore [26]. The relatively small contribution of offshore wind in

Portugal highlights both the challenges and opportunities in this sector, especially in light of the country's ambitious targets for expansion.

Portugal has achieved significant milestones in wind energy, contributing crucially to its renewable energy production. In 2023, the country set a new record by exceeding its electricity needs for six consecutive days, a testament to the substantial role of wind energy [27]. Historically, wind energy has been a significant part of Portugal's renewable energy portfolio. The WindFloat Atlantic Project marked a key development in offshore wind energy [28].

Wind energy contributed 24% to Portugal's total energy generation, placing it among the top European countries for wind energy utilization [29]. The Portuguese government has set ambitious targets for 2030, aiming for renewable energy to cover 80% of the country's electricity needs, with 31% expected from wind power. This objective underscores the strategic importance of wind energy in Portugal's energy policy [26].

To meet these ambitious goals, Portugal plans to launch its first offshore wind power auction by the end of 2024, aiming for an installed capacity of 10 gigawatts (GW) by 2030. This plan demonstrates the country's commitment to expanding its wind energy capabilities, particularly in offshore wind [30]. Despite currently having only one operational offshore wind farm – Windfloat Atlantic, located off Viana do Castelo – the development of offshore wind energy remains a key aspect of Portugal's strategy to enhance its renewable energy portfolio. The environmental impacts of this development, including potential effects on wildlife and ecosystems, are important considerations in this expansion [31]. These developments indicate Portugal's strong commitment to enhancing its energy sector through wind power, contributing to a sustainable and environmentally friendly energy landscape.

2.4 Future of Wind Energy in Europe

To predict the future, two scenarios are used: the Central Scenario and the 2030 Targets Scenario. The Central Scenario presents the projected installed capacity in Europe for the next five years, depicting the latest changes in national policies, EU regulations, project development schedules, and the wind power industry's ability to secure additional capacity in upcoming auctions and tenders. In this scenario, Europe will install 129 GW at a rate of 25.8 GW installed annually on average [26].

The 2030 Targets Scenario encompasses the installation targets required to meet the 2030 goals of non-EU nations, which include the UK (50 GW), Turkey (18 GW), Norway (12 GW), Switzerland (0.2 GW), as well as the REPowerEU target (440 GW) (extracted from [26]). New installations for both scenarios are presented in Figure 2.3 (extracted from [26]).

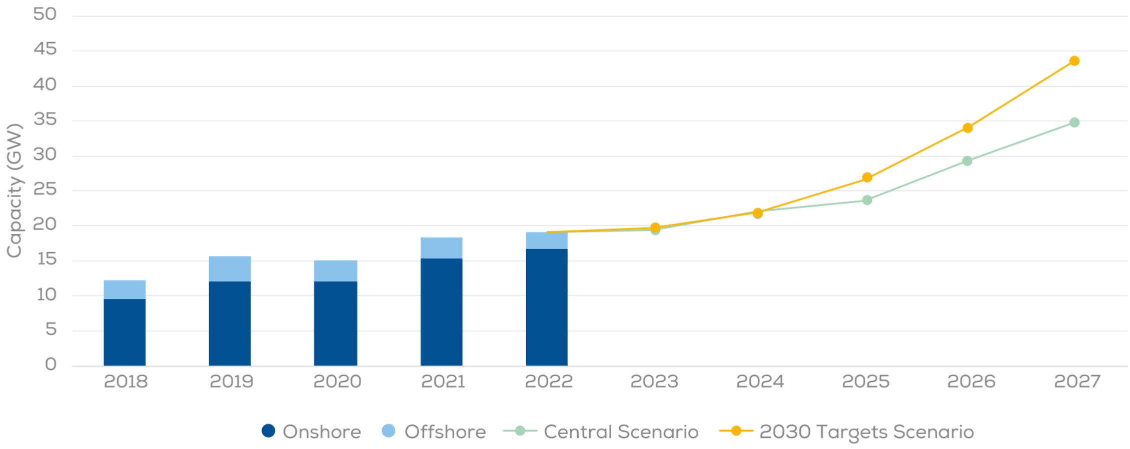


Figure 2.3: New installations in Europe – WindEurope's scenarios.

Technological advancements are set to play a pivotal role in the future of wind energy in Europe. Emerging technologies are revolutionizing wind turbine design and efficiency, with turbines now reaching heights of up to 200 meters and blades spanning over 100 meters. These larger turbines capture stronger winds at higher altitudes, significantly increasing energy production. According to the Wind Energy Technology Development Report 2020, these technological advancements are projected to continue until 2050, focusing on efficiency and integration with energy storage solutions [32].

The policy and regulatory environment in the EU, including the European Green Deal and the Renewable Energy Directive, is driving wind energy development. The EU aims for carbon neutrality by 2050, with an updated target of 32% renewable energy in EU final energy consumption by 2030. These goals are supported by legislation and financial incentives, promoting wind energy development [33].

Economic factors significantly impact wind energy development in Europe. The wind energy sector contributed €36.1 billion to the EU's GDP in 2016 and supported 240,000 to 300,000 jobs in 2020. In 2022, Europe installed 19.1 GW of new wind capacity, a 4% increase from the previous year, demonstrating continued growth and investment in the sector [26].

Environmental and social considerations are increasingly prioritized in wind energy expansion. In 2019, wind energy in Europe saved 118 million tonnes of CO_2, with technological advancements reducing noise and visual impacts, thereby increasing social acceptance [34].

International collaboration and trade are integral to the advancement of the wind energy sector in Europe. The European Wind Power Action Plan and the European Wind Charter highlight the importance of technology exchange, research collaboration, and cross-border trade in supporting wind energy development [35].

2.5 Neural Networks

2.5.1 Introduction to Neural Networks

Neural networks play a central role in artificial intelligence, drawing inspiration from the complex structure and functionality of the human brain to process information [36]. These networks are designed to mimic the human ability to learn and adapt from data, using processing units known as artificial neurons. Organized in layers, these units work together to perform complex tasks, ranging from pattern recognition to decision-making, based on the analysis of large volumes of data [37].

The structure of a neural network comprises several layers, each playing a specific role in the learning process. The input layer is the starting point, where raw data are introduced into the network [38]. This data is then processed through one or more hidden layers, which are the heart of the neural network. These intermediate layers use weights and activation functions to transform the input data, allowing the network to identify complex patterns and features [39]. Finally, the output layer produces the final result of the network's processing, which can vary depending on the application, from image classification to time series forecasting [39].

The operation of a neural network is based on the propagation of input data through its layers. This process, known as forward propagation, involves the transformation of data through the weights and activation functions associated with each connection between neurons [39, 38]. Adjusting these weights is crucial for the network's learning and is achieved through a process called backpropagation [38]. During training, the network iteratively adjusts the weights of the connections to minimize the difference between the produced output and the expected output, thus improving its ability to make accurate predictions or classifications [38].

Neural networks find applications in a wide range of fields, thanks to their versatility and ability to learn from examples. They are used in computer vision for image recognition and object detection, in natural language processing for automatic translation and text comprehension, in data analysis for market forecasts and consumer trend analysis, and in robotics for motion control and autonomous navigation [38]. The continuous development of neural network technology, along with the increase in data availability, promises significant advances in many areas, making them an essential tool for the progress of artificial intelligence.

2.5.2 Comparison of RNNs, LSTMs, and GRUs

Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and Gated Recurrent Units (GRUs) play pivotal roles in processing time series, each with unique characteristics suited for specific applications such as time series analysis or natural language processing.

RNNs provide a foundational framework for sequence processing, effectively capturing short-term dependencies. They are challenged by long-term dependencies due to vanishing and exploding gradient problems. This limitation has spurred the development of LSTMs and GRUs, which introduce mechanisms to better manage gradient flow and enhance memory retention for extended sequence learning [40, 41].

LSTMs Enhance RNNs, featuring a sophisticated architecture with forget, input, and output gates. This design allows for efficient long-term dependency management, positioning LSTMs as highly effective for understanding complex temporal patterns in applications such as machine translation and time series forecasting [42]. LSTMs excel in scenarios requiring nuanced understanding of time-based patterns, thanks to their robustness against the vanishing gradient problem.

GRUs offer a streamlined alternative to LSTMs by consolidating the forget and input gates into a single update gate, thereby reducing model complexity and enhancing computational efficiency. This makes GRUs particularly useful in scenarios with computational constraints, without significantly compromising on performance. GRUs have shown to be effective in tasks requiring sequential data processing, similar to LSTMs, but with potentially faster training times [43, 44]. While GRUs are versatile and efficient, the choice between GRUs and LSTMs often depends on the specific requirements of the task and dataset complexity.

LSTM networks stand out in the forecasting of wind energy data due to their superior ability to model the nonlinear and complex temporal dependencies characteristic of wind speed and power generation data. Their architecture, which effectively retains and manages information over prolonged periods, is particularly beneficial in the energy sector where accurate forecasting impacts operational efficiency and the integration of renewable energy sources [45, 46].

Moreover, LSTMs have propelled advancements in time series forecasting across various domains, demonstrating their capability to handle sequence length variability, anomalies, and non-stationarities. This adaptability highlights their indispensable role in modern forecasting solutions, catering to diverse data types and forecasting needs [47].

Technical Clarification: LSTMs and GRUs address the vanishing and exploding gradient problems that challenge traditional RNNs through their innovative gating mechanisms. These mechanisms allow for selective retention and discarding of information, ensuring the preservation of relevant data over long sequences. This is particularly vital for modeling the intricate temporal dynamics seen in energy forecasting and similar applications.

2.6 LSTM Model

The fundamental strength of LSTM networks is rooted in their cells, which are equipped with three types of gates (input, output, and forget gates) within their hidden layers [48]. These gates, along with two state vectors - the cell state and the hidden state - enable the LSTM to effectively regulate the flow and retention of information. The cell state functions as a sort of residual block, accumulating information progressively over the sequence. Conversely, the hidden state, which is actively propagated through the network, is updated alongside the cell state at every time step, ensuring continuous data flow and stability in learning.

The LSTM cell is composed of three primary functional blocks: the forget, input, and output blocks. Each block employs a Fully Connected (FC) layer, which plays a pivotal role in processing the data. In these blocks, the FC layer first concatenates the current input vector with the previous hidden state, creating a combined input. It then transforms this combined input to facilitate various operations within the LSTM.

The input block of an LSTM cell plays a crucial role in updating the cell state with new information. It operates on the current input data and the previous hidden state. This block consists of two main components: a sigmoid activation function and a tanh activation function. The sigmoid function determines how much of the new information is relevant and should be allowed into the cell state. Meanwhile, the tanh function processes the current input and previous hidden state, normalizing these values to a range between -1 and 1. The outputs of these functions are then multiplied together, which results in a filtered version of the input data. This filtered input is used to update the cell state, ensuring that only relevant and normalized information is added, thus maintaining the efficiency and accuracy of the network.

The forget block is essential for regulating the information retained in the LSTM cell's memory. It decides which information should be kept and which should be discarded from the cell state, thus preventing the accumulation of irrelevant data over time. This block also uses a sigmoid function, which acts on the current input and the previous hidden state. The output of this sigmoid function ranges from 0 to 1, with values closer to 1 indicating that the information should be retained, while values closer to 0 signify that it should be forgotten. The result of this function is then multiplied with the cell state. By applying this selective filtering, the forget block helps in reducing the network's complexity and enhances its ability to model long-term dependencies without being hindered by unnecessary information.

The output block determines the next hidden state, which is crucial for the LSTM's output at the current time step and also for the next time step. This block involves two key operations. First, it processes the current input and the previous hidden state through a sigmoid function to decide which parts of the cell state will be used to generate the output. Simultaneously, the cell state is passed through a tanh function to normalize its values, and this normalized cell state is then multiplied with the output of the sigmoid function. The resultant product forms the final output of the LSTM cell. This output becomes part of the next hidden state and is also used for making predictions or as input to subsequent layers in a more complex neural network architecture. The output block thus plays a pivotal role in controlling the flow of information from the cell state to the outside world, ensuring that the LSTM generates relevant and context-aware outputs.Figure 2.4 shows the structure of an LSTM.

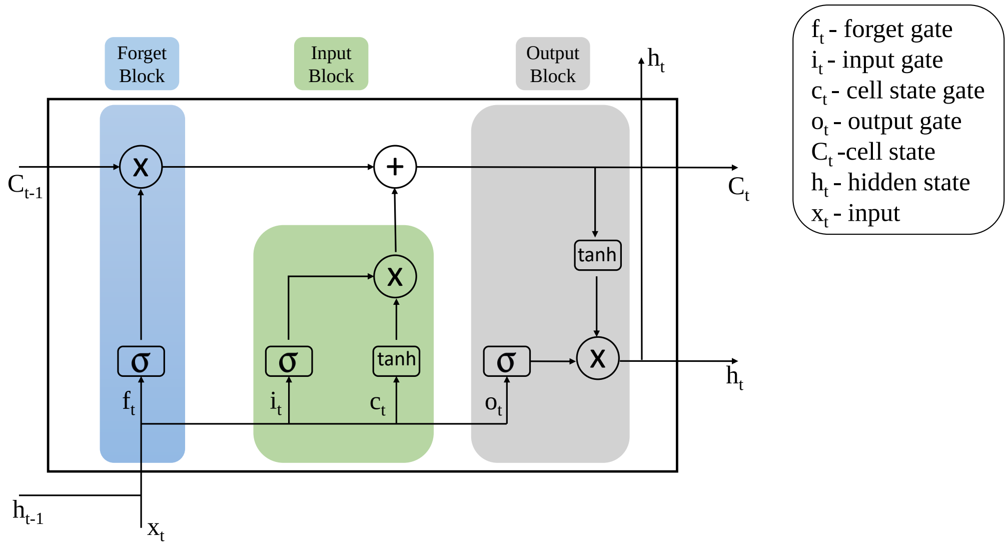


Figure 2.4: The arquitecture of LSTM unit.

To carry out predictive analysis using the LSTM model, a Python script utilizing the PyTorch library was developed. For LSTM training in PyTorch, data must be in tensor format. The model can be built using *torch.nn.LSTM* for a simplified, high-level interface that automatically manages the LSTM's gates, sequence processing, and internal state updates [49]. Alternatively, *torch.nn.LSTMCell* offers more detailed control, requiring manual iteration and updates for sequence processing, but only creates the cell unit [49].

Although the LSTM module provides ease of use with its straightforward approach, there was a preference for greater control over how sequences are managed, along with the possibility of incorporating extra functionalities such as warm-up. Consequently, the decision was made to utilize the LSTMCell module, which offers more detailed control over these aspects.

The forward function is responsible for producing the outputs of the neural network. It involves initializing the initial cell states of the LSTM with zeros and then "warming up" the LSTM. This "warm-up" process is a crucial step where the LSTM is gradually exposed to and processes the initial part of the input sequence [50]. During this phase, the LSTM updates its internal states (both the cell state and the hidden state) based on these early inputs. This allows the LSTM to accumulate contextual information and dependencies present in the sequence, which is vital for making accurate predictions or analyses later in the sequence. Essentially, the warm-up phase helps the LSTM to 'get ready' and be more effective in handling the sequential data it will encounter in subsequent steps of the forward function [50].

In the "fit" training loop, each epoch begins with the model set to training mode. An epoch represents one full pass through the entire training dataset [51]. During each epoch, data batches from the DataLoader are transferred to the computing device and input into the model. This process results in the generation of predictions and enables the calculation of training and validation losses. The crucial optimization step happens within this loop. The optimizer's gradients, which accumulate by default, are reset to zero at the start of each iteration to prevent compounding errors. After this, the loss is computed based on the model's predictions.

Backpropagation then occurs, where gradients are calculated and stored based on this computed loss. The optimizer, specifically the *Adam* optimizer in this scenario, uses these gradients to update the model's parameters. The "fit" function is responsible for recording and averaging the training and validation losses for presentation purposes. While the training loss is computed directly within this loop, validation loss calculation is a bit different, requiring the "evaluate" function. This function has a similar operational framework to both "fit" and "predict" functions, with a major distinction being the model's state. In "evaluate" mode (*.eval()*), unlike in training mode (*.train()*), no gradient computations take place. However, the data feeding process and overall structure remain consistent with the training loop.

The loss function calculates the loss value, which evaluates how closely the model's predicted values match the target values for the same input. A lower loss value signifies more accurate predictions. In regression tasks such as this, *torch.nn.MSELoss* is often utilized. This particular loss function computes the mean squared error, essentially the average of the squares of the differences between each element in the predicted output and the corresponding target value [52].

After running the training and evaluation processes, one can determine the model's performance based on the train and validation losses. Subsequently, the "predict" function is used to generate the actual model forecasts. Similar to the "evaluate" function, it sets the model to evaluation mode, disabling gradient calculations. However, the "predict" function is simpler, as it does not involve loss computation and directly produces predictions for a specified number of time points.

2.7 Data

To maximize the accuracy of the forecast, it is imperative that the data utilized is exceptionally precise. Essentially, the data is the raw material; high-quality raw materials will form a precise forecast. The data to be used have been treated in Section 3, where the entire process can be consulted in detail.

The data was processed using Python, and subsequently, it was processed using PyTorch's *torch.utils.data Dataset* [53]. Essentially, this step serves to divide the data into three different DataFrames: train, validation, and test. It is also at this stage that the number of inputs and targets to be used is specified, as well as which are the input variables and the target variables. Following this, the data will be transformed into tensors using *torch.FloatTensor*, resulting in the formation of two different tensors, one for the inputs and another for the targets [54].

The tensors are then processed by the PyTorch *torch.utils.data DataLoader* [53]. This DataLoader essentially creates a loop that traverses the dataset, allowing easy access to each data piece. Moreover, there is a shuffling of data in each epoch to improve learning, as it prevents the risk of the model memorizing the data order. This approach also enhances the model's ability to handle various types of data, thus improving its overall performance.

In the transformation of data into tensors, it is necessary to define the 'input_width'. The 'input_width' is the number of data points that will be used in the forecast. For example, if an 'input_width' of 5 is set, it means that 5 pairs of inputs and targets will be used. In other words, the model will look at 5 previous data points to make a prediction for the next point.

There is a need to define the 'batch size' in the DataLoader, where the batch size is the amount of new data provided to the network in each iteration. Using batches is important for the efficiency of the network as it optimizes the processing needs, and the fact that the data segments are different from each other allows for better learning [55]. It should be noted that during the preparation phase, the model only had access to training and validation data. The test data were not seen by the model, ensuring that the forecast is realistic.

Chapter 3

Database

3.1 Available Files

In this chapter, the available data for the execution of this analysis is verified. The analysis encompasses the examination of available file types, an assessment of the data contained within each file, and a determination of the data relevant to the development of this thesis. The data contains daily information from 2010 until 2021.

3.1.1 File Types

The available data is presented in five distinct file types: WSD, UCD, MDD, MWD, and UDD. Additionally, there is a PES file containing monthly information. The files are not consistently available at all times. Table 3.1 provides information regarding the availability of these files.

