Sustainable Wind Turbine Predictive Maintenance: Detecting Mexican Hats in Wind Speed Data for Efficient Operations

Maryam Marasuli

Thesis submitted for the degree of
Master in Applied Computer and Information Technology - ACIT
(Data Science)
30 credits

Department of Computer Science
Faculty of Technology, Art and Design

Oslo Metropolitan University — OsloMet

Spring 2023

Sustainable Wind Turbine Predictive Maintenance: Detecting Mexican Hats in Wind Speed Data for Efficient Operations

Maryam Marasuli

Sustainable Wind Turbine Predictive Maintenance: Detecting Mexican Hats in Wind Speed Data for Efficient Operations

http://www.oslomet.no/

Printed: Oslo Metropolitan University
— OsloMet

Abstract

Wind turbines are critical for renewable energy generation, but unexpected maintenance issues can hinder their efficient operation. Predictive maintenance can mitigate these issues by detecting early signs of turbine degradation. Wind gusts, sudden changes in the wind speed in turbulent wind fields, can cause significant fatigue loads on wind turbines, leading to a shorter lifespan of their components. In addition, power generation oscillations or ramping can result in rapid grid voltage changes, creating further challenges for the stability of the electric grid. In this study, we present an innovative and novel approach for sustainable wind turbine predictive maintenance that leverages a time-series analysis-based AI model to accurately detect specific types of wind gusts, called Mexican Hats, and their characteristics in wind speed data, which may indicate potential turbine performance anomalies. Our approach has the potential to optimize maintenance schedules, reduce downtime, and extend turbine lifespan for more sustainable and efficient operations. The results of this study can contribute to the development of more robust control systems and enhance the reliability of wind turbine operations in the renewable energy sector. The research uses a multivariate time series dataset collected at different heights from offshore platforms in the North Sea (FINO1 platform), which has been made available to OsloMet for research purposes.

Acknowledgements

I would like to express my heartfelt gratitude and appreciation to all those who have contributed to the completion of my master's thesis.This accomplishment would not have been possible without their support, encouragement, and assistance.

First and foremost, I am deeply indebted to my thesis supervisor, Pedro Lind, for his invaluable guidance, expertise, and patience throughout the entire process. His unwavering commitment, constructive feedback, and insightful suggestions have been instrumental in shaping the quality and direction of my research. I am genuinely grateful for his mentorship and encouragement.

I must convey my profound gratitude to my external supervisor from SLB (Schlumberger), Justo Matheus, who provided me with help and advice during my research project. His help, passion, and prompt responses to my questions and requests were invaluable to the success of my research.

Furthermore, I want to express my deepest gratitude to my family, including my husband Hassan, my beloved daughters Nika and Kiana, and my parents. Their staunch belief in my abilities has been a driving force behind my success. During this challenging journey, they have been my pillars of strength, constantly motivating me to achieve this significant milestone in my academic career. I am immensely thankful to Hassan for his endless emotional support, time management assistance, and valuable feedback on my research. I feel truly fortunate to have such a supportive partner by my side, and I thank him from the bottom of my heart.

Maryam Marasuli – Norway – Spring 2023

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Of all the forces of nature, I should think the wind contains the largest amount of motive power—that is, power to move things. Take any given space of the earth's surface— for instance, Illinois; and all the power exerted by all the men, and beasts, and running-water, and steam, over and upon it, shall not equal the one hundredth part of what is exerted by the blowing of the wind over and upon the same space. And yet it has not, so far in the world's history, become proportionably valuable as a motive power. It is applied extensively, and advantageously, to sail-vessels in navigation. Add to this a few windmills, and pumps, and you have about all. . . . As yet, the wind is an untamed, and unharnessed force; and quite possibly one of the greatest discoveries hereafter to be made, will be the taming, and harnessing of it."

— *Abraham Lincoln*

1.1 Background

Humans have been harnessing different energy sources, mainly fossil fuels, for centuries. However, due to the rapid increase in global energy demands and the aim to achieve net zero emissions by 2050, (Bouckaert et al., 2021) low-carbon energy sources, including wind power, have garnered significant attention and investment in recent years. The main objectives are to enhance the efficiency of power generation, decrease generation costs, and increase the scalability of these systems. As such, thoroughly comprehending the factors that affect the functioning of these systems is essential.Digitalization has further optimized wind energy output and improved the design of turbines and wind farms. This has made it easier to maintain equipment and extend its lifetime, contributing to greater efficiency and cost savings.

Wind energy is a clean energy source with almost no CO2, SOx, NOx, or PM emissions. The CO2 footprint of wind turbines is negligible, with a turbine paying off its lifecycle emissions within 6-9 months of operation.Compared to electricity generated from gas and coal, wind power emits approximately 95% and 98% less CO2, (Tremeac & Meunier, 2009)respectively, making it

an excellent option for sustainable energy production. The European Union (EU) recognizes offshore wind as a critical sector for the future and has set ambitious targets for renewable energy generation. Offshore wind has the potential to significantly contribute to the EU's renewable energy goals and reduce greenhouse gas emissions. By delivering the wind volumes outlined in Europe's energy security strategy, REPowerEU, Europe can save 65 bcm of gas, which will reinforce its energy independence. Additionally, every new turbine helps reduce Europe's reliance on energy imports and lowers exposure to volatile fossil fuel prices, which helps minimize power bills.('WindEurope - Wind Energy', 2023-04-27)

By 2050, wind energy is projected to satisfy half of the global electricity demand.(Association et al., 2011) However, this also implies a significant need to maintain and repair offshore wind turbines (OWT) in the coming decades. Therefore, it is crucial to consider the impact of OWT maintenance on the environment and develop a suitable strategy that prioritizes profitability and sustainability.(Z. Ren et al., 2021) Achieving this balance is vital to the long-term development of offshore wind energy. Given the challenges associated with OWT maintenance, it is essential to implement effective maintenance practices to minimize downtime and energy output losses.

When it comes to wind power, the primary factors that influence the performance and expected power output of the systems are extreme weather conditions and wind speed fluctuations (Albadi & El-Saadany, 2010).

The importance of understanding extreme events in wind data lies in the highly fluctuating nature of wind power, which leads to frequent occurrences of extreme events. While wind turbines can convert wind speed into power, there are limits to their ability to do so. The wind turbine's saturation curve and physical strain impose a constraint on the quantity of wind energy that can be transformed into power. Therefore, creating a model capable of identifying and predicting such extreme conditions would be incredibly advantageous for both energy suppliers and users.

The economic and security risks caused by gusts are widespread, affecting various human activities and the environment, such as transportation, electric power systems, building structures, and weather forecasts (Brasseur, 2001). This is due to the fact that wind power is proportional to the square of the wind speed. Suomi and Vihma (Suomi & Vihma, 2018) emphasized the practical significance of estimating the return levels of wind gusts for construction planning and insurance companies. Accurate full gust forecasts can provide valuable information for disaster preparedness planning and monitoring in various operation fields, including shipping, navigation, land transportation, aviation, forestry, and energy over the next day to a few weeks (Hewston & Dorling, 2011) and (Ágústsson & Ólafsson, 2009).

By developing a statistical understanding of the frequency and nature of extreme events, it will be possible to estimate the expected extreme loads that wind turbines will experience in the future. This knowledge would be invaluable to engineers designing wind turbines, allowing them to create more efficient and durable turbines capable of withstanding extreme events. While

this goal may be challenging, the scientific importance of solving this problem cannot be overstated.

The accurate prediction and detection of extreme events, such as sudden increases in wind speed and power, pose significant challenges in the field of wind energy. To mathematically describe these abrupt changes, the Mexican hat function is commonly used due to its ability to account for the physical constraints of the atmosphere, which behaves as a fluid system capable of supporting diverse motions.

1.2 Motivation

The rationale behind detecting Mexican hats is rooted in the fact that these extreme weather events can put significant stress on wind turbines, causing wear and tear on their components and reducing their lifespan. Moreover, power fluctuations and abrupt changes in grid voltage can pose further difficulties for the electric grid and impact the amount of energy generated by wind turbines and their overall dependability and resilience. Consequently, a comprehensive grasp of these extreme events, including their patterns and frequencies, would allow for the prediction of maintenance needs for wind turbines, better planning for energy distribution, and fewer operational surprises. Additionally, it would enhance confidence in wind power as a trustworthy and sustainable energy source.

Figure 1.1: A view of multiple wind turbines in the Alpha Ventus Offshore Wind Park (FINO1 Platform) (Lehmann-Matthaei, 2023)

1.3 Problem Statement and Scope of Thesis

One of the key challenges in sustainable wind energy operations is securing efficient maintenance of wind turbines to minimize downtime and maximize power generation. Predictive maintenance involves detecting potential issues before they become critical failures, which is crucial for cost-effective and

sustainable operations.

However, current methods to detect maintenance requirements in wind turbines need to be more accurate and can result in unnecessary maintenance or costly breakdowns. Additionally, optimizing wind turbine operations based on wind speed data can be challenging due to distinct patterns called Mexican hats, which require sophisticated data analytics techniques for detection and interpretation. Addressing these challenges and improving the accuracy of predictive maintenance and Mexican hat detection can significantly enhance wind energy operations' efficiency, reliability, and environmental impact, making it a crucial area of research and innovation in renewable energy.

This thesis seeks to advance the performance and reliability of wind turbines, with the ultimate goal of contributing to the development of a more sustainable energy future with the following research targets:

1. Develop an AI-assisted methodology for detecting Mexican hat events in wind turbine datasets with a high level of accuracy. This methodology should take into account the specific characteristics of Mexican hat events, such as their shape, frequency, and duration, and should be able to distinguish them from other types of events and background noises.

2. Analyze the behavior and characteristics of the detected Mexican hat events as input to the algorithm to predict the maintenance needs. This analysis should encompass various factors, including wind speed, the widths, and amplitudes of the detected Mexican hats, the wind velocity at their peak, and other relevant variables.

This study's findings could aid in creating effective control mechanisms and improve the dependability of wind turbine functions within the renewable energy industry.

This thesis employs a rigorous research methodology that combines data analysis, statistical techniques, and machine-learning algorithms to enhance wind turbine predictive maintenance. This methodology consists of the following steps:

- **Data Collection**: The first step in the research method was accessing a wind turbine dataset. We managed to get access to a multivariate time series dataset collected at different heights from offshore platforms in the North Sea (FINO1 platform), which was made available to OsloMet for research purposes.

