Wind Speed and Curtailment Modelling

Chetali Jain

August 31, 2022

Contents

List of Figures

List of Tables

Dissertation Summary

Wind-based power projects are subject to a significant amount of uncertainty because of the underlying natural resource (wind) is irregular. Due to this reason, when wind speed is very high, wind energy is produced in excess. When the excess wind energy can not be stored for future use, it needs to be discarded. This process of discarding excess wind energy produced is known as Curtailment. For this research, the goal is to build a model for wind speed and curtailment that replicates the main statistical features of the available data and it is processed for the Gwynt y Môr site, which is one of the largest offshore wind farms in the UK.

To achieve the wind speed model and curtailment model, there were several challenges to overcome. The lack of available wind speed data for the Gwynt y Môr location was one of the main challenges for the investigation. The closest onshore wind farms' (Crosby, Speke, and Woodvale) wind speed data were utilised to estimate the site's data. This was achieved by establishing correlation between the wind speed data for the sites. Later, two methods, Principal Component Analysis method and Inverse Distance Weighting method, were used and compared to analyse the estimated wind speed for the Gwynt y Môr site.

Once the wind speed for the year 2019 was estimated for the Gwynt y Môr site, next step was to model wind speed data for the year 2020. Another hurdle in the research was that the wind speed data was only estimated for the year 2019, whereas the curtailment data was available for the years 2019 and 2020. Instead of forecasting, wind speed data were modelled. The prime reason for this was a lack of adequate data for forecasting, which could led to inaccurate forecasting outcomes. After studying various methods to model wind speed data, Stochastic Difference Equations were used to model the data.

The next challenge was combining the estimated wind speed data and the curtailment data for the Gwynt y Môr location. While curtailment data was only provided on a half-hourly basis, wind speed data was estimated on an hourly basis. The interpolation method was employed to resolve the problem. Data on wind speed was

interpolated for every half hour. The curtailment model was subsequently constructed once the final wind speed data and curtailment data were combined. A Markov Chain method which takes into account the transitioning probabilities, (i.e., probability of shifting from one bin to another), were used with some modifications to develop the curtailment model. Once the model was created, it was examined by comparing the statistical properties, like the transitioning probabilities.

Acknowledgement

I want to express my gratitude to my supervisor, Prof. Neil P. Strickland, for all of his support with this study. I would also like to thank Rosie Archer, my dissertation support worker, for her thoughtful criticism of my presentation. I also wish to thank the module leader, Prof. Caitlin Buck, for her advice.

Finally, thanks to my parents and sister for their prayers and support throughout this time. I appreciate your support.

Chapter 1

Introduction

1.1 Background

The capacity to harness energy outside of that produced by humans and animals continues to impact civilization. An expanding portion of the global population can now heat and light their homes, fertilize and irrigate their crops, connect to another, and travel anywhere in the world largely owing to a succession of industrial and agricultural revolutions. Our ability to identify, extract, and utilize energy with ever-increasing dexterity is what drives all of this advancement. The development of a sustainable future based on clean energy production, transmission, and distribution, the storage of electrical and chemical energy, energy efficiency, and improved energy management systems are aided by research in the field of materials science.

1.1.1 Wind Energy and Curtailment

A growing number of nations have revealed their goals to cut carbon dioxide emissions and create clean, alternative energy sources including wind, solar, and tidal energy, among others. Due to its capacity to be exploited globally and the quick technological advancement of wind turbines (WTs), clean and renewable wind energy has recently taken up a sizable portion of the energy industry. The amount of wind energy produced in Europe in 2016 was 13489.9 MWh (megawatt $\times$ hour). As per Ma et al. (2018) there is a high chance that wind power may supply about 30% of the EU's electricity needs by 2030.

In comparison to the current total annual world electricity consumption of roughly 1.5×10^{10} MWh, the capacity of onshore wind energy is quite large: 2×10^{10} - 5×10^{10} MWh. Wind energy production is influenced by variables such as average wind speed, theoretical wind speed distribution, turbulence intensities, and wind turbine system costs (Joselin Herbert et al. 2007).

The design specifications and technical uniqueness of the wind turbine technology are distinctive. Exceptional advancements in wind power design have been made because of recent technological innovations. The energy yield of the turbines has increased 5% annually since 1980 as a result of developments in aerodynamics, structural dynamics, and "micrometeorology". The turbine blades are becoming stronger, lighter, and more effective due to advanced research methods (Joselin Herbert et al. 2007).

For power system operators and planners, the inherent characteristics of wind farms' output power—namely, intermittency and being "non-dispatchable" (i.e., wind farms cannot adjust the power generation as per the electricity demand) in wind turbine output generation—raise numerous new, significant issues. The operator of the wind farm may be required to discard the additional wind power because the output generation of wind farms cannot be managed, and because of transmission limitations, oversupply during periods of low load, and faulty wind speed forecasts (Saber et al. 2019). There is no standardised method to quantify curtailment because it has no universal definition. For this study, curtailment is defined as the disposal of excess wind energy generated due to insufficient storage space.

1.1.2 Wind Farms

It is crucial to initially study wind farms to comprehend how wind energy operates. A wind farm, wind park, wind power station, or wind power plant is a group of wind turbines that are used to produce electricity in one place. The Gwynt y Môr site, which is located off the coast of North Wales, comprises 160 offshore turbines of the 3-6 MW class working over a total site area of 79 km at a planned capacity of 576 MW. This research is based on this location. This facility is built to generate as much electricity as is used by 400,000 typical UK households. Due to its size, Gwynt y Môr is strategically significant for the nation in terms of producing renewable energy, preserving energy security, adhering to carbon dioxide reduction obligations, and reducing climate change (Drew & Rigby 2013).

1.1.3 Modelling Techniques

This research's main goal is to model wind speed and curtailment data for the Gwynt y Môr location. It is crucial to first estimate the wind speed data for the Gwynt y Môr site before modelling the data. Principal Component Analysis (PCA) is used in the process. The goal is to identify a single random variable S, or the wind speed at the Gwynt y Môr location, which is a linear combination of the wind speed at the closest wind farms and can be roughly estimated to be $\Lambda S'$ for any constant Λ. The PCA approach, which presents the wind speeds as a linear combination, employs the eigenvector for the biggest eigenvalue. To get the wind

speed for the Gwynt y Môr site, the outcomes of the linear combination are later multiplied by a constant. In the long run, the value of Λ will not significantly affect the wind speed value and can be chosen at random. For building the wind speed model, Stochastic Difference Equations (SDEs) are used. SDEs takes into account the series of random variable y_m, $(m \in \mathbb{N})$, with a standard normal distribution, and $\text{corr}(y_m, y_{m+r}) = \alpha^r$ (for some chosen parameter $\alpha \in (0, 1)$). The wind speed model thus can be defined with the help of x_i which follows standard normal distribution and $y_i = \alpha y_{i-1} + \sqrt{1 - \alpha^2} x_i$. The method is discussed in more detail in later chapters.

The Markov Chain technique is utilised for the final step, which is defining the curtailment model. With the transitioning probability, this strategy is effective. It is possible to define transitioning probabilities as the probability of shifting from one category to another. Since transitioning probabilities in this research only consider curtailment data, modifications are made to the model to incorporate wind speed data as well. Neural networks are utilised to achieve this. Although the model in this study is created for the Gwynt y Môr site, the same process can be used if data is available for any other site.

1.2 Research Objective

Due to the intermittent nature of the underlying natural resource, wind-based power projects are exposed to a great deal of unpredictability. The control of turbine generator torque and blade pitch, turbine maintenance frequency, production planning, and supply contract amounts in the power markets are only a few examples of how this uncertainty affects both economic and engineering decisions. All of these applications depend heavily on wind condition forecasts, which can be needed a few minutes or up to several hours or even days in advance, depending on the task (Holland & Ikeda 2014).

For this research, the goal is to build a model for wind speed and curtailment that replicates the main statistical features of the available data and it is processed for the Gwynt y Môr site as the curtailment data was available for this site. On further research, it turns out that this site is not only the world's second-largest wind farm but also one of the main sources of wind energy for the United Kingdom. To build a curtailment model for the site it is crucial to first evaluate the wind speed at this location because curtailment is mostly dependent on the amount of wind energy produced which is not consistent due to uncertainty in wind speed. The biggest obstacle in this research is the lack of wind speed data for this location, hence it is critical to first model wind speed for the Gwynt y Môr location. Even though the wind speed data at Gwynt y Môr is not publicly available, the problem is solved as an interesting data science exercise by estimating the wind speed data for this site

using the wind speed data from the neighbouring weather stations. The data from the three closest onshore wind farms are used to estimate the wind speed at Gwynt y Môr. The three nearest onshore wind power stations are:

1. Speke
2. Crosby
3. Woodvale

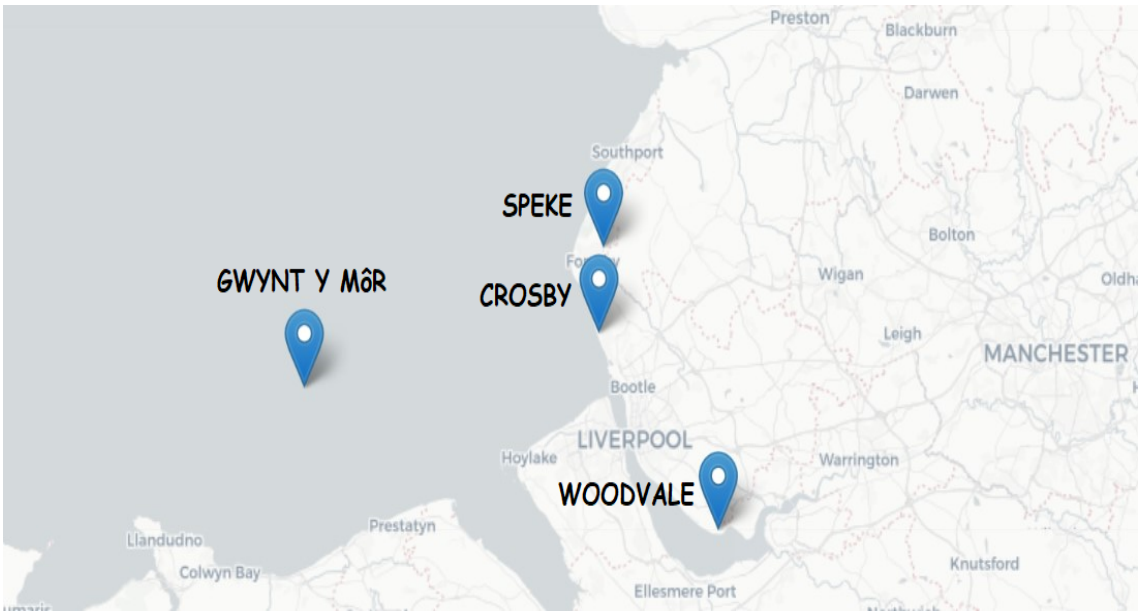

Figure 1.1: Map showing the selected onshore wind farms and the Gwynt y Môr site.

These three locations are found on Great Britain's western coast (Irish Sea) and are highlighted in Figure 1.1. These sites have names based on where the weather stations are (Met Office ID 17309, 1101, and 25069, respectively). Their relative latitudes and longitudes are (53.497 N, 3.0563 W), (53.5674 N, 3.0495 W), and (53.3316 N, 2.8455 W). These sites are all located in the same county, Merseyside. In Table 1.1, the distance (in km) between each location is displayed. It can be seen that the offshore Gwynt y Môr site is closest to the Crosby site. Additionally, the site Woodvale is the furthest away from the offshore site Gwynt y Môr and has the majority of the data missing.

Site	Gwynt y Môr	Speke	Crosby	Woodvale
Gwynt y Môr	0.00	37.43	35.05	50.45
Speke	37.43	0.00	7.84	29.50
Crosby	35.05	7.84	0.00	23.10
Woodvale	50.45	29.50	23.10	0.00

Table 1.1: Distance (in km) amongst stations. The closest onshore site to the Gwynt y Môr site is Crosby, while the farthest site is Woodvale.

Once the wind speed is modelled and using the curtailment data for this site, a relation between them can be estimated using a standard Machine Learning pipeline. Therefore, leading to build a curtailment model for the Gwynt y Môr site.

1.3 Dissertation structure

For the Gwynt y Môr site, a model for wind speed and wind energy curtailment will be developed in this dissertation. Introduction and preliminary analysis of the wind speed data for the onshore sites (Speke, Crosby, and Woodvale) as well as the curtailment data available for the Gwynt y Môr site are presented in Chapter 2. The primary goal of the data analysis is to understand the structure of the wind speed and curtailment data, which will later benefit the model's development.

In Chapter 3, the characteristics of wind speed for onshore sites are introduced. These properties will be used to estimate the wind speed at the Gwynt y Môr site. Principal Component Analysis technique and Inverse Distance Weighting approach are utilised to achieve this. Based on the literature review, the distribution for the wind speed data is later finalised. The Weibull distribution is chosen to fit the wind speed data across all sites, and it provides a good fit.

Chapter 4 discusses the wind speed modelling after the distribution has been decided upon and the Gwynt y Môr site's wind speed data has been estimated. To do this, many strategies like the Time Series Model, Artificial Neural Network, and Support Vector Machine are examined. Stochastic Difference Equations approach is employed to model wind speed data since the data is insufficient to accurately predict data on wind speed.

In Chapter 5, a curtailment model is created using the Markov Chain method, which considers the probabilities of transitions. However, a few changes were made to the model, and a neural network approach is employed to do this.

Finally, Chapter 6, the final chapter, summarises all the conclusions. This chapter also discusses potential future work.

1.4 Summary

A brief explanation of terms like wind farm, wind energy, and curtailment is provided in this chapter. The chapter also discusses the dissertation's objective as well as its structure.

Chapter 2

Exploratory Data Analysis

The data that will be used in this research are introduced in this chapter. In addition, a discussion of the available data is conducted. To acquire a quick understanding of the data, some interim analysis is done on it.

2.1 On-shores Wind Speed Data

Most of the wind speed data from the wind farm are not accessible to the general world and the same is for the offshore wind farm Gwynt y Môr. Although North Hoyle and Rhyl are the closest offshores, but due to the data restriction, data available for the closest onshore (i.e., Crosby, Speke and, Woodvale) are used in this study to predict the wind speed for the desired site. On the MIDAS website, https://archive.ceda.ac.uk/, wind speed data for the three sites are accessible. The raw data file is available in *.txt* format and contains a lot of variables which are not required for this study. Therefore, the data file used involve only the important variables as discussed later.

For the year 2019, the wind speed data available is for every hour, starting from 1 January'2019 till 31 December'2019. The data contains information about the site id, wind speed (in knots) and also the direction of wind speed (in degrees). Another data file that contains descriptions of several wind farms is used to find out more information about these sites, such as the name, county, latitude, longitude, closing, and starting dates. The chosen sites were all founded around the 1970s and are all still operational.