Table 3.1: Files availability

Year and Month	Files
Jan 2010 - Fev 2010	UCD WSD
Mar 2010 - Sep 2012	UCD WSD UDD MWD MDD
Oct 2012 - Jul 2013	UCD WSD UDD
Aug 2013 - Dec 2021	UCD WSD UDD MWD MDD

3.1.2 Files Data

Table 3.2 presents an overview of all the data provided by WSD files. The data represents averages calculated every ten minutes. Information regarding the date, hour, minutes, and seconds of when the data was saved is available. These files also contain information about the wind turbines (identified by plant number); there are a total of twelve distinct wind turbines. The error message provides information about whether there were any issues in acquiring the average data during the 10-minute collection interval. Information is available on the average, maximum, and minimum wind speeds in the nacelle, as well as the average, maximum, and minimum rotor rotation and power. The nacelle's orientation can vary between -760º and +760º, representing two and a half complete rotations in either direction. Lastly, data regarding the duration of energy production by the wind turbine in terms of hours and minutes, along with corresponding energy output in kilowatt-hours (kWh) at each instance, is possessed.

Table 3.2: WSD file information

Date
Hour
Minute
Second
Plant Number
Error Message
Average Speed Nacelle
Max Speed Nacelle
Min Speed Nacelle
Average Rotation
Max Rotation
Min Rotation
Average Power
Max Power
Min Power
Nacelle orientation
Hours of Work
kWh produced at each instant
Minutes of Work

The UCD files originate from a specific plant responsible for storing control data for the entire park, which comprises the 12 wind turbines mentioned earlier. This control data encompasses various parameters, including maximum, minimum, and average power, as well as maximum, minimum, and average reactive power. Additionally, the files contain information about the minimum, maximum, and average frequency, voltage, and current for the three distinct phases. Lastly, the files include data concerning both the active energy produced and consumed, the induced reactive energy produced, and the capacitive reactive power produced. Furthermore, information is available on the reactive energy induced, consumed, and capacitive energy consumed. All of this information is presented in detail in Table 3.3.

Table 3.3: UCD file information

Date
Hour
Minute
Second
Plant Number
Error Message
Average Power
Max Power
Min Power
Average Reactive Power
Max Reactive Power
Min Reactive Power
Average Frequency
Max Frequency
Min Frequency
Average Phase 1, 2, 3 Voltage
Max Phase 1, 2, 3 Voltage
Min Phase 1, 2, 3 Voltage
Average Phase 1, 2, 3 Current
Max Phase 1, 2, 3 Current
Min Phase 1, 2, 3 Current
Produced Accumulated Active Energy
Consumed Active Energy
Produced Induced Reactive Energy
Produced Capacitive Reactive Power
Consumed Induced Reactive Power
Consumed Capacitive Reactive Power

The MDD files contain data acquired from a weather station situated in the wind farm site under study. These variables are listed in table 3.4. The station records data on air temperature, pressure, and humidity, along with information on rainfall and sunlight exposure.

Table 3.4: MDD file information

Date
Hour
Minute
Second
Plant Number
Error Message
Average Temperature
Average Air Pressure
Average Air Humidity
Average Rainfall
Average Sunbeam

Similar to MDD files, the MWD files, table 3.5, contain data acquired from the wind farm's weather station. Within these files, data is provided for two different heights: 30 meters and 65 meters. This data includes the maximum, minimum, and average wind speeds for both heights, as well as the wind directions. This wind-related information holds significant importance due to its minimal susceptibility to errors, unlike the wind data acquired within the nacelle. However, it's essential to examine whether the wind turbines within the wind farm might influence wind patterns in specific directions.

Table 3.5: MWD file information

MWD
Date
Hour
Minute
Second
Plant Number
Error Message
Average Windspeed 65m
Max Windspeed 65m
Min Windspeed 65m
Average Windspeed 30m
Max Windspeed 30m
Min Windspeed 30m
Average Horizontal Wind Direction 65m
Average Vertical Wind Direction 65m
Average Horizontal Wind Direction 30m
Average Vertical Wind Direction 30m

The data within the UDD files, table 3.6, is gathered from a control unit substation and contains information regarding the voltage for the three different phases, as well as the averages for power and reactive power.

Table 3.6: UDD file information

UDD
Date
Hour
Minute
Second
Plant Number
Error Message
Average Voltage 1 Phase 1
Average Voltage 1 Phase 2
Average Voltage 1 Phase 3
Average Voltage 2 Phase 1
Average Voltage 2 Phase 2
Average Voltage 2 Phase 3
Average Power
Average Reactive Power

Within the PES files, the data is no longer recorded at 10-minute intervals. These files capture information about various events, including failures and operational restarts. As depicted in table 3.7, the data encompasses the date and time of the wind turbine's occurrence, its state and substate. In 3.7, it is possible to consult all states available in PES files.

Table 3.7: PES file information

Date
Hour
Minute
Second
Plant Number
State
Sub-State

3.2 Preliminary data analysis

In this section, an initial analysis of the available data is conducted. The process begin with data cleaning and then proceed to interpret the data. Data cleaning is indispensable for analysis as it directly impacts the accuracy and reliability of the results. By identifying and rectifying errors and inconsistencies in data, it ensures that the conclusions drawn from the analysis are based on accurate and valid information. This process not only enhances the efficiency of data analysis by reducing errors and anomalies, but it also maintains the integrity of the data, particularly when integrating multiple sources.

3.2.1 Raw Data

In Figure 3.1, all available data points for Turbine 1 in 2011 are depicted using the SCADA data. Both the selection of the turbine and the year were made randomly. In this case, the graph comprises approximately 52 000 points, with the average speed for each 10-minute interval on the x-axis and the average power on the y-axis. This graph provides insight into the potential power curve of the turbine under analysis. To obtain a more accurate curve, it is necessary to eliminate data points that disrupt the power curve, thereby bringing it closer to the true power curve.

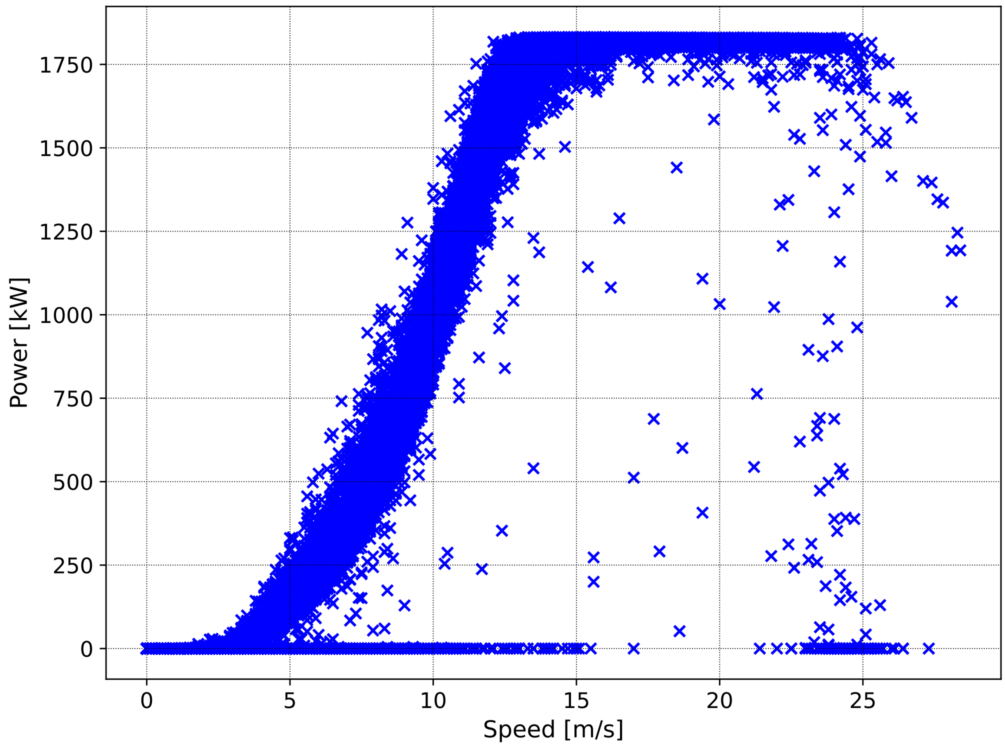


Figure 3.1: Power curve of all data for Turbine 1 in 2011.

3.2.2 First stage of Data Cleaning

There are points that are known not to accurately represent the power curve. For example, there are instances where the average wind speed exceeds 3 m/s, yet the turbine's average power output is recorded as 0 W. Wind turbines typically start generating electricity at wind speeds known as the "cut-in speed" or "start-up speed," which is usually around 3 to 4 m/s [56]. Below this wind speed, the energy captured by the turbine blades is not sufficient to produce electricity effectively. Wind turbines are designed to operate within a range of wind speeds. Beyond the cut-in speed, there are two more critical parameters: the rated wind speed, at which the turbine produces its maximum rated power, in this case 1800 kW, and the cut-out speed, at which the turbine shuts down to avoid damage during very high wind conditions. The cut-out speed is usually around 25 m/s [56]. Thus, while wind turbines require a minimum wind speed to start generating power, they can operate and generate electricity across a wide range of wind conditions until the wind speed becomes too high, at which point they shut down for protection.

Additional instances of this nature involve points where the minimum power produced is zero, despite the minimum wind speed being not less than 3 m/s. This indicates instances where the turbine was halted and not generating power during the 10-minute period, seemingly without any discernible reason. The flowchart used is in Figure 3.2.

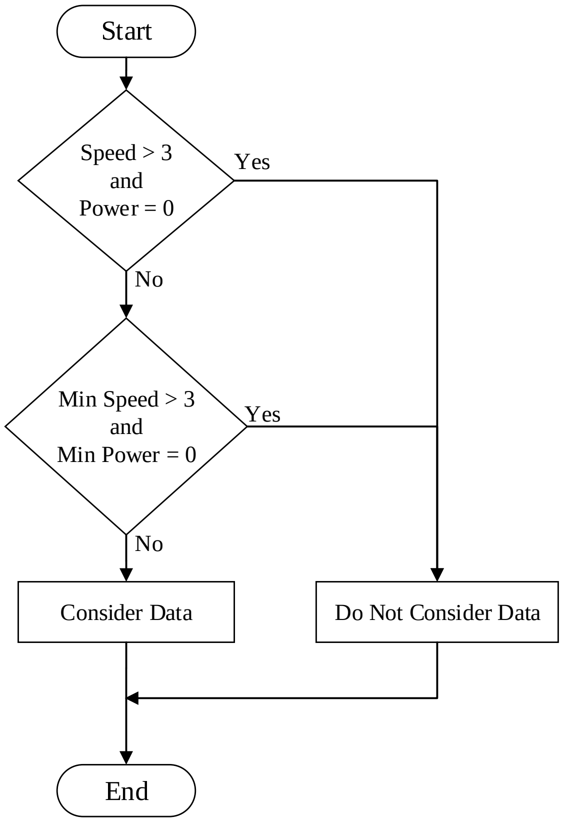


Figure 3.2: Flowchart of first data filtering.

In Figure 3.3, points exhibiting the aforementioned characteristics are excluded. Additionally, points corresponding to speeds above 20 m/s were removed due to their irregular behavior. Occasionally, it is necessary to temporarily limit the turbines. Figure 3.4 shows a power curve for Turbine 9, where these limitations are evident. In such instances, the specific day and time of the limitation are identified, and these data points are removed to prevent them from impacting the accuracy of the error data.

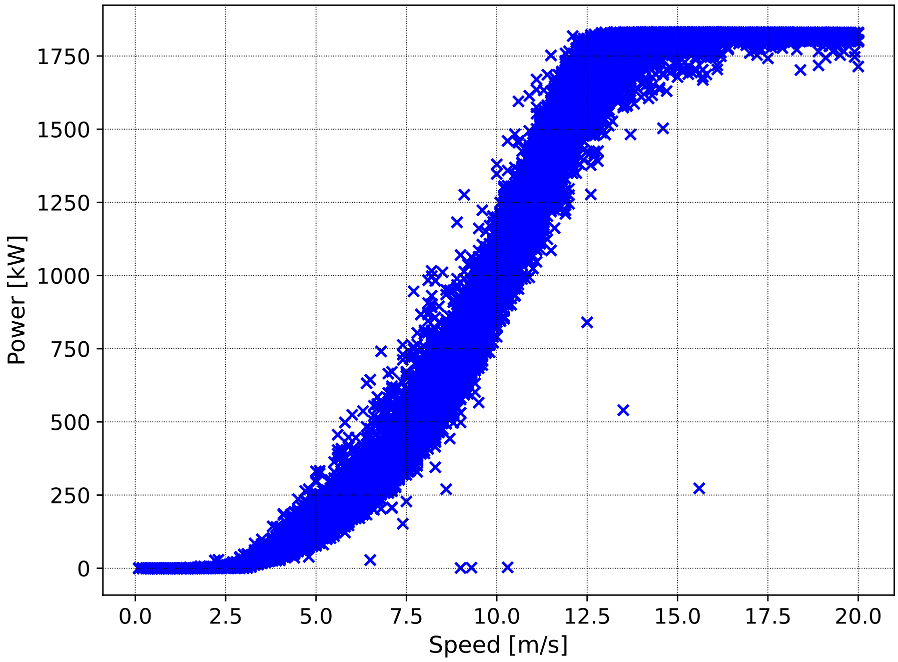


Figure 3.3: Power curve in first stage of data cleaning for Turbine 1 in 2011.

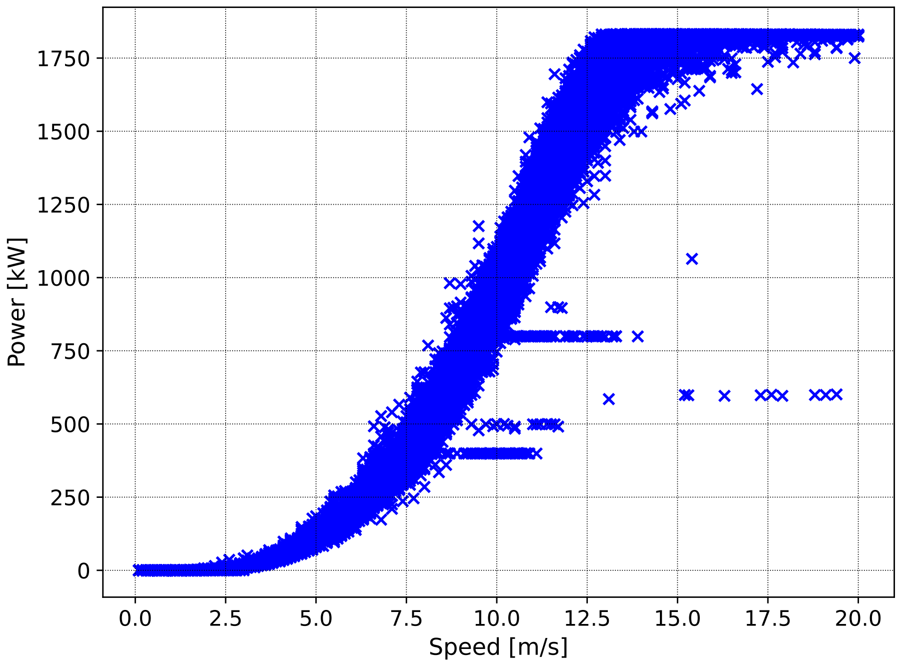


Figure 3.4: Power curve of Turbine 9 with power limitations in 2011.

3.2.3 Last stage of Data Cleaning and Data Visualization

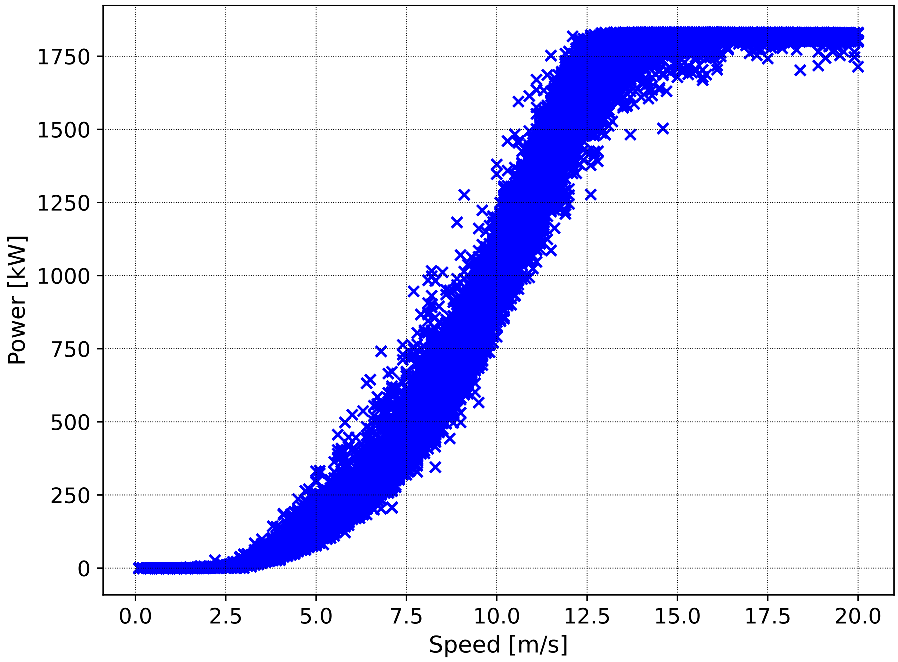
To reduce the number of points outside the turbine's characteristic power curve, consultation with the PES files for turbine 1 was undertaken. As explained earlier, these files contain data on turbine actions, such as whether the turbine is starting or undergoing maintenance, in annex A is all the possible actions. With this information, it was possible to identify 10-minute intervals that may be affected by errors, as the turbine was not operating normally.

In Figure 3.5a, the graph depicts the results after consulting the PES Files. Through this process, a significant number of points were successfully removed, resulting in a cleaner power curve. In annex B are three more examples of power curves for different turbines. After obtaining the cleanest graph possible, the average power for each speed was calculated. This enabled the drawing of a line that facilitates the examination of point distribution on the power curve, providing a more accurate representation. The average power curve is illustrated in Figure 3.5b.

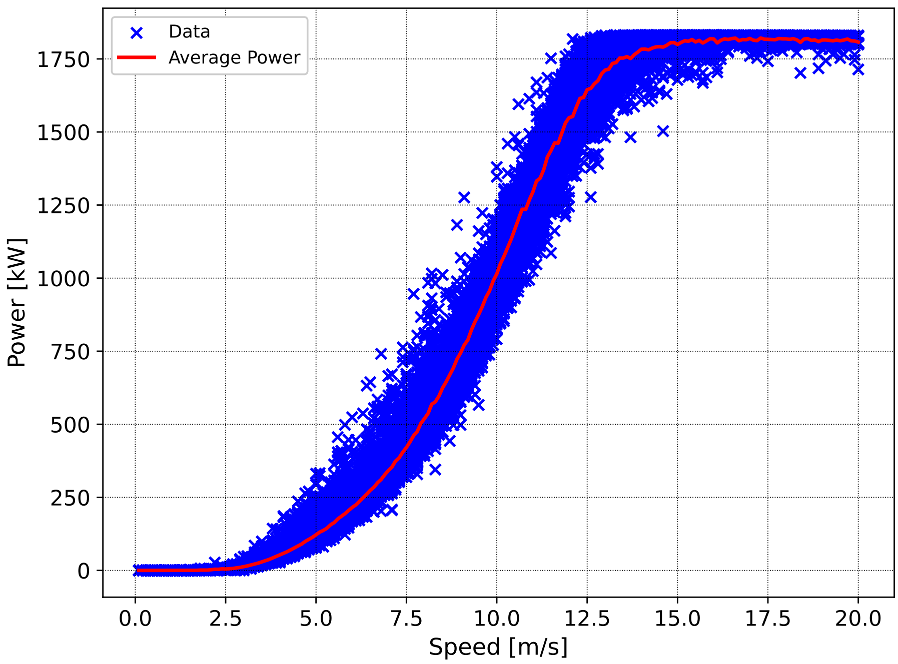
Observing figure 3.5b, it is clear that for some speed values, there is a large range of power values; for example, for 7.5 m/s, the power produced ranges between 300 W and 750 W, approximately. One of the reasons could be the differences in temperature, and, in turn, the variation in air density throughout the year. In Figure 3.6, the power curves for the different seasons of the year can be compared. One immediate conclusion that can be drawn is that the average winter power curve does not achieve the same maximum power in the stabilization zone at high speeds, unlike other power curves. This pattern is not consistent across all turbines, as can be seen in the data provided in Annex B. However, for the rest of the curve, specifically in the 0 to 12.5 m/s range, the differences between the curves are not significant. Although not visible, in the summer, the power curve only shows average values below 17 m/s. This is to be expected, as it is anticipated that there will be less wind during the summer months. The figure 3.7 illustrates the average summer power curve.