- **Literature Review**: A comprehensive literature review was performed to gain insight into the current state of understanding of wind turbine predictive maintenance, Mexican hat detection, offshore wind energy, and related topics. This review helped identify research gaps and refine the research approach.

- **Event Detection**: The next step in the research process was to develop an

innovative and robust algorithm to identify and characterize the Mexican hat events with high accuracy. This step involved using signal processing techniques, wavelet analysis, and cross-correlation to identify the patterns associated with these events.

- **Evaluation and validation**: The event detection algorithm's accuracy, effectiveness, and reliability were assessed through an extensive manual quality check of the output. These measures were to ensure the results are a reliable foundation for the subsequent steps and further analysis and conclusions.

- **Statistical Analysis**: Once the Mexican hat events were identified, statistical analysis techniques, such as descriptive statistics (e.g., mean, median, and standard deviation) and data distribution analysis, were used to understand the characteristics and behavior of the identified Mexican hat events. In addition, the best distribution functions fitting the data distributions were determined.

- **Conclusions**: Finally, the research findings were carefully examined and analyzed to draw valid and reliable conclusions. Based on the research outcomes, recommendations were provided for future research directions, practical applications, and policy implications.

Chapter 2

Literature Review

Detecting Mexican hat events in wind speed data is a challenging problem due to their sudden and intermittent occurrence and the presence of noise and measurement errors in wind speed datasets.

Therefore, to better understand these events and their impact on wind turbine performance and load generation, we conducted an extensive literature review of relevant research areas such as wind speed data analysis, extreme events, gusts, Mexican hats, and predictive maintenance of wind turbines. By reviewing the literature in these areas, we were able to refine our detection model and improve its ability to accurately identify Mexican hats and their characteristics in the FINO1 dataset.

This chapter provides an overview of the key statistical and time-series data analysis terms and techniques we used in our data processing operations and a summary of our fundamental findings in the literature review.

2.1 Definitions and Statistical Learning

2.1.1 Time-Series Analysis and Statistics

Time-series data refers to a collection of measurements or observations that are recorded at regular intervals, such as daily, weekly, monthly, or annually. The data is ordered chronologically, and each measurement is associated with a specific time point. Figure 2.1 illustrates a sample time-series data (FINO1 wind speed data).

Time-series analysis is a statistical technique that involves examining and modeling the patterns in the data to identify underlying trends, seasonal effects, or other patterns. This analysis can be used to conduct regression analysis and forecasting, which is essential for predicting future events, tracking changes over time, and providing input for decision-making processes.

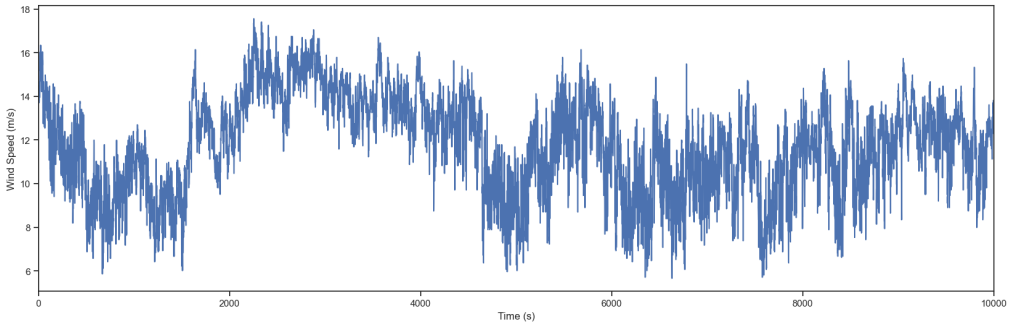


Figure 2.1: Graphical representation of time-series data

The application of time-series analysis is widespread across various domains, including economics, finance, engineering, environmental science, and many others. In economics, for example, time-series data is used to track and analyze economic indicators such as inflation rates, GDP, and unemployment rates. In finance, time-series analysis is used to forecast stock prices and other financial metrics. In engineering, it is used to monitor and predict the behavior of complex systems such as machines, buildings, and transportation networks. In environmental science, time-series analysis is used to study long-term trends in climate data, atmospheric pollutants, and natural resource consumption.

Time-series analysis enables us to extract meaningful insights from complex and often noisy data, allowing us to make informed decisions and predictions. By identifying patterns in the data, we can better understand the underlying factors contributing to changes over time and use this knowledge to anticipate future trends and events. As such, time-series analysis is a critical tool for businesses, governments, and researchers alike, enabling them to leverage historical data to make informed decisions and plan for the future.

Time series data models come in various shapes and can depict stochastic processes. The auto-regressive (AR), integrated (I), and moving average (MA) models (Pasari & Shah, 2020) are three major types that are important for describing changes at a process's level. These three classifications are linearly correlated with earlier data points (X. Liu et al., 2021) . Auto-regressive moving average (ARMA) and auto-regressive integrated moving average (ARIMA) models are created by combining these concepts. The auto-regressive fractionally integrated moving average (ARFIMA) model generalized the previous three. Multivariate time-series models are an extension of these classes that deal with vector-valued data. Occasionally, the last acronyms are also extended by adding an initial "V" for "vector," as in "VAR" for "vector auto-regression" (S. Singh, Mohapatra et al., 2019).

It is interesting when a series' level depends non-linearly on earlier data points, partly because it raises the prospect of a chaotic time series. However, more crucially, empirical studies can show that employing predictions from non-linear models has an advantage over predictions from linear models, as is the case, for instance, with non-linear auto-regressive exogenous models. Several forms of motivation and data analysis are suitable for time series for vari-

ous applications. Time series analysis may be used for clustering, classification, query by content, anomaly detection, and forecasting in the following applications: data mining, pattern recognition, and machine learning. Time series data can be segmented into distinct parts with unique characteristics. The two main objectives of time-series segmentation are finding the segment boundary points and describing the dynamic aspects of each segment. This issue can be solved through change-point detection or modeling the time series as a more complex, such as a Markov jump linear system.

To facilitate understanding, we explain some fundamental concepts below (Woolf et al., 2004):

Trend: A trend is a long-term pattern in a time series. It represents the overall direction of the data over time. A trend can be upward, downward, or flat.

Histogram: Histogram is a visualization tool that displays the distribution of a set of data through a bar graph. The x-axis shows the range of values, and the y-axis indicates the frequency or count of observations within each range.

Standard deviation: The standard deviation provides information about the dispersion of the dataset from the mean value. When a dataset has a slight standard deviation, it means that the data points are closely grouped around the mean, indicating a high level of precision. Conversely, when a dataset has a large standard deviation, it indicates that the data points are spread over a wide range of values, suggesting high variability. The mathematical formula for the standard deviation is:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{i=n}(X_i - \bar{X})^2}$$

Mean: The mean, also called the average, is calculated by dividing the sum of all the observed values by the total number of observations, denoted as 'n'. Although some data points may fall above or below the mean, it is often a valuable estimate for predicting future data points. The formula for calculating the mean is:

$$\bar{X} = \frac{\sum_{i=1}^{l=n} X_i}{n}$$

Nevertheless, equation above applies only under the condition that the error linked with each measurement is either identical or unknown.

Skewness: Skewness is a measurement of the asymmetry of the Skewness measures the degree of asymmetry in a distribution and is used to evaluate the distribution's symmetry with respect to its mean. When the skewness value is positive, the tail of the distribution is longer on the positive side, whereas a

negative skewness score indicates that the tail is longer on the opposite side:

$$\tilde{\mu}_3 = \frac{\sum_i^N (X_i - \bar{X})^3}{(N-1) * \sigma^3}$$

Kurtosis: Kurtosis measures the degree of peakedness or flatness of a distribution. High kurtosis values indicate a more peaked distribution with more values in the tails, while low kurtosis values suggest a flatter distribution with more values around the mean:

$$\text{Kurt } = \frac{\mu_4}{\sigma^4}$$

Autocorrelation: Autocorrelation is a statistical measure that can also be used to gauge the correlation between a time series and a lag version of itself. Its purpose is to detect patterns and dependencies in time series data. If the autocorrelation value is positive, it indicates that values tend to be similar to their lagged counterparts. Conversely, if the autocorrelation value is negative, it suggests that values tend to deviate from their lagged counterparts.

These are just some basic concepts in time-series analysis. There are many statistical techniques and methods available for analyzing time-series data, and the choice of technique will depend on the characteristics of the data and the research question being addressed.

2.1.2 Wind Speed Data Analysis

Wind speed data analysis is a fundamental aspect of scientific research, enabling a deeper understanding of atmospheric dynamics and their influence on various environmental processes. The collection of accurate wind speed data involves the deployment of anemometers and other specialized instruments at designated locations, such as weather stations, offshore platforms, or wind farms (Association et al., 2012). These instruments record the speed at which air passes a specific point, typically measured in meters per second (m/s) or miles per hour (mph). The quality control measures applied during data collection ensure the reliability and consistency of the recorded wind speed data.

Exploratory data analysis (EDA) techniques are employed to gain initial insights into wind speed data (St. Amant & Cohen, 1997). Statistical measures, such as standard deviation, mean, and median, provide summary statistics that describe the central tendency and variability of the data. Histograms, box plots, and time series plots offer visual representations of wind speeds' distribution and temporal patterns. EDA helps identify outliers, gaps in data, and potential data quality issues, allowing for necessary data preprocessing steps.

The analysis of wind speed data goes beyond descriptive statistics and delves into examining the connections with other meteorological variables.

Correlation analysis is often employed to assess the strength and direction of the association between wind speed and factors such as temperature, humidity, or atmospheric pressure. Additionally, temporal analysis techniques, including time series and spectral analysis, reveal long-term trends, seasonal variations, and periodic patterns in wind speeds. (Quan et al., 2016)These analyses contribute to understanding the underlying mechanisms driving wind patterns and their implications for weather forecasting, climate studies, and renewable energy applications.

Wind speed data analysis often involves fitting a suitable statistical model to the data, such as the Weibull distribution. The Weibull distribution is widely used for modeling wind speed data due to its ability to capture the characteristics of both weak or low values and maximum values. While the fundamental derivation of the Weibull distribution focuses on the minimum or most inadequate values, it is an empirical model that can also effectively fit the maximum data points (Castillo et al., 2005).