2.1.1 Missing Data

The wind speed data missing for the Speke site is 6.28% and that for the site Crosby is about 0.27%. The Woodvale site has a critical challenge with missing wind speed

data. For this site, 78.46% of the data is missing. Since the majority of the data for this site is missing, it can be argued that the wind speed data for this site shouldn't even be taken into consideration; however, the fact that there is still 21.54% of the data available means that if the decision is made to remove all of the data, then 21.54% of the data will be lost. Therefore, taking into account the amount of missing data, various methods can be thought of for predicting wind speed for the Gwynt y Môr site. Now talking about missing data in the other two sites, several methods might be used. One method is to use the mean wind speed of the relevant site to replace the missing value. The issue with this method is that since the data are time-varying, it is not optimal to substitute the mean for each missing value. The other option is to interpolate the missing number, but this strategy has the drawback that the gap left by the missing value makes it unreliable for solving the issue. Additionally, it is not possible to eliminate the missing values, since doing so would cost a lot of information as discussed earlier.

Wind speed data for the curtailment model must be available for the years 2019 and 2020 at intervals of 30 minutes, but is only available for the year 2019 for each hour. To address these issues, it will be necessary to create a wind speed model that will estimate wind speed data and then interpolate for intervals of 30 minutes. Upcoming chapters will go into more detail on this.

2.2 Curtailment Data

Wind curtailment is the process of lowering the quantity of power generated below the maximum amount that a system of properly working wind turbines is capable of producing. It results in a considerable loss of economic and energy effectiveness. The offshore site, or Gwynt y Môr offshore wind farm, is where the data on wind curtailment is available for this research. For the years 2019 and 2020, the data gives information about the settlement period, curtailment percentage, and energy produced (in %) at half-hourly intervals.

Figure 2.1 shows that in one year, the average curtailment percentage significantly increased from less than 0.2% to 1.2%. When taking a closer look at this, Figure 2.2 further shows that the highest energy reduction occurred in October for the year 2019 however for the year 2020 maximum energy was curtailed in June, and also minimal energy is curtailed in December for the year 2020. Additionally, during the calendar year 2019, there were no energy curtailed in January, February, May, July, and November.

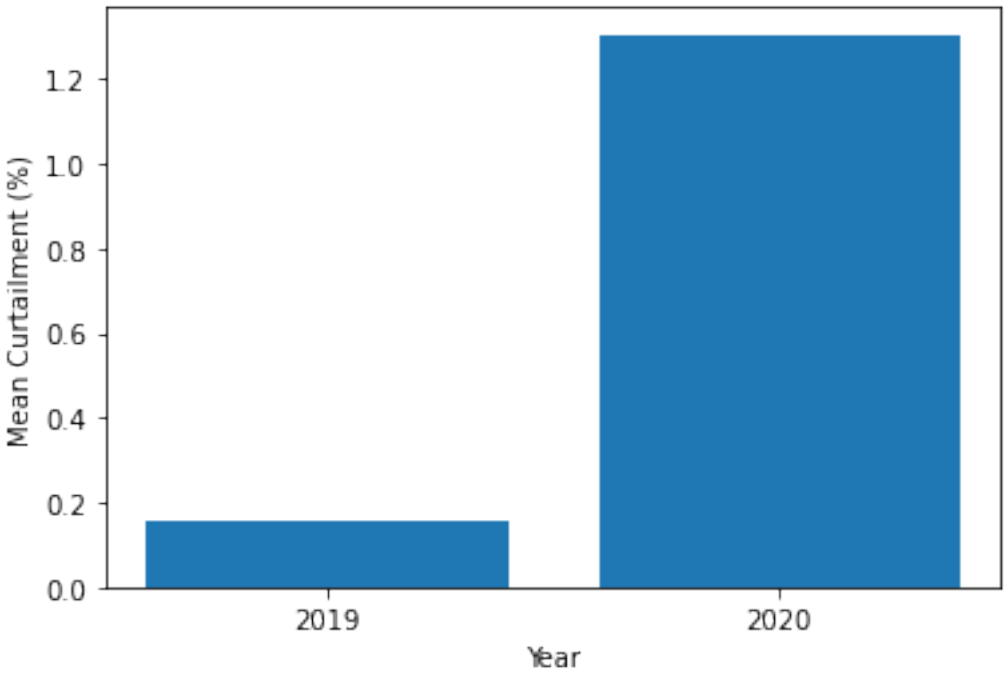

Figure 2.1: Average energy curtailed (%) at Gwynt y Môr wind farm in 2019 and 2020. For 2020 the energy curtailed (%) approximately 6 times the energy curtailed (%) in 2019.

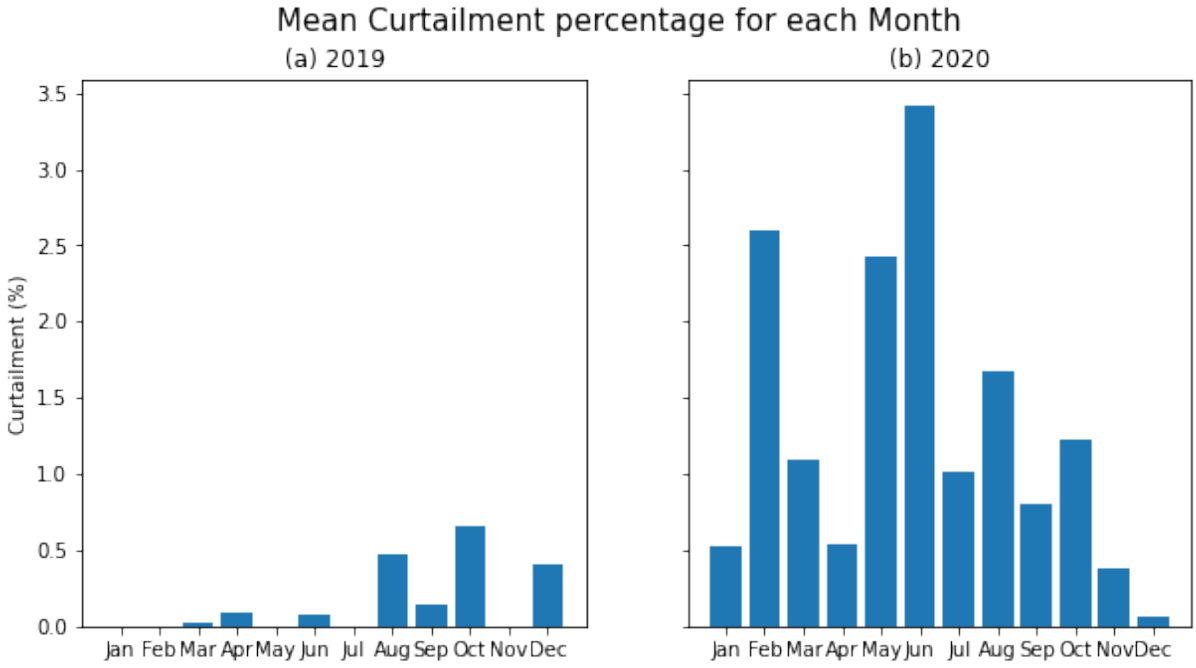

Figure 2.2: The mean percentage of energy curtailed each month in 2019 and 2020 in Gwynt y Môr offshore. Energy is curtailed for each month in 2020 whereas there are many months where there was no curtailment for the year 2019.

From Figure 2.3, it is evident that there is almost no energy reduction when only 0 - 40 % of power is produced. Maximum curtailment of energy occurs when approximately 70% of the power is generated by the wind farm. It is expected that all of this characteristics should be reflected in the curtailment model.

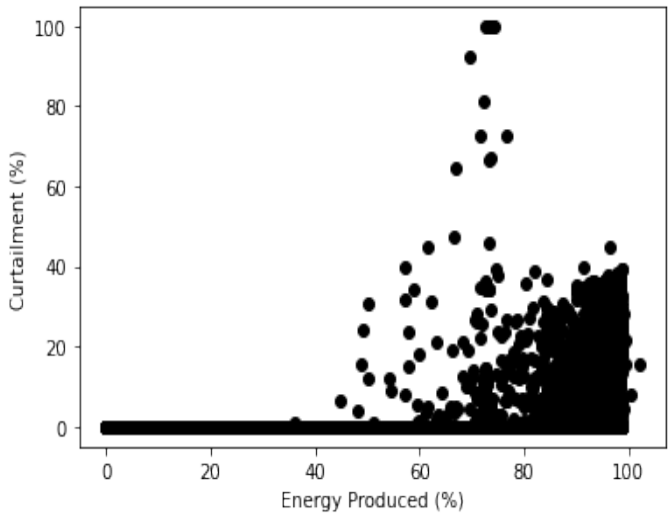

Figure 2.3: Percentage of energy curtailed with respect to the percentage of energy produced in Gwynt y Môr site. There is almost no curtailment when less than 40% of the energy is produced, but there is a maximum energy curtailment when around 70% of the energy is produced.

2.3 Summary

This chapter contains an interim analysis of the data that is accessible. Figure 2.3 illustrates how the percentage of power curtailed depends on the amount of energy produced by the wind farm, which in turn depends on wind speed. Therefore, wind speed for the Gwynt y Môr site is needed and should be combined with the curtailment data to develop a curtailment model. It should be noticed that the curtailment data is for every 30 minutes, whereas the accessible wind speed data are hourly data. The interpolation approach will be employed to integrate these two data sets.

Chapter 3

Wind Speed

The statistical properties of wind speed data for the accessible sites, namely Speke, Crosby, and Woodvale, will be covered in this chapter. A few methods for fitting wind speed data will also be briefly addressed.

3.1 Wind Speed Characteristics

The statistical properties of the wind speed, such as mean, which describes the average wind speed, and standard deviation, which describes the variation in the wind speed over the period, have an impact on the production of wind farms. The signal indicating the power output of wind turbines is constantly changing because it is widely known that wind speed is unreliable and an uncontrollable source of energy. This is another key motivation for developing a wind speed model.

To understand the structure of the wind speed data, a preliminary analysis was carried out. The mean monthly wind speeds for the sites of interest are determined using the hourly wind speed data. Figure 3.1 shows that the mean wind speed varies from 8.45 knots to a maximum of 14.99 knots. Additionally, it can be seen that Crosby typically experiences stronger winds than the other two places. Besides that, the three sites' maximum wind speeds are recorded in March. However, the minimum wind speed varies for across all sites.

In Figure 3.2, the average daily wind speeds for the year 2019 during twenty-four hours (hourly change) are shown. At all the sites, the average wind speed throughout the day is greater than 8 knots. The graph demonstrates that the daily average wind speed is not consistent over the hours. Additionally, it can be observed from the hourly graph that the mean wind speed in Crosby is usually higher than the other 2 sites. Apart from that, the graph for the site Woodvale is a bit short because data for many hours is missing for this site. However, this is not the main

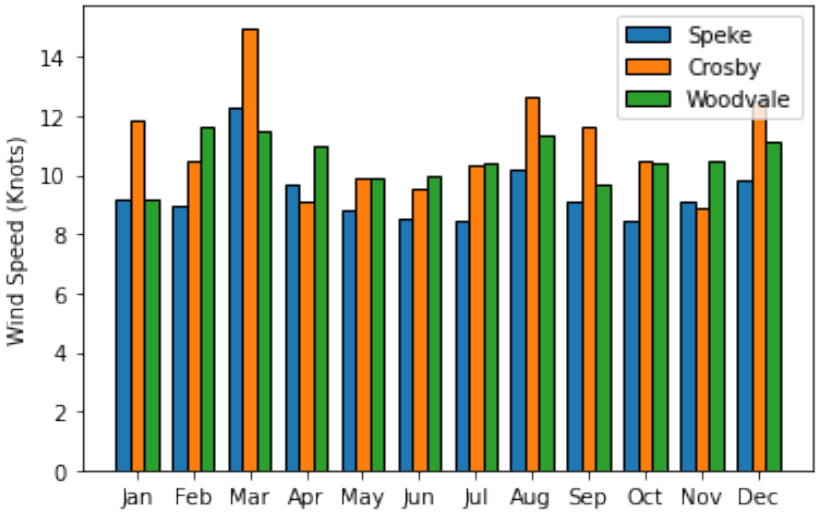

Figure 3.1: Monthly mean wind speed (in knots) for all sites. In March, the mean wind speed is at its highest across all sites, although the minimum wind speed varies across all sites.

concern as later in the chapter various methods will be discussed which overcomes the issue of missing data.

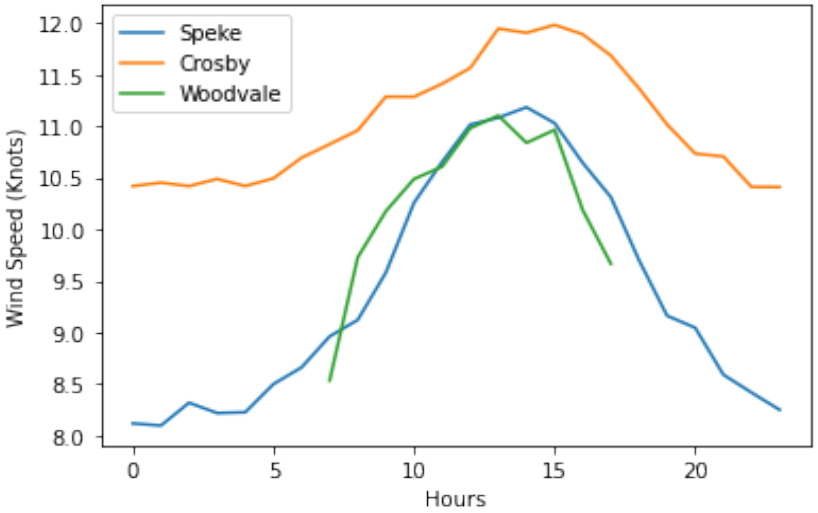

Figure 3.2: Hourly mean wind speed for all sites. All sites have hourly mean wind speeds greater than 8 knots, with the Crosby site having the highest.