Another analysis involves understanding how the power curve varies with direction. For this analysis, the speed and power data is categorized according to different directions. This analysis is represented in Figure 3.8. Referring to Figure 3.8, a notable difference is observed between the power curves for this turbine. The power curve when facing north consistently surpasses all others at equivalent wind speeds. For instance, at a wind speed of 12 m/s, the variation in power production exceeds 250W.

Analyzing wind behavior involves a valuable examination. In Figure 3.9a, the distribution of the number of wind occurrences is presented as a percentage per sector of the wind rose. In red, a line is displayed that allows for an understanding of the occurrences for each direction and to see the distribution of occurrences within the sector. This graph facilitates observation of the predominant wind directions. Similar to the previous analysis, the focus is on turbine 1 in the year 2011.



(a) Last stage of data cleaning.



(b) Average power.

Figure 3.5: Power curves for Turbine 1 in 2011.

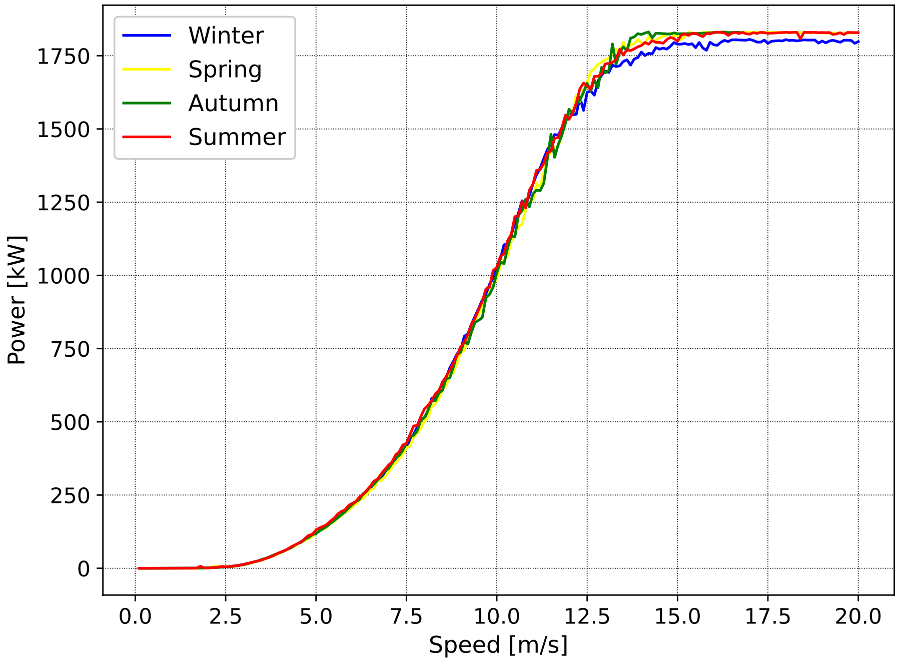


Figure 3.6: Power curve for average power by season for Turbine 1 in 2011.

It is possible to create a similar graph, but this time, instead of representing occurrences by sector, the average wind speed by sector can be depicted. This allows for an understanding of which sector has the highest kinetic energy. In Figure 3.9b, the distribution of average wind speed by sector of the wind rose can be observed.

Wind energy production doesn't solely depend on the average speed per sector or the number of occurrences per sector. It thrives when both the number of occurrences and the average wind speed in a given sector are high. Understanding which sector possesses the best characteristics is crucial. In Figure 3.9c, both the number of occurrences and the average speed for each sector are presented in the same chart. This provides a comprehensive view, allowing us to better gauge the potential energy production in each sector. In annex B three more examples can be seen.

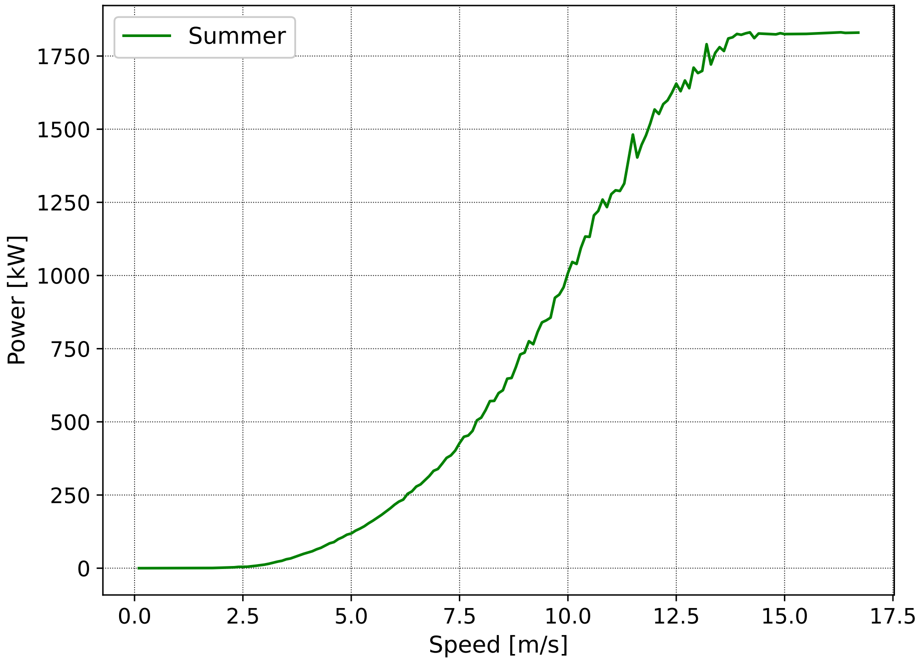


Figure 3.7: Summer power curve for average power for Turbine 1 in 2011.

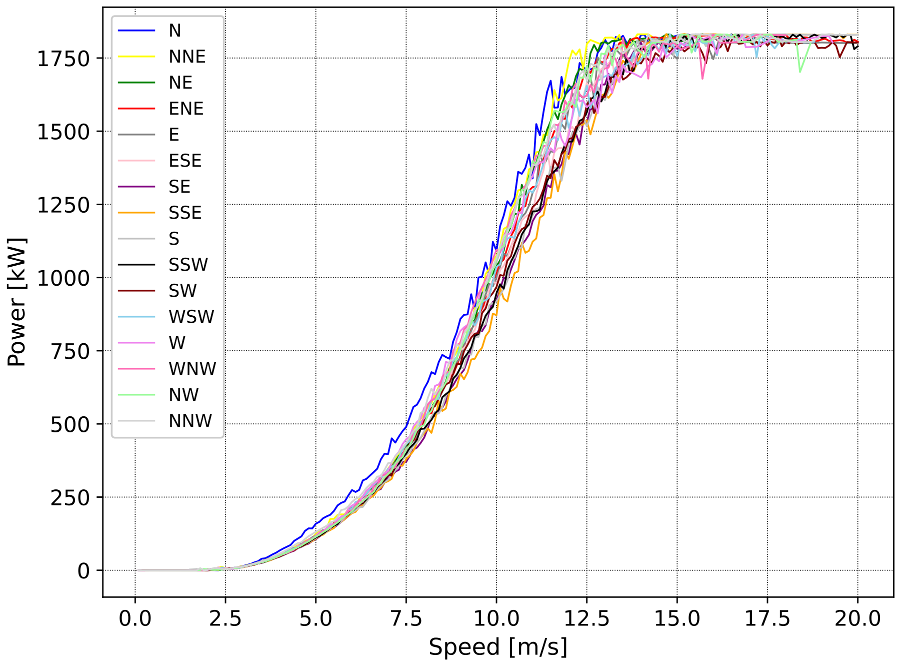
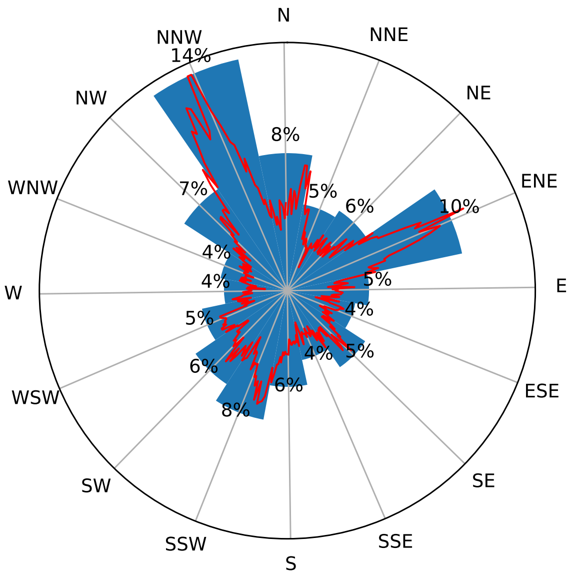
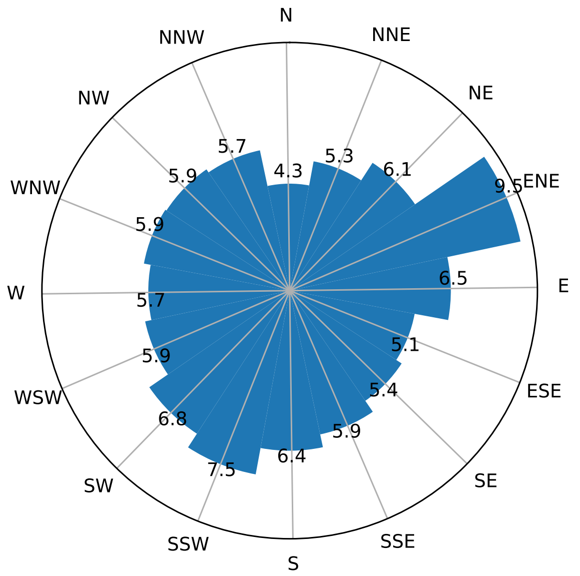


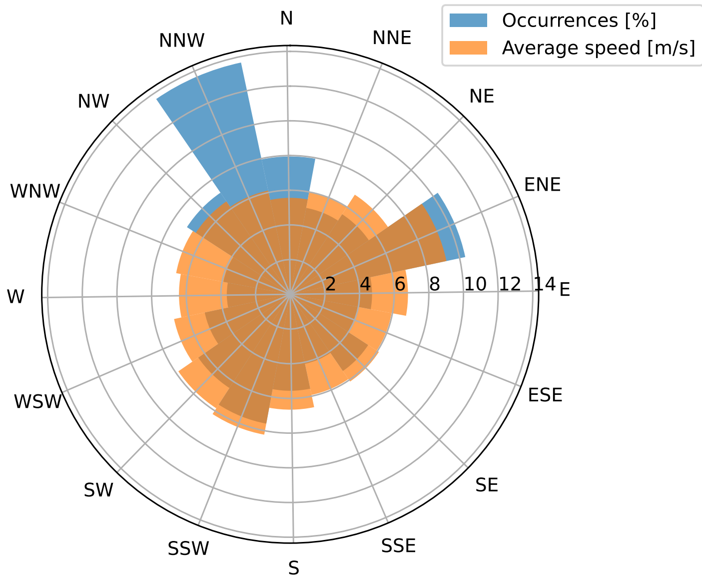
Figure 3.8: Power curve of average Power relative to direction for Turbine 1 in 2011.



(a) Occurrences per sector.



(b) Average speed per sector [m/s].



(c) Average speed and occurrences per sector.

Figure 3.9: Detailed Analysis of Wind Distribution and Speed for Turbine 1 in 2011.

3.3 Wind Farm

In this section, a brief analysis of the wind farm as a whole is conducted, rather than focusing on individual turbines. Special attention is paid to the influence of orography on the performance of the turbines, as well as the impact of neighboring turbines.

3.3.1 Orography

The efficiency and performance of wind turbines are significantly influenced by their interaction with natural elements. This subsection delves into how orographic features such as hills, mountains, and valleys can create turbulence and affect the incident angles of wind, thereby influencing the performance of wind turbines.

Turbulence in the context of wind turbines is a significant factor that impacts both their mechanical integrity and operational performance. The presence of turbulent winds, often caused by complex terrain, subjects turbine blades, towers, and structural components to fluctuating loads. This variability not only increases wear and tear but may also necessitate more frequent maintenance, potentially reducing the lifespan of the turbines [57]. Moreover, the turbulent conditions bring about changes in wind speed and direction as they reach the turbine, leading to fluctuations in power output. These fluctuations pose challenges for grid integration and efficiency optimization, highlighting the complex nature of managing wind energy resources in varying environmental conditions [58].

The concept of incident angles of wind plays a crucial role in the efficiency and effectiveness of wind turbines. Wind turbines are engineered to achieve optimal efficiency when the wind strikes the blades perpendicularly, with the maximum effective incident angle for many turbines being around 8 degrees. Achieving this optimal alignment is essential for maximizing energy capture [59]. However, the terrain surrounding the turbines can significantly alter wind directions, affecting the incident angle at which wind interacts with the turbine blades. This variability in incident angles, often caused by non-optimal alignment due to terrain influences, can lead to reduced turbine efficiency [59]. To mitigate this issue, turbines equipped with a yaw mechanism can adjust to changing wind directions. Despite these adjustments, maintaining optimal alignment remains a challenge in areas characterized by variable wind conditions, ultimately impacting the overall energy production.

The influence of orography on wind turbine performance is an essential factor in their design and siting. Understanding and managing the effects of turbulence and incident wind angles is crucial for maximizing efficiency and longevity of wind turbines in complex terrains.

3.3.2 Influence of Neighboring Turbines

The wake effect, a phenomenon caused by upstream turbines, significantly impacts the efficiency and longevity of wind farms. When wind passes through an upstream turbine, it decreases in speed and increases in turbulence before reaching downstream turbines. This altered wind condition leads to a reduction in efficiency and causes power output to become more variable [60]. Furthermore, the increased turbulence attributed to the wake effect places higher mechanical stress on the turbines. This stress necessitates more frequent maintenance and could potentially shorten the lifespan of the turbines [60]. The implications of the wake effect highlight the challenges in optimizing wind farm layouts to minimize its adverse impacts on turbine performance and durability. Understanding and managing the wake effect is crucial for maximizing the efficiency of wind farms and ensuring sustainable energy production.

For confidentiality reasons, it's not possible to share specific information about the location of the turbines. However, by observing the contour lines, the positions of neighboring turbines, and the terrain in three dimensions, some comments can be made. In the study of the wind farm, it is observed that the turbines are situated on significantly sloped terrain, reaching an altitude of approximately 1300 meters above sea level and surrounded by a large number of neighboring turbines between the northwest and east.

After this presentation, it is interesting to see if there are examples of reduced efficiency due to orography or neighboring turbines. Figure 3.10 shows the power curves of Turbine 5. It is evident that to the north and northwest, the terrain is very steep, as indicated by the closely spaced contour lines (contour lines can't be shown for confidentiality reasons). Notably, it is in these sectors (NW, WNW, NNW, W) that the turbine exhibits power curves shifted to the right, Figure 3.11, representing lower power output at each wind speed.

Just like Turbine 5, Turbines 1, 4, 9, 10, and 11 also show less efficient curves in sectors with steeper slopes. As for the influence of neighboring turbines on efficiency, it is not possible to analyze this in the current case. In areas without steep terrain issues, the turbines are surrounded by neighboring ones. For instance, Turbines 6, 7, 8, and 12, despite being located in flatter areas, are affected by the proximity of other turbines in various directions, precluding a credible analysis.

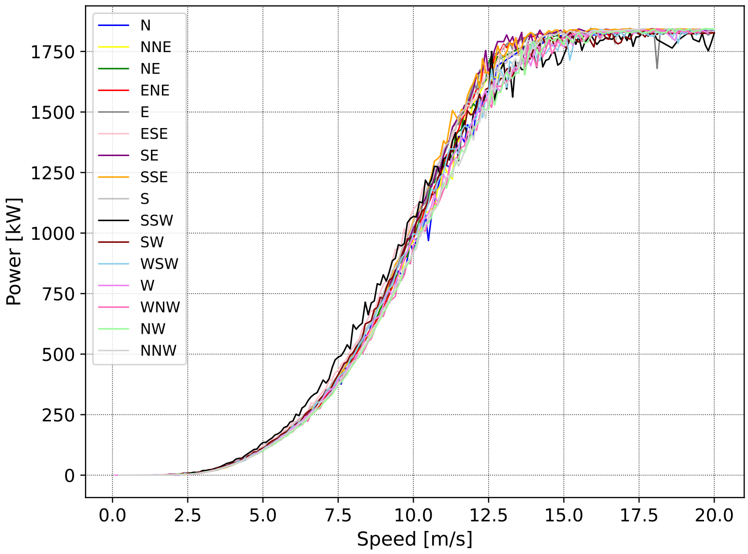


Figure 3.10: Power curve showing average Power Output relative to each sector for turbine 5.

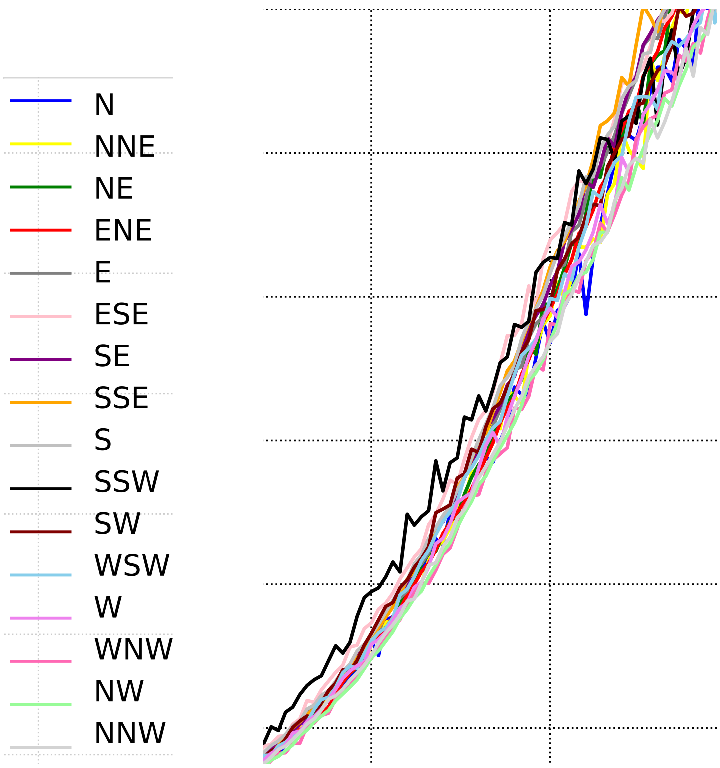


Figure 3.11: Zoomed Power curve for turbine 5.