2.1.3 Extreme Events in Wind Data

Countless studies have consistently emphasized the significance of precisely comprehending wind gusts. The main meteorological parameter widely employed in assessing wind-induced damage is wind gusts.

So an extreme event (gust) is a sudden change in wind speed within a turbulent wind field. Sharp increases characterize gusts and decrease wind speed, quickly returning to the average wind speed. The shape of the gust can take on various forms, but special attention is often given to cases where the wind speed increases rapidly (Milan et al., 2013). Positive extreme events, which are faster and potentially more harmful, are often the focus of attention. Damaging gusts can also cause significant load fluctuations. The "Mexican hat" form is an extreme and distinctive gust shape (Garcia et al., 2019)that can result in considerable load fluctuations due to its effect on wind speed composition. Several parameters characterize this gust type, including amplitude, rise time, maximum gust variation, and lapse time. According to this concept, wind speed should return to its typical pace after a gust. It is possible to establish a general critical mean gust shape that could lead to ultimate thrust loading by understanding the load response characteristics of the wind turbine in terms of aerodynamics and pitch action.

The IEC standard (International Electrotechnical Commission), highlights an important design factor: a strong gust combined with a change in wind direction.(TC88-MT, 2005) This event involves a gust with a speed of 15 m/s that lasts for 10 seconds and coincides with a shift of 30° in the wind's direction. According to engineers, this extreme event generates the highest loads in simulations, and turbines that have experienced similar conditions in real-world scenarios have suffered damage.

2.1.4 Definition of Mexican hat

Wind turbines experience Extreme Operating Gusts (EOGs), also known as Mexican hat gusts, which are caused by atmospheric turbulence and last for approximately 20 seconds. These gusts are characterized by a sudden and steep increase in wind speed, followed by a rapid drop back to the actual average speed. The "Mexican hat" form, which incorporates the gust amplitude and duration, is used to simulate the velocity fluctuation during an operating gust. This is an illustration of the wind speed variation during an intense operational gust. Figure 2.1.4 illustrates several Mexican hats with the same amplitude but varying durations (widths).

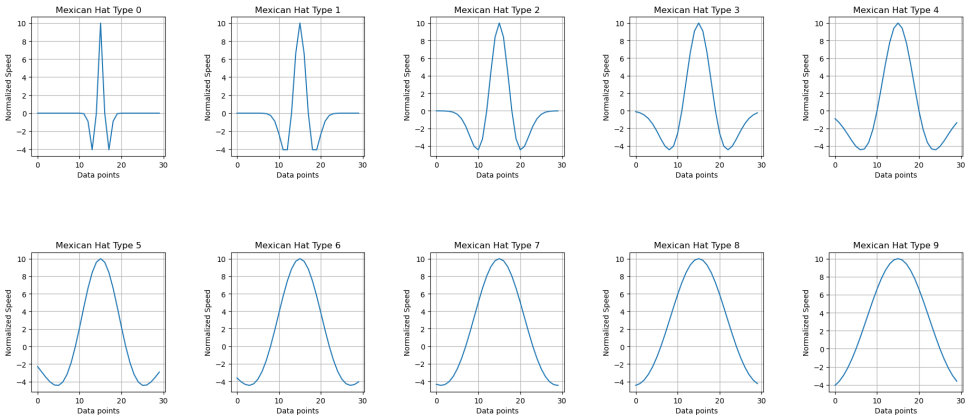


Figure 2.2: Demonstration of several Mexican hats with a constant amplitude and differing widths

The Mexican hat wavelet is a specialized wavelet transform used for time series analysis to identify patterns and fluctuations in data. Its unique shape, similar to that of a Mexican hat, makes it especially adept at detecting changes in the mean and standard deviation of data. (Leys et al., 2013) In analyzing wind speed data, the Mexican hat wavelet can be utilized to identify changes in wind resources and optimize the control and performance of wind turbines. This wavelet can detect changes in wind speed patterns that may impact turbine performance, allowing for proactive measures to be taken to mitigate any negative effects on energy production. The Mexican hat wavelet can also identify opportunities to optimize energy production by adjusting blade pitch or yaw to capture more wind energy.

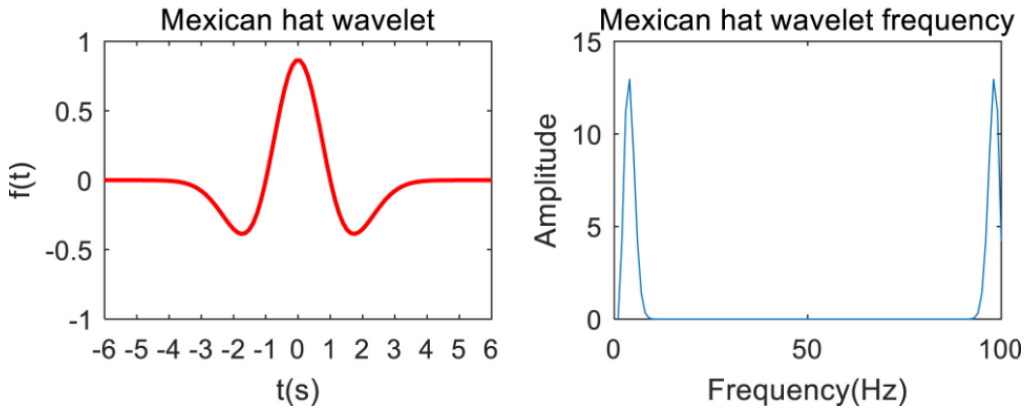


Figure 2.3: Time frequency characteristics of Mexican hat wavelet function (Meng et al., 2022).

Overall, the Mexican hat wavelet is an important tool in time series analysis that can identify significant changes in data patterns, particularly in wind speed data. Its application in detecting fluctuations in the wind resource can improve wind turbine performance and energy production, leading to more significant economic and environmental benefits of wind power. By utilizing the Mexican hat wavelet in wind energy production, operators can optimize energy output while mitigating potential disruptions, making wind energy a more viable and sustainable power source.

For instance, wind turbine operators can use the Mexican hat wavelet to identify changes in wind speed patterns that could impact turbine performance. By detecting these changes, operators can take proactive measures to reduce the negative impact on energy production, such as adjusting blade pitch or shutting down the turbine if necessary. The Mexican hat wavelet can also help identify opportunities to optimize energy production by adjusting blade pitch to capture more wind energy(Carcangiu et al., 2014). With the Mexican hat wavelet, wind turbine operators can achieve greater efficiency, thereby contributing to the long-term sustainability of wind power.

The mathematical formula for the Mexican hat wavelet is given by the following expression (Ryan, 1994):

$$F_b(t) = \frac{2}{\sqrt{3b}\pi^{\frac{1}{4}}}\left(1 - \frac{t^2}{b^2}\right)\exp\left(-\frac{t^2}{2b^2}\right)$$

where $F_b(t)$ is the wavelet function, **t** is the time parameter, **b** is the Mexican hat rise time. In the context of wind turbines, the Mexican hat wavelet can be used to analyze wind speed data and to identify changes in the wind resource that may impact the performance of the turbine.

It is commonly acknowledged that gusts and turbulence are visible for periods of time ranging from one second to ten minutes. The following is the clearest explanation of what a gust means: In a turbulent wind field, a gust is

a brief change in wind speed. A classic example, the so-called "Mexican-hat" gust, is taken from this. But we'll get various shapes from our examination of the mean gust shape. We classify a gust as a front when the wind acceleration is significant and results in a high wind speed in a brief period of time. Longer incidents known as Squalls can cause confusion. Eventually, the focus shifts to the pleasant yet swift gusts A strong rise in loads might also be implied by a negative gust.

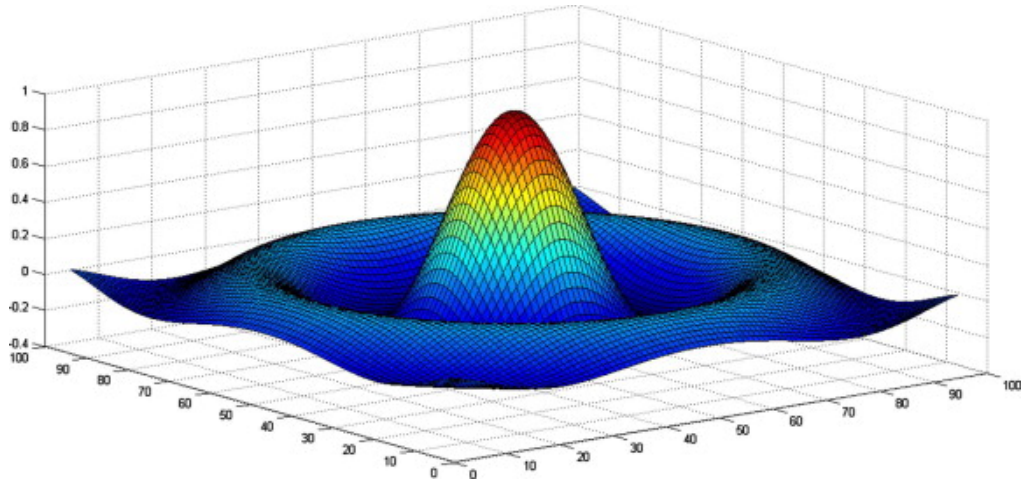


Figure 2.4: 3D demonstration of Mexican hat wavelet function, (Taghavifar & Mardani, 2014).

extreme event is determined by exploring a small temporal window around the instant in which the correlation exceeds the threshold. The Mexican hat wavelet is a symmetric function that looks like a bell curve with a sharp peak and a broad base. It is centered at zero and has zero mean, which makes it useful for analyzing signals with zero mean. The second derivative of the Gaussian function is used to create the Mexican hat wavelet, which makes it sensitive to edges and details in the signal.

2.2 Related Work

The field of extreme events in wind speed data has received significant attention and undergone notable advancements in recent years. Researchers have made remarkable strides in improving wind power technologies through various approaches, including modeling, forecasting, design, and analysis. These efforts have led to significant progress in addressing challenges related to extreme wind events, such as wind turbine failures and power grid disruptions.