Wind speed is a two-point time-correlated variable, which is another attribute that may be inferred from the wind speed data. The wind speed autocorrelation function often exhibits an exponential decrease over hours (Zárate-Miñano et al. 2013) as

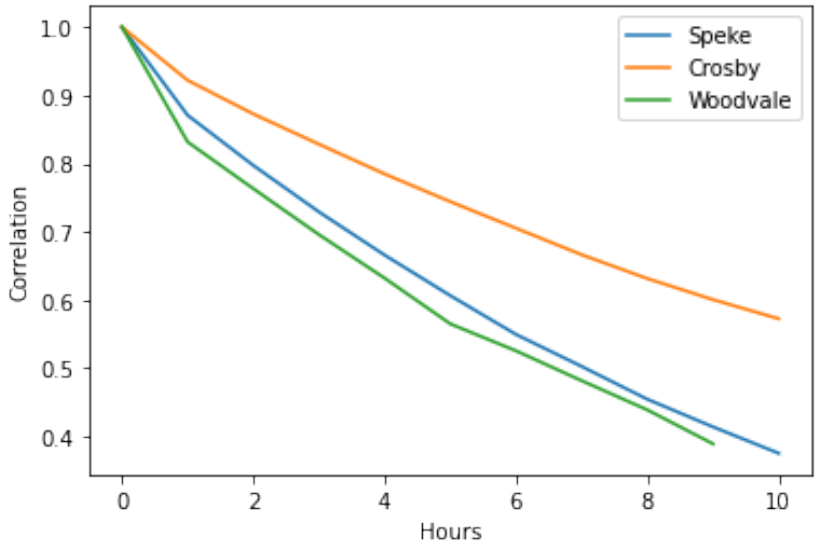

Figure 3.3: Correlation between wind speed over first 10 hours. The graph displays the sites' exponential decay characteristics in correlation.

can be seen in Figure 3.3. This is one of the most important characteristics of wind speed and it should be reflected in the wind speed model.

Skewness and kurtosis measurements can be used to identify the distribution's characteristics. To determine whether the distribution is normal, left-tailed, or right-tailed, skewness is used. On the other hand, kurtosis is used to assess the distribution's height. The measurements of skewness and kurtosis are initially looked at while choosing the distribution which makes statistical analysis crucial in the field of energy generation (Aslam 2021).

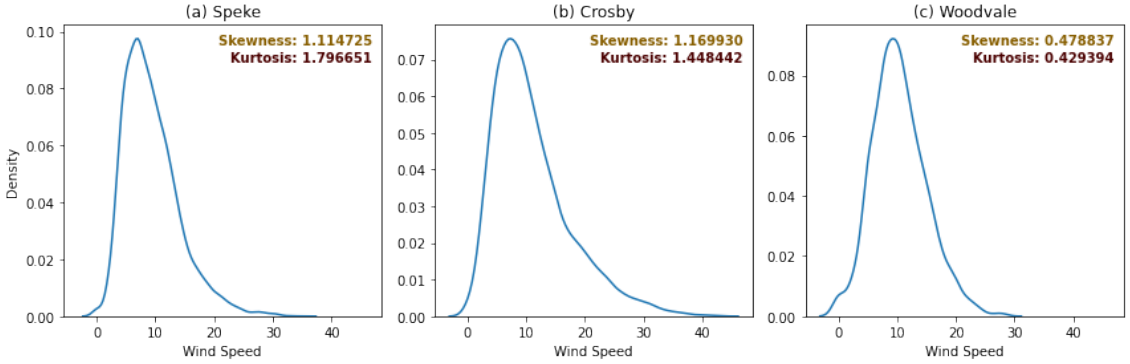

Figure 3.4: Density plot showing skewness and kurtosis for each site. Skewness for Speke and Crosby indicates that the wind data is positively skewed, whereas it is nearly symmetrical for Woodvale. According to kurtosis, the data will have a Platykurtic distribution or a flat tail for all the sites.

According to Figure 3.4, the skewness for the locations Speke and Crosby are greater than 1, indicating that the data is positively skewed. The fact that the skewness for the Woodvale site is within a range of -0.5 and 0.5 suggests that the Woodvale wind speed data is almost symmetrical. However, because so much of the site's data is missing could be the reason for this discrepancy, and for further analysis, it should be noted. When taking into account the kurtosis for all of the sites, it can be seen that the kurtosis value is all less than 3, which denotes Platykurtic Distribution. This distribution suggests flat tails, further demonstrating that there will be minor outliers in the data.

The wind speeds at each site were compared, and the correlation coefficients between them were calculated, as shown in Table 3.1. It can be suspected that the distance between the farms is influencing their correlation. The sites at Woodvale and Crosby have the highest correlation, whereas Speke and Woodvale have the lowest. The correlation, however, is strong enough for all the sites to suggest that it can be used to project the wind speed at Gwynt y Môr offshore.

Site	Speke	Crosby	Woodvale
Speke	1.00	0.76	0.73
Crosby	0.76	1.00	0.82
Woodvale	0.73	0.82	1.00

Table 3.1: Correlation coefficients of wind speeds amongst stations. The wind speed data at Speke and Woodvale have the minimum correlation, whereas Woodvale and Crosby have the maximum.

3.1.1 Wind Direction Analysis

When developing a wind speed model, it is feasible to take into account the direction of the wind speed for the designated sites. The wind rose displays the direction, speed and frequency of wind in a particular location. A wind rose diagram is therefore useful and depicted for every site in Figure 3.5. The percentages mentioned in the circles are the showing the wind frequencies and the colour codes are indicating the wind speed (in knots) for each site. As can be seen, the diagrams for the locations Speke and Crosby are relatively similar, both in terms of frequency (i.e., 12.6 and 12.8 respectively) and wind direction, however, the diagram for the site Woodvale shows a comparable direction to the other two sites but an increased magnitude of frequency and wind speed. Apart from that, for all the sites it can be interpreted that the majority of the time wind is travelling from West to East.

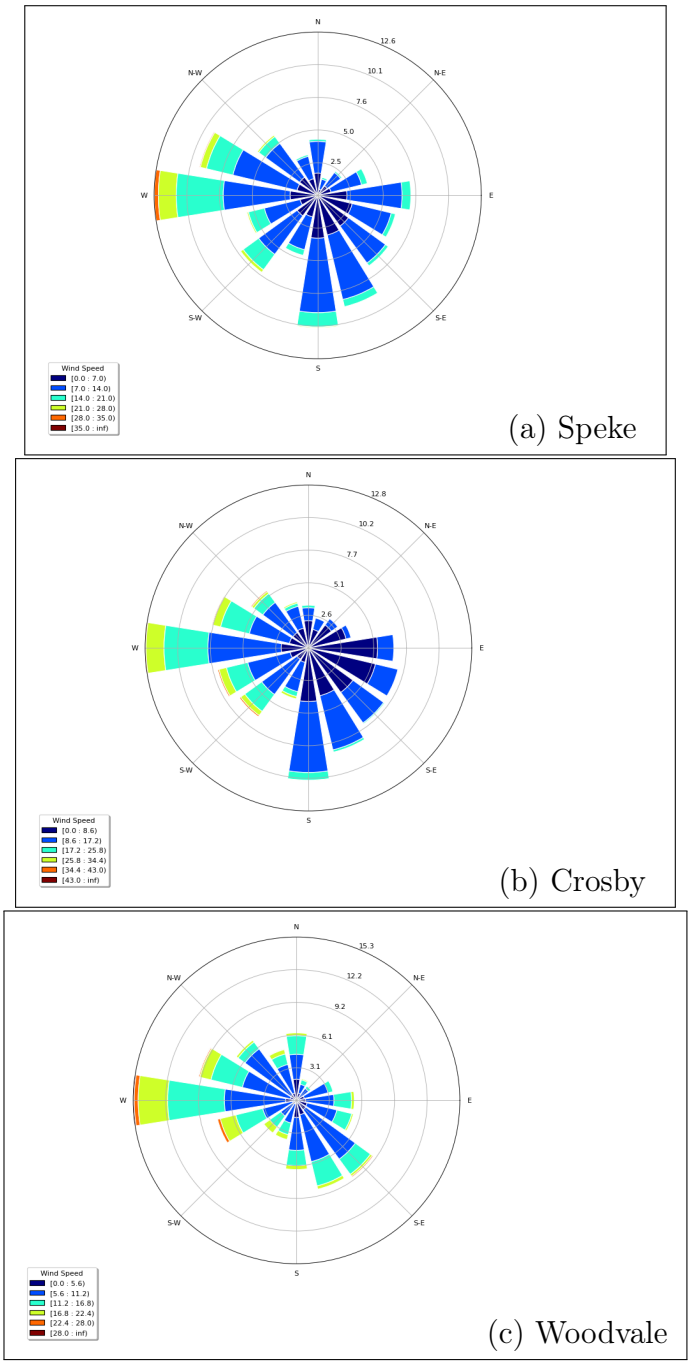

Figure 3.5: Wind Rose diagram showing direction, speed and frequency of wind for each site. The frequency and direction of the wind are fairly comparable in the wind diagrams for the sites Speke and Crosby. In contrast to the other two sites, Woodvale's graphic indicates a similar direction but a higher magnitude of frequency and wind speed. Most of the time, the direction of the wind is West to East.

3.2 Wind Speed Estimation for Gwynt y Môr Offshore

Now that the relationship between existing wind farms and wind speed characteristics is recognised, the next phase of this research will be to estimate the wind speed for the Gwynt y Môr site. There is not a lot of prior work that can be utilised in this context. As a result, the estimation is carried out using the eigenvalue, eigenvector and the correlation between the wind farms (Strickland n.d.). There are different approaches for estimating wind speed for an unknown location based on the missing wind speed data.

One of the simplest methods is to assume that the wind speed at Crosby and Gwynt y Môr sites are the same. This site's close to the Gwynt y Môr site and the fact that it is the location with the fewest discontinuities in the wind speed data, all point to this idea. The major disadvantage, however, is that this method is unreliable since, despite a high correlation between wind farms, certain variations in wind speed characteristics still occur, as can be seen in Section 3.1.

Utilizing all of the information at hand to calculate wind speed is another strategy. This method requires a lot of mathematical computation, which makes it a little challenging.

Step I : First, the mean wind speed for each site is computed. It is denoted by $\bar{s_i}$ where s_i is wind speed at site i ($i = 0, 1, 2$; 0 represents Speke, 1 represents Crosby, and 2 represents Woodvale).

Step II : Considering only times at which speeds are recorded at all the three sites, let h_{ij} be the average value of (s_i - $\bar{s_i}$)(s_j - $\bar{s_j}$) (where i,j = 0,1,2).

Step III : This gives a matrix h, which will have an orthonormal basis of eigenvectors with real eigenvalues. Let λ be the largest eigenvalue. Let u $\in \mathbb{R}^3$ unit eigenvector with eigenvalue λ and with non-negative entries. Compute the first principal component of the speeds at three sites, i.e.,

$$s^{'} = u_0 s_0 + u_1 s_1 + u_2 s_2. \tag{3.1}$$

The aim is to find a single random variable $s^{'}$ which is a linear combination of s_0, s_1, s_2 as represented in Equation 3.1. Remember this is only defined at times at which speeds are recorded at all three sites. It is discussed above that the majority of the data for site Woodvale is missing and because of this reason, estimating wind speed for Gwynt y Môr is done by estimating wind speed using data from site Speke and Crosby which is a more reliable approach. Therefore, Equation 3.1 will be used when wind speed for the three sites is provided and Equation 3.2 will be used when data is missing from any one of the sites.

$$s^{'} = \bar{s}^{'} + \begin{bmatrix} s_0 - \bar{s_o} & s_1 - \bar{s_1} \end{bmatrix} \begin{bmatrix} h_{00} & h_{01} \\ h_{01} & h_{11} \end{bmatrix}^{-1} \begin{bmatrix} \lambda u_0 \\ \lambda u_1 \end{bmatrix} \tag{3.2}$$

The goal is to estimate wind speed at the Gwynt y Môr site as a constant multiple of $s^{'}$, and therefore, it does not make any real difference if it is multiplied by a constant. If all the random variables, i.e., wind speed from nearest on-shore sites, are the same, the above procedure will give $S^{''} = s^{'}/\sqrt{\Lambda}$, where Λ is some constant. Note that for further analysis the value of Λ will not affect the analysis significantly and therefore, it can be chosen randomly. For this research value of Λ is taken to be 3 since it is assumed that the actual wind speed at the Gwynt y Môr site shares the same characteristics as the wind speeds at the other three locations. Now the resulting value, i.e., $S^{''}$ can be used as the estimated wind speed for the Gwynt y Môr site.

Since this method has not yet been studied, it is compared to one of those in the Luo et al. (2008) i.e., Inverse Distance Weighting (IDW). IDW interpolation blends the advocated notion of closeness with a trend surface's steady shift. In comparison to measured values further away, those nearer to the prediction site will have a greater impact on the anticipated value. Numerous attempts have been made to interpolate climatic data using the distance-decay method. According to IDW, every measured point has a little local influence that gets smaller as it gets further away. The expression for IDW is:

$$\hat{Z}(s) = \frac{\sum_{m=1}^{N} W(d_m) Z(s_{im})}{\sum_{m=1}^{N} W(d_m)} = v_0 s_0 + v_1 s_1 + v_2 s_2, \tag{3.3}$$

where $v_i = \frac{W(d_i)}{\sum_j W(d_j))} (j = 0, 1, 2)$, $Z(s)$ is the estimated wind speed at Gwynt y Môr site, s_{im} is the value of wind speed at site i and at timestamp m. N is the total number of observations. $W(d_m)$ is the weighting function defined as:

$$W(d_m) = d_i^{\gamma}(1 + D_{im}), \tag{3.4}$$

where d_i is the distance component i.e., the distance between site i and Gwynt y Môr, γ is the weighting power and D_{im} is the directional component i.e., the direction of wind speed at time m for site i. The optimal value for weighting power as suggested by Legates (1987) and is chosen as 0.95. It is important to keep in mind that most of the data for the site $i = 2$, or for the Woodvale site, is missing. Therefore, when the data for Woodvale is missing, data from only two locations, namely Speke and Crosby, will be utilised to estimate the wind speed for the Gwynt y Môr site.

It should be noted that the weights given to the wind speed for each site are the sole change between Equations 3.1 and 3.3. As a result, if the results from the two methods differ when being analysed, the weights can be further explored.

3.2.1 Results

By examining the statistical properties of wind speed, the estimations for wind speed at the Gwynt y Môr site can be analysed. The monthly variation in wind speed is depicted in Figure 3.6. The wind speed pattern is the same for both approaches, it should be noted, however, that the Principal Component Analysis (PCA) approach yields far higher wind speed values than the Inverse Distance Weighting (IDW) method. The hourly mean wind speed estimates for both the methods, as shown in Figure 3.7, share the same features. In addition, Figure 3.7 shows that the wind speeds are higher than 8 knots, as they were for the three sites that were provided. Additionally, there is a noticeable disparity between the estimations of the wind speed in the early hours, which decreases in the middle of the day before increasing once more after nightfall. The reason for the divergence can be further examined by examining the weights assigned to wind speed for each site, but that is outside the scope of this dissertation. Finally, it can be noted that Figures 3.6 and 3.7 show that the wind patterns at the three sites are essentially the same as the one estimated for the Gwynt y Môr site.