3.4 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistical method widely used in data analysis and machine learning for reducing the dimensionality of large data sets [61]. By converting a set of possibly correlated variables into linearly uncorrelated variables known as principal components, PCA simplifies the complexity of multi-dimensional data [62]. The process involves identifying the maximum variance direction in the data (the first principal component) and the subsequent orthogonal directions. These components are transformed versions of the original variables, organized in a way that retains most of the variation in the data [61].

PCA is employed in various fields, including finance, bioinformatics, and engineering, for purposes like exploratory data analysis, enhancing predictive models, and data visualization. While effective, PCA assumes linearity and that the most significant directions are those of maximum variance, which may not be accurate in non-linear relationships. Despite these limitations, PCA's utility in simplifying and interpreting complex datasets is facilitated by its incorporation in many data analysis and machine learning tools [61]. Next, a PCA analysis performed on the data is presented. This analysis aims to determine whether there is a correlation between variables.

The following sections detail the steps involved in conducting PCA, aiming to elucidate the correlation between variables and simplify complex datasets.

3.4.1 Procedure of Principal Component Analysis

1. **Data Normalization:** Normalize the data to ensure each variable contributes equally to the analysis, equation 3.1, crucial for PCA's sensitivity to variances [63].

$$X_{\text{norm}} = \frac{X - \mu}{\sigma} \tag{3.1}$$

   where X is the original data matrix, μ is the vector of column means, and σ is the standard deviation of the variables.

2. **Covariance Matrix Calculation:** Compute the covariance matrix, equation 3.2, to capture the variance and covariance among variables [63, 64].

$$\Sigma = \frac{1}{n-1} X_{\text{norm}}^{\top} X_{\text{norm}} \tag{3.2}$$

   where n is the number of observations.

3. **Eigenvalues and Eigenvectors Calculation:** Derive eigenvectors and eigenvalues from the covariance matrix, equation 3.3 to determine the directions and magnitudes of maximum variance [63, 64].

$$\Sigma v = \lambda v \tag{3.3}$$

where v are the eigenvectors and λ are the eigenvalues.

4. **Principal Components Ordering:** Order eigenvectors by descending eigenvalues to prioritize the components with the highest variance.

5. **Principal Components Selection:** Select the first k eigenvectors, corresponding to the k largest eigenvalues, to reduce dimensionality while retaining significant variance.

6. **Data Transformation:** Transform the data onto the new basis formed by the selected principal components, equation 3.4.

$$X_{\text{PCA}} = X_{\text{norm}} v_k \tag{3.4}$$

where v_k are the selected eigenvectors, forming the transformed data matrix X_{PCA}.

7. **Calculation of Loadings:** Compute loadings, equation 3.5 to interpret the contribution of each original variable to the principal components [65].

$$\text{Loadings} = v_k \times \sqrt{\lambda_k} \tag{3.5}$$

which helps in understanding the relationship between original variables and components.

3.4.2 First PCA

For the analysis, columns representing average wind speed, average power produced, nacelle orientation (Position), and time (including year, month, day, hour, minute, and second) were utilized. Table 3.8 displays the variances for the three principal components. PC1 accounts for 49.47% of the variance, capturing nearly half of the total variation in the data. PC2 contributes an additional 25.04% to the variance. Combined, PC1 and PC2 explain approximately 74.51% of the total variance. PC3 further adds 24.38% to the explained variance. Altogether, PC1, PC2, and PC3 account for about 98.89% of the total variance in the data.

Table 3.8: Proportion of variance for the three principal components

	PC1	PC2	PC3
Proportion	0.4947	0.2504	0.2438

The cumulative variance explained by the first three components is notably high, reaching nearly 99%, which is exceptionally favorable. This implies that these three principal components effectively capture the majority of the information contained in the original dataset. To gain insights into what each principal component represents, it is recommended to examine the weights (or 'loadings') of each original variable on the components. Analyzing these loadings will provide a clearer understanding of which variables are most strongly associated with each component.

Table 3.9: 'Loadings' of each original variable on the components

	PC1	PC2	PC3
Average Speed	0.6989	0.0892	-0.0594
Position	-0.1344	0.4991	-0.8560
Time	-0.0737	0.8564	0.5110
Average Power	0.6986	0.0971	-0.05135

From Table 3.9, it can be observed that the 'Average Speed' has a high loading on PC1 (0.6989) and relatively low loadings on PC2 and PC3. This suggests that this variable exerts a strong positive influence on the first principal component but has little influence on the next two components.

Regarding the component 'Position', it exhibits a negative loading in PC1, a moderately positive loading in PC2, and a strong negative loading in PC3. The significant negative influence on PC3 indicates that this variable makes a substantial and inverse contribution to the third principal component.

The variable 'Time' has relatively low loadings on PC1 but a very high and positive loading on PC2, along with a moderately positive loading on PC3. This indicates that this variable exerts a strong positive influence on the second and third principal components.

At 'Average Power', there is a high loading in PC1 and low loadings in PC2 and PC3, which is identical to the 'Average Speed'. This suggests that this variable also exerts a strong positive influence on the first principal component. Consequently, it can be concluded that the variables 'average speed' and 'average power' are the most influential on PC1.

The 'Time' variable holds the most influence in PC2 and also contributes significantly to PC3. On the other hand, the 'Position' variable has a strong inverse influence on PC3. Given the high and nearly identical values of 'average speed' and 'average power' in PC1, this may indicate a correlation between these components, which is expected since power production depends on wind speed.

3.4.3 PCA incorporating meteorological data

In the previous analysis, it was not possible to establish a correlation between most of the variables. Only the average speed and power demonstrated a significant relationship. Therefore, it cannot be concluded that there is any significant correlation between the nacelle orientation or time and the power produced. For this reason, a PCA (Principal Component Analysis) using meteorological data is conducted.

In energy generation systems utilizing wind turbines, the efficiency never reaches 100%. This indicates that some energy loss is inevitable when converting one form of energy to another.

Specifically, for the total energy harnessed from wind, only about 59% can be effectively converted into electrical power. This phenomenon is recognized as the Betz limit [66].

Equation 3.6 represents theoretical power, showing that power depends on air density (ρ). While direct measurements of air density at the study site are not available, it can be inferred from existing data on air temperature, relative humidity, and atmospheric pressure found in the MDD files.

$$P = \frac{1}{2}\rho A V^3 \tag{3.6}$$

To calculate the air density, the Equation for the Determination of the Density of Moist Air was employed [67]. Utilizing the collected data on temperature, humidity, atmospheric pressure, and the calculated air density, another PCA (Principal Component Analysis) analysis was conducted.

In this analysis, the same columns as in the previous analysis were incorporated, supplemented with additional columns for Temperature, Air Pressure, Humidity, and Air Density. The calculation of air density was performed using the previously described method. Table 3.10 illustrates the variations in the principal components. Notably, PC1 is responsible for over 34% of the variance, indicating that it captures more than a third of the total variance in the data. The cumulative contribution of the first six principal components accounts for 95% of the total variance. This suggests that they effectively reduce the dimensionality of the data while retaining most of its informational content.

Table 3.10: Proportion of variance for the principal components incorporating meteorological data

	PC1	PC2	PC3	PC4	PC5	PC6
Proportion	0.3431	0.1723	0.1445	0.1122	0.1082	0.0791

Table 3.11 in the PCA analysis indicates that PC1 is primarily influenced by variations in Temperature (with a significant negative loading of -0.5005), and positively by Air Density (0.4420), Average Speed (0.4261), and Average Power (0.4253). This suggests that PC1 is a composite indicator of thermal conditions, air density, and movement-related energy. PC2 is characterized by a marked negative loading from Air Pressure (-0.5200) and a positive loading from Position (0.5816), hinting that PC2 may represent the interplay between atmospheric pressure and spatial positioning. PC3's loadings are more evenly spread among the variables, yet the positive loading from Air Density (0.5088) stands out, although the component's overall interpretation is less clear. Collectively, these components delineate distinct and coherent patterns within the meteorological and operational parameters of the dataset.

Table 3.11: 'Loadings' of each original variable on Principal Components in Analysis incorporating meteorological data

	PC1	PC2	PC3	PC4	PC5	PC6
Average Speed	**0.4261**	-0.3018	-0.4442	0.0434	0.0362	0.1601
Position	0.0167	**0.5816**	-0.114	0.1423	-0.1966	0.7346
Time	-0.1083	-0.0678	-0.1553	0.6095	-0.7108	-0.2854
Average Power	**0.4253**	-0.3061	-0.4376	0.0674	0.0472	0.1678
Temperature	**-0.5005**	-0.1088	-0.3423	-0.0121	0.0669	0.1640
Air Pression	-0.1318	**-0.5200**	0.4290	0.1048	-0.1670	0.5323
Humidity	0.4027	0.4187	0.0557	0.0342	-0.0797	-0.1145
Air Density	**0.4420**	-0.1111	**0.5088**	0.0501	-0.1267	0.0334

The PCA analysis shows a prominent loading for Temperature and Air Density on both PC1 and PC3. The negative loading of Temperature on PC1, contrasted with its positive loading on PC3, alongside the positive loadings of Air Density on both components, suggests an inverse relationship between these two variables. This is in alignment with physical laws that dictate an inverse correlation between air temperature and density; as temperature increases, air density typically decreases, and vice versa. The PCA loadings reinforce this fundamental atmospheric principle within the dataset.

In the first principal component (PC1), a notable inverse relationship is observed between 'Temperature' and 'Air Density'. Temperature is negatively loaded, while 'Air Density' is positively loaded, suggesting that as temperatures rise, air density falls, which is consistent with the physical behavior of air. This relationship is significant for wind turbines, as power generation is directly proportional to air density, equation 3.6.

Additionally, both 'Air Density' and 'Average Power' exhibit positive loadings on PC1, indicating a direct correlation between these two variables. Higher air density, which implies more mass in the wind, correlates with increased power output, given that more kinetic energy is imparted to the turbine blades. This is a crucial factor in locations with temperature variations, as it affects the optimal operation and efficiency of wind turbines.

The orientation of the turbine, denoted by 'Position', shows a strong positive loading on PC2, juxtaposed with a strong negative loading for 'Air Pressure'. This indicates that turbine orientation has a substantial relationship with atmospheric pressure in this dataset. It implies that different orientations may experience variations in wind pressure, which can influence the efficiency and power output of the turbine.

The positive loading of 'Average Speed' on PC1 correlates with 'Average Power', reinforcing the well-understood relationship that wind speed is a primary determinant of power output in wind turbines. The more substantial the wind speed, the greater the kinetic energy available for conversion to electrical energy. Moreover, the negative loading of 'Temperature' on PC1, when related to power output, indicates that cooler temperatures -often associated with higher air densities — could potentially enhance turbine power output, assuming wind speeds remain constant.

The consistent appearance of 'Air Density' as a significant variable on multiple principal components, including PC1 and PC3, points to its integral role in wind turbine performance. This underscores the complex interplay between environmental conditions and operational efficiency, with air density serving as a critical factor across various scenarios.

The PCA loadings from the wind turbine data suggest that 'Humidity' has a moderate and complex influence on the principal components, contributing positively to both PC1 and PC2. Humidity's correlation with 'Average Speed' and 'Average Power' within PC1 may point to a role in power generation, possibly mediated by its effect on air density. However, its influence appears less significant than that of 'Temperature' and 'Air Density',variables with more substantial loadings on PC1, reflecting their more pronounced effect on kinetic energy and power potential as per fundamental physical laws. In PC2, the positive association of 'Humidity' with 'Position' may imply an interplay between turbine orientation and ambient moisture levels, potentially pertinent to operational efficiency and component longevity. Although not the primary driver in the dataset, the consistent presence of humidity in multiple principal components underscores its integral, albeit less pronounced, role in the complex dynamics of wind turbine performance and efficiency.

In summary, the PCA highlights the intricate dynamics that determine wind turbine performance, emphasizing the significant impact of air density, temperature, turbine orientation, and wind speed. The role of humidity, while not as dominant as these factors, is nonetheless present and contributes to this complex interplay, affecting power generation through its impact on air density and potentially influencing turbine efficiency and maintenance in relation to orientation. These insights are crucial for deepening the understanding of how to tailor environmental and positional factors to optimize energy output. The analysis advocates for the development of sophisticated, multi-faceted optimization models that integrate a comprehensive range of meteorological and technical parameters, to fully harness the capabilities of wind energy technology.

Chapter 4

Forecast

4.1 Model

In this section, some details of the model are explained. The objective is to make a 6-hour forecast. Since the data are time series with 10-minute intervals, 36 points will be predicted. To make the 6-hour forecast, the previous 12 hours will be used. Thus, in terms of data points, 72 points will be used to predict the next 36 time points.

In total, approximately 9 years of data were used for the 12 turbines. The years 2020 and 2021 were not included because consumption patterns changed due to COVID-19. As shown in Table 3.1, there are periods when no MDD data (meteorological data) were recorded, so these periods were also excluded from the analysis. The data were split as follows: 70% for training, 20% for validation, and 10% for testing. The Mean Validation Loss was used to evaluate the performance of the forecast; the smaller this loss, the better the performance.

Hyperparameters are parameters that are not learned during the training process and therefore must be predefined. In the current scenario, the objective isn't to identify the optimal hyperparameters for the network, but rather to ensure that there is no overfitting and that the model's performance is sufficiently robust for the goal of predicting a specific variable of the wind farm under study 6 hours in advance. To this end, monitoring the training and validation loss is essential. This allows us to detect any signs of overfitting and to evaluate the model's performance accurately.

The hyperparameters to be defined include learning rate, batch size, input width, number of epochs, and hidden size. The learning rate controls how much the model is adjusted in response to the estimated error at each update of the weights. Typical values for a neural network with standardized inputs are usually between 1 and 10^{-6} [68]. A learning rate that is too high can increase the training error, while one that is too low can slow down the training and even get permanently stuck with a high training error. An initial learning rate of 0.001 was chosen, with the possibility of adjusting it as needed. As it is stated, "When the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error. [...] When the learning rate is too small, training is not only slower, but may become permanently stuck with a high training error" - Page 429, Deep Learning, 2016 [37].

The batch size has already been briefly explained in section 2.7. By providing data in batches instead of all at once, the model is enabled to process a manageable set of samples at a time [69]. This approach not only helps in efficient memory utilization but also allows the model to update its weights more frequently, potentially leading to faster convergence. Furthermore, using batches helps the model to react better to different data types within the dataset, as it can adjust to the variability and diversity present in smaller subsets of data [69]. In observing the batch as a whole, the model can better identify common features across the samples. For this model, a batch size of 32 was chosen, a common practice in LSTM models that often represents a good balance between performance and memory usage.

In the context of training a single-layer LSTM model with PyTorch, the initial setup of weight initialization is a critical step. This process involves setting the initial values for the weights of the LSTM layer before training begins. PyTorch typically employs default initialization methods that are suitable for LSTM architectures [70]. For a single-layer LSTM, this initialization is crucial as it sets the stage for the network's ability to learn from sequential data effectively [71]. This step ensures that the model starts training from a stable and effective point, leading to efficient learning in tasks involving sequential data.

The number of epochs required depends on the specific case. In this particular instance, 100 epochs will be used, and as can be observed in the following section, 100 epochs were sufficient to achieve a very low loss, with careful monitoring of both training and validation loss to avoid overfitting. Regarding the Input Width, the previous 12 hours, representing 72 data points, will be used, which equates to twice the duration aimed to be predicted. This data granularity, with readings every 10 minutes, is chosen to ensure comprehensive coverage of diurnal patterns, crucial for accurate meteorological and wind predictions in wind turbines. While it might be possible to adjust this value, in this specific case, the selected duration effectively captures significant variations.

The hidden size determines the size of the hidden vector, and a larger hidden size allows the network to capture and retain more information. With more units in the hidden layer, the network can learn more complex patterns in the data. Considering the significant variability over time in the dataset, a larger value for the hidden size seems appropriate, while being mindful of the risk of overfitting. Therefore, a hidden size of 320 was chosen, an intermediate number that aligns with common practices, as identified through research and forum discussions where values up to about 1000 are often considered. This decision complements the other hyperparameters previously discussed, forming a cohesive set of configurations tailored to the model's specific requirements and the nature of the data.

For documentation purposes, it should be noted that the predictions were produced using a Windows 10 Pro (Build 19045.3803) computer. This system was powered by an AMD Ryzen 5 5600H 16-Core Processor and supported by 32 GB of DDR4 RAM with a speed of 3200 MHz. Additionally, it featured a 12GB NVIDIA GeForce RTX 3060 graphics card, running on driver version 31.0.15.3623.

4.2 Algorithm

In this section, the developed algorithm is presented, starting with the retrieval and processing of data, leading up to the presentation of forecast data. Figure 4.1 depicts a flowchart illustrating the entire process, which includes data loading, preprocessing, forecasting, and the presentation of results.

The loaded data refers to the already processed data; the data cleaning is described in section 3.2. The forecasting aspect was extensively covered in sections 2.7 and 4.1 However, there is a preprocessing step that has not yet been mentioned. After the data is split into training, validation, and testing sets, a normalization process is applied. This process requires the calculation of a mean and standard deviation, which in this case was computed using only the training data. After normalization, variables with a standard deviation of 0.0001 or less were removed, thereby reducing noise and eliminating unnecessary parameters.