The literature reviewed in this section provides an extensive overview of the various methods utilized to study wind behavior and its impact on wind turbines, which are related to the focus of my research. These scholarly articles cover a wide range of topics related to wind energy, including wave

height modeling, wind speed forecasting, wind turbine design, peak detection, wind gust analysis, wind power prediction, and time-frequency analysis. Furthermore, the articles delve into specific areas, such as identifying extreme wind gusts, designing and forecasting wind speed, estimating anticipated extreme loads, constructing wind power output time series, and applying Fourier transformations for manifold analysis:

2.2.1 Wind Speed & Time Series

Burg and Williams' article provides a high-level overview and classification of existing CPD (Continuous Professional Development) work (van den Burg & Williams, 2020).

Gandhi et al. concluded that a related problem of online time series approximation, which is the same as wind behavior, is to summarize the data in a single pass and build an approximate representation that can enable a range of time series queries while keeping worst-case error limitations (Gandhi et al., 2010).

Assuming a constant and contemporary distribution of wind farms, Cannon et al. (Cannon et al., 2015) utilized a state-of-the-art, 33-year reanalysis data set (MERRA, from NASA-GMAO) to construct an hourly time series of wind power output in Great Britain (GB) on a national scale. They found that the estimates are well correlated with the recorded data during the recent timeframe and quantified the frequency, seasonal, and inter-annual variability of several severe GB-wide wind power production events using this 33-year data series.

Kanev et al., in their study (Kanev & van Engelen, 2010), proposed an algorithm to recognize extreme wind gusts and direction changes in wind turbines, which can cause fatigue and unnecessary shutdowns. The algorithm uses a non-linear observer and detection algorithm to identify extreme events and activate the EEC algorithm to prevent rotor over speed and reduce blade loads. The method was tested on a complex turbine model.

To estimate the expected extreme load on wind turbines, Fitzwater et al., (Fitzwater et al., 2003), introduced the environmental contour method as a useful tool for estimating design loads on wind turbines by providing information about the combination of joint environmental variables that are most critical to each specific turbine. The contours were developed from design code descriptions and measured data for site-specific applications and used to make implicit FORM estimates for the turbine response. The study concluded that the environmental contour method was a reliable method for estimating the expected extreme load and presented the background of the development of environmental contours as applied to wind energy systems.

In a study focused on nowcasting rain and wind speed near Malpensa airport, Chkeir et al., (Chkeir et al., 2023), merged different datasets and used a Long Short-Term Memory Encoder Decoder (LSTM E/D) approach to develop a machine learning model for nowcasting meteorological variables. The analysis was based on four different datasets configurations, providing a nowcast for 1 hour with a time step of 10 minutes. The results showed a high probability of detection for extreme wind speeds (over 90%), low false alarms (less than 2%), and good performance in detecting extreme rain for the first 30 minutes.

To consistently integrate the resulting severe wind gust in a (coherent) stochastic wind field, Gunner C. Larsen and Kurt S. Hansen, (Larsen et al., 2004), established a logical way for calibrating the wind speed gust magnitudes specified. The approach used a wavelet expansion in conjunction with extreme statistics to produce consistent estimates of wind speed gust magnitudes for any return period, mean wind speed, and gust time scale.

Haselsteiner and Thoben, (Haselsteiner & Thoben, 2020), examined the modeling of significant wave height in marine structures like offshore wind turbines using the translated Weibull distribution, which often underestimates extreme loads. The study analyzed wave data from six offshore wind turbine locations and found that the exponentiated Weibull distribution provided a better overall fit. The study used weighted least squares estimation to prioritize high wave height observations and improve model fit at the tail. The proposed method accurately estimated the wave height's 1-year return value and could be used to calculate design loads for offshore wind turbines. The study concludes that the proposed method provides better results at the upper tail than traditional methods.

Singh et al., (S. Singh, Mohapatra et al., 2019), investigated the contribution of different decomposed time series on wind speed forecasting error, focusing on popular statistical methods such as ARIMA and hybrid methods like WT-ARIMA. The study proposed a new Repeated WT-based ARIMA (RWT-ARIMA) model with improved accuracy for very short-term wind speed forecasting. The study compared the proposed model with the benchmark persistence model, ARIMA model, and WT-ARIMA model for various forecasting time scales ranging from 1 to 10 minutes. The results demonstrated the superiority of the proposed RWT-ARIMA model over other models in very short-term wind speed forecasting.

In their paper, Kumar and Sengupta, (Kumar & Sengupta, 2022), proposed a new method for peak detection that improves upon the limitations of traditional single FBG peak detection. The proposed algorithm uses the Mexican hat wavelet function, and the Hilbert transform to find zero-crossing points, allowing for the easy detection of peak wavelengths. The method was found to improve the speed and accuracy of peak detection for both single

FBG and multi FBG, as demonstrated through theoretical, experimental, and simulation results.

In their article, Hu et al., (W. Hu et al., 2018), used data from 3D sonic anemometers placed at 3 and 65 meters in moderately complex terrain in the northeastern United States to analyze ten descriptors of wind gusts and determine the parent distributions that best describe them. The study found that the parent distributions were consistent across different descriptors of the gust climate. Gust intensity parameters fit the two-parameter Weibull distribution, while unitless ratios were described by log-logistic distributions and other properties were lognormally distributed. The study also found that gust factors scale with turbulence intensity, and gusts have different length scales than the average eddy length scale. Gust periods at the lower measurement height were consistent with shear production, while those at 65 meters were not. Although there was only a weak directional dependence of gust properties on-site terrain and land cover variability, large gust length scales and gust factors were more likely to occur in unstable atmospheric conditions.

Soman et al.'s paper (Soman et al., 2010) provided insights into the foremost forecasting techniques associated with wind power and speed based on numeric weather prediction (NWP), statistical approaches, artificial neural networks (ANN), and hybrid techniques over different time scales. It also discussed the comparative analysis of various available forecasting techniques. Moreover, emphasis was given to the major challenges and problems associated with wind power prediction.

Chandra et al. (D Chandra et al., 2014) proposed a 30-hour wind speed profile forecast using an Adaptive Wavelet Neural Network (AWNN) with Mexican hat and Morlet mother wavelets for decomposition. The wind speed data was collected from the NREL website for hourly averaged 10-min data sets in the Midwest ISO region for 2004. The data sets were normalized in the range of [-1, 1] to improve the training performance of forecasting models. A total of 8760 samples were used for forecasting analysis. The system's accuracy was evaluated by calculating statistical parameters after the forecasting phase and comparing different configurations.

Singh et al. (A. Singh et al., 2022) presented a unique method for time-frequency analysis, which provided a centralized way to represent discrete and continuous time-frequency. The Mexican hat wavelet was considered one of the primary wavelet functions formulated by the second derivative of the Gaussian function to define the Mexican hat wavelet transform. The theory of MHWT was implemented to obtain the Mexican hat wavelet Stieltjes transform of a bounded variation function, and some convenient properties of MHWST were presented. A standard method was introduced for representing functions of class B(m, n), and an integral transform was constructed using the Fourier

summation kernel. This construction resulted in a flexible way to present some necessary and sufficient conditions for a function of class B(m, n) to be a Mexican hat wavelet and MHWST.

Hu et al. developed an algorithm for uncertainty quantification analysis using the Monte Carlo technique for extreme value statistics (X. Hu et al., 2023). The study focused on three commonly used probability distributions and developed simplified models of the standard error for extreme value quantiles as a function of various factors. The results showed that the standard error increased with the return period and distribution parameters but decreased with the sample size. The simplified models were then applied to extreme event analyses of wind speed and precipitation at several typical sites and to develop a wind hazard map for non-typhoon winds in mainland China. The paper demonstrated the effectiveness and practicality of the proposed method.

In a study (X. Liu et al., 2021), the researchers proposed a Seasonal Auto-Regression Integrated Moving Average (SARIMA) model for predicting hourly wind speeds in Scotland's coastal/offshore area. They used three wind speed time series collected at different elevations from a coastal met mast. To compare SARIMA's performance, they also developed deep-learning-based algorithms of Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM). Despite the recent advancements in machine learning algorithms, the researchers found that SARIMA outperformed both GRU and LSTM in accuracy for forecasting future lags of offshore wind speeds over time. The SARIMA model was found to be the most accurate and robust among the tested predictive models for the corresponding datasets and forecasting horizons.

Branland investigated how gusts propagated through the row of turbines at the ECN test farm (Branlard, 2009). The analysis had never been conducted before, so the researcher took several steps before arriving at the propagation results. First, they defined gusts and computed several methods of gust detection in line with the different conventions recommended by the literature. Next, they established a database of gusts and compared the experimental mean gust shape with the theoretical gust shape. Finally, they selected specific gusts and studied their propagation.

Overall, these studies demonstrated the ongoing efforts to improve wind power technologies through various methods such as modeling, forecasting, design, and analysis.

2.2.2 Extreme Events & Mexican hat

A study of the measurements taken over a decade at a Danish coastal site, Høvsøre, found that wind speed events from offshore exceeded the prescribed extreme turbulence model for wind turbine safety (Hannesdóttir et al., 2019).

The variance was caused by ramp-like increases in wind speed from larger-scale meteorological processes, not severe turbulence. Filtering reduced the 50-year turbulence levels below the International Electrotechnical Commission (IEC) standard, but ramp-like events at certain wind speeds led to extreme tower-base loads.

Rapella et al. estimated the behavior of extreme winds on the European panorama from 1950 to 2020 (Rapella et al., 2023). This was done to examine the large-scale weather regimes associated with them and how they impact the availability of offshore wind energy. The study identified substantial alterations in the frequency of high and low extreme wind events, which proved that climate change or long-term internal climate variability has already impacted offshore wind power output. Additionally, the investigation of weather regimes revealed that high and low extreme wind events may occur concurrently across Europe.

In their paper, Mi et al. compared the relationship between scale and period in ecological pattern analysis and wavelet analysis, adapting the Morlet wavelet to ecological pattern analysis (Mi et al., 2005). Monte Carlo assessments and statistical significance tests were applied to wavelet analysis, and the properties of the Morlet and Mexican hat wavelets were investigated and compared using field data and artificial transects. They concluded that the Mexican hat provided better detection of patch and gap events, while the Morlet was better at detecting scale. Combining the advantages of both wavelets was the best approach. The characteristics of the wavelets affected their performance, and ecologists should combine information from different wavelets for a broader view and more precise pattern information.