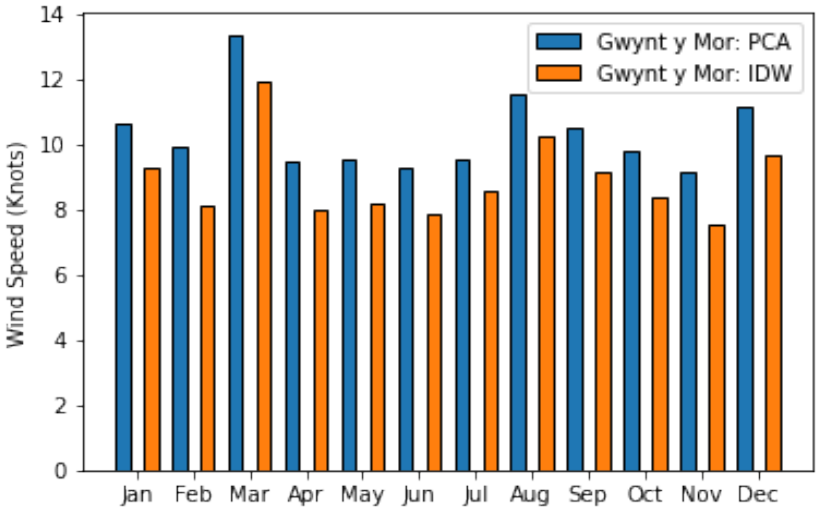

Figure 3.6: Gwynt y Môr site's monthly mean wind speed using two methods. Using both techniques, the maximum mean wind speed is observed in March, and it exhibits the same pattern as it does for onshore sites.

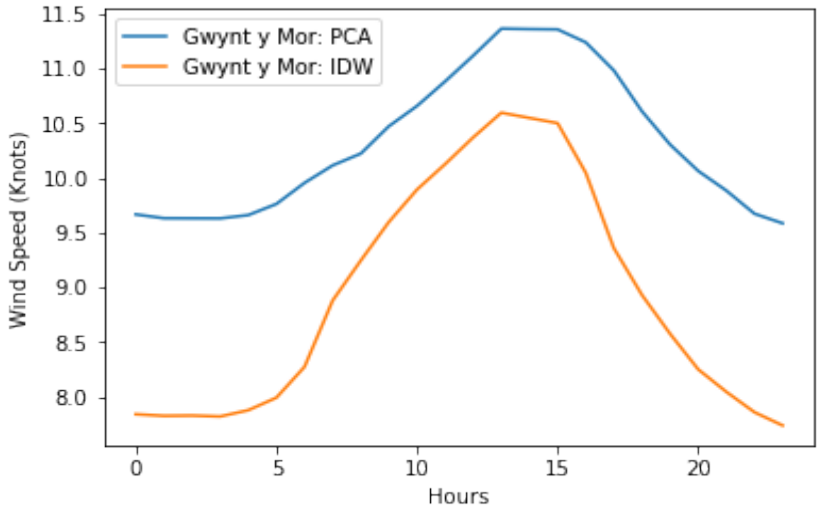

Figure 3.7: Hourly mean wind speed estimates for Gwynt y Môr site using two methods. The estimated hourly mean wind speeds are shown on the graph to be greater than 8 knots, and they follow the same pattern as they do for onshore sites.

3.3 Wind Speed Probability Distributions

This section will cover the various distributions that can be used for fitting wind speed data before actually fitting the optimal distribution to the data given. The statistical properties of the wind speed distribution have been the subject of extensive investigation during the past few years (for example, Morgan et al. (2011), Jamil et al. (1995), Dookie et al. (2018)).

Numerous possible distributions were examined using the wind speed data from four places in the research by Morgan et al. (2011). The bimodal Weibull mixture distribution (BIW), which is a combination of two Weibull (two-parameter) distributions, provides the highest R^2, according to the given data. In addition, the generalizations of various distributions i.e., Kappa (KAP) and Wakeby (WAK) distributions also provide a good fit to the wind speed data. Later, a comparison between BIW and W2 was conducted, and it was discovered that BIW was leading in terms of goodness of fit at all locations. The paper's conclusion refers to the fact that various aims call for the best distribution in that application and additional analysis are required.

Another study by Carta et al. (2009) examines and analyses the effectiveness of 12 distribution models. The study concluded that W2 is one of the main choices for wind speed because it is simple to fit and only requires two parameters for the data. Using the W2 distribution is made simpler by the fact that it can be represented in a closed form (i.e., parameters are finite). The study also discussed some of the

limitations of applying W2, such as the inability to use it with large proportions of missing wind speed data. It was further mentioned that the three-parameter Weibull distribution can be used to get around the W2 distribution's disadvantage.

The W2 and Rayleigh distributions are compared and fitted to the wind speed data in a paper by Jamil et al. (1995), and the authors concluded that while both distributions provide a respectable fit to the data, they are unreliable when the mean wind speed values in the actual wind speed data are low. Garcia et al. (1998) also supported the problem raised in this research. In this study, W2 and Lognormal distributions were examined. Two sites with low mean wind speeds were tested, and it was found that the Lognormal distribution performed better there than it did at the other sites with higher means, where the W2 distribution provided a better fit.

Contrary to all of these papers, the W2 and Rayleigh distributions did not provide a satisfactory fit when they were applied to two locations in Trinidad and Tobago by Dookie et al. (2018). Even though Rayleigh outperformed W2, the results were still not consistent.

One of the main drawbacks of W2 is that it performs poorly at low wind speeds, however, this is not an issue in our scenario because curtailment only occurs at extremely high wind speeds, leading to excess power generation. The results of all the investigations considered further demonstrate that, despite certain limitations, W2 provides a better fit for the wind speed data and has many advantages over competing models. Therefore, it can be stated that W2 distribution will be used in this research.

3.3.1 Weibull Distribution

In statistical analysis, the Weibull distribution is frequently utilised. It is used to describe the wind speed probability density function, also known as the wind speed distribution, in wind energy analysis (Seguro & Lambert 2000). Researchers working on wind speed analysis around the world almost universally use the Weibull distribution function, and the corresponding Weibull model has also been widely used in wind power analysis (Shu et al. 2015). The two-parameter Weibull distribution function is used in this research to evaluate wind speed data. The Weibull distribution function is given by

$$f(s_i, k, c) = \left(\frac{k}{c}\right)\left(\frac{s_i}{c}\right)^{k-1} \exp\left[-\left(\frac{s_i}{c}\right)^{k}\right]; \qquad (s_i > 0, k, c > 0) \tag{3.5}$$

where: s_i represents the wind speed at site i (i = 0, 1, 2), where $i = 0$ represents Speke, $i = 1$ represents Crosby, and $i = 2$ represents Woodvale. $f(s_i)$ represents the probability of the observed wind speed (s) at site i, c represents the Weibull scale parameter (in knots), and k represents the dimensionless Weibull shape parameter.

The related mathematical representation of the cumulative distribution function is

$$F(s_i, k, c) = 1 - \exp\left[-\left(\frac{s_i}{c}\right)^k\right], \tag{3.6}$$

and the quantile (inverse cumulative distribution) function can be defined as

$$Q(p, k, c) = c(-\ln(1-p))^{\frac{1}{k}}, \tag{3.7}$$

for $0 \leq p < 1$.

3.3.2 Estimation of Weibull Parameters

The aforementioned articles made use of a variety of techniques to estimate the Weibull distribution's parameters. Finding an effective method to estimate the parameters is not the core focus of this research because the main objective is to construct a curtailment model. The Maximum Likelihood (ML) strategy, which is effective in earlier investigations (Seguro & Lambert 2000), is applied in this study.

The ML approach can be used to fit the Weibull distribution to time-series wind data. The two equations below are used to estimate the scale factor c and the shape factor k.

$$k = \frac{\sum_{m=1}^{n} (s_{im})^k \ln(s_{im})}{\sum_{m=1}^{n} (s_{im})^k} - \frac{\sum_{m=1}^{n} (s_{im})^k}{n}, \tag{3.8}$$

$$c = \left[\frac{1}{n}\sum_{m=1}^{n} (s_{im})^k\right]^{\frac{1}{k}}, \tag{3.9}$$

where n is the number of non-zero wind speeds, s_{im} is the wind speed value at site i for the time stamp m.

Although it is possible to manually solve these equations by first solving Equation 3.8 iteratively, and then explicitly solving Equation 3.9. In this research it is done using the python function *stats.weibull_min.fit* from the *scipy* library (Virtanen et al. 2020).

3.3.3 Results

The Probability Distribution Function (PDF) calculated using the ML approximation is shown plotted against the wind speed in Figure 3.8. The data is shown as a histogram, whilst the fitted distribution is shown as a line plot. The line plot in the

figure shows the precise pattern generated by the real wind speed data for each site except for the Woodvale site. As was already indicated, one of the problems with the two-parameter Weibull distribution is that it cannot effectively fit data with a high percentage of missing data. This disadvantage can be overlooked because the W2 distribution offers a suitable fit for the Gwynt y Môr location.

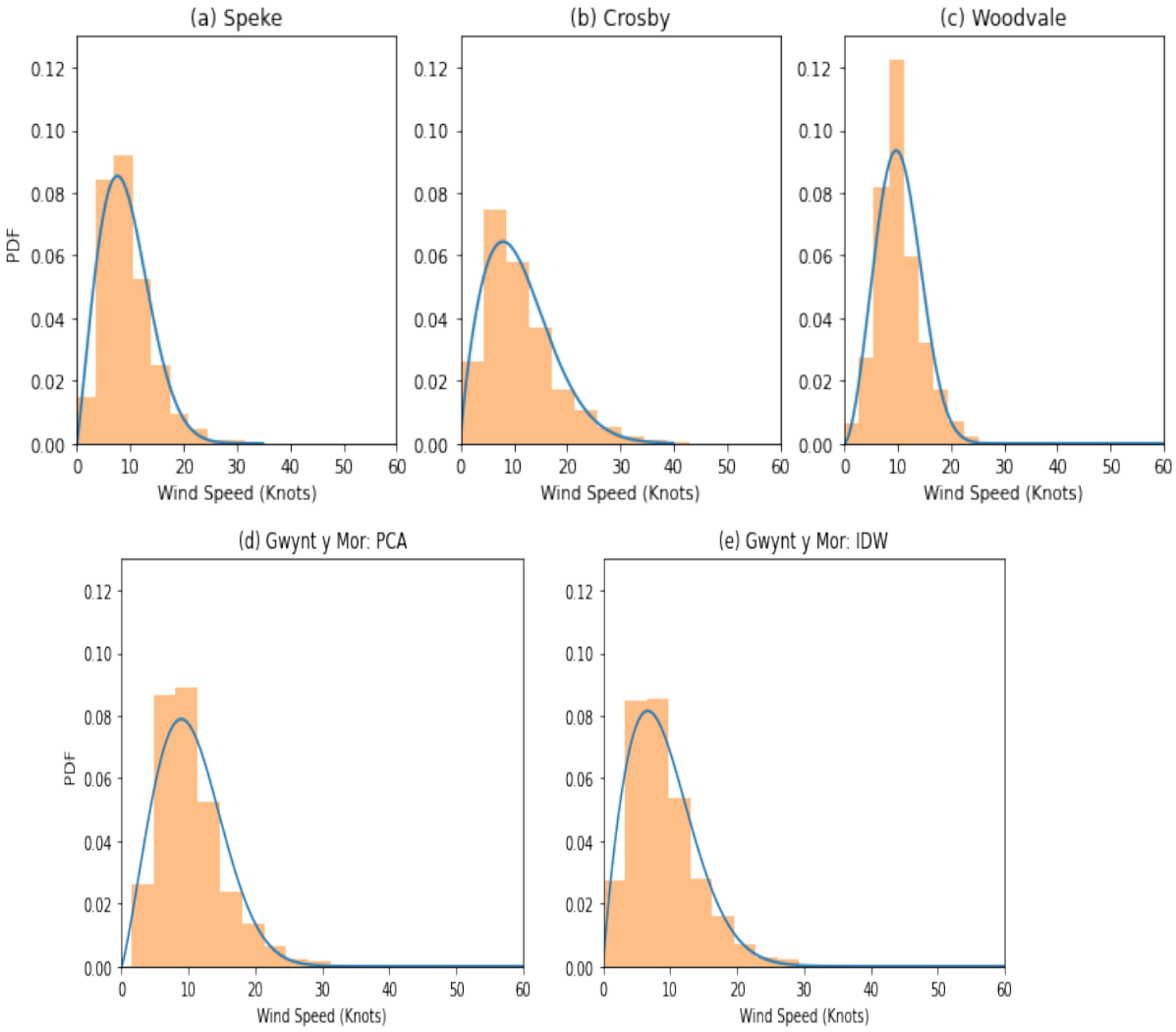

Figure 3.8: Histogram is representing the data and the fitted distribution is shown as a line plot. Weibull distribution appears to be a good fit for each site.

The estimated values for the Weibull parameters are shown in Table 3.2. Estimates of the shape and scale of Weibull distribution do not differ considerably between the various sites. In Table 3.3, the mean and standard deviation for the actual wind speed data and the fitted distribution are compared to further assess the accuracy of the parameter estimates. Both descriptive statistics are similar across all sites.

Therefore, it may be claimed that the chosen distribution can be regarded as a good fit.

Site	Shape (k)	Scale (c)
Speke	2.09	10.35
Crosby	1.80	12.50
Woodvale	2.72	11.57
Gwynt y Môr: PCA	2.22	11.69
Gwynt y Môr: IDW	1.88	10.23

Table 3.2: Weibull parameter estimates for all sites. Between different sites, estimates of the shape and scale of the Weibull distribution do not significantly change.

Site	Observed		Theoretical	
	Mean	Std. Dev.	Mean	Std. Dev.
Speke	9.10	4.68	9.20	4.61
Crosby	11.02	6.59	11.13	6.35
Woodvale	10.26	4.02	10.29	4.07
Gwynt y Môr: PCA	10.32	4.93	10.38	4.93
Gwynt y Môr: IDW	9.06	5.08	9.05	5.02

Table 3.3: Mean and Standard Deviation for the observed wind speed data and the theoretical data. Across all sites, both descriptive statistics are similar.

3.4 Summary

The statistical characteristics of wind speed for every site were examined in this chapter. The wind speed data that was used as the analysis's foundation and the numerous distributions looked at both seem to correspond well with the Weibull distribution. In addition, two methods—the Principal Component Analysis technique and the inverse distance weighting approach—are now used to estimate the wind speed at the Gwynt y Môr location (IDW). The wind speed estimate for the Gwynt y Môr site appears to follow the same trend as the wind speed for the given locations, despite the slight differences in figures between the two methods.

Chapter 4

Wind Speed Modelling

The parameters of the Weibull distribution have been established so far after an extensive study of the distribution of wind speed data. The following step in the research is to create a wind speed model, which will be used to predict the wind speed data for the year 2020 and addressed in this chapter.