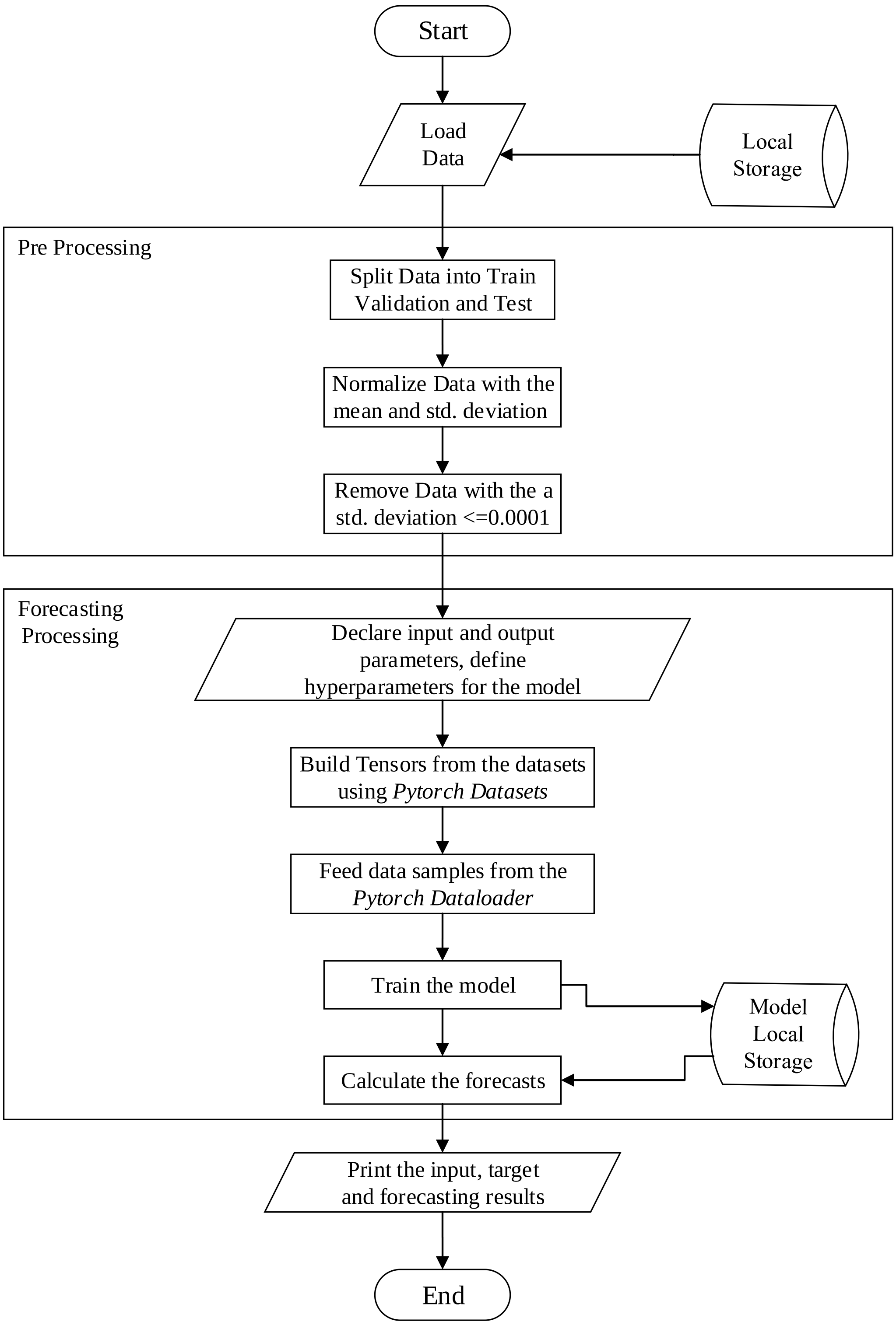


Figure 4.1: Data flux, from clean data to the forecasting results.

4.3 Forecast

This section addresses the forecast itself, including the associated losses and error. The goal is to predict several parameters: the average rotor rotation speed of the turbine, the average power output, maximum and minimum power, as well as the orientation of the nacelle. The input variables are as follows (18 in total): average, minimum, and maximum wind speed; average, minimum, and maximum rotor rotation; average, minimum, and maximum power produced; accumulated power; hours and minutes of turbine operation; nacelle orientation; date in seconds (with the zero second being on January 1, 2010, at 00:00:00); air temperature; atmospheric pressure; humidity; and air density.

The most important predictions for the dissertation are those of average power and nacelle orientation. The remaining predictions are used to attest to the quality of these forecasts. For instance, if there is a prediction of high power production but a very low rotor rotation, it is an indication that there might be an error in the prediction. The predictions of minimum and maximum power are also interesting, as for each point, the average power must be less than the maximum, and the minimum must be lower than the average.

As previously mentioned, the forecast aims to predict a 6-hour period using the previous 12-hour data, meaning it will use 72 data points to predict 36. The hyperparameters include a Learning Rate of 0.001, Hidden Size of 320, Batch Size of 32, and Number of Epochs of 100, along with the Warm-up feature. Table 4.1 presents the total number of entries (data points), the Loss (MSE), and the training time for the 12 turbines and the respective average values.

Table 4.1: LSTM model performance per turbine.

Turbine	Number of Entries	Train Loss	Validation Loss	Test Loss	Train Time [s]
1	321005	0.002280	0.000558	0.000056	365
2	342929	0.002156	0.000780	0.000030	389
3	344486	0.002058	0.000753	0.000017	389
4	343973	0.002175	0.000971	0.000122	383
5	333769	0.001945	0.000664	0.000027	380
6	342395	0.002075	0.000735	0.000035	401
7	340470	0.002263	0.000929	0.000060	408
8	344114	0.001970	0.000785	0.000065	406
9	338631	0.002265	0.000913	0.000228	399
10	339739	0.002314	0.000936	0.000038	408
11	326027	0.002155	0.000961	0.000072	398
12	340549	0.002094	0.000790	0.000052	416
Average	**338174**	**0.002146**	**0.000815**	**0.000067**	**395**

Tables 4.2 and 4.3 present the RMSE values for three distinct time points: at 1 hour (time step 78), at 3 hours (time step 90), and at 6 hours (time step 108), with each time step representing 10

Table 4.2: RMSE for Average Power predictions at 1, 3, and 6 hours.

Turbine	RMSE (1h)	RMSE (3h)	RMSE (6h)
1	2.79930	1.70505	13.11172
2	6.51013	6.31799	6.55566
3	0.98590	1.12811	1.23694
4	0.75635	0.89636	3.04321
5	6.27753	8.87421	5.64954
6	0.89777	1.00342	1.62329
7	3.22574	4.07645	5.08710
8	2.84766	2.93979	3.63300
9	1.67952	4.63904	5.21025
10	2.17868	3.73166	3.00473
11	1.41388	2.86450	9.48291
12	3.19604	5.88147	4.03113
Average	**2.73071**	**3.67150**	**5.13912**

Table 4.3: RMSE for Nacelle Orientation Predictions at 1, 3, and 6 hours.

Turbine	RMSE (1h)	RMSE (3h)	RMSE (6h)
1	0.17874	0.03273	0.87045
2	0.06879	0.02542	0.27350
3	0.34749	0.26228	0.41368
4	0.88098	1.19776	2.20176
5	0.18420	0.35114	0.35941
6	0.60925	0.42484	0.23042
7	0.26971	0.51259	0.77979
8	0.54320	0.28563	0.17556
9	0.34061	0.57225	0.31979
10	0.31342	0.05457	0.52779
11	0.44112	0.01143	1.01260
12	0.41940	0.14771	0.33478
Average	**0.38308**	**0.32320**	**0.62496**

minutes. Table 4.2 presents the errors for average power and Table 4.3 for nacelle orientation. The tables for the remaining errors can be found in Annex C. Figure 4.2 illustrates the evolution of the training and validation loss over the epochs for Turbine 1. Figures 4.3 and 4.4 display the graphs for each variable of interest, including the inputs, predictions, and targets, using Turbine 1 as an example. The graphs for the other turbines are located in Annex D.

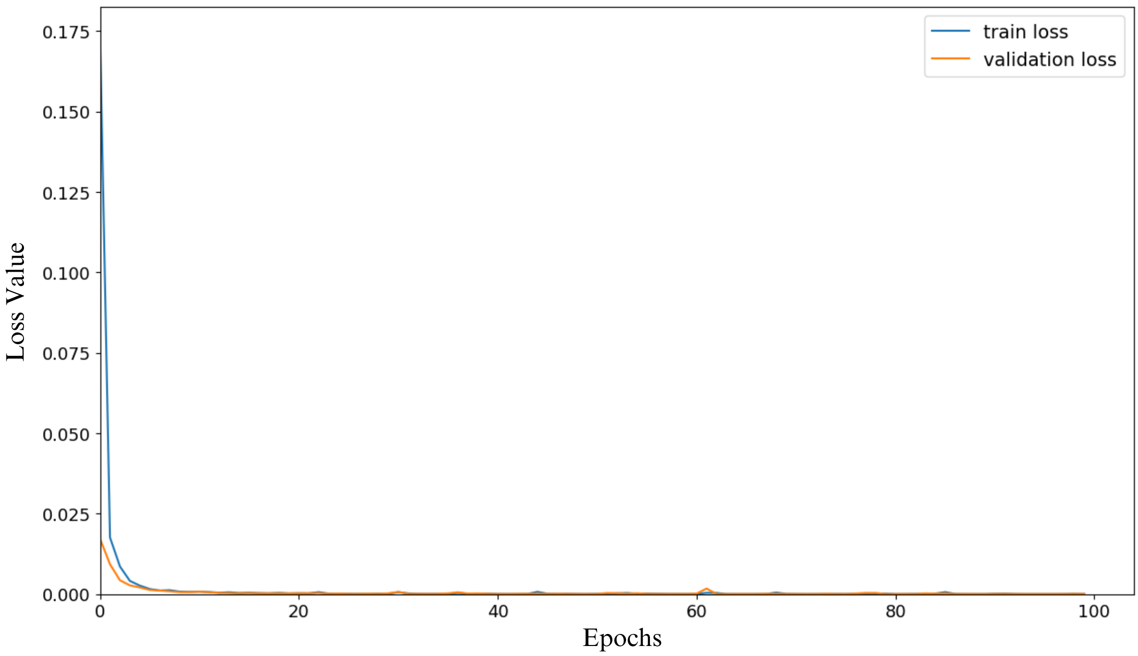


Figure 4.2: Evolution of the training and validation Loss values.

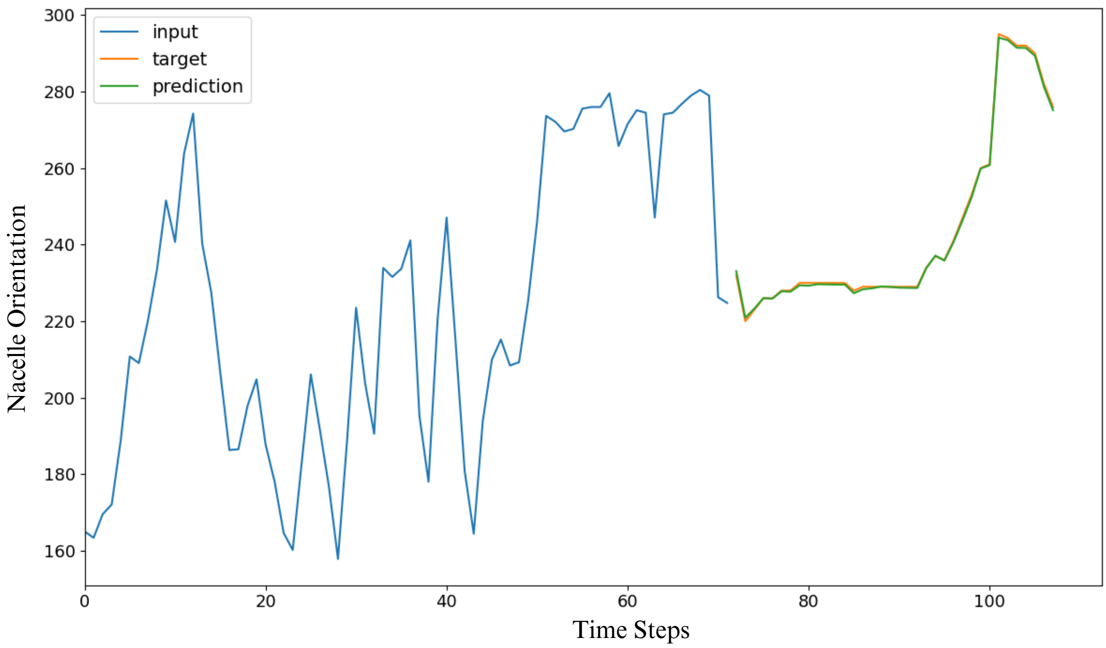
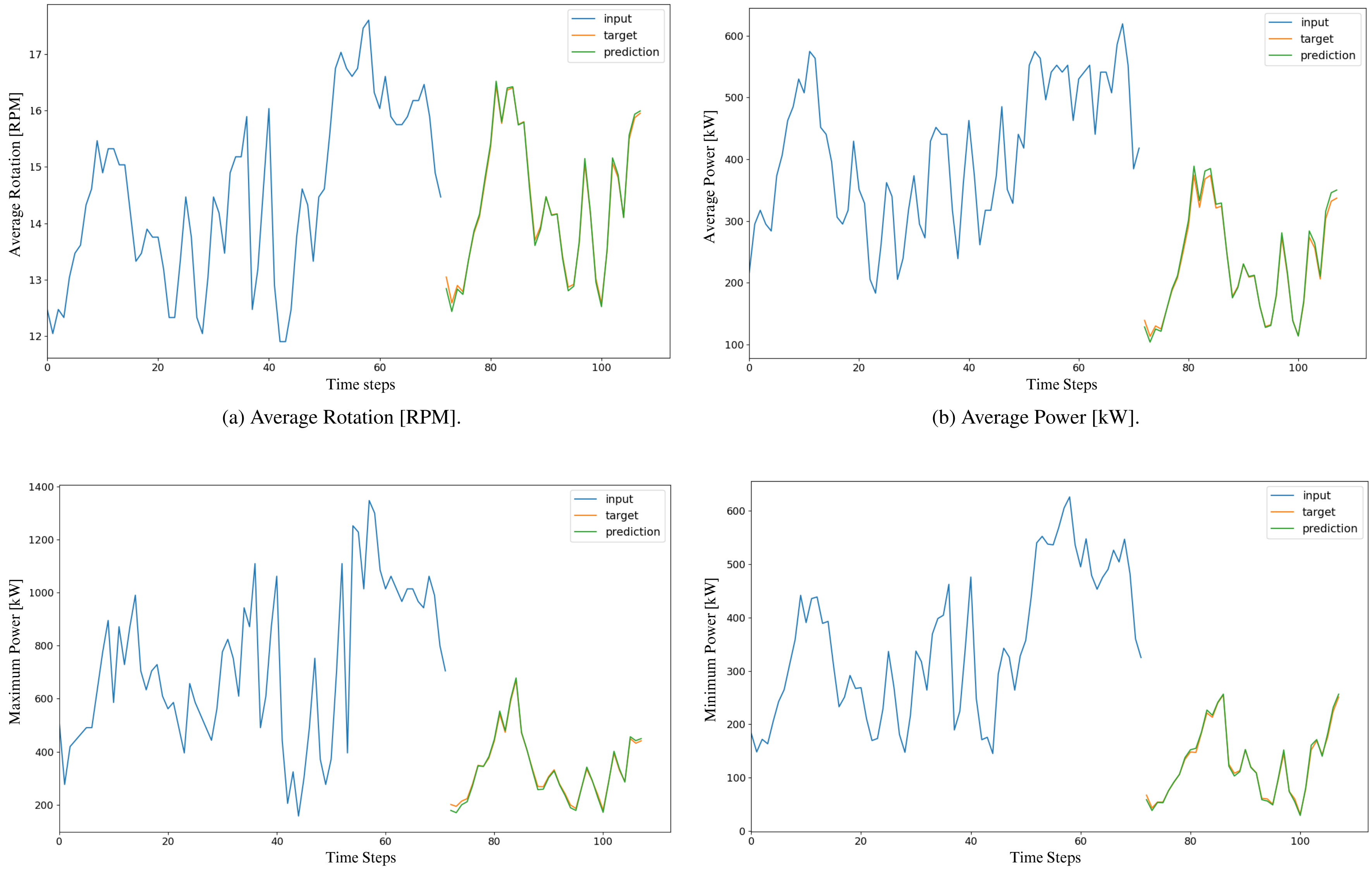


Figure 4.3: Nacelle Orientation.



(a) Average Rotation [RPM].

(b) Average Power [kW].

(c) Maximum Power [kW].

(d) Minimum Power [kW].

Figure 4.4: Illustration of the forecasting results with a representation of the given inputs and targets, and predictions obtained.

4.4 Discussion

After presenting the results in the previous section, the discussion now delves into their detailed analysis. Initially, the focus is on evaluating the quality of the results obtained, assessing whether they make sense within the proposed context, the magnitude of forecast errors, and the appropriateness of the chosen hyperparameters for the model. Subsequently, an interpretation and extrapolation of these results are conducted. The aim here is to understand how the developed model can be integrated into the real world and what the benefits of such integration would be. Finally, a SWOT analysis is carried out, where an examination of the strengths, weaknesses, opportunities, and threats associated with the model is made. This analysis seeks to provide comprehensive insight into the potential impact of the model in practical applications.

4.4.1 Discussion Among Results

It can be stated that the results obtained were quite good. Regardless of the turbine, its location, exposure to neighboring turbines, and the type of terrain, all turbines achieved very satisfactory results. Observing turbine 5, which is one of the turbines with fewer neighbors and located in a northern area with a steep terrain, turbine 7, which has many aerogenerators nearby but is situated on flatter land, or even turbine 11, which is the furthest south of all and on moderately inclined terrain, for all of them, regardless of the conditions, the target line is always very close to the forecast line for all the variables intended to predict.

The errors in the predictions were very low, both in the loss calculation, where the Mean Squared Error (MSE) metric was used for all 6 hours of prediction, and also using the Root Mean Squared Error (RMSE) metric, in which point errors were calculated at 1 hour, 3 hours, and 6 hours. The aim was to understand the evolution of the error over the course of the prediction. The results showed errors very close to 0. For the MSE in test data, the maximum error was 0.000228 and the minimum was 0.000017, with an average error for all turbines of 0.000067. In the average power prediction using RMSE, the maximum error at 1 hour was 6.27753 and the minimum was 0.89777, with an average of 2.73071. At 3 hours, the maximum was 8.87421, the minimum was 0.89636, with an average of 3.67150. For 6 hours, the maximum was 13.11172, the minimum was 1.23694, with an average of 5.13912. It is evident that the trend is for the error to increase with the extension of the prediction interval; however, even for the 6-hour mark, the associated error is considerably low.

In contrast, the average RMSE values for the nacelle orientation at 1h, 3h, and 6h were 0.38308, 0.32320, and 0.62496 respectively, being even lower than the errors associated with the average power prediction, which was expected since the RMSE represents the error in the same scale as the data. The average power can vary between 0 and 1800, however, the orientation of the nacelle can only vary between 0 and 360, therefore, it is expected that the error associated with the nacelle orientation would be lower. An interesting detail is that, unlike all other variables, the 3-hour error for the nacelle orientation is lower than the 1-hour error. All the aforementioned values can be found in Tables 4.1, 4.2 and 4.3.

This discussion also extends to the quality of the data used. When low-quality data, containing noise or irrelevant information, are utilized, it can confuse the network during training and hinder the learning of patterns. Overfitting is another potential issue, where the model achieves high accuracy on training data but fails to make quality predictions on new, previously unseen data. This issue arises due to the model's focus on noise or irrelevant details. As previously discussed, the errors associated with the predictions are very low.

Another challenge when using low-quality data is the inability to achieve convergence. Throughout the training process, the model may fail to reduce the loss as epochs progress. Figure 4.2 illustrates that the model converged with relative ease and maintained this convergence with both the training and validation data. Low-quality data may inadequately represent the population, leading to biased or imprecise learning. In this particular case, the absence of such issues suggests that the data treatment and cleaning were conducted effectively.

It is also crucial to evaluate whether the hyperparameters were appropriately selected. Inadequate selection of hyperparameters can lead to several issues, including overfitting, which occurs with the excessive use of layers or neurons. Conversely, underfitting may arise when the model is overly simplistic and fails to comprehend the complexity of the data. As previously noted, such issues were not evident in this instance.

The learning rate that was chosen can be considered optimal, as a suboptimal rate might have resulted in slow or nonexistent convergence. Conversely, excessively high learning rates could cause disproportionately large jumps during the learning process, leading to the loss of crucial information. Although the loss decreased significantly before reaching 20 epochs, suggesting that the number of epochs could be reduced, maintaining a higher number of epochs was instrumental in preventing overfitting and ensuring no divergence post-convergence. The results indicate that the decision to use data from the preceding 12 hours to predict the subsequent 6 hours was effective, given the minimal errors observed in the predictions.