Hou and Qin systematically studied the Mexican hat wavelet (MHW) on manifold geometry, including its derivation, properties, transforms, and applications (Hou & Qin, 2012). The MHW was rigorously derived from the heat kernel by taking the negative first-order derivative with respect to time. Its explicit expression in the Fourier domain was obtained by analyzing it and its transforms from a Fourier perspective, which is a scaled differential operator continuously dilated via heat diffusion. The MHW was localized in both space and frequency, enabling space-frequency analysis of input functions. Continuous and discrete transforms were defined as convolutions of bivariate kernels, and a fast method to compute convolutions by Fourier transform was proposed. The MHW was applied to graphics problems of feature detection and geometry processing to broaden its application scope.

Liu et al. proposed a new wind turbine de-noising method based on the Genetic Algorithm (GA) (W. Liu et al., 2016). First, the filtering characteristic of the Mexican hat wavelet was studied and improved with two shape-changing parameters. Then, the GA was used to optimize the shape parameters through the selection, genetic, and replacement process. The wavelet transform (WT)

scale factor was also optimized using the same arithmetic. Finally, the optimal Mexican hat wavelet and the scale factor were applied in the continuous wavelet transform to extract the valuable features of the analyzed signal. The experiment results, using an actual wind turbine gearbox vibration signal, proved the effectiveness and validity of the new method.

In another study, a de-noising method was proposed that was based on a Mexican hat wavelet with optimized parameters (W. Liu & Han, 2013). The researchers chose the Mexican hat wavelet as the mother wavelet because of its similarity in shape to mechanical shock vibration signals. They improved the wavelet's shape parameters through optimization to better filter out noise in raw vibration signals. The improved Mexican hat wavelet was then used in a continuous wavelet transform (CWT) to filter out noise jamming. The wavelet's shape parameters were optimized using the cross-validation method (CVM), and the optimal scale factor was obtained through the same method. Using the optimal shape parameters and scale factor in the CWT process enabled the extraction of useful components. Experimental results showed that the proposed method effectively de-noises and accurately extracts fault features.

Ren et al. proposed a new wind turbine weak feature extraction method using Cross Genetic Algorithm (CGA) optimized Mexican-Hat Wavelet (MHW) (H. Ren et al., 2017). The researchers started by improving the MHW by adding two parameters to alter the mother wavelet's waveform. Then, they used the CGA to optimize the parameters for establishing the mother waveform and continuous wavelet transform. They could extract weak features with the optimal parameters in the presence of substantial background noise interruption. The proposed method was tested on real wind turbine transmission system vibration signals, and the results showed that it effectively suppressed strong background noise and extracted wind turbine weak features accurately.

Tian focused on analyzing the dynamic behavior of wind power time series and its impact on wind power grid systems (Tian, 2023). Various methods, such as power spectral density analysis, autocorrelation function analysis, and information entropy method, were used to study the chaotic dynamics of wind power time series at different time scales. The research involved the examination of actual wind power data from a wind farm. The findings of the study revealed that wind power time series is non-stationary and non-white noise and that wind power is unpredictable in the long term. The output level of wind power is dependent on the time scales, and the wind power time series obeys fractal Brownian motion with chaotic and fractal characteristics. The study concluded that the maximum prediction horizon changes with the time scales and that wind power at large time scales is more chaotic. The research has practical implications for improving the prediction accuracy of wind power and understanding its fluctuation law.

The selected papers cover various aspects of extreme events and Mexican hat wavelets, including signal processing, pattern analysis, and prediction models. The use of Mexican hat wavelets in these studies demonstrates their effectiveness in detecting and filtering noise from signals, which is essential in various applications such as wind turbine systems and ecology. The studies also show the importance of combining different wavelets and methods to obtain more accurate and comprehensive results. Additionally, the prediction models presented in the papers highlight the challenges in forecasting extreme events and the need for robust and precise models. Overall, these papers contribute to advancing research on extreme events and Mexican hat wavelets in various fields.

2.2.3 Wind Power Predictive Maintenance

Developing an efficient maintenance strategy can significantly minimize the energy cost produced by offshore wind farms, reducing the expenses incurred for operations and maintenance. This highlights the crucial role of maintenance strategy design in achieving cost reduction objectives. The impact of supervision on the life cycle of offshore wind farms is challenging to determine due to its complex and uncertain nature.

Zhou and Yin (Zhou & Yin, 2019) proposed a dynamic opportunistic condition-based maintenance strategy for offshore wind farms using predictive analytics. The strategy considered varying maintenance lead time and economic dependence between wind turbines and components. A maintenance model was developed and demonstrated to be effective in reducing annual maintenance costs compared to widely employed and simple strategies.

In a study, Tchakoua et al. explored a comprehensive review and classification of wind turbine condition monitoring (WTCM) methods and techniques, with a particular focus on current trends and future challenges (Tchakoua et al., 2014). Condition monitoring, diagnosis, and maintenance analysis are examined to elucidate the relationships between these concepts and related theories, and emerging trends and challenges in the WTCM industry are identified. The findings from this research offer valuable insights into the strengths and weaknesses of the current WTCM industry and highlight research priorities needed to address the wind energy industry's evolving technological and market demands.

Li et al. proposed a new double-rotor wind turbine generation system to overcome the traditional rigid connection wind turbine's shortcomings in rejecting impact load disturbance (Li et al., 2021). The study involved carrying out a concept design of the new system, deriving the balance equations of speed, torque, and power, and analyzing the system's characteristics under different wind conditions through simulation to evaluate its anti-impact load

capacity. The simulation results showed the superiority of the proposed novel double-rotor wind turbine generation system.

Chapter 3

Methodology

3.1 Data Description and Processing

The wind data used in this master's thesis report was collected from a meteorological station located at FINO1 at the North Sea platform. The data set covers a period of 15-month, ranging from October 2015 to December 2016.

The wind speed measurements were made every one second at eight different heights, starting at 30 meters above sea level and going up every ten meters to a height of 100 meters.

FINO research platforms are located in the North Sea and Baltic Sea (Figure 3.1). The FINO1 research platform is positioned near places where wind farms are being constructed or already used. The Federal Ministry for Economic Affairs and Energy (BMWi) is sponsoring research to lower wind energy costs, achieve economies of scale, and improve wind turbine reliability. (Lehmann-Matthaei, 2023) Essential outcomes for optimizing offshore wind farms are derived using the measurement data gathered at FINO1.

A 32 Mb/s directional radio connection to the island of Borkum is used to send data that was captured on the platform to the mainland. Data are transmitted via the landline system on Borkum, where they are subsequently relayed to the German research network. The representatives of measurement institutions can then obtain and process the raw data from this location. The majority of the findings are entered into the FINO-database, which is open to the public.

However, it is important to note that our data set does not include wind speed measurements at sea level or close by, which may restrict our capacity to properly understand the impact of drag along the wind profile. Nevertheless, this study offers insightful information about how wind speeds behave at various heights and could be a great resource for future studies in the area of wind energy.

Figure 3.1: FINO1 Platform in the North Sea (Lehmann-Matthaei, 2023)

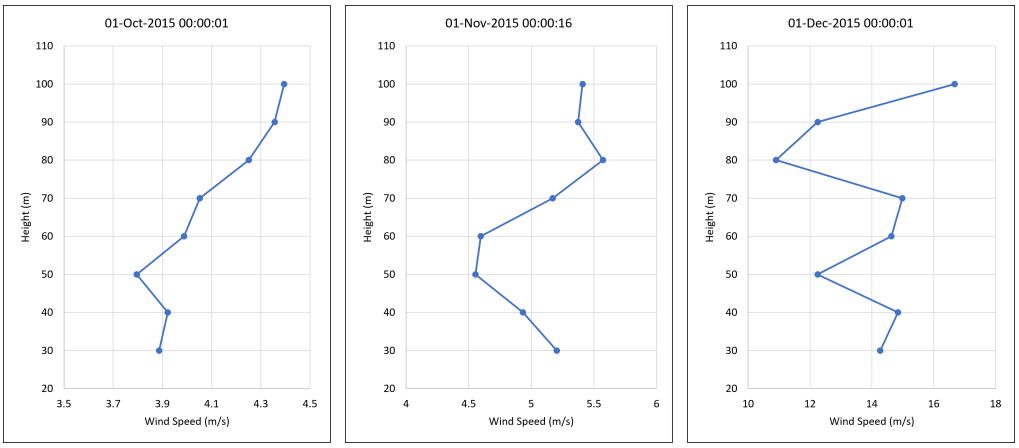


Figure 3.2: Illustration of the fluctuation and heterogeneity of wind patterns at FINO1 at three distinct timestamps.

In addition, the theory of wind profiles in the atmospheric boundary layer suggests that wind speed is typically lower when closer to the ground and increases with altitude, following a log-law or power-law, depending on elevation (Abubaker et al., 2018). However, it is essential to note that in practical situations, this theory may only sometimes hold true in practice, especially when studying high-frequency wind data. In our dataset, we observed that the wind behavior along the vertical axis was often non-uniform. Figure 3.2 illustrates three examples of this behavior. This non-uniformity can be attributed to various factors, such as changes in atmospheric

stability and topographical features. It is crucial to consider and account for such non-uniform behavior when studying wind patterns to ensure accurate predictions and design of wind-dependent structures, including wind turbines. Ignoring this non-uniform behavior could result in inaccurate predictions and suboptimal designs, which can negatively affect the performance and safety of wind turbines.

3.2 Methodology Definition

In order to complete the project, we adopted a sequence of methods that were selected based on their effectiveness in current research and the resources available to us. These methods were chosen in a way that allowed for a smooth progression from more straightforward tools to more sophisticated ones, in keeping with the increasing complexity of our analyses. By utilizing these techniques, we achieved our goals and drew meaningful conclusions from our data.

3.2.1 Detrending

In many applications, trends in the data can mask the underlying patterns that are of interest. Detrending is the process of removing a trend from a time series to isolate the underlying patterns of the data. It is essential in analyzing time series data, as it allows us to focus on the underlying patterns in the data and ignore any long-term trends that may obscure the patterns. Trends in time series can arise due to various factors such as seasonality, cyclic patterns, or changes in the overall trend of the data. By removing these trends, analysts can more accurately identify the patterns in the data and make better predictions. Detrending is commonly used in fields such as finance, meteorology, and economics to study the behavior of variables over time.