4.1 Wind Speed Model

To develop a wind speed model, researchers have looked at a wide range of approaches over the years, some of which are discussed below. The prediction approach, according to Lei et al. (2009), may be broadly divided into two categories: the *physical method*, which makes extensive use of physical factors to achieve the best prediction precision, and the *statistical method*, such as the ARMA model. The results are reviewed by studying the characteristics of the projected wind speed because, in both situations, the prediction cannot be made with absolute certainty due to the intermittent nature of wind speed. The paper also discusses recent advances in artificial intelligence techniques, such as the widely used artificial neural network (ANN), as well as hybrid models, which are of course more sophisticated and error-free than earlier techniques. This section will evaluate various strategies, and the best method will then be applied to model the wind speed data for the Gwynt y Môr location in 2020.

4.1.1 Time Series Model

Due to the interest in wind as a potential alternative source of energy, time series modelling of wind speed has been the focus of numerous discussions. Box Jenkins has classified the time series model into the following categories: moving average model (MA), autoregressive moving average model (ARMA), and autoregressive

integrated moving average model (ARIMA). Based on historical data, the model accounts for the high autocorrelation and enables the creation of a time series that maintains all of the key features of the original data (Kamal & Jafri 1997).

The time series model can be defined as:

$$s_t = \sum_{i=1}^{n} \varphi_i s_{t-i} + \sum_{j=1}^{m} \theta_j \alpha_{t-j} \tag{4.1}$$

where φ_i is the autoregressive parameter, θ_j is the moving average parameter, α_t is the normal white noise, and s_t is the value of wind speed at time t. Equation 4.1 defines a typical ARMA model, and if m is equal to zero, it will reflect the AR model. A model for the ARIMA algorithm will be defined by the ordered differential equation of Equation 4.1. The series stationarity is used as the basis for choosing the model. If the series is stationary, the ARMA model is chosen; otherwise, the series is differentiated d times to achieve stationarity, in which case the ARIMA model is applied.

There are several benefits to employing the Box Jenkins technique for modelling wind speed data, according to Blanchard & Desrochers (1984). One of them is that it doesn't presume a probability distribution for the wind speed data and is easily adaptable to data from other places. However, Wang et al. (2018) points out several drawbacks of this approach. It claims that when the data is non-linear, the ARIMA model is unable to effectively predict wind speed. In addition, this technique presupposes that the interference sequence is white noise and is only appropriate for steady data. However, according to Sfetsos (2000), the ARIMA model when compared to Artificial Neural Networks (ANNs), its performance will increase the accuracy.

4.1.2 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) have grown in significance for modelling wind speed data throughout time. Numerous processing components known as neurons which are also the fundamental processing unit of ANNs are arranged in successive layers in Artificial Neural Networks. The input layer is where data enter the network, while the output layer is where the finished result is displayed. The output of an ANN as a function of its inputs is portrayed as:

$$y_i = \phi \sum_{j=1}^{n} w_i x_j - \theta \tag{4.2}$$

where x_j represents the j^{th} input signal, w_j represents the weight assigned to it,θ

represents a threshold, and $\phi(.)$ represents a sigmoid activation function. The neurons in a multilayer perceptron (MLP) network are arranged in layers without any lateral or feedback connections. Hidden layers refer to layers of neurons that are not part of the output layer. A group of sensors that merely offer input signals and no computations make up the input layer. By changing the weights of the connections between neurons, MLP is learned. Until the error function reaches a predetermined value or the weights stop changing, an input-output pair is presented and the weights are updated iteratively (Mohandes et al. 2004).

It is stated in Bilgili et al. (2007) that neural networks learn from training data (which in this case is previous values of hourly wind speed) and that no mathematical model, or a priori, needs to be specified. Besides that, neural networks have a higher tolerance for data errors and are better able to respond to real-time measurements. However, the paper also notes that it requires a training process and a substantial amount of training data, making the method complex. When Mohandes et al. (2004) compares the method to the Support Vector Machine (SVM) algorithm, it turns out that SVM outperforms the accuracy of MLP.

4.1.3 Support Vector Machine

The Support Vector Machine for Regression (SVR) belongs to a group of learning models that can be trained using convex optimization methods. It works by performing a nonlinear mapping of the data x into a high-dimensional feature space and then doing a linear regression in this feature space. In the end, the technique produces a sparse prediction function by picking just a few support vectors, or training points. It is thus very appropriate for modelling and forecasting wind speed. The SVM regression function is defined as:

$$y = w^T \varphi(x) + b \tag{4.3}$$

where $\varphi(x)$ is the nonlinear mapping function and the w and b are coefficients that are estimated by minimizing the regularised risk function.

In Yu et al. (2018), the Support Vector Machine for Regression technique is compared to numerous hybrid methods (i.e., combinations of two or more models), and the results show that while SVM outperforms ANNs, it performs best when combined with other models, such as neural networks. Despite having a good generalization performance, according to Tascikaraoglu & Uzunoglu (2014), SVM has several drawbacks, including a need for careful parameter adjustment, a difficult optimization procedure, and a lengthy training period. In order to get accurate results, Mohandes et al. (2004) divided 12 years of previous data into three parts: one for training the model, one for validating the variables, or choosing which variables to use that best define the wind speed, and one for testing data, which was

used to assess the model's accuracy. Only one year's worth of wind speed data is available for this study, and aside from that, additional factors like wind direction, site height, and numerous other variables are also not available. Due to all of these constraints, using SVM to forecast or model the data on wind speed is challenging.

One thing to keep in mind is that all of the approaches discussed above are often used to forecast wind speed for brief periods. These techniques can be used to forecast wind speed in this study, although for this study, forecasting wind speed has the drawback of just having data for 2019, which is insufficient for reliable conclusions. Due to this data shortage, modelling wind speed is favoured in this research instead of making predictions for the Gwynt y Môr site in 2020. For modelling wind speed data, Stochastic Difference Equations approach will be used in this paper.

4.2 Stochastic Difference Equations

A broad and adaptable modelling framework is needed to account for any potential non-linearity in the wind speed data and benefit from the randomness. Stochastic Differential Equations are generated as another popular stochastic modelling framework to model wind speed. Stochastic differential equations are frequently used in a range of scientific, technical, and financial sectors to model stochastic occurrences. It can be thought of as extensions of deterministic differential equations in which the main dynamics are combined with stochastic processes or random parameter evolution. The development of wind speed trajectories, wind speed forecasts, Monte Carlo simulation, and wind speed scenarios are some of the methods utilised in these fields (Zárate-Miñano et al. 2013). For this paper, instead of using the Stochastic Differential Equations, Stochastic Difference Equations (SDEs) is used.

With the use of the proposed model, wind speed trajectories with comparable statistical characteristics to those shown in the wind speed data recorded at the Gwynt y Môr location are expected to be produced. The created model accounts for the wind speed's autocorrelation function and marginal distribution. The two-parameter Weibull distribution is implemented in the model creation to demonstrate the suggested approaches.

The starting point is SDE:

$$y_m = \alpha y_{m-1} + \beta x_m, \qquad m \in [0, T] \tag{4.4}$$

where $x \sim N(0, 1)$ (is a sequence of random variables representing environmental stochasticity), $y_0 = x_0$ and y_m is seen at discrete time intervals, $m \in [0, T]$. Also, α and β are the drift and diffusion terms of the SDE, respectively. Autocorrelation parameter θ is used to obtain the values of α and β. They can be defined as:

$$\alpha = e^{-\theta\tau} \tag{4.5}$$

and

$$\beta = \sqrt{1-\alpha^2} \tag{4.6}$$

To obtain the estimate of θ, it is important to first estimate the parameters of Weibull distribution for the wind speed data at the Gwynt y Môr site. Using these parameters, the next step is to define the cumulative distribution function (CDF), i.e., $F(s_i)$ and then substitute $g_i = \Phi^{-1}(F(s_i))$. This will help in obtaining the sample correlation between g_{im} and $g_{i(m+r)}$, where $g_{im} = \Phi^{-1}(F(s_{im}))$ and s_{im} is wind speed for site i at timestamp m, for some range of values of r. Now the value of θ can be obtained by equating the resulting sample correlation to $e^{-\theta\tau r}$, i.e.,

$$cor(g_{im}, g_{i(m+r)}) = e^{-\theta\tau r} \tag{4.7}$$

where ideal value of $cor(g_m, g_{m+r})$ can be approximated using the following equation:

$$cor(g_{im}, g_{i(m+1)}) = cor(g_{im}, g_{i(m+r)})^{r/2} \qquad r = 2, 3, 4, ..., 10 \tag{4.8}$$

Note: In order to obtain $\Phi^{-1}(.)$, some approximations are required as shown in A.

To obtain an autocorrelated Weibull distributed process, the following transformation is used to the trajectories y_i:

$$s_i = z[y(m_i)] = F_w^{-1}[\Phi(y_i)], \tag{4.9}$$

where $F_w[.]$ represents the Weibull Cumulative Function, i.e.,

$$F_w[u] = 1 - exp\left[\left(\frac{\mu}{c}\right)^k\right], \forall \mu > 0, \tag{4.10}$$

$c > 0$ and $k > 0$ are the Weibull distribution's scale and shape parameters, respectively, and $\Phi[.]$ is the Gaussian cumulative distribution function, i.e.,

$$\Phi\left[\frac{\mu - E[\mu]}{\sqrt{Var(\mu)}}\right] = \frac{1}{2}\left(1 + erf\left[\frac{\mu - E[\mu]}{\sqrt{2Var(\mu)}}\right]\right), \forall \mu \in \mathbb{R} \tag{4.11}$$

The resulting process s_i is the simulated wind speed at the Gwynt y Môr site, and can also be defined as Weibull distributed autocorrelated stochastic process at a time is τ with the following statistical properties:

$$E[s_i] = \mu_w = c\Gamma\left[1 + \frac{1}{k}\right], \tag{4.12}$$

$$Var[s_i] = \sigma_w^2 = c^2\Gamma\left[1 + \frac{2}{k}\right] - \mu_w^2, \tag{4.13}$$

$$Aut[s_i, s_j] \approx \exp\left[\alpha|m_j - m_i|\right], \qquad \forall m_i, m_j \in [0, T], \tag{4.14}$$

where μ_w and σ_w^2 are the mean and standard deviation of Weibull distribution with parameters c and k respectively.

4.3 Results

The convergence error, as well as consistency and numerical stability, of a numerical integration method for SDEs, depends on the integration step length r, much like in the deterministic case. To produce a good approximation to the real trajectories and the statistical features of s_i while retaining a manageable computational cost, a suitable integration step length r must be determined.

Usually, hourly or averaged hourly measurement data are used to calculate the marginal distribution and the autocorrelation of the wind speed. Autocorrelation values are therefore normally calculated for time delays stated in hours. However, in this instance, the autocorrelation components must be adjusted to the half-hourly period. The reason for this is that 30-minute intervals of data on curtailments are available. The proposed approaches directly implement this adaptation by scaling the associated hourly-based value of the parameter α appropriately.

The parameters of a Weibull fit and the 30-minute autocorrelations of the Gwynt y Môr site's estimated wind speed data (using the *Principal Component Analysis method* mentioned in Section 3.2) serve as examples of how well the generated models function. Parameters k, c, and α in Equations 4.4 and 4.9 for the model are set according to the data in Table 4.1.

Location	c	k	α	Correlation
Gwynt y Môr	12.07	2.19	0.9034	0.9505

Table 4.1: Weibull parameters and 30-M correlation of modelled wind speed data.

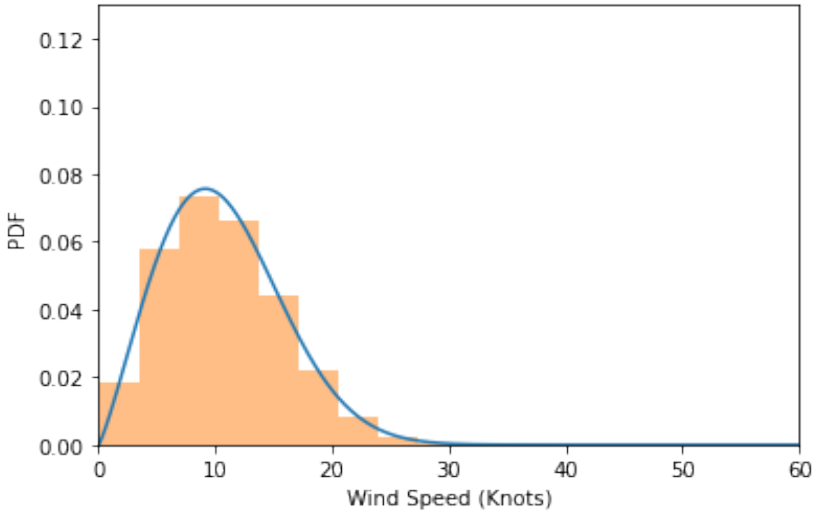

Figure 4.1: Histogram is representing the data and the fitted distribution is shown as a line plot. Graph shows that the Weibull distribution is a good fit for the modelled wind speed data.

Figure 4.1 shows the Weibull Probability distribution functions and the probabilities of the process generated by the developed model. As is observed in the figure, the wind speed data generated by the developed model follow the corresponding Weibull distribution. This is also corroborated by Table 4.2, where the mean and the standard deviation of the wind speed trajectories generated by the developed model are compared with the expected theoretical mean (Equation 4.12) and standard deviation (square root of Equation 4.13) of the Weibull distribution, respectively. Figure 4.2 displays the correlation between the wind speeds produced by the developed model (the orange line). To illustrate the exponential decay of the correlation, the graphic also displays the equation $dN/dt = -\lambda N$ (the blue line), where N is a number from 1 to 11, and λ is a positive rate equal to 0.1. It can be observed that the established models' correlation of the data on wind speed is close to exponential decay.

	Mean	Standard Deviation
Weibull Distribution	10.7560	5.1546
Generated Process (s_i)	10.6883	5.1438

Table 4.2: Descriptive statistics of the Weibull distribution compared to the descriptive statistics of the generated process. The statistics are approximately same for both the series.