4.4.2 Interpretation and Extrapolation

At this stage, it is critical to interpret the results obtained and examine how the model developed in this thesis can be implemented in the energy sector to enhance the efficiency of wind power production. The precision achieved by the model, with an error margin of less than 1% for both the prediction of average power output and the orientation of the nacelle, marks a substantial advancement. Having access to this information in advance provides significant benefits for energy producers: it not only optimizes the planning and operation of turbines but also greatly reduces the risk of unforeseen discrepancies that can affect grid stability and profitability.

The use of the LSTM neural network model developed in this thesis for predicting the performance of wind turbines, particularly over a short-term horizon like 6 hours, offers significant advantages in the operation and maintenance of wind farms. With accurate predictions of energy production, operators can optimize the use of turbines, quickly adapting to changes in wind conditions. This allows for more effective management of wind farm resources, minimizing unnecessary downtime and maximizing energy production.

The ability to accurately forecast the energy output in advance significantly reduces the likelihood of congestion in the electrical grid. If all producers can reliably predict the amount of energy they will generate and share this information with the grid operator, the latter can manage the energy distribution much more efficiently, ensuring a balance between supply and demand and thus avoiding the undesirable fluctuations that can lead to disruptions.

> "In recent years, the costs of photovoltaic solar and onshore wind have significantly decreased, making these technologies the most competitive way to generate electricity in Portugal today. Not only are these renewable sources cheaper in terms of total costs compared to conventional fossil technologies, but the levelized costs of photovoltaic solar and onshore wind production are now lower than the marginal costs of conventional fossil technologies. Therefore, it no longer makes sense to promote new investments in these technologies with feed-in tariffs, as they no longer require subsidies and should be encouraged to participate in a competitive market environment."
>
> Renewable Energy Forum in Portugal 2020 [72].

As discussed in the Renewable Energy Forum in Portugal 2020, the decision to end feed-in tariffs for photovoltaic solar and onshore wind energy reflects the decreasing costs of these technologies, which no longer require subsidies. This move encourages energy producers to operate efficiently in the competitive electricity market.

Deviations in wind power production, whether as surplus or deficit, occur frequently due to the variable nature of wind. In the case of excess production, producers are typically compensated at a value below the market rate, as it represents an unplanned energy supply. Conversely, if a producer fails to deliver the agreed amount of energy, they may incur penalties. Both compensations for surplus and penalties for deficit vary according to legislation, making the assessment of their real impacts challenging. However, an analysis is conducted using data from the National Electricity Network (REN) to provide insights into these dynamics.

An analysis is conducted focusing on year 2023, to understand the impact in last year. REN provides information on deviations from five different producers: EDP GEM Portugal, S.A., Endesa Generación, S.A., Iberdrola Generación, S.A. - Branch in Portugal, MOVHERA - Northern Hydroelectric, S.A., and REN - Trading, S.A. Without specific data on the origin of the deviations, it is possible to assume that wind energy constitutes about 60% of the deviations. This assumption is based on the consideration that solar energy, which contributes less than 8% of the electrical production in Portugal, and wind energy, accounting for about 24% in 2023, are the main sources of deviations.

There was a total deviation of approximately 183070.35 MWh in surplus and 208545.98 MWh in deficit for the year 2023. Given that market prices vary daily, average values will be assumed for this analysis throughout the year. The average compensation price for surplus was 66.44 €/MWh, and the average penalty for deficit was -109.38 €/MWh, while the average daily market price was 88.13 €/MWh. Therefore, the compensation for surplus represents a loss of 21.69 €/MWh compared to market values, resulting in a total 'loss' of approximately 3.97 million euros. Similarly, by applying the average penalty to the total deficit, a penalty of about 22.8 million euros is observed. This analysis underscores the importance of more accurate forecasts, a challenge that the model developed in this thesis aims to address.

With the ability to forecast up to 6 hours in advance, it's possible to effectively manage both surplus and deficit of energy. For instance, in the case of a predicted surplus, this information allows for the selling of energy in the intraday market or in the continuous market. The intraday market has a gap of about 3 hours between market closure and the injection of energy into the grid (Table 2.1). In other words, except for the first session, the traded energy is 3 hours away from the market closure. With the 6-hour forecasts, it's possible to negotiate 3 hours of energy with high certainty. However, the continuous market, as explained in subsection 2.2.2, is even more advantageous for the model. This market has sessions every hour, allowing energy injection in the following hour, thus enabling constant energy trading with precise forecasts.

In cases where the forecast indicates that it will not be possible to inject the previously negotiated amount of energy into the grid, resulting in a deficit, it is possible to resort to buying energy in the intraday or continuous market to avoid penalties. Typically, penalty costs are higher than market prices, as observed by the average values in 2023, although this can vary depending on the period. Therefore, by adjusting the position in the market based on forecasts, significant financial penalties that would occur due to the energy deficit can be avoided. However, to fully ascertain the real-world impact of the model, collaboration with an actual wind energy producer is essential. Only through such a partnership can the true benefits of implementing the developed tool be confidently determined.

After exploring the potential impact of forecasting average power production, it becomes pertinent to investigate the advantages of predicting the nacelle position. Recently, there has been a significant shift in research approach towards wind farms. Traditionally, each turbine was viewed as an isolated entity, with the goal of maximizing its individual energy output. However, with the advancement of neural networks, there arises an opportunity to view wind farms as an integrated system, where the focus shifts from the individual production of each turbine to the efficiency of the entire park. This renewed perspective promises to optimize not only energy production but also the overall management and sustainability of wind farms.

With the ability to predict the orientation of the nacelle for the next 6 hours, it also becomes possible to anticipate the impact that each turbine has on its neighbors. For example, in situations where the wind blows in a specific direction, a turbine can create significant 'drag' for the turbines positioned behind it, negatively affecting their efficiency. In these cases, the question arises: to what extent would it be beneficial to adjust the nacelle's orientation, even if this results in a slight loss of efficiency for that turbine, to reduce drag and, consequently, increase the overall efficiency of the turbine group? This is a critical consideration, as such an adjustment could optimize the overall performance of the wind farm, a benefit that the 6-hour forecast model could make feasible.

The model developed can serve as the foundation for a more complex and comprehensive system, aimed at optimizing the efficiency of an entire wind farm. To achieve this, it would be necessary to reprogram the turbines to operate in a coordinated manner, in contrast to the current independent mode. This reprogramming would allow the turbines to collectively respond to the nacelle orientation forecasts and wind conditions. Subsequently, extensive testing would be required to determine the optimal orientation of the nacelles for each wind direction, aiming to maximize the overall efficiency of the wind farm. This optimization process would represent a significant advancement in wind energy management, increasing the total production of the park in a more sustainable way.

4.4.3 SWOT

To conclude the discussion, a SWOT analysis is conducted, in which the strengths, weaknesses, opportunities, and threats associated with the developed project are carefully examined. This analysis is crucial for understanding not only the potential of the model within the current context of the wind energy industry but also for identifying challenges and opportunities that could influence its success and future adoption. Through this evaluation, clear strategic directions aim to be highlighted and valuable insights provided for the project's continuous evolution, aligning it with the needs and emerging trends of the sector.

Strengths

One of the primary strengths of the project is the ability to provide accurate forecasts up to 6 hours in advance, with a very low associated error margin. This capability enables highly effective actions in the electricity market, avoiding penalties and offering the opportunity to trade energy in both the intraday and continuous markets in cases of unexpected surplus. Additionally, the capacity to predict nacelle orientation opens doors to a holistic view of wind farms, enhancing overall efficiency. Short-term forecasting also allows for better management of maintenance and operational activities. A yet unmentioned strength is the model's adaptability to forecast other types of data, such as solar energy. This adaptability is particularly important for hybrid systems that generate energy through both solar panels and wind turbines, making it possible to predict the combined energy output effectively.

In addition to the strengths already mentioned, the model is distinguished by its ease of implementation and adaptability. Requiring pre-processed data, it benefits significantly from the vast amount of data available, thanks to the continuous monitoring of wind turbines, which occurs at 10-minute intervals. This characteristic ensures that the model can be effectively trained and applied to any wind turbine, in any location, provided that quality data are available for training. Such flexibility underscores the versatility of the model, making it a valuable tool for optimizing wind energy production across various operational settings.

Weaknesses

One of the weaknesses is the need for data to feed and train the model. When a new wind farm is constructed, it is not possible to make predictions from the moment the farm starts operating; it is necessary first to collect data and then develop a specific model for that farm. The effectiveness of the model is directly linked to the quality of the data. If the model is fed with low-quality data, there is a possibility that the model may perform poorly.

Another weakness is the need for constant model updates to maintain its accuracy and effectiveness. Although the model was initially trained with data from previous years, in a real-world scenario, it would need to be continuously fed with new data. This implies a continuous requirement for data handling, including the acquisition, cleaning, and integration of updated information, to make the model increasingly robust and intelligent. While this need poses a challenge, it can be mitigated by automating the data treatment process. By implementing automated data acquisition and cleaning systems, it is possible to reduce the operational burden associated with model maintenance, turning this weakness into an efficient and self-sustaining operation.

In addition to the previously mentioned challenges, it is important to recognize that the model, while operational in its current form, is not entirely finalized for long-term applications without additional adjustments. To maximize its utility and accessibility, the development of an intuitive user interface is essential. Such an interface would facilitate the interaction of wind farm operators with the model, allowing for more efficient usage and informed decisions. However, creating this interface requires the mobilization of specialized human resources and financial investments, constituting a significant operational and financial challenge.

Another limitation of the model lies in its inability to predict unpredictable events that can significantly impact wind energy production, such as natural disasters or animal interference, for example, bird flocks. These events can adversely affect the accuracy of predictions, highlighting an area where predictive modeling has clear limits. Dependence on historical data and known patterns implies that the model may not be able to anticipate or quickly adapt to such anomalies, which can compromise the accuracy of forecasts in extraordinary circumstances. However, with technological advancements and the accumulation of more comprehensive databases in the future, it is possible that these unpredictable events may become predictable. Currently, this capability is not a reality, but it represents a promising area for future research and development, pointing towards the continuous evolution of predictive modeling in wind energy.

Opportunities

The renewable energy market is witnessing substantial growth, with various projects in progress to establish new wind farms and upgrade existing ones. This expansion indicates a burgeoning demand for precise forecasting. Continuous advancements in sensor technology and data acquisition capabilities present a significant opportunity to develop even more sophisticated models with narrower error margins. Furthermore, the potential shift away from feed-in tariffs opens up an opportunity, as energy producers may need to sell their energy on the market, making efficient market operations essential.

The increasing digitalization of the energy industry and the expansion of Big Data offer opportunities to further enhance the accuracy of forecasts through the analysis of larger and more varied data sets. This could enable the identification of complex patterns and the optimization of renewable energy production.

Hybridizing wind farms with solar energy systems represents a strategic opportunity. Electrical grids have a capacity limit for energy injection, and hybrid parks that combine solar and wind energy allow the grid to operate close to its capacity limit, maximizing the utilization of available resources. Accurate short-term energy production forecasts are crucial for this integration to be effective, underscoring the importance of the developed prediction models.

Threats

There are several threats to the developed project. One of them is unpredictable events. Although it may be possible to predict these events in the future, they currently pose a significant threat. Another threat is the development of similar technology capable of making short-term predictions but with less data dependence. This could enable the implementation of highly efficient models with a smaller set of training data. Additionally, there is the potential threat of legal regulations to control the use of neural networks. While it is an unlikely threat, it cannot be entirely dismissed.

Moreover, the rapid technological evolution is both an opportunity and a threat, as emerging technologies can quickly render existing solutions obsolete. Staying up-to-date with the latest technologies is crucial to maintaining the project's competitiveness.

Another potential threat is the increasing competition in the wind energy prediction market. As the demand for accurate forecasts grows, more companies and institutions may enter the market, intensifying competition.

Finally, the dependence on IT infrastructure to run prediction models poses a threat, as any failure or interruption in the IT infrastructure can impact the ability to make accurate and timely predictions. With the growing reliance on IT infrastructures and large-scale data collection, cybersecurity becomes a significant threat. Cyberattacks targeting critical infrastructures, such as energy networks and prediction systems, can compromise the integrity and availability of services.

Chapter 5

Conclusions and Future Works

In conducting this study, it was observed that Portugal and Europe have already made significant investments in renewable energies, particularly in wind energy, and are committed to further increasing the energy produced from these sources. The goal is to minimize CO_2 emissions to the atmosphere, thereby necessitating a decrease in fossil fuel consumption. The global trend is moving towards making energy production 100% green, paving the way for innovation in this field.

The analysis of data acquired from the moment a turbine is installed represents a rich source of information. Once cleaned and processed, this data can be used to analyze the wind, including the sectors with the most occurrence, average speed per sector, and the energy produced by the turbine, among many other things. This data allows for the characterization of the wind and the location of the farm. By relating it to the dispersion of turbines throughout the farm or to the orography of the site, it is possible to understand the influence that external variables have on energy production. The principal component analysis also made it possible to see how the variables are related, notably the influence of meteorological variables.

Access to 9 years of data from 12 neighboring turbines in a park enabled the development of an LSTM model capable of making power and nacelle orientation predictions up to 6 hours in advance with a negligible error margin. This model, which worked equally well for all turbines analyzed, addresses some existing problems, particularly the issue of deviations in the electricity market. It was observed that short-term forecasts allow for a reduction in costs, especially with compensation below market value in the case of excess electricity injection or the imposition of penalties in the case of an energy injection deficit.

During the discussion of results, it was also understood that this model opens the door to something greater, namely the possibility of starting to view wind farms as a whole rather than as a set of individual turbines. This approach can increase the overall efficiency of wind farms.

Finally, it was concluded that, in addition to all the strengths already mentioned, there are some weaknesses in the model, particularly the need for quality data for good learning, which subsequently forms good predictions and some threats such the dependence on IT infrastructure.

For future work:

- Real-World, Real-Time Model Application: Using the model in real-world, real-time scenarios to understand how it behaves.
- Automatic Data Processing: It would be interesting to develop an algorithm capable of automatically cleaning and processing data and then feeding the model with the newly processed data.
- User Interface Development: Developing an interface to make it easier and more intuitive for general use as a support tool in the operation of electricity markets.
- Predicting over timeframes longer than 6 hours: Extending the prediction horizon of the LSTM model to enable negotiation in the electricity market with greater certainty over electricity production for timeframes longer than 6 hours.

References

[1] Apache, "Batik *SVG* Toolkit Architecture," 2005. Available in http://xml.apache.org/batik/architecture.html#coreComponents.

[2] H. Ritchie, M. Roser, and P. Rosado, "Co2 and greenhouse gas emissions," *Our World in Data*, 2020. Available in https://ourworldindata.org/co2-and-greenhouse-gas-emissions.

[3] G. Kalra, J. Muppaneni, M. Bertasi, M. Proudfoot, and N. Sharma, "Technical co2 removals market:present and future," tech. rep., Tuck School of Business, May 2022.

[4] "Portugal's renewable energy production hit new record in 2023," January 2024. Available in https://www.reuters.com/article/idUSKBN1Z10VR.

[5] "This Nation of 10 Million People Just Ran Entirely on Renewable Energy for 149 Hours," Dec. 2023. Section: Green Tech.

[6] "Portugal targets 85% renewables in electricity mix by 2030," July 2023. Available in https://www.renewablesnow.com/news/portugal-targets-85-renewables-in-electricity-mix-by-2030-674832/.

[7] Portugal Energia, "Setor energÉtico," 2024. Available in https://www.portugalenergia.pt/setor-energetico/#.

[8] E. R. dos Serviços Energéticos, "Produção," 2020. Available in https://www.erse.pt/eletricidade/funcionamento/producao/.

[9] S. Eletricidade, "Quem somos 2020 março 2020," 2024. Disponível em https://sueletricidade.pt/pt-pt/page/586/quem-somos.

[10] P. Energia, "Agentes." Available in https://www.portugalenergia.pt/agentes/#.

[11] P. Energia, "Setor energÉtico." Available in https://www.portugalenergia.pt/setor-energetico/.

[12] ERSE, "The sector." Available in https://www.erse.pt/en/electricity/the-sector/.

[13] ERSE, "Tarifa regulada de eletricidade." Available in https://www.erse.pt/consumidores-de-energia/eletricidade/tarifa-regulada-de-eletricidade/.

[14] C. et al., "Descrição do funcionamento do mibel.," tech. rep., MIBEL, 2009.

[15] OMIE, "Electricity market." Available in https://www.omie.es/en/mercado-de-electricidad.

[16] OMIP, "Roles." Available in https://www.omip.pt/en/roles-omip.

[17] OMI, "Forward market omip." Available in https://www.grupoomi.eu/en/forward-market-omip.

[18] OMIP, "Modelo de mercado." Available in https://www.omip.pt/index.php/en/node/632.

[19] OMIE, "Aggregate supply and demand curves mibel," 2024. Available in https://www.omie.es/en/market-results/daily/daily-market/aggragate-suply-curves?scope=daily&date=2023-12-01&hour=1.

[20] OMIE, "Intraday auction market." Available in https://www.omie.es/en/mercado-de-electricidad.

[21] OMIE, "Intraday continuous market." Available in https://www.omie.es/en/mercado-de-electricidad.

[22] C. Correia, "Mercado multi-agente de eletricidade: Comercialização de energia renovável em bolsa e por contratos bilaterais com gestão dinâmica de preço e volume," Master's thesis, Faculdade de Ciências da Universidade de Lisboa, 2016.

[23] M. L. C. A. Gouveia, "Estratégias de participação de um parque eólico no mercado de eletricidade," Master's thesis, Faculdade De Engenharia Da Universidade Do Porto, 2020.

[24] DGEG, "Mercado de serviço de sistemas." Available in https://www.dgeg.gov.pt/pt/areas-transversais/mercados-e-mecanismos-de-capacidade/outros-mercados/mercado-de-servico-de-sistemas/.

[25] REN, "Market information." Available in https://mercado.ren.pt/EN/Electr/MarketInfo/Pages/default.aspx.

[26] G. Costanzo, G. Brindley, and P. Cole, "Wind energy in europe, 2022 statistics and the outlook for 2023-2027," tech. rep., WindEurope, February 2023.

[27] REVE, "Portugal sets new record for solar and wind energy and exceeds electricity needs for six consecutive days," November 2023. Disponível em https://www.evwind.es/2023/11/09/portugal-sets-new-record-for-solar-and-wind-energy-and-exceeds-electricity-needs-for-six-consecutive-days/94818.

[28] "World's first semi-submersible floating offshore wind farm." Available in https://www.windfloat-atlantic.com/.

[29] APREN, "Electricity generation by energy sources in mainland portugal in 2023)," 2023. Available in https://www.apren.pt/en/renewable-energies/production/.

[30] Reuters, "Portugal to launch first offshore wind auction, eyes 10 gw by 2030," January 2023. Available in https://www.reuters.com/business/energy/portugal-launch-first-offshore-wind-auction-eyes-10-gw-by-2030-2023-01-23/.

[31] "How is portugal preparing to accelerate offshore wind energy production?." Available in `https://www.lisbonenergysummit.com/accelerate-offshore-wind-energy-production`.

[32] T. Thomas, "Wind energy technology development report 2020," *Publications Office of the European Union*, 2020.

[33] C. of the EU, "Renewable energy: Council adopts new rules," October 2023. Available in `https://www.consilium.europa.eu/en/press/press-releases/2023/10/09/renewable-energy-council-adopts-new-rules/`.