In Python, detrending can be done using various methods, including linear regression, moving averages, and the scipy.signal.detrend function. In our study, we used the detrend function from the scipy.signal module which is one of the most popular methods. This function uses a linear least-squares fit to remove the linear trend, or higher-order polynomial trends by specifying the order parameter, from the data. It subtracts the best-fit line from the data and returns the detrended time series. Figure 3.3 presents the wind speed data of the FINO1 dataset for a certain timeframe and how the downward trend of the original data (blue line) is removed using the scipy.signal.detrend function (orange line).

The detrended data can then be analyzed to identify patterns or features that are not obscured by the overall trend. After detrending, analysts can then compare the similarity of the underlying patterns in different time series using methods such as cross-correlation or wavelet analysis.

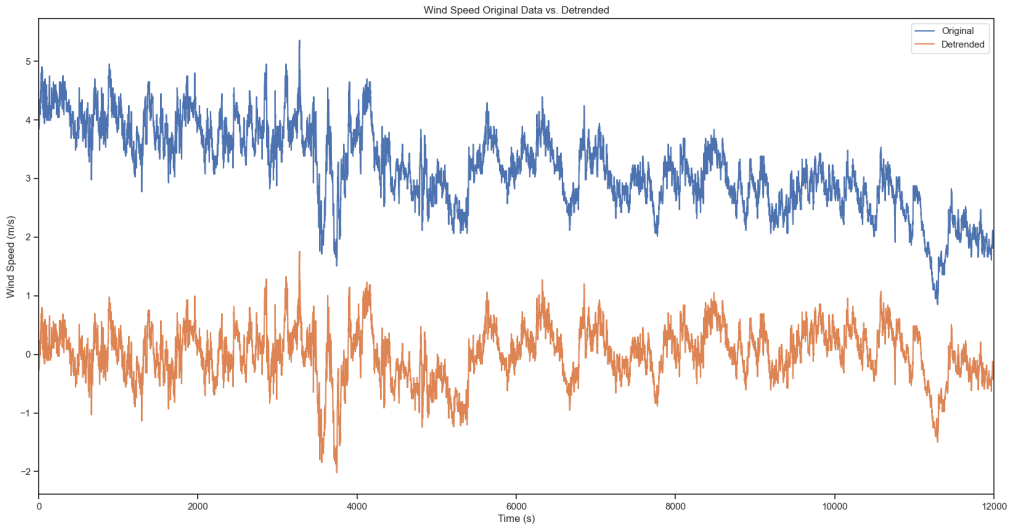


Figure 3.3: Detrending the FINO1 wind speed data using scipy.signal.detrend function: original data (blue) vs. detrended (orange)

3.2.2 Rolling-Window Technique

The Rolling-Window technique is a commonly used approach in data analysis and time-series analysis for identifying patterns and similarities in sequential data (Zivot et al., 2003).

In this technique, a sliding window of a fixed size is moved across the time-series data, and for each window position, a set of features is extracted from the data within the window. The features can include statistical measures such as mean, variance, and correlation or more advanced features such as wavelet coefficients or Fourier transforms. Figure 3.4 illustrates this concept on the dataset of size N samples and the rolling window size of m. The difference between successive windows is one period; therefore, the technique divides the entire dataset into N- m +1 subsamples.

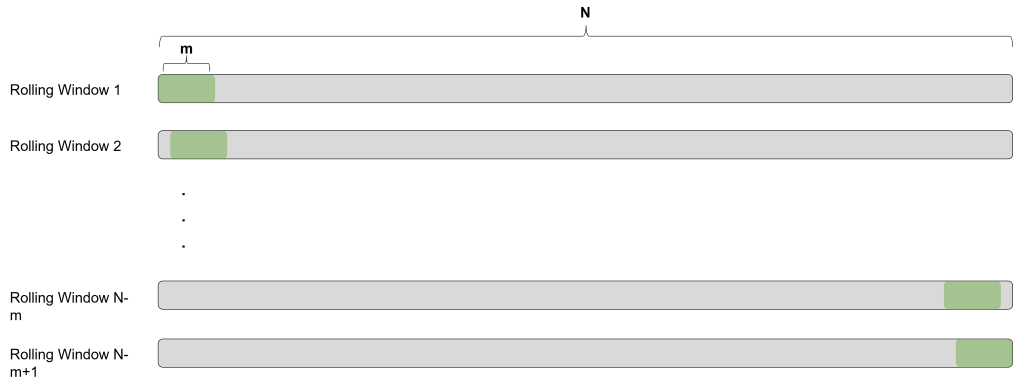


Figure 3.4: Illustration of Rolling Window Technique

Once the feature vector is obtained for each window position, the similarity between different time series can be evaluated by comparing the corresponding feature vectors. Various similarity metrics can be used for this purpose, such as Pearson Coefficient, Euclidean distance, cosine similarity, and Dynamic Time Warping (DTW).

The rolling-window technique is frequently deployed in various applications, including speech recognition, image processing, and financial data analysis. It can help identify patterns and trends in the data and can also be used for anomaly detection and prediction.

3.2.3 Pearson's Product Moment Correlation Coefficient(r)

The Pearson product-moment correlation coefficient (PPMCC) is an extensively used statistical approach that measures the strength and direction of a linear association between two variables, represented by **r**. it can range from -1 to 1 (Puth et al., 2014) , where a score of -1 implies a perfectly negative correlation, 0 implies no correlation, and 1 indicates a perfect positive correlation. Calculating the PPMCC involves computing the covariance between the two variables and dividing it by their standard deviations' product. Social sciences, finance, and engineering are some of the fields where PPMCC is commonly used to assess the degree of association between two variables. When it comes to time series data, PPMCC is particularly useful for comparing the similarity of time series. To compute the r coefficient for two time series, you can use the following formula:

$$r = \frac{\sum((X_i - \overline{X})(Y_i - \overline{Y}))}{\sqrt{\sum(X_i - \overline{X})^2 \sum(Y_i - \overline{Y})^2}}$$

where:

X_i and Y_i are the data points
$\overline{X}$ is the mean of the X-values
$\overline{Y}$ is the mean of the Y-values

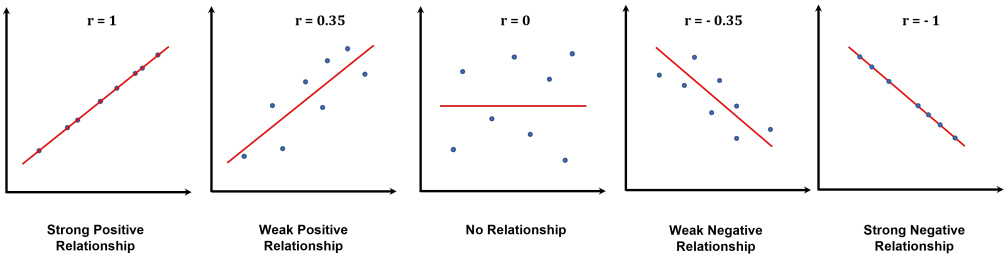


Figure 3.5: Pearson's Product Moment Correlation Coefficient

This method can help determine the extent and direction of correlation between two time series. A PPMCC close to +1 suggests a strong positive correlation, indicating that both series tend to move in the same direction.As illustrated in Figure 3.5, a PPMCC close to -1 implies a strong negative correlation, indicating that the two time series tend to move in opposite directions. A PPMCC close to 0 implies that there is no correlation between the two time series.

3.2.4 Weibull Distribution

The **Weibull Distribution** , which is named after the Swedish mathematician Waloddi Weibull, (Hallinan Jr, 1993) is a continuous probability distribution widely applied in reliability engineering and survival analysis for modeling the time-to-failure of systems or components. This distribution has two parameters, a Shape parameter (k) and a Scale parameter (λ) , and the probability density function (PDF) is given by:

$$f(x;k,\lambda) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$$

where **x** is the random variable.

The Weibull distribution can be utilized to model various phenomena, such as the lifetime of light bulbs, failure times of electronic components, and the time between customer arrivals at a service center.
The Shape parameter (k) determines the distribution's shape.If k<1, the distribution is called decreasing or left-skewed, with a long tail to the right. A k value of 1 results in an exponential distribution with a constant hazard rate, while a k>1 produces an increasing or right-skewed distribution with a long tail to the left.
The Scale parameter λ determines the scale of the distribution. It represents the typical or average time-to-failure of the component or system. As λ increases, the distribution becomes more spread out, while as λ decreases, the distribution becomes more concentrated around the origin.

The Weibull Distribution is also utilized to model wind speeds in most parts of the world. By analyzing this distribution, we can determine the frequency at which winds of different speeds occur at a specific location with a certain average wind speed. This information is useful when selecting a wind turbine, as it helps to identify the optimal cut-in and cut-out speeds of the turbine.

The shape of the Weibull Distribution is controlled by its Shape parameter. Figure depicts 3.6 five Weibull distributions, each with an average wind speed of 6 m/s but varying Weibull k values. Broader distributions are associated with lower k values, as illustrated in the graph(HOMER-Energy, 2021).

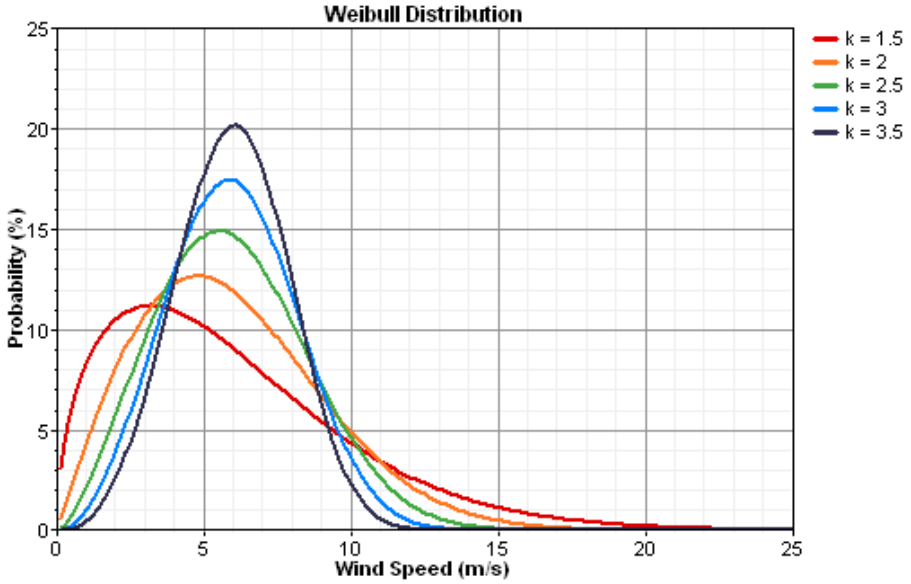


Figure 3.6: Effect of Shape parameter(k) on Weibull distribution graph, (HOMER-Energy, 2021).