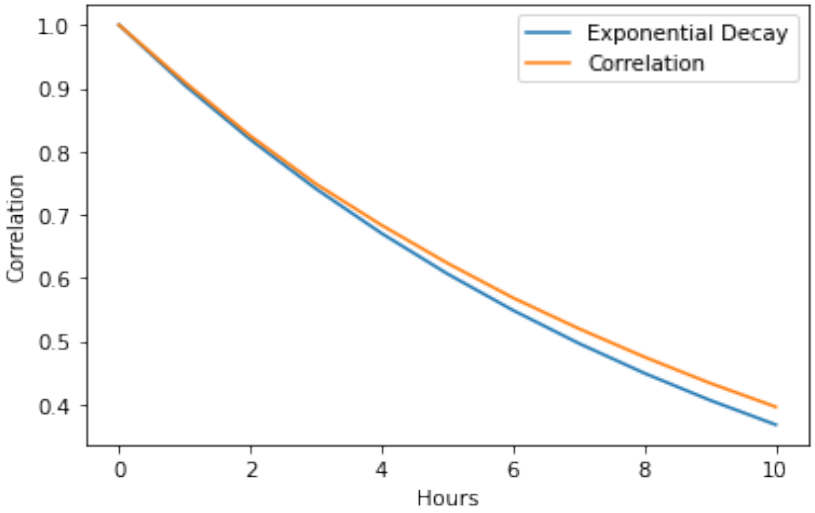

Figure 4.2: Correlation of the created process throughout the first 10 hours compared with the exponentially decaying series. The graph demonstrates that the modelled wind speed data followed the characteristics of exponential decay of correlation.

Now that the wind speed model's functionality has been established, it can be further examined to see if the generated series exhibits the same wind-speed characteristics as those predicted for the Gwynt y Môr sites. Hourly and monthly mean wind speeds can be investigated to establish this. The estimated wind speed is compared to the modelled wind speed at the Gwynt y Môr site, across the months and the hours in a day to check if they are maintaining the same trend. Therefore, to illustrate the resemblance, Figures 4.3 and 4.4, respectively, show the monthly mean and the hourly mean for the estimated wind speed and the modelled wind speed at the Gwynt y Môr site. (Note: Here, estimates made using the Principal Component Analysis method are compared with the modelled wind speed.)

Figure 4.3 shows some variances in the values, but they are relatively infrequent and the pattern as a whole is identical. The hourly mean wind speed for the modelled data in Figure 4.4 is not as smooth as it is for the estimated data for the Gwynt y Môr site. It is debatable, though, if the general pattern used here is the same. Although it might be argued that the modelled wind speeds are not particularly realistic, they are nonetheless intended to represent the Gwynt y Môr site's wind speed structure.

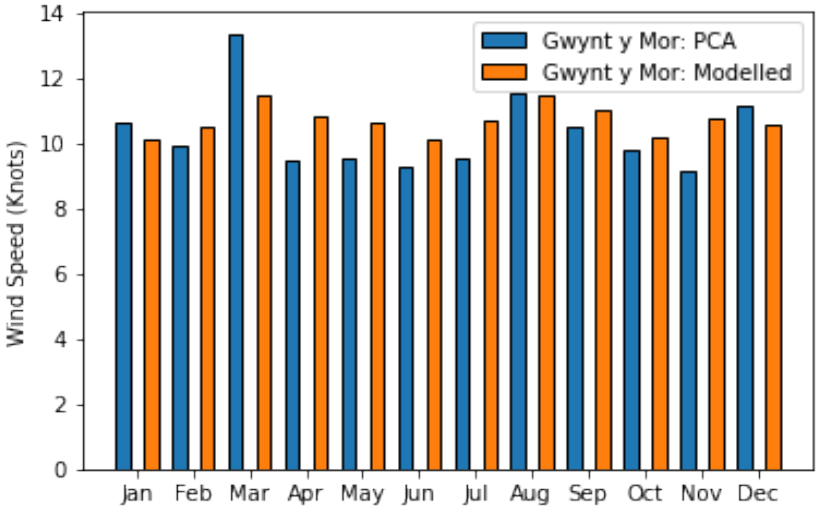

Figure 4.3: The monthly mean wind speed model for the Gwynt y Môr site was compared with the estimated wind speed. For both series, the pattern appears to be similar.

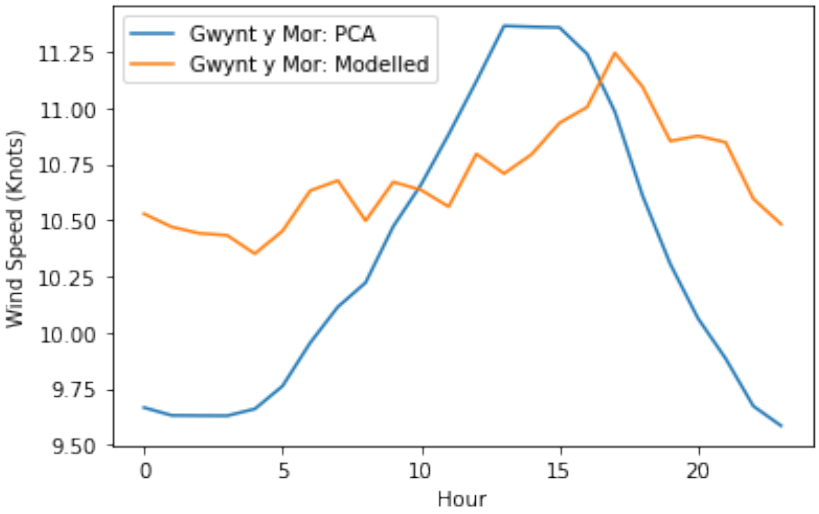

Figure 4.4: The monthly mean wind speed model for the Gwynt y Môr site was compared with the estimated wind speed. The hourly mean wind speed for the Gwynt y Môr site is not as smooth as it is for the estimated data.

4.4 Summary

This chapter compared different approaches to predicting and modelling wind speed. Modelling wind speed is preferable to forecasting since for this study, the year 2019 is the only year for which data on wind speed is available. Stochastic Difference

Equations are employed to achieve this. Although there is some variation between the estimated and modelled wind speeds, it is acceptable because the wind speed model simply replicates the characteristics of the data and not the actual numbers.

Chapter 5

Curtailment Model

Earlier chapters introduced the ideas of wind energy, wind power, and wind energy curtailment. Thus far, the wind speed for the Gwynt y Môr site has been calculated, and a wind speed model employing Stochastic Difference Equations is also created. The Gwynt y Môr site's wind speed data will be combined with the existing curtailment data in this chapter. These data were also examined in Chapter 2. Additionally, as the final step in this dissertation, the new combined data will aid in developing the curtailment model for the Gwynt y Môr site.

5.1 Curtailment

The balancing system is a collection of economic and technological organizations that keeps the electricity system's active power balance over seconds to hours. To keep the system frequency steady, supply and demand must always match with extreme precision. A frequency deviation that exceeds tolerance thresholds caused by an imbalance of more than a few per cent over a short period may harm generating machinery and cause system failure. The wholesale market, in theory, maintains this equilibrium since generators only create as much electricity as they have sold to suppliers, who contract electricity based on demand. But due to inaccurate projections and unforeseen circumstances like the shutdown of power plants or lines, deviations from agreed-upon positions are unavoidable. The system operator, who contracts for reserve capacity to be used in real-time, manages and corrects the resulting imbalances. Grid fees and/or imbalance prices (penalties for deviating from the schedule) are typically used to recover the cost associated with these procedures (Joos & Staffell 2018).

In the UK, wind farms can "bid" to lower their energy production through the Balancing Mechanism (BM), which regulates wind energy curtailment. If these offers are accepted, the wind farm will be reduced in size and produce less energy over-

all. Offshore wind farms are referred to in the BM as Balancing Mechanism Units (BMUs). To participate in the balancing mechanism, wind farms must volunteer to reduce their output at the expense of lost revenue from the sale of electricity and Renewable Obligation Certificates (ROCs). Onshore wind farms only receive 0.9 ROCs per MWh, whereas offshore wind farms receive 2 ROCs per MWh, which allows them to demand larger compensation and experience fewer curtailments. Gwynt y Môr is depicted in the BM, with the volume of accepted curtailment and the energy produced for each BMU provided for each 30-minute settlement period for the years 2019 and 2020.

The Balancing Services Use of System (BSUoS) charge is used to communicate all expenses associated with managing the power system, including reserve availability and usage payments, reactive power and black start capabilities, and accepted bids and offers in the balancing mechanism. By dividing the expenditures incurred by the volume of electricity used and fed during the settlement period, the charge is computed for each settlement period. The BSUoS charges in Britain have been rising steadily in recent years, from £746 million in 2010 to £1207 million in 2016, which is an increase of 62%. In addition, there are environmental costs associated with reductions. There is no method to reintroduce the curtailed energy when there is a positive residual load if there is an excess of power available. As a result, curtailment raises fuel consumption and emissions from generation at traditional power plants. The running costs of the system as well as external costs like NOx emissions that are not covered by the EU Emissions Trading System are negatively impacted by this. And therefore, it is important to model wind power curtailment to reduce the cost.

5.1.1 Data Combination

Joos & Staffell (2018) suggests several ways to lower balancing expenses, which can be classified as follows: (i) better forecasting; (ii) better spot markets; and (iii) better balancing management and markets (including European integration). The latter two points cannot be worked on by statisticians, although forecasting can be enhanced by modelling curtailment using the data that is currently available for the Gwynt y Môr site. To do this, the anticipated wind speed data and the curtailment data for the Gwynt y Môr site must first be combined. The indirect dependence of wind power curtailment on the erratic behaviour of wind speed makes this step crucial. The only problem in combining the data is that curtailment data is available at half hourly whereas the wind speed data is available hourly. To overcome the issue, the linear interpolation method is used with the help of the *interpolate* function present in the *scipy* library in *Python*. Later the data is combined with the curtailment data using the *merge* function from the *pandas* module.

5.2 Curtailment Model

In this section, the work of the curtailment model will be explained in detail. However, before starting with the model, it is important to analyse the parameters that will be used in the model and also introduce some of the terms that will be used.

5.2.1 Energy Produced (%) and Wind Speed (Knots)

The relationship between curtailment and wind power as well as the relationship between curtailment and wind speed can now be examined since the data for the Gwynt y Môr site have been combined. The wind speed data is an important factor in developing a curtailment model. It is necessary to analyse their relationship before creating an accurate model. For analysing the relationship between the factors, the scatter plots are displayed in Figures 5.1 and 5.2. This was created using the *scatter* function present in the *seaborn* module in *Python* (Waskom 2021).

Figures 5.1 and 5.2 show a positive relation between curtailment with both the variables i.e., the energy produced (in %) and wind speed (in knots). Figure 5.1 shows that the curtailment (%) increases more than 20% when the energy is produced more than 60%, below that, curtailment (%) is approximately 0% only. From Figure 5.2, it can be seen that when wind speed is less than 10 knots, there is very little curtailment. The curtailment (%) is maximum when wind speed lies between 10 knots to 30 knots. When the wind speed is more than 30 knots there is no evidence of curtailment (%).

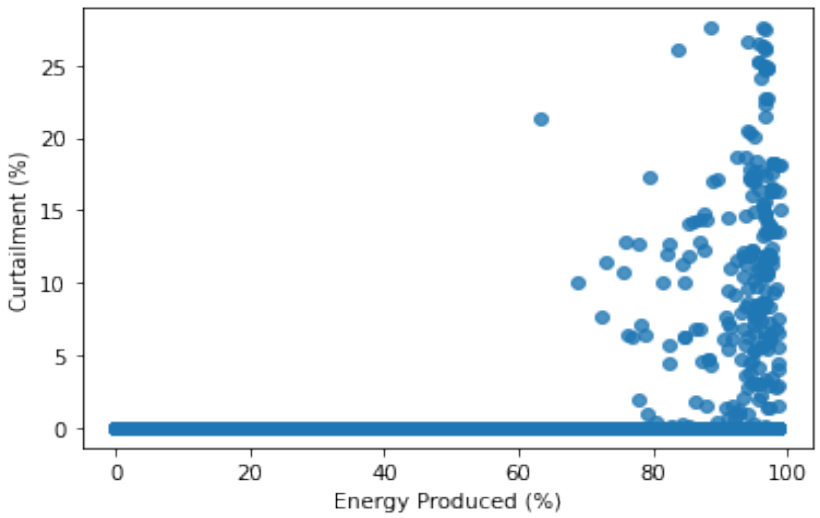

Figure 5.1: Curtailment (%) against energy produced (%). Curtailment (%) increases when the energy produced is more than 60%.

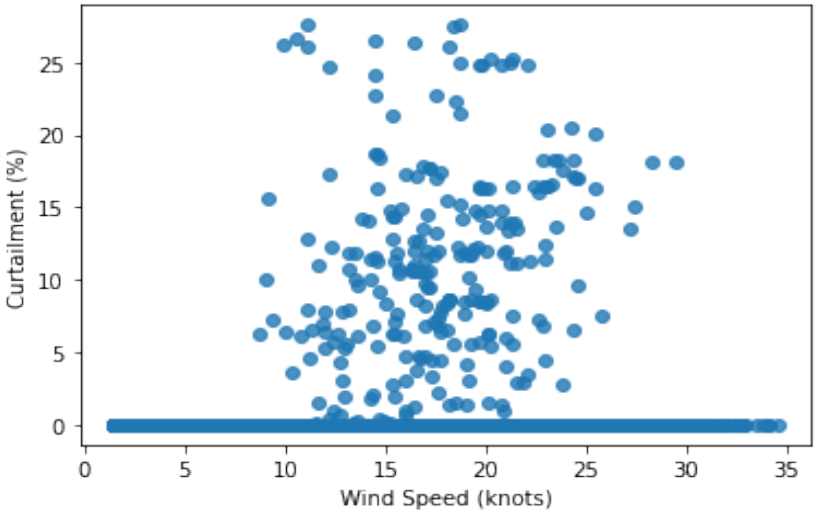

Figure 5.2: Curtailment (%) against wind speed (knots). Curtailment is usually high when wind speed is between 10 to 25 knots.

5.2.2 Curtailment Bins

The examination of the curtailment data is given in Section 2.2. The investigation revealed that the data was frequently not constrained at all as can also be seen in Figure 5.3. To ease the curtailment model, the curtailment percentage can be divided into 4 bins: 0%, 1%-10%, 11%-20% and 20%-100%. The bins are numbered as 0,1,2,3 and are denoted by b_t, i.e., curtailment bin at time t. Figure 5.4 displays the bar plot showing the frequency of each bin. As it can be seen in the figure, bin 0 i.e., curtailment at 0% is still maximum whereas the rest of the categories are consistent.

5.2.3 Transitioning Probability

Transitioning probability can be defined as the observed probability of shifting from bin d to bin l where d is curtailment bin at time t, i.e., b_t and l is curtailment bin at time $t+1$, i.e., b_{t+1}. This is denoted by p_{dl}. For example, p_{12} defined the probability of transitioning from curtailment bin 1 to curtailment bin 2 where curtailment bin 1 covers 1% to 10% of curtailment whereas curtailment bin 2 contains curtailment from 11% to 20%.

To compute the transitioning probability, another term is introduced that will take into account the number of times a value in bin d is followed by one in the bin l. This can be denoted by n_{dl}. For example, n_{12} is computed by calculating the number of times the curtailment bin is shifted from 1 to 2.