[34] G. Msigwa, J. O. Ighalo, and P.-S. Yap, "Considerations on environmental, economic, and energy impacts of wind energy generation: Projections towards sustainability initiatives," *Science of The Total Environment*, vol. 849, p. 157755, 2022.

[35] E. Commission, "Eu wind energy - european commission," 2023.

[36] IBM, "What is a neural network?," 2024. Available in `https://www.ibm.com/topics/neural-networks`.

[37] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. `http://www.deeplearningbook.org`.

[38] Amazon Web Services (AWS), "What are neural networks?," 2024. Available in `https://aws.amazon.com/what-is/neural-network/`.

[39] C. C. Aggarwal, *Neural Networks and Deep Learning: A Textbook*. Springer Publishing Company, Incorporated, 2018.

[40] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.

[41] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International Conference on Machine Learning (ICML)*, 2013.

[42] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[43] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, 2014.

[44] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *CoRR*, 2014.

[45] A. Gensler, J. Henze, B. Sick, and N. Raabe, "Deep learning for solar power forecasting — an approach using autoencoder and lstm neural networks," *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2858–2865, 2016.

[46] X. Shi, Q. Li, F. Meng, and L. Qu, "Machine learning for wind turbine blades maintenance management," *Journal of Physics: Conference Series*, vol. 1064, no. 1, p. 012070, 2018.

[47] H. Hewamalage, C. Bergmeir, and K. Bandara, “Recurrent neural networks for time series forecasting: Current status and future directions,” *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, 2021.

[48] Wikipedia contributors, “Long short-term memory — Wikipedia, the free encyclopedia,” 2024. [Online; accessed 25-January-2024].

[49] Kion Kim, “Lstm vs lstmcell on gluon,” 2018. Available in `https://medium.com/@kion.kim/lstm-vs-lstmcell-and-others-on-gluon-aac33f7b54ea`.

[50] G. Lambrechts, F. De Geeter, N. Vecoven, D. Ernst, and G. Drion, “Warming up recurrent neural networks to maximise reachable multistability greatly improves learning,” *Neural Networks*, vol. 166, pp. 645–669, 2023.

[51] Simplilearn, “What is epoch in machine learning?,” 2023. Available in `https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-epoch-in-machine-learning`.

[52] PyTorch 2.1, “Mseloss,” 2024. Available in `https://pytorch.org/docs/stable/generated/torch.nn.MSELoss.html`.

[53] PyTorch 2.1, “Torch.utils.data,” 2024. Available in `https://pytorch.org/docs/stable/data.html#torch.utils.data.DataLoader`.

[54] PyTorch 2.1, “Torch.tensor,” 2024. Available in `https://pytorch.org/docs/stable/tensors.html`.

[55] MEDIUM, “Why small batch sizes lead to greater generalization in deep learning,” 2022. Available in `https://medium.com/geekculture/why-small-batch-sizes-lead-to-greater-generalization-in-deep-learning-a00a32251a4f`.

[56] C. M. St. Martin, J. K. Lundquist, A. Clifton, G. S. Poulos, and S. J. Schreck, “Wind turbine power production and annual energy production depend on atmospheric stability and turbulence,” *Wind Energy Science*, vol. 1, no. 2, pp. 221–236, 2016.

[57] A. Ismaiel and S. Yoshida, “Study of turbulence intensity effect on the fatigue lifetime of wind turbines,” *Evergreen*, vol. 05, pp. 25–32, 03 2018.

[58] Y. A. Tadesse Jemal, Samuel Shimels and S. O. Fatoba, “Impact of turbulent flow on h-type vertical axis wind turbine efficiency: An experimental and numerical study,” *International Journal of Heat and Technology*, vol. 41, 2023.

[59] N. Mittelmeier and M. Kühn, “Determination of optimal wind turbine alignment into the wind and detection of alignment changes with scada data,” *Wind Energy Science*, vol. 3, pp. 395–408, 06 2018.

[60] J. Lundquist, K. Duvivier, D. Kaffine, and J. Tomaszewski, “Costs and consequences of wind turbine wake effects arising from uncoordinated wind energy development,” *Nature Energy*, vol. 4, 01 2019.

[61] I. T. Jolliffe and J. Cadima, “Principal component analysis: a review and recent developments,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.

[62] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[63] I. T. Jolliffe, *Principal Component Analysis*. Springer, 2002.

[64] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Pearson, 2007.

[65] M. Ringnér, "What is principal component analysis?," *Nature Biotechnology*, vol. 26, no. 3, pp. 303–304, 2008.

[66] K. H. Bergey, "The lanchester-betz limit (energy conversion efficiency factor for windmills)," *J. Energy*, 1979.

[67] R. S. Davis, "Equation for the determination of the density of moist air (1981/91)," *Metrologia*, vol. 29, p. 67, jan 1992.

[68] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," *CoRR*, vol. abs/1206.5533, 2012.

[69] MEDIUM, "Selecting optimal lstm batch size," 2020. Available in `https://medium.com/@canerkilinc/selecting-optimal-lstm-batch-size-63066d88b96b`.

[70] PyTorch 2.1, "Torch.nn.init," 2024. Available in `https://pytorch.org/docs/stable/nn.init.html`.

[71] N. Vecoven, D. Ernst, and G. Drion, "Warming-up recurrent neural networks to maximize reachable multi-stability greatly improves learning," *CoRR*, vol. abs/2106.01001, 2021.

[72] LNEG, "Fórum energias renováveis em portugal 2020," tech. rep., National Laboratory of Energy and Geology, 2020.

Appendix A

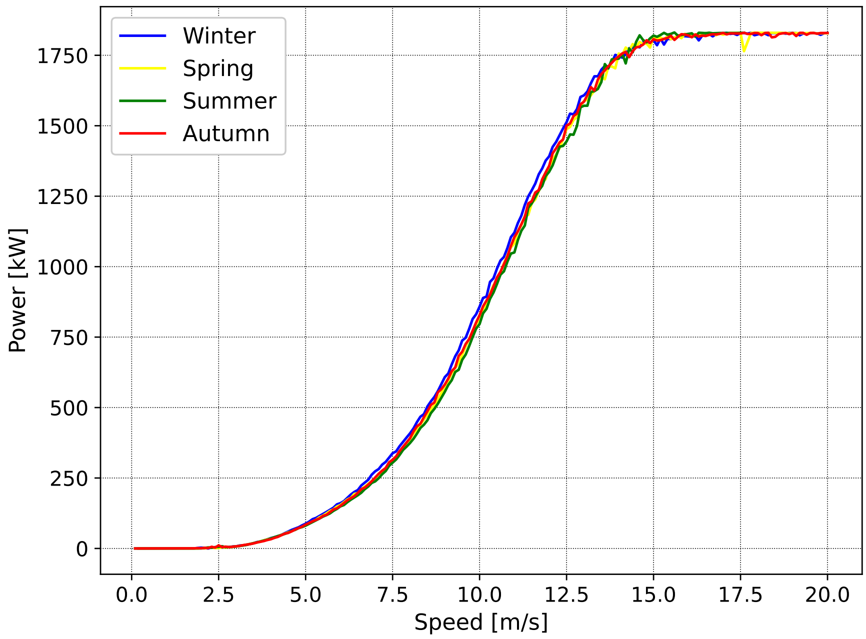
Annex A - Files information

Table A.1: State Information in PES Files for Wind Turbines

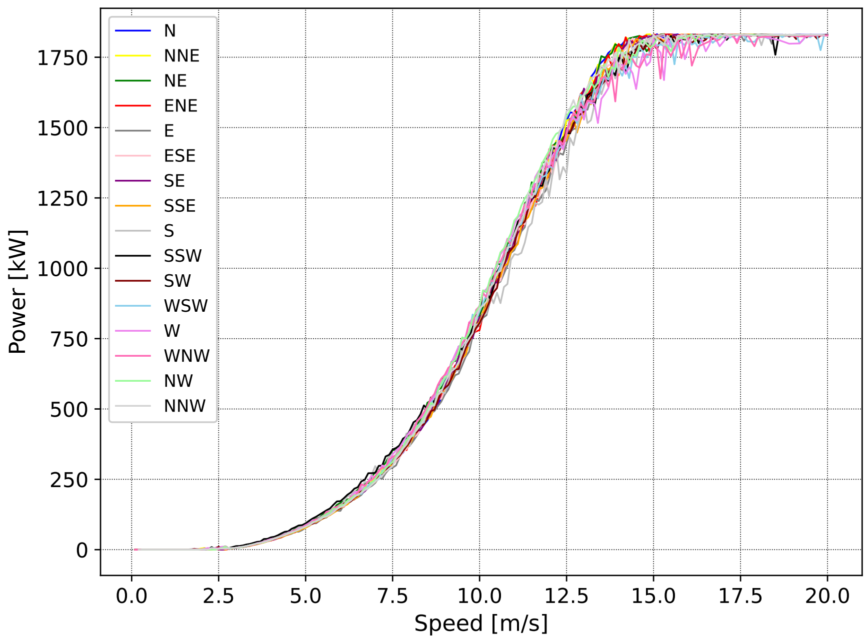
State		
Turbine in operation	Rotor brake activated manual	Fault yaw inverter
Turbine stopped	Rotor lock	Anemometer interface
Lack of Wind	Ice detection	Vibration sensor
Storm	Turbine moist	Tower oscillation
Shadow stop	Overspeed-switch test	Rotor overspeed
Blade defroster	Test security system	Rotor overspeed switch
Battery heat up time	Wind measurement fault	Pitch control error
Unauthorized access	Cable twisted	Main security circuit fault
Maintenance	Yaw control fault	Emergency stop battery fault
Emergency stop actuated	Fault yaw brake	Battery charging error
Protection circuit breaker tripped	Initialize EEPROM	Supply excitation
Semiconductor fuse blown	Program incompatible	Supply I/O-board
Error temperature-measurement	No turbine-id	Supply display
Smoke detector	Supply MPU	Supply blades
Fault ENERCON-UPS	Turbine control bus error	Fault inverter control
Data bus error	Feeding control bus error	MPU error
Data ring error	No data from MPU	Malfunction data bus control
Converter bus error	No data from power control	no data from
Battery test fault	Supply yaw brake	Air gap monitoring
Fault security system	Supply rectifier	Torque monitoring
Speed sensor error	Process reset	Bearing Temperature
Monitoring switch	Watchdog reset	Excitation error
Bladeheating faulty	Turbine reset	Supply power control
Fault lubrication system	Wrong status	Supply anemometer interface
Mains failure	Overvoltage in converter	Hazard light
Mains breakdown	Fault rectifier	Fault UPS WEC
Feeding fault	Overtemperature	Timeout warnmessage
Overcurrent in converter	Transform overtemperature	General warning message
Supply inverter control	Acoustic sensor	Tower door
Supply costumer interface	Generator overtemperature	Remote control PC

Appendix B

Annex B - Power curves

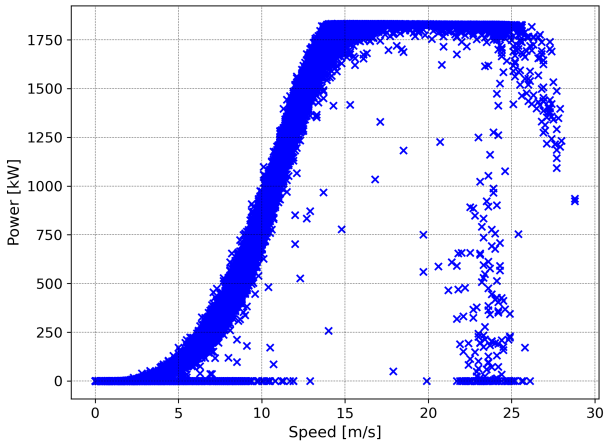


(a) Power curve for average power by season.

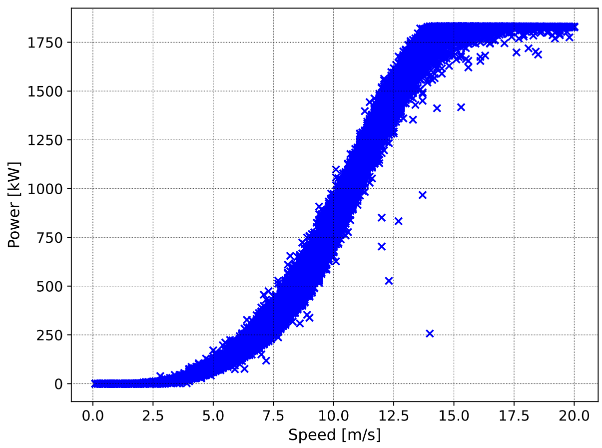


(b) Power curve of average Power relative to direction.

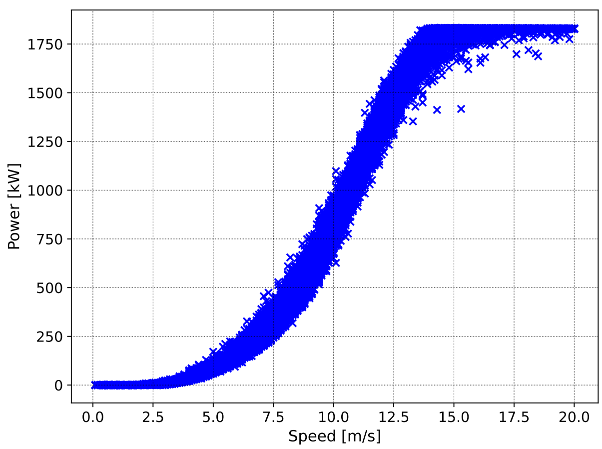
Figure B.1: Power curves for turbine 2.



(a) All Data.

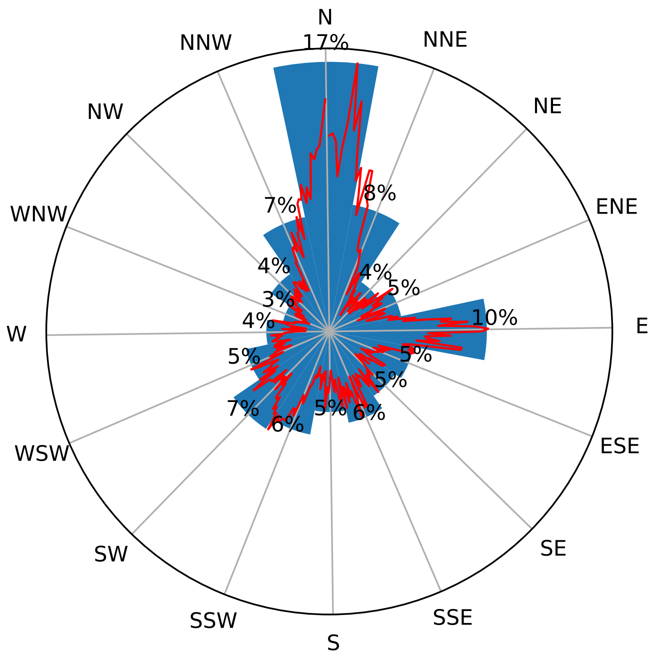


(b) First stage of data cleaning.

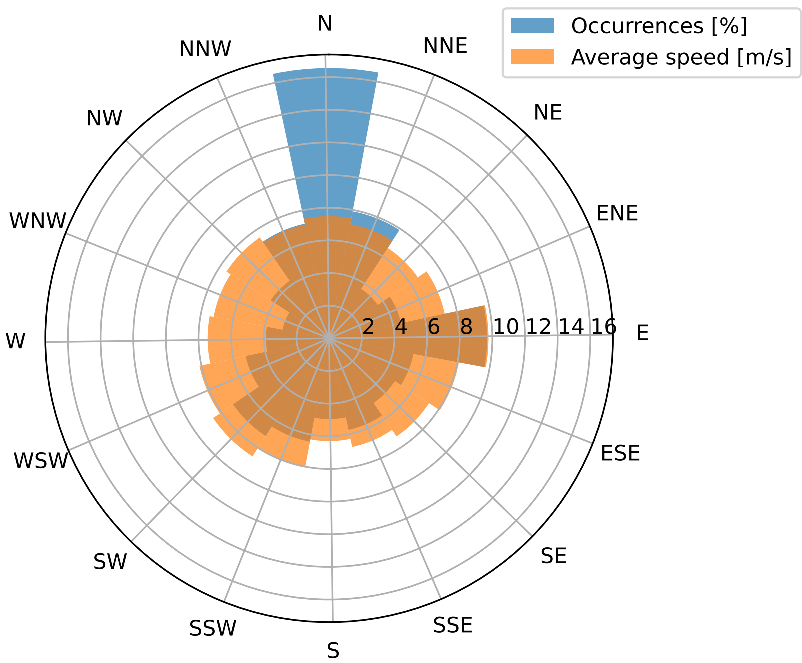


(c) Last stage of data cleaning.

Figure B.2: Data cleaning for turbine 2.

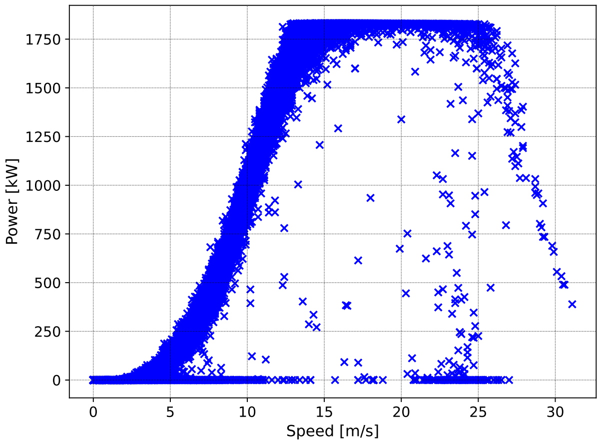


(a) Occurrences per sector.

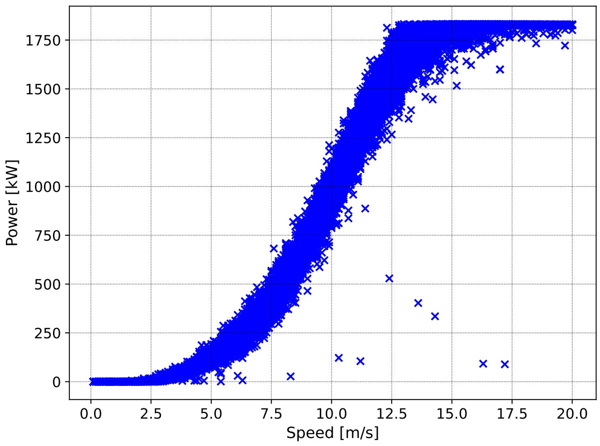


(b) Occurrences and average speed per sector.

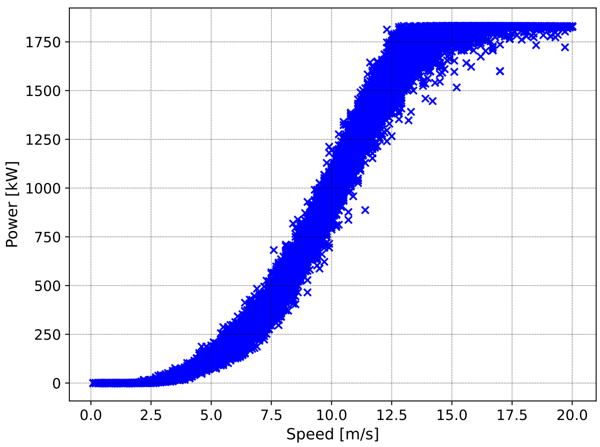
Figure B.3: Wind information for turbine 2.