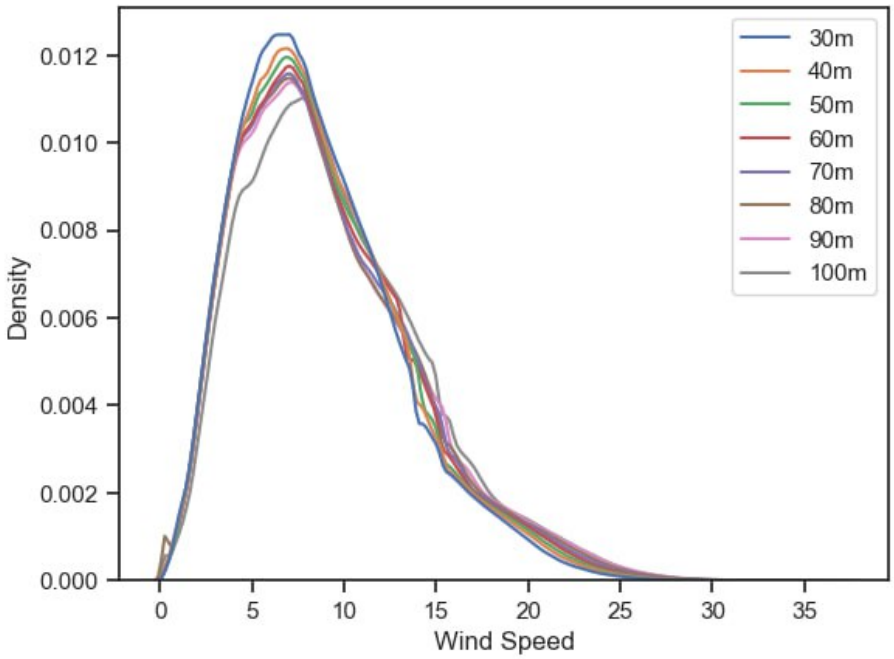


Figure 3.7: Wind speed distribution of the FINO1 dataset at different elevations

In most locations around the world, including Northern Europe, the Shape value is around 2; therefore, wind turbine manufacturers, as a rule of thumb, use a Shape value of 2 in their performance figures. A higher Shape value corresponds to a higher median wind speed. Locations with a lot of low wind

speeds and some very strong winds would have a Shape value below 2, while locations with reasonably constant wind speeds close to the median would have a Shape value of 3. We can calculate the probability of a particular wind speed occurring at a specific location using the Weibull Distribution. This information can be utilized to estimate the likelihood of the occurrence of certain wind speeds in hours per year and hence, the overall power output of a wind turbine in a year.

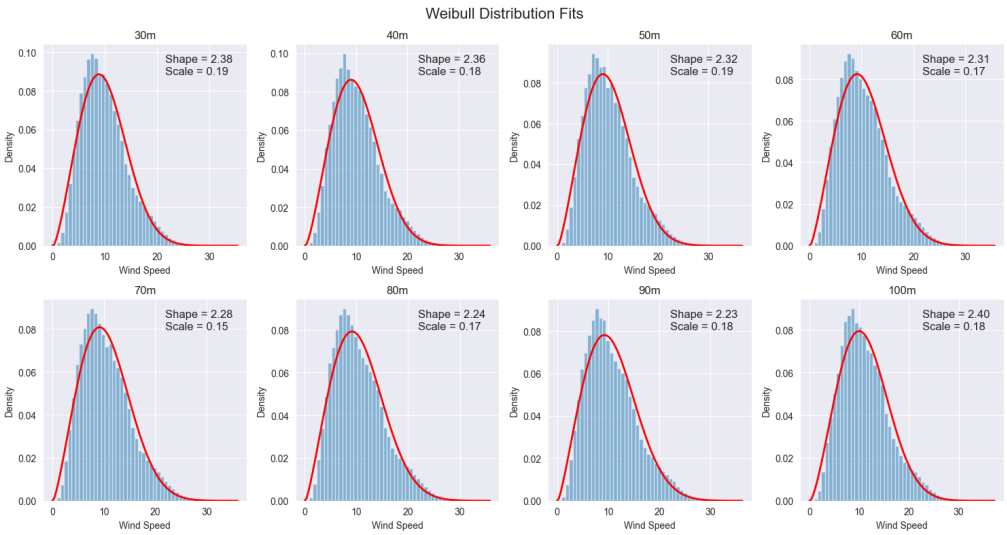


Figure 3.8: Distribution of the wind speed data for the FINO1 project at different elevations, their Weibull fit, and associated shape and scale parameters

Figures 3.7 and 3.8 illustrate the distribution of the wind speed data for the FINO1 project at different elevations, their Weibull fit, and associated shape and scale parameters.

3.2.5 Mexican hat Function Definition

For the definition of the Mexican hat function in our model, we utilized the Mexican hat wavelet, also known as the Ricker wavelet formula (Ryan, 1994). Additionally, to enhance the flexibility of this function and allow for the creation of Mexican hats with various amplitudes, we introduced a new variable, denoted by **A**. This variable controls the amplitude of the Mexican hat without affecting other properties such as width. Incorporating the A variable empowered the creation of Mexican hats with various amplitudes, which gave us more flexibility in our analyses and, consequently, higher precision in detecting the Mexican hat events in our wind dataset:

$$f(t, A, b) = \frac{2A}{\sqrt{3b}\pi^{\frac{1}{4}}} \left(1 - \frac{t^2}{b^2}\right) \exp\left(-\frac{t^2}{2b^2}\right)$$

where:
t is the time
A is a variable that controls the amplitude (a in Figure 3.8) of the Mexican hat; and b is the rise time of the Mexican hat.

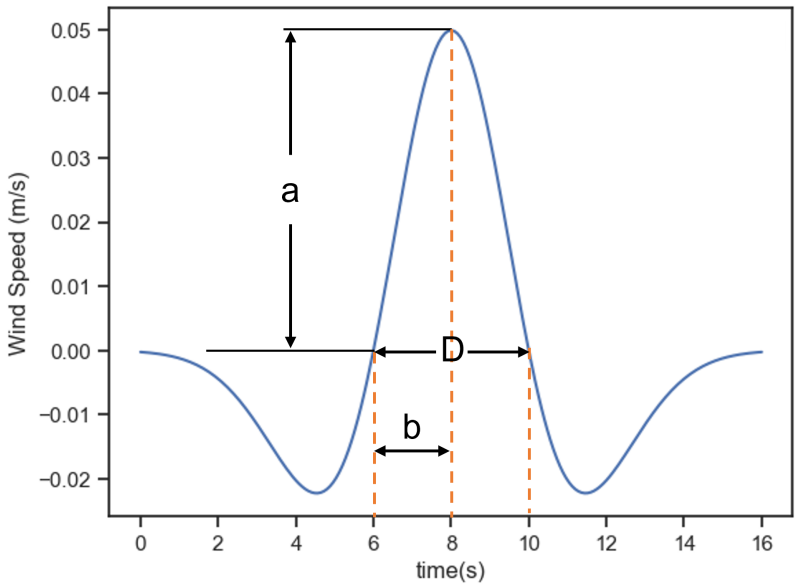


Figure 3.9: Illustration of Mexican hat parameters. a is the amplitude, b is the rise time; and D is the width of the Mexican hat.

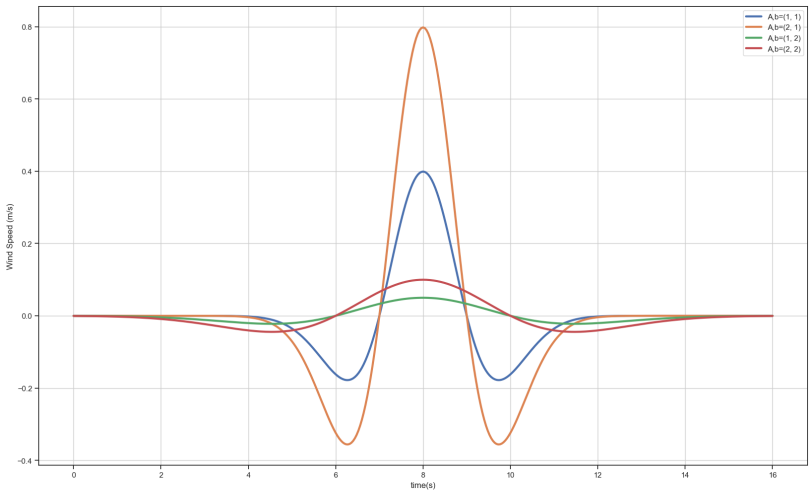


Figure 3.10: The Mexican hat function for several combinations of A and b

Figure 3.9 illustrates the definition of Mexican hat parameters, and Figure 3.10 depicts the impact of A and b on the amplitude and width of the Mexican

hat, respectively. Since the shape of a Mexican hat is symmetric, the width, D, equals 2b.

3.3 Methodology Plans

As the first step, we had to preprocess the data to ensure our analysis was accurate. The first step was to identify and remove any outliers that could potentially affect the results, such as extremely high or low wind speeds that were unlikely to be accurate. Additionally, we checked for duplicated entries and measurement unit errors. If the interval for these errors was short, less than five seconds, we used the linear interpolation technique to calculate the missing values. However, if there were longer intervals of invalid measurements, we removed them from the study. Once we had a clean dataset, we applied various statistical methods and analyses to identify patterns or trends in the data.

Our main objective was to detect the Mexican hat events in the wind speed data. To achieve this, we employed the rolling window technique, whereby a subset of the data was selected. Then, for this subset, i.e., window, we first detrended the wind speed data and then compared the similarity of the detrended wind speed data with the Mexican hat function for a given A and b, using the Pearson product-moment correlation coefficient (PPMCC). This comparison generated a match score for the central point of the window (peak), which we recorded. We repeated this process by moving the window forward by one second until we reached the end of the dataset.