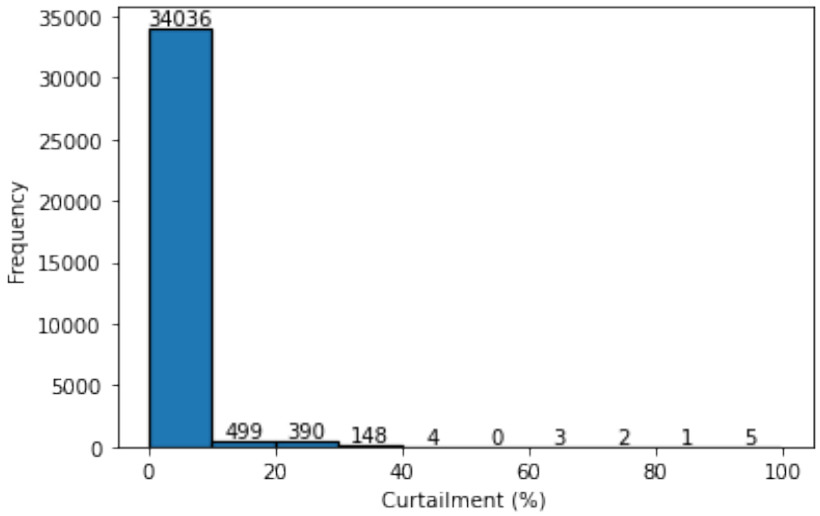

Figure 5.3: Histogram exhibiting the frequency of curtailment (%). Most of the time, there is no energy curtailed.

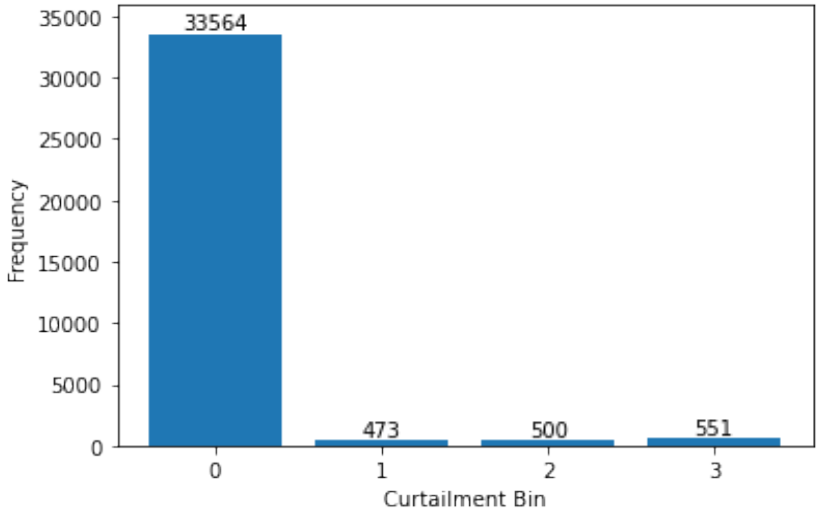

Figure 5.4: Barplot representing the counts of each curtailment bin. The graph shows that the bin 0, or curtailment at 0%, is now maximum, while the other categories are currently at least consistent.

Therefore, transitioning probabilities can be defined as

$$p_{dl} = \frac{n_{dl}}{\sum_{k=0}^{3} n_{dk}}, \tag{5.1}$$

and the transitioning matrix will be

$$\pi = \begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} \\ p_{10} & p_{11} & p_{12} & p_{13} \\ p_{20} & p_{21} & p_{22} & p_{23} \\ p_{30} & p_{31} & p_{32} & p_{33} \end{bmatrix}. \tag{5.2}$$

The Markov Chain approach, which incorporates the transitioning matrix (π), can be used to develop a curtailment model, although it has the limitation of not considering site-specific wind speed. Prof. Neil Strickland proposed certain changes that have been made to the model and are explained below to address this problem.

To take wind speed into account, the curtailment model can be defined as

$$\mu_l = a_{dl} + b_{dl} \log(s), \qquad \forall d, l = 0, 1, 2, 3, \tag{5.3}$$

where $\log(s)$ is the log of wind speed values for the site, the current curtailment bin is d i.e., the value of curtailment bin at time t (b_t), a_{dl}, and b_{dl} are parameters defining the weights associated to curtailment bin and wind speed respectively. Later p_l is defined as the probability of transitioning to bin l at the next time step, i.e., at time $t + 1$, as defined in Equation 5.4.

$$p_l = \frac{\exp(\mu_l)}{\sum_{k=0}^{3} \exp(\mu_k)} \tag{5.4}$$

For the model defined in Equation 5.3, it is important to choose the optimum value of the weights, i.e., a_{dl} and b_{dl}. To achieve this, adjustment of parameters is necessary to maximise the likelihood of the observed data.

5.2.4 Neural Networks

In this paper, Artificial Neural Networks (ANNs) are utilised as the method for maximising the parameters in Equation 5.3. Subsection 4.1.2 provides a basic overview of ANNs.

The first step in defining the ANNs is to define the input layer. For this model, the input layer will be defined as a combination of values of curtailment bin and wind speed value. This can be achieved by encoding the state of the observed data at

time step t as a vector, $x_t \in \mathbb{R}^8$. If the curtailment bin is d and the wind speed is s at time t, then x_t should have 1 in position d and $\log(s)$ in position $d+4$ and should otherwise be zero. For example, if at time $t = 1$, curtailment is 0%, which implies $d = 0$, i.e., curtailment bin is 0 at time 1, and (if say) speed at that time is, $s = 15$ knots, then in this case $x_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & \log(15) & 0 & 0 & 0 \end{bmatrix}$.

Once the input layer is defined, the next step is to encode the desired output at time step t at a vector $y_t \in \mathbb{R}^4$. If the curtailment bin is l at time $t+1$, then y_t should have 1 in the position of l and should be zero otherwise. For example, if at time $t = 2$, curtailment is 6% then curtailment bin is 1 at time 2 which in other words can be said as $l = 1$ at time $1+1$. In this case y_t will be $\begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}$.

The last step is to set up a neural network with a single dense layer with 8 inputs, i.e., $x_t \in \mathbb{R}^8$ and 4 outputs, i.e., $y_t \in \mathbb{R}^4$ with no bias term and softmax activation. The reason for choosing softmax activation is that in this process, layers will convert inputs to a probability vector of outputs as it is described in Equation 5.3. Once the model is fitted using the observed inputs x_t and outputs y_t, prediction can be done to generate the curtailment bin at time $t+1$ from the curtailment bin and speed at time t. However, in this research, instead of using any prediction function, weights will be obtained that will contain the parameters a_{dl} and b_{dl}, and thus compute the probabilities of transitioning from d to l as mentioned in Equations 5.3 and 5.4.

5.3 Results

Using the Equations 5.1 and 5.2, transitioning matrix for the year 2019 is computed and are denoted by π_{2019}. It will be compared with the probabilities obtained for the year 2020 for different values of wind speed using the Equations 5.3 and 5.4.

$$\pi_{2019} = \begin{bmatrix} 0.996 & 0.224 & 0.024 & 0.004 \\ 0.003 & 0.605 & 0.164 & 0.009 \\ 0.001 & 0.144 & 0.716 & 0.102 \\ 0. & 0.027 & 0.096 & 0.886 \end{bmatrix}$$

The next step is to use the curtailment model as defined in Equations 5.3 and 5.4 to compute the transitioning probabilities and check whether the change in wind speed affecting the probabilities or not. For that, first it is important to establish neural network model mentioned above to compute the weights, i.e., a_{dl} and b_{dl}. *Tensorflow* and *Keras* which are *Python* libraries, are employed for this. Model is defined as:

```
model = tensorflow.keras.Sequential([
    keras.Input(shape= (8)),
    keras.layers.Dense(4,use_bias=False,activation = "softmax")])
```

The model is fitted using data on curtailment and wind speed for the Gwynt y Môr site in 2019. This apart, the number of times the curtailment bin is 0 is extreme compared to other bins, as seen in Figure 5.4. As a result, it is feasible that after the model has been fitted, it will consistently forecast curtailment bin 0 for 2020. This is referred to as the Imbalanced Class Distribution problem. The simplest oversampling technique, known as Random Oversampling, is used to solve the problem by randomly replicating samples from the minority class in the training dataset.

Once the data is fitted, *get_ weights* function is used to get the values of a_{dl} and b_{dl}. The weights are:

$$a_{dl} = \begin{bmatrix} 0.4463 & -0.6140 & -1.2538 & -0.5639 \\ 0.0435 & 0.9022 & 0.3906 & -0.0872 \\ -0.5417 & -0.2987 & 0.8013 & -0.426 \\ -1.0508 & -0.7645 & -0.2286 & 0.8777 \end{bmatrix}$$

$$b_{dl} = \begin{bmatrix} 1.2913 & -0.6176 & -0.1473 & -0.1318 \\ 0.0991 & 0.1787 & -0.222 & -0.269 \\ 0.0178 & 0.4106 & 0.6152 & -0.0861 \\ -0.652 & 0.1048 & 0.0798 & 0.2711 \end{bmatrix}$$

The change in transitioning probabilities as the wind speed changes are seen in Figure 5.5. The transition of the curtailment bin from bin 0 to other bins is shown in Figure 5.5.a. This graph illustrates the fact that, regardless of the wind speed, there is a very high probability of remaining in bin 0 and that the probability of shifting to other bins from 0 is approximately 0. The transitioning probability from bin 1 to subsequent bins is again shown in Figure 5.5.b. It is noticeable from this graph that the probability of moving from bin 1 to other bins does not align with the wind speed. The probabilities of shifting from bin 1 to bin 0 are falling and the probabilities of remaining in the same bin are increasing as wind speed increases. The probability of transitioning from bin 1 to bin 2 ranges from 0.0 to 0.2, while the probability of shifting from bin 1 to bin 3 is nearly zero.

As the wind speed increases, there is a substantial increase in the probability of staying in the curtailment bin 2, as seen in Figure 5.5.c. Additionally, the probability of transitioning from curtailment bins 2 to 3 has decreased considerably when the wind speed is increasing. When the wind speed increases from 0 knots to 5 knots, the probability of shifting from bin 2 to bin 1 first increases, but it later decreases as the wind speed increases. On the other hand, regardless of the wind speed, the probability of shifting from bin 2 to 0 is quite low. Concerning wind speed, the transitioning probabilities from bin 3 to other bins are highly stable. The probability of remaining in bin 3 undisturbed, or not moving from there to

another bin, is maximum, as shown in Figure 5.5.d.

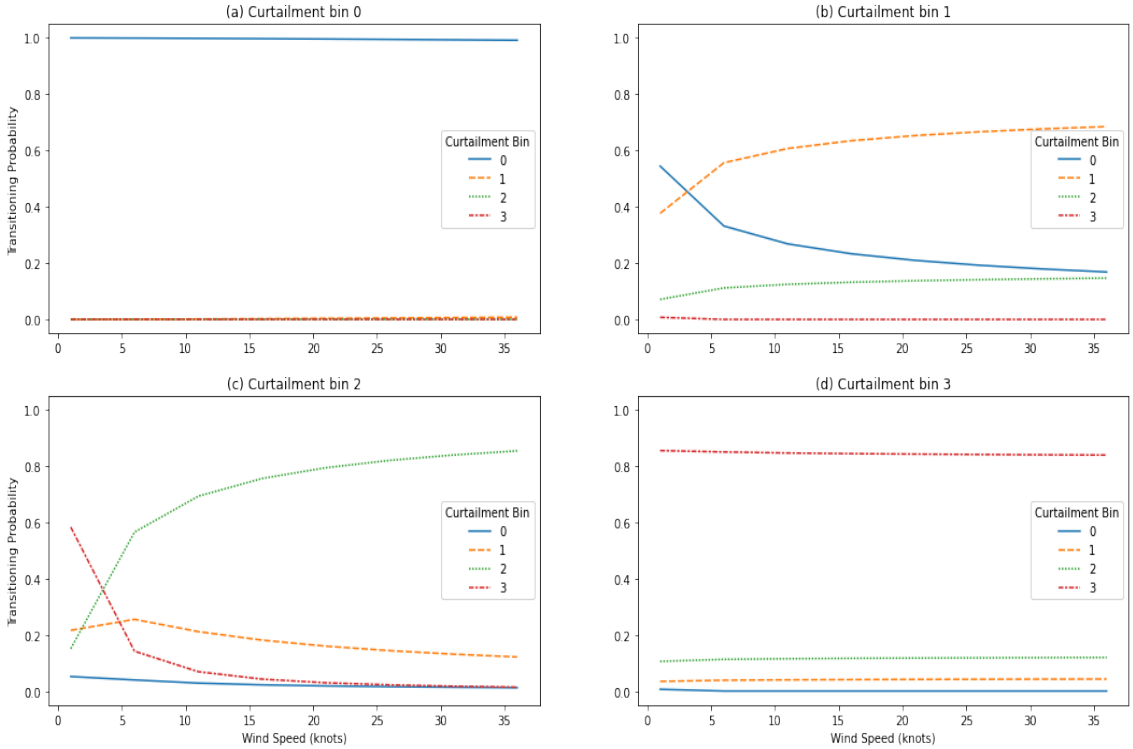

Figure 5.5: Variations in the probabilities of transitions in relation to variations in wind speed. (a) The probability of shifting from bin 0 to another bin. (b) The probability of transitioning from bin 1 to another bin. (c) The probability of shifting from bin 2 to another bin. (d) The probability of transitioning from bin 3 to another bin.

Based on the comparison of the transitioning probabilities for the years 2019 and 2020, some probabilities appear to be relatively comparable, while other values do differ. As was previously stated, statistical models are used to illustrate the data's characteristics rather than to predict outcomes. It appears that the features of the transitioning probability are met in this instance since one crucial point to note is that the probability of transitioning from bin d to bin l is quite high when $d = l$ for both years.

To learn the curtailment bin at time $t + 1$, one can use use the *predict* function with the 2020 curtailment bin data at time t. However, as was already mentioned, there are some drawbacks to that. For example, it will take a lot of time to predict the curtailment bin at time $t + 1$ given the curtailment bin at time t for the entire year 2020. In addition, since the training data, i.e., curtailment data for 2019, is not significant, and therefore, the prediction function might not produce accurate results.

5.4 Summary

For the Gwynt y Môr site, the curtailment data and the modelled wind speed data were combined in this chapter. Due to the fact that wind power was typically not curtailed, the data on curtailment (in %) was later categorised into groups. Furthermore, the curtailment model was established using transitioning probabilities, although some adjustments were made because the site's wind speed is not taken into account. In order to calculate the transitioning probability, weights were assigned to the curtailment bin at time t and wind speed data at time t. The weights were calculated using artificial neural networks.