(a) All Data.

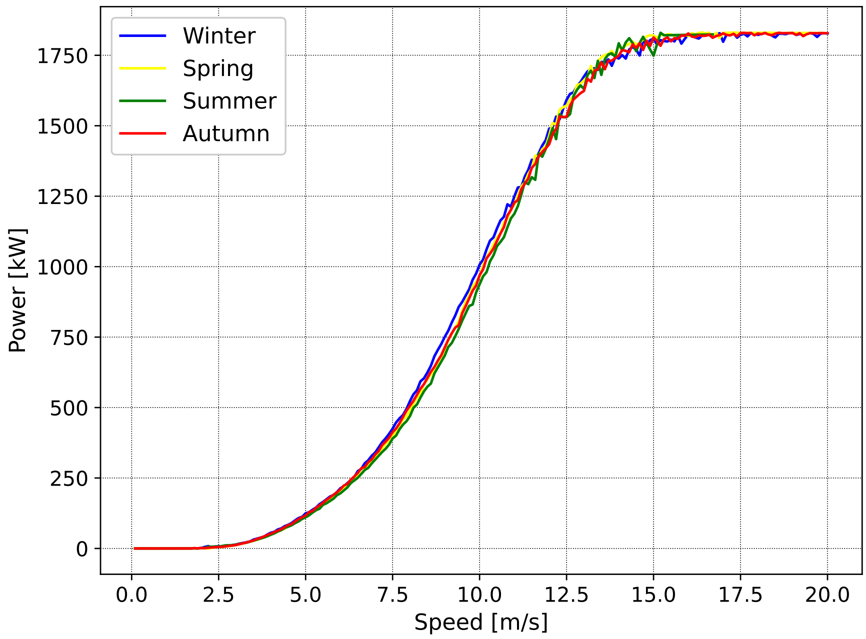


(b) First stage of data cleaning.

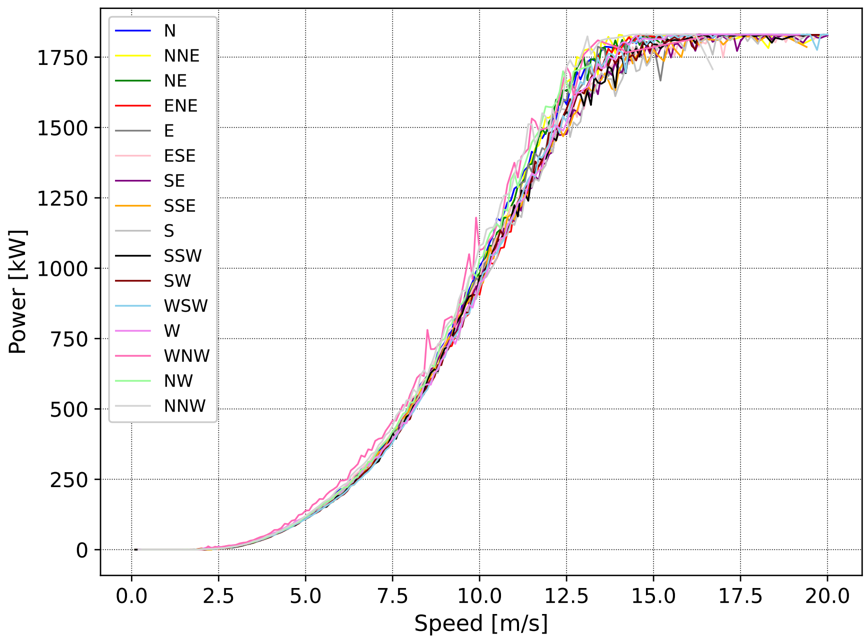


(c) Last stage of data cleaning.

Figure B.4: Data cleaning for turbine 3.

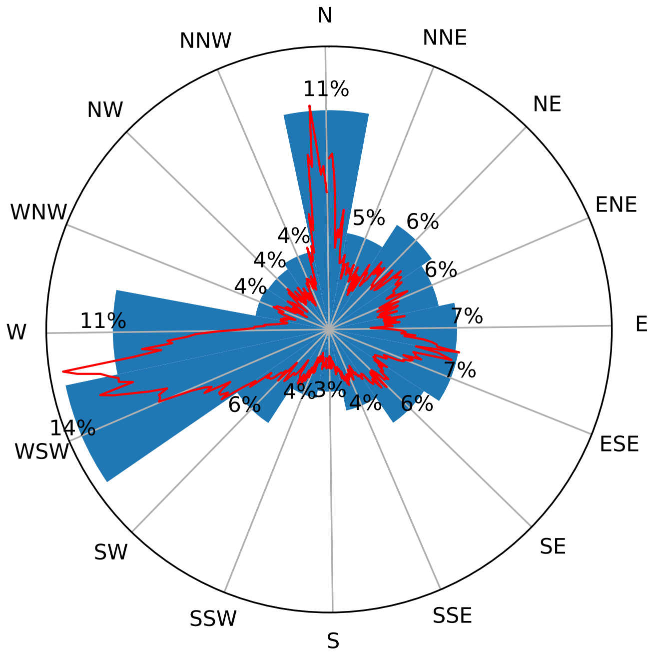


(a) Power curve for average power by season.

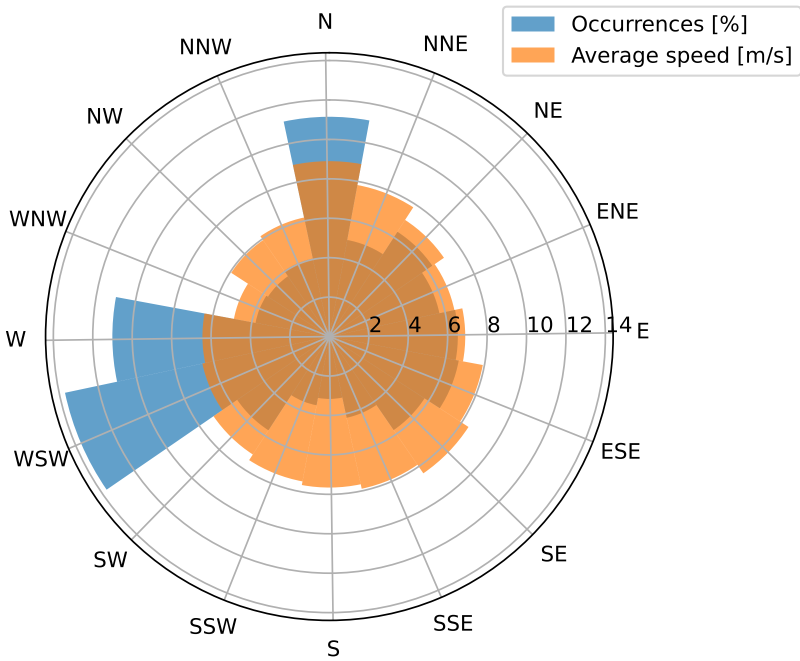


(b) Power curve of average Power relative to direction.

Figure B.5: Power curves for turbine 3.

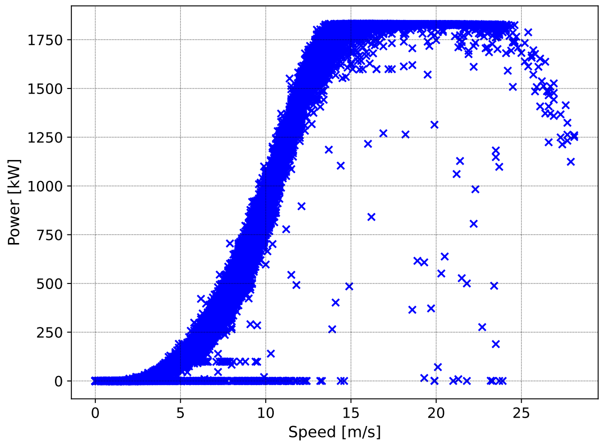


(a) Occurrences per sector.

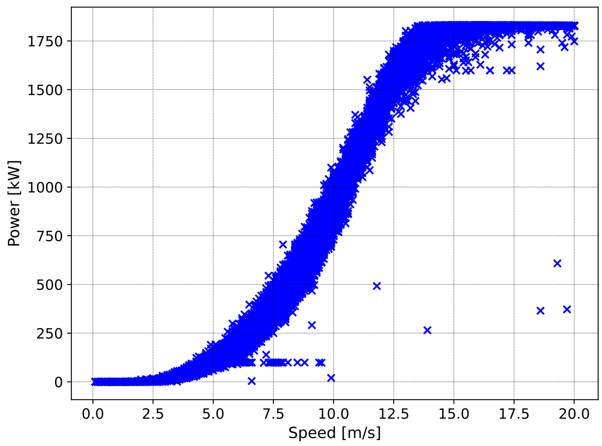


(b) Occurrences and average speed per sector.

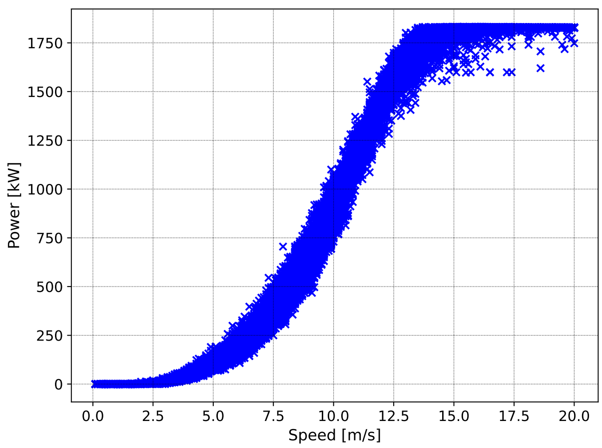
Figure B.6: Wind information for turbine 3.



(a) All Data.

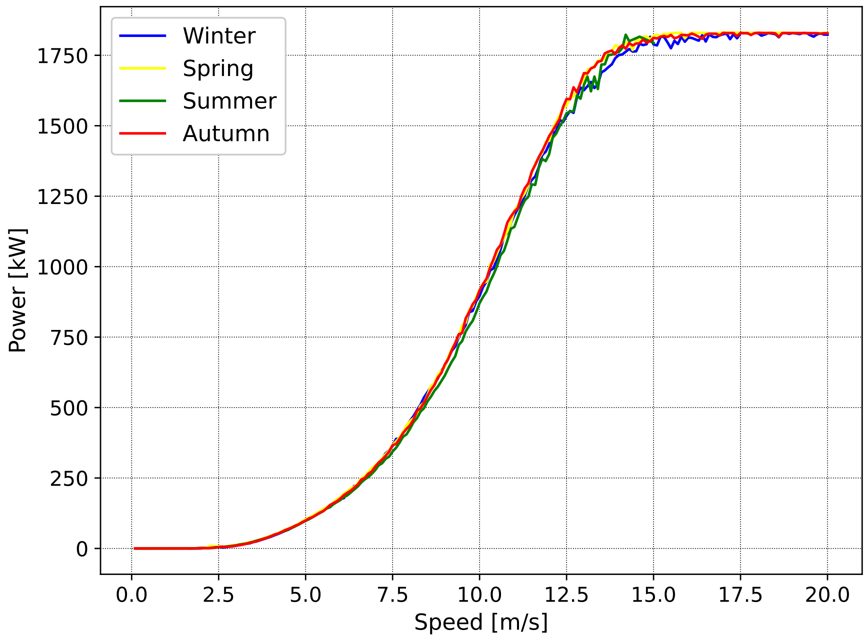


(b) First stage of data cleaning.

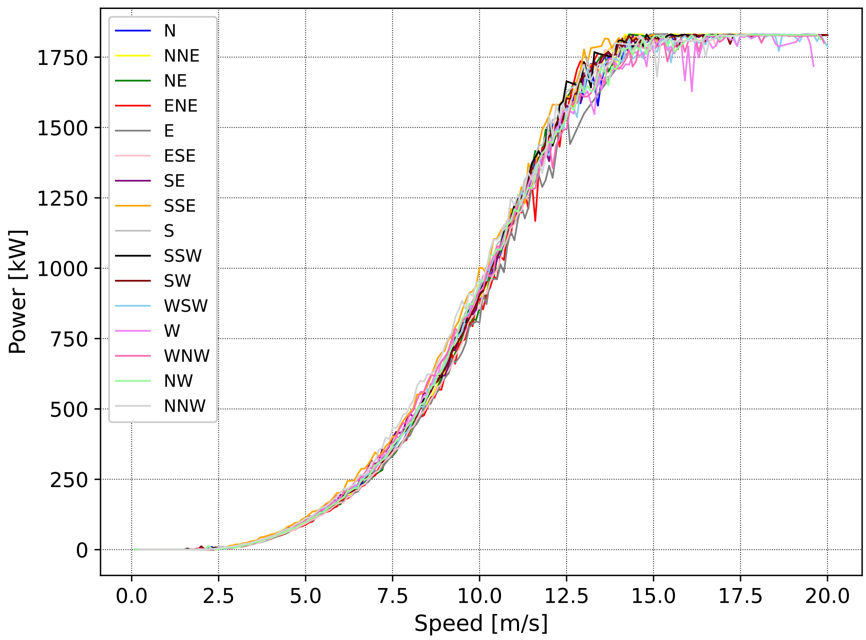


(c) Last stage of data cleaning.

Figure B.7: Data cleaning for turbine 4.

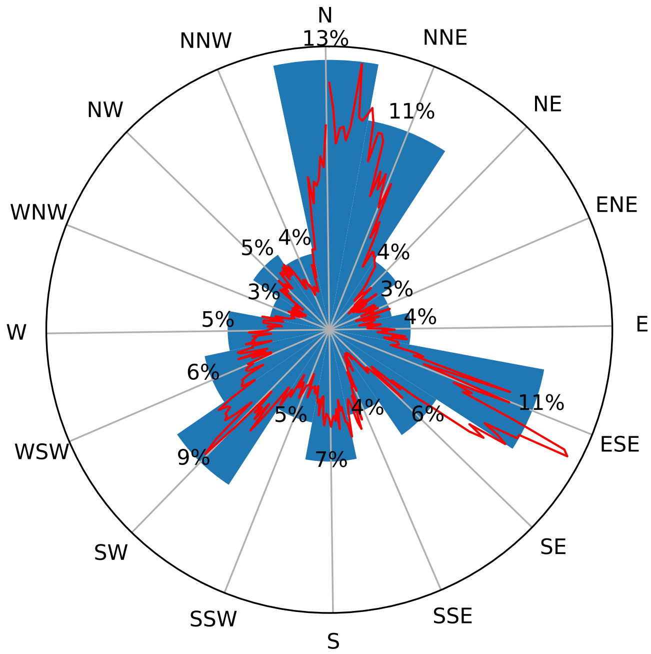


(a) Power curve for average power by season.

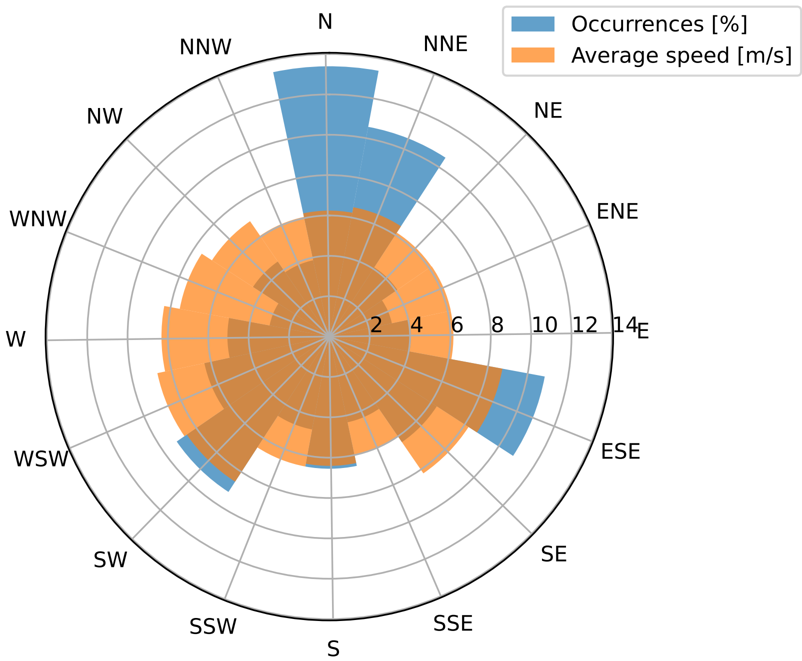


(b) Power curve of average Power relative to direction.

Figure B.8: Power curves for turbine 4.



(a) Occurrences per sector.



(b) Occurrences and average speed per sector.

Figure B.9: Wind information for turbine 4.

Appendix C

Annex C - Forecast Errors

Table C.1: RMSE for average Rotation Predictions at 1, 3, and 6 hours.

Turbine	RMSE (1h)	RMSE (3h)	RMSE (6h)
1	0.03594	0.04848	0.04047
2	0.02503	0.01634	0.04428
3	0.03449	0.03539	0.00646
4	0.00011	0.00744	0.03454
5	0.01700	0.00844	0.01651
6	0.03956	0.00560	0.09417
7	0.00467	0.00864	0.01904
8	0.01676	0.03003	0.01548
9	0.01982	0.07589	0.09009
10	0.00784	0.00773	0.01103
11	0.01165	0.00844	0.03066
12	0.00544	0.01240	0.01585
Average	**0.01819**	**0.02207**	**0.03488**

Table C.2: RMSE for Maximum Power Predictions at 1, 3, and 6 hours.

Turbine	RMSE (1h)	RMSE (3h)	RMSE (6h)
1	2.58292	9.36044	8.38705
2	7.33765	5.12134	8.65454
3	0.34033	0.07959	0.87439
4	1.97510	2.15625	0.76318
5	4.14880	3.77313	2.39429
6	1.79053	0.38257	2.25525
7	4.66418	1.85785	20.00073
8	0.45215	3.90405	0.80527
9	2.12384	3.67252	4.49048
10	0.00784	0.00773	0.01103
11	0.64838	3.11731	3.20514
12	3.16113	0.01782	2.39600
Average	**2.43607**	**2.78755**	**4.51978**

Table C.3: RMSE for Minimum Power Predictions at 1, 3, and 6 hours.

Turbine	RMSE (1h)	RMSE (3h)	RMSE (6h)
1	0.71800	1.96185	5.33527
2	4.88599	2.63721	3.10742
3	0.78180	3.36731	0.67291
4	0.96472	0.94077	1.49072
5	4.02475	3.53833	3.07486
6	0.38226	0.81970	1.17300
7	1.74402	5.07477	4.64621
8	0.53885	1.11531	0.30568
9	2.95723	1.09636	2.19202
10	2.31126	2.31747	2.50439
11	0.10339	0.34946	4.87183
12	2.32495	1.42804	6.88086
Average	**1.88144**	**2.05388**	**3.02126**

Appendix D

Annex D - Forecast Graphs

The graphs can be consulted in
https://drive.google.com/drive/folders/1v7ECehQQCRSstrWSpo1LPeSg2S4qN76U?usp=sharing

Rot Med - Average Rotation[RPM]

Pow Med - Average Power [kW]

Pow Max - Maximum Power [kW]

Pow Min - Minimum Power [kW]

Corrected Pos - Nacelle Orientation