Chapter 6

Conclusion

This study will come to an end by summarising the key findings in respect to the research's goals and the questions. The importance and contribution of the results will also be examined. In this chapter, the study's limitations and recommendations for further research will also be discussed.

6.1 Conclusion

The study's primary goal was to create a wind speed and curtailment model. This was built for the Gwynt y Môr location. Initially, the closest onshore wind farms, Crosby, Speke, and Woodvale, were used to estimate the wind speed data for the Gwynt y Môr location. Before that, however, the wind speed characteristics for the accessible sites were examined. After doing so, it was discovered that the data on the wind speed for the three sites, namely Crosby, Speke, and Woodvale, vary monthly. The month of March was seeing the highest wind speeds at all of the sites, however, the lowest wind speeds vary from site to site. Additionally, it was clear from this research that the wind speed data correlation for all of the sites followed the exponential decay characteristics. The wind speed data for the Gwynt y Môr location was computed using the Principal Component Analysis while taking into account all of these features of wind speed. An additional comparison was made between this strategy and the Inverse Distance Weighting method. Results from both approaches were consistent with the wind speed characteristics for the three sites that were considered. After the data for the Gwynt y Môr site had been estimated, modelling the data was the next step, but it was crucial to identify the probability distribution for the wind speed data first. To accomplish this, a variety of distributions including the Weibull two-parameter distribution, the Kappa (KAP) and Wakeby (WAK) distribution, the Rayleigh distribution, and the lognormal distribution were examined. Weibull's two-parameter distribution was

chosen in the end because, despite a few drawbacks, it offered a good fit for the wind speed data.

Once the data for the Gwynt y Môr site was computed and the distribution of the data was finalised, the next step in the dissertation was to build the wind speed model. To do so, many approaches, like Time Series Analysis, Artificial Neural Networks and Support Vector Machine were analysed. Later, it was seen that these approaches were using huge amount of previous data, which helped in getting accurate predictions of the wind speed data. Due to the limitation of data, Stochastic Difference Equations approach was used model the data. The outcomes of this method were consistent with the data on actual wind speeds. However, the pattern of the hourly mean wind speed data showed some variations, but they were reasonable. Furthermore, because the modelled data can only show a similar pattern and not the actual data, some fluctuation in the data is acceptable.

The last step of the dissertation was to build the curtailment model for the Gwynt y Môr site. To do so, Markov Chain approach was used. In this approach transitioning probabilities were computed for the provided data. The only limitation with this approach was that it did not used the information of the wind speed data. To overcome the issue, modifications were done. In the new model, weights were assigned to curtailment and wind speed. The weights were computed using the Artificial Neural Networks. Later, the transitioning probabilities were computed for different values of wind speed data. The findings show that when the wind speed variable were not taken into account, the chance of remaining in the same curtailment bin was generally higher than shifting to another bin. Only when the curtailment bin was either 0 or 3, resembled the modified technique. As the wind speed changed, the probability differed greatly when the curtailment bin was either 1 or 2. It can be said that the improved approach, which involved modelling the curtailment for the Gwynt y Môr site using both wind speed and curtailment data, produced better results than just using the curtailment data.

6.2 Further Work

Due to the limited resources available, there are a few things that can be done in the future to enhance the model. Time constraints hindered a full examination of the data on hourly and monthly variation in wind speed. More time would allow for a better consideration of these variations. Other than that, it was not possible to acquire the wind speed information for the Gwynt y Môr location. With improved resources, a model can be created for the location where wind speed and curtailment data are both accessible. If actual wind speed data for the site were available, the model could be more precise. Furthermore, additional variables that might affect curtailment, could be taken into account when determining how transitioning

probabilities depend on them.

Appendix A

ERF Approximation

For this research, Prof. Neil Strickland has provided a detailed explanation to approximate $erf(x)$ to compute $\Phi^{-1}(.)$. The Cumulative Distribution Function (CDF) for the standard normal distribution is

$$\Phi(y) = \frac{1 + erf(y/\sqrt{2}}{2}) \tag{A.1}$$

and the CDF for the Weibull distribution is

$$F_{kc}(x) = 1 - \exp\left(-\frac{x}{c}^{k}\right) \tag{A.2}$$

If Y is a standard normal random variable, then the variable $X = cq(y)^{2/k}$ will be distributed according to F_{kc}. Equivalently, if X is distributed according to F_{kc}, then the variable $Y = q^{-1}((X/c)^{k/2})$ will have standard normal distribution. Thus, in the wind model, there will be empirical measurements x_i and will be able to calculate $y_i = q^{-1}((x_i/c)^{k/2})$, where k and c are estimates of the shape and scale parameters. One way to do this is to use the formula

$$q^{-1}(x) = \sqrt{2}erf^{-1}(1 - 2\exp(-x^2)) \tag{A.3}$$

However, this is not numerically stable. When x is reasonably large it can be seen that $(1 - 2\exp(-x^2))$ is extremely close to 1, and the standard algorithms for calculating erf^{-1} behave poorly near 1 and may give up and return an infinite result. It is better to perform some mathematical simplification of the above formula for erf^{-1} rather than using it directly.

For this, Winitzki's approximation can be used, i.e.,

$$erf^{-1}(z) = sgn(z).\sqrt{\sqrt{\left(\frac{2}{\pi a} + \frac{\ln(1-z^2)}{2}\right)^2 - \frac{\ln(1-z^2)}{a}} - \left(\frac{2}{\pi a} + \frac{\ln(1-z^2)}{2}\right)} \tag{A.4}$$

where $a = 8(\pi - 3)/(3\pi(4 - \pi))$. On comparing Equations A.3 and A.4, it can be stated that $z = 1 - 2\exp(-x^2)$. This gives

$$1 - z^2 = 4\exp(-x^2) - 4\exp(-2x^2), \tag{A.5}$$

and so

$$\ln(1 - z^2) = 2\ln(2) - x^2 + \ln(1 - \exp(x^2)) \tag{A.6}$$

It is convenient to write things in terms of the variable $w^2 = -\ln(1 - \exp(-x^2))$, which is the unique positive solution to $\exp(-x^2) + \exp(-w^2) = 1$. Many things are symmetrical between x and w. In particular, the expression above becomes

$$\ln(1 - z^2) = 2\ln(2) - x^2 - w^2 \tag{A.7}$$

If now put

$$c = (1 - 2/\pi)/a - \ln(2) \tag{A.8}$$

$$d = (4/\pi - 1)/a^2 \tag{A.9}$$

$$u = (w^2 + x^2)/2 + c \tag{A.10}$$

$$v = \sqrt{(u^2 + d)} + u - 1/a \tag{A.11}$$

it works out that $q^{-1}(x) \approx \sqrt{2v}sgn(x - \sqrt{\ln(2)})$.

Appendix B

Ethics

Form 1: Student declaration
(for research that does not involve human participation or analysis of secondary data)

School of Mathematics and Statistics

Research Ethics Review for Postgraduate Taught Students

Dissertation title: **Machine Learning for Electrical Engineering**

In signing this student declaration I am confirming that:

My project does not involve people participating in research either directly (e.g. interviews, questionnaires) and/or indirectly (e.g. people permitting access to data).

My project does not therefore require an ethics review and I have not submitted a Research Ethics Application Form.

Name of student: Chetali Jain

Name of supervisor: Neil Strickland

Signature of student: Chetali Jain

Signature of Supervisor: Neil Strickland

Date: 31 April' 2022

Date: 31 April' 2022

References

Aslam, M. (2021), 'A study on skewness and kurtosis estimators of wind speed distribution under indeterminacy', *Theoretical and Applied Climatology* **143**(3-4), 1227–1234.
URL: *https://link.springer.com/article/10.1007/s00704-020-03509-5*

Bilgili, M., Sahin, B. & Yasar, A. (2007), 'Application of artificial neural networks for the wind speed prediction of target station using reference stations data', *Renewable Energy* **32**(14), 2350–2360.
URL: *https://www.sciencedirect.com/science/article/pii/S0960148106003429*

Blanchard, M. & Desrochers, G. (1984), 'Generation of autocorrelated wind speeds for wind energy conversion system studies', *Solar Energy* **33**(6), 571–579.
URL: *https://www.sciencedirect.com/science/article/pii/0038092X84900136*

Carta, J., Ramírez, P. & Velázquez, S. (2009), 'A review of wind speed probability distributions used in wind energy analysis: Case studies in the canary islands', *Renewable and Sustainable Energy Reviews* **13**(5), 933–955.
URL: *https://www.sciencedirect.com/science/article/pii/S1364032108000889*

Dookie, I., Rocke, S., Singh, A. & Ramlal, C. J. (2018), 'Evaluating wind speed probability distribution models with a novel goodness of fit metric: a trinidad and tobago case study', *International journal of energy and environmental engineering* **9**(3), 323–339.
URL: *https://link.springer.com/article/10.1007/s40095-018-0271-y*

Drew, J. & Rigby, A. (2013), 'Gwynt y môr: Britain's world-leading offshore wind farm takes shape', *Proceedings of the Institution of Civil Engineers - Energy* **166**(3), 120–126.
URL: *https://doi.org/10.1680/ener.12.00015*

Garcia, A., Torres, J., Prieto, E. & de Francisco, A. (1998), 'Fitting wind speed distributions: a case study', *Solar Energy* **62**(2), 139–144.
URL: *https://www.sciencedirect.com/science/article/pii/S0038092X97001163*

Holland, M. J. & Ikeda, K. (2014), 'Forecasting in wind energy applications with site-adaptive weibull estimation', pp. 2184–2188.
URL: *https://ieeexplore.ieee.org/abstract/document/6853986*

Jamil, M., Parsa, S. & Majidi, M. (1995), 'Wind power statistics and an evaluation of wind energy density', *Renewable Energy* **6**(5), 623–628. Solar Electricity: Photovoltaics and Wind.
URL: *https://www.sciencedirect.com/science/article/pii/096014819500041H*

Joos, M. & Staffell, I. (2018), 'Short-term integration costs of variable renewable energy: Wind curtailment and balancing in britain and germany', *Renewable and Sustainable Energy Reviews* **86**, 45–65.
URL: *https://www.sciencedirect.com/science/article/pii/S1364032118300091*

Joselin Herbert, G., Iniyan, S., Sreevalsan, E. & Rajapandian, S. (2007), 'A review of wind energy technologies', *Renewable and Sustainable Energy Reviews* **11**(6), 1117–1145.
URL: *https://www.sciencedirect.com/science/article/pii/S136403210500095X*

Kamal, L. & Jafri, Y. Z. (1997), 'Time series models to simulate and forecast hourly averaged wind speed in quetta, pakistan', *Solar Energy* **61**(1), 23–32.
URL: *https://www.sciencedirect.com/science/article/pii/S0038092X97000376*

Legates, D. (1987), 'A climatology of global precipitation. publ', *Climatol* **40**(1), 1–85.

Lei, M., Shiyan, L., Chuanwen, J., Hongling, L. & Yan, Z. (2009), 'A review on the forecasting of wind speed and generated power', *Renewable and Sustainable Energy Reviews* **13**(4), 915–920.
URL: *https://www.sciencedirect.com/science/article/pii/S1364032108000282*

Luo, W., Taylor, M. C. & Parker, S. R. (2008), 'A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from england and wales', *International Journal of Climatology* **28**(7), 947–959.
URL: *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.1583*

Ma, J., Fouladirad, M. & Grall, A. (2018), 'Flexible wind speed generation model: Markov chain with an embedded diffusion process', *Energy* **164**, 316–328.
URL: *https://www.sciencedirect.com/science/article/pii/S0360544218317481*

Mohandes, M., Halawani, T., Rehman, S. & Hussain, A. A. (2004), 'Support vector machines for wind speed prediction', *Renewable Energy* **29**(6), 939–947.
URL: *https://www.sciencedirect.com/science/article/pii/S0960148103003860*

Morgan, E. C., Lackner, M., Vogel, R. M. & Baise, L. G. (2011), 'Probability distributions for offshore wind speeds', *Energy Conversion and Management* **52**(1), 15–26.

URL: *https://www.sciencedirect.com/science/article/pii/S019689041000227X*

Saber, H., Moeini-Aghtaie, M., Ehsan, M. & Fotuhi-Firuzabad, M. (2019), 'A scenario-based planning framework for energy storage systems with the main goal of mitigating wind curtailment issue', *International Journal of Electrical Power Energy Systems* **104**, 414–422.
URL: *https://www.sciencedirect.com/science/article/pii/S014206151731092X*

Seguro, J. & Lambert, T. (2000), 'Modern estimation of the parameters of the weibull wind speed distribution for wind energy analysis', *Journal of Wind Engineering and Industrial Aerodynamics* **85**(1), 75–84.
URL: *https://www.sciencedirect.com/science/article/pii/S0167610599001221*

Sfetsos, A. (2000), 'A comparison of various forecasting techniques applied to mean hourly wind speed time series', *Renewable Energy* **21**(1), 23–35.
URL: *https://www.sciencedirect.com/science/article/pii/S0960148199001251*

Shu, Z., Li, Q. & Chan, P. (2015), 'Statistical analysis of wind characteristics and wind energy potential in hong kong', *Energy Conversion and Management* **101**, 644–657.
URL: *https://www.sciencedirect.com/science/article/pii/S019689041500535X*

Strickland, N. (n.d.), 'A method of esimation involving pirincipal components'.
URL: *https://github.com/NeilStrickland*

Tascikaraoglu, A. & Uzunoglu, M. (2014), 'A review of combined approaches for prediction of short-term wind speed and power', *Renewable and Sustainable Energy Reviews* **34**, 243–254.
URL: *https://www.sciencedirect.com/science/article/pii/S1364032114001944*

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P. & SciPy 1.0 Contributors (2020), 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python', *Nature Methods* **17**, 261–272.
URL: *https://docs.scipy.org/doc/scipy/tutorial/stats.html*

Wang, J., Niu, T., Lu, H., Guo, Z., Yang, W. & Du, P. (2018), 'An analysis-forecast system for uncertainty modeling of wind speed: A case study of large-scale wind farms', *Applied Energy* **211**, 492–512.
URL: *https://www.sciencedirect.com/science/article/pii/S0306261917316719*

Waskom, M. L. (2021), 'seaborn: statistical data visualization', *Journal of Open Source Software* **6**(60), 3021.

URL: *https://doi.org/10.21105/joss.03021*

Yu, C., Li, Y., Bao, Y., Tang, H. & Zhai, G. (2018), 'A novel framework for wind speed prediction based on recurrent neural networks and support vector machine', *Energy Conversion and Management* **178**, 137–145.
URL: *https://www.sciencedirect.com/science/article/pii/S0196890418311087*

Zárate-Miñano, R., Anghel, M. & Milano, F. (2013), 'Continuous wind speed models based on stochastic differential equations', *Applied Energy* **104**, 42–49.
URL: *https://www.sciencedirect.com/science/article/pii/S0306261912007969*