To increase the model's ability to detect Mexican hats of various shapes and sizes, we employed a mesh grid of A and b combinations. Also, due to the large size of the dataset, we used vectorized variables rather than a nested loop. Furthermore, as illustrated in Figure 3.8, the total duration of a Mexican hat, i.e., from the moment the value becomes non-zero till the final point reaches zero again, is 8b; therefore, in our similarity detection analysis between the FINO1 wind speed data and the Mexican hat function, instead of using a fixed rolling window size, we used dynamic window sizes of 8b.

After we had all the match scores, we reviewed them to find the combination that generated the highest score and dropped the rest.

Finally, To be able to analyze the frequency of Mexican hat events as a function of elevation and review the changes with height, we repeated this process for all eight elevations.

Chapter 4

Analysis of the Results

In this chapter, we will provide a comprehensive analysis of the performance of this algorithm across a range of different experimental conditions. We will begin by presenting the data obtained from our extensive simulations, which were conducted to evaluate the accuracy and robustness of the algorithm under different levels of noise and Mexican hat sizes. Then, we will present the results of our experiments on the FINO platform data and at various elevations. In addition to presenting the raw data, we will also provide a detailed analysis of our findings.

4.1 Quality Check of the Model Outputs

As the first step of the analysis process, the wind speed data from a selected dataset and at various elevations were examined using the Mexican hat detection model. This stage involved comparing the wind speed data with predefined Mexican hat models through cross-correlation. Figure 4.1 illustrates detected Mexican hat events over a sample dataset.

The model output included the following:

- A match score
- **A** and **b** values
- The amplitude of the (potential) Mexican hat event
- The velocity of the wind speed data for each point.

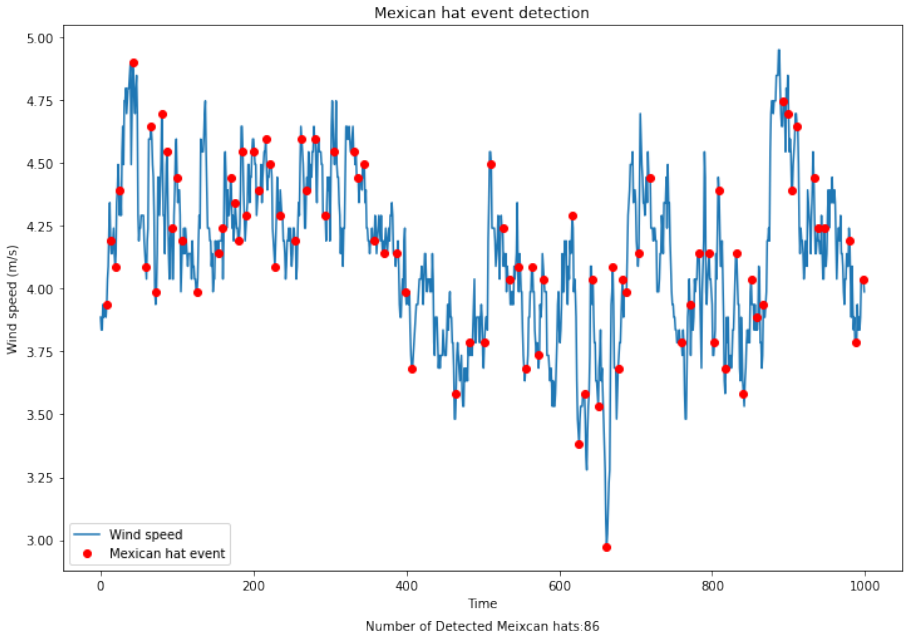


Figure 4.1: Detected Mexican hat events over a smaple dataset

The match score indicated the level of similarity between the Mexican hat models and the detrended wind speed data at a specific point. The 'A' and 'b' values represented the amplitude and width of the predefined Mexican hat templates that best matched the wind speed data, respectively. Figure 4.2 visually represents a sample of the model result, showcasing different outputs for each data point.

Time Index	MH Events Time	Wind Speed (m/s)	Optimum D (MH Width)	Optimum A (MH Amplitude)	Measured Amplitude	Match Score
6	2015-10-01 00:00:07	4.19167	1.0	19.5	0.17634714285710718	0.878489850589355
13	2015-10-01 00:00:14	4.59751	2.8	19.5	0.5453447198978874	0.8901477518067505
20	2015-10-01 00:00:21	4.49605	1.5999999999999999	28.5	-0.24847057342656886	-0.8787408260093634
25	2015-10-01 00:00:26	4.49605	4.399999999999995	27.5	-0.30661364084204467	-0.8255353987185451
31	2015-10-01 00:00:32	4.64824	3.999999999999999	17.5	0.4645198547071827	0.86404387884434
32	2015-10-01 00:00:33	4.59751	2.5999999999999996	19.5	0.5592312596556509	0.8883934408840556
33	2015-10-01 00:00:34	4.59751	2.3999999999999995	13.5	0.4476703501850079	0.8576661481187704
35	2015-10-01 00:00:36	4.69897	1.0	7.5	0.15218999999997354	0.8793791773490157
42	2015-10-01 00:00:43	4.85116	2.5999999999999996	2.5	0.10123769539970875	0.8286608669354575
45	2015-10-01 00:00:46	4.7497	2.9999999999999996	12.5	-0.20485427749602202	0.8261529699624421
50	2015-10-01 00:00:51	5.05408	1.9999999999999998	5.5	0.19805064300407543	0.8140249254865124
53	2015-10-01 00:00:54	5.40919	2.3999999999999995	1.0	0.10022278969045284	0.8342217468191039
54	2015-10-01 00:00:55	5.257	2.1999999999999997	3.5	0.15809903619745413	0.8510175849154966
57	2015-10-01 00:00:58	4.85116	1.5999999999999999	20.5	-0.14687910530637482	-0.9081813494335303
58	2015-10-01 00:00:59	4.85116	2.3999999999999995	12.5	0.18158976588568246	0.911444163378033
61	2015-10-01 00:01:02	4.80043	1.0	29.5	0.10870714285713423	0.9434516392498774
62	2015-10-01 00:01:03	4.85116	3.1999999999999993	19.5	-0.09071719578033875	-0.8675789108888395
63	2015-10-01 00:01:04	4.85116	1.7999999999999998	18.5	0.19173642328348603	0.8483850615406021
64	2015-10-01 00:01:05	4.64824	3.1999999999999993	1.5	0.4891375243033409	0.8800568620249436
66	2015-10-01 00:01:07	4.69897	2.9999999999999996	16.5	0.338665212196505	-0.8069978380323102
71	2015-10-01 00:01:12	4.44532	1.5999999999999999	4.5	0.17961049362402143	-0.8127139389118343
72	2015-10-01 00:01:13	4.64824	2.9999999999999996	3.5	0.401862289761727	0.9359845954619213
74	2015-10-01 00:01:15	4.2424	3.399999999999995	11.5	-0.11026350272700088	-0.8544892231993575

Figure 4.2: Sample output of the Mexican hat detection model

In order to certify the precision and dependability of the output data, we performed a meticulous manual quality control process. This approach involved carefully examining the model output to confirm the successful implementation of the model.

To provide a visual representation of this process, we have included Figure 4.3, which displays the manual quality check of four data points with varying

match scores. The Figure clearly illustrates the degree of alignment between the wind speed data and the Mexican hat models at different match scores. As depicted, the highest match score marks the closest resemblance to Mexican hats. Overall, this quality check underscores the accuracy and reliability of our data output, giving us confidence in the validity of our findings.

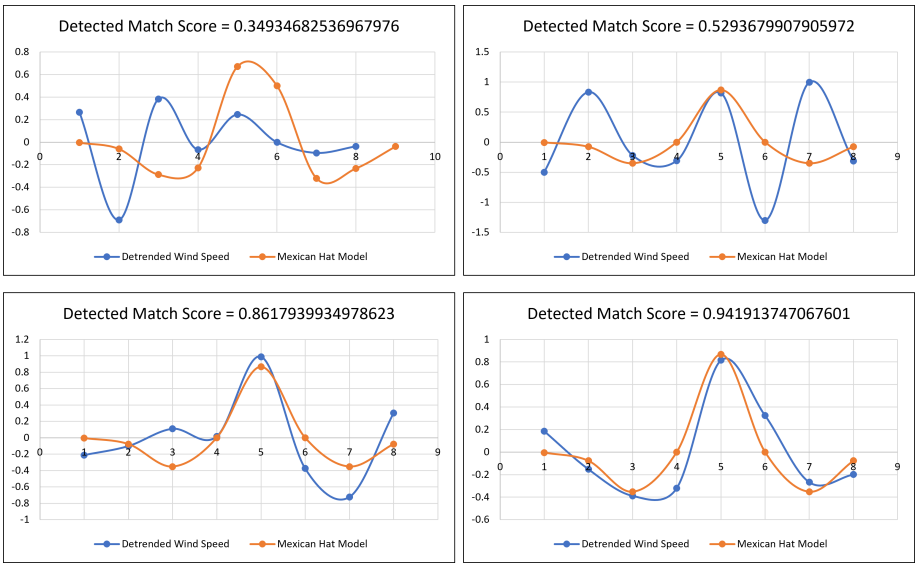


Figure 4.3: Manual quality check of four points with different match scores

4.2 Statistical Analysis of the Model Outputs

After gaining confidence in the accuracy and reliability of our output data in the previous steps, we proceeded to execute our model on a much larger dataset. We then conducted a comprehensive analysis of the output data distribution for three critical parameters: the amplitudes and widths of the Mexican hats, as well as the wind speed at their peak points. As we anticipated the possibility of poor match data points, we established six distinct thresholds that ranged from 0.5 to 0.9 match scores. These thresholds helped us filter out any imprecise data points, thus ensuring that our analysis would be based on proper matches only. To summarize our findings, we present a statistical summary in Appendix A that outlines the properties of the identified Mexican hats based on the specified thresholds. From our analysis, we were able to draw some key insights:

- According to the data presented in Figure 4.4, the number of detected Mexican hats was slightly lower at elevations of 50m, 60m, and 70m, with the 60m height having the lowest number of detected Mexican hats. Also, the results showed that while the 100m elevation had a lower Mexican hat count than the other elevations at lower thresholds, it had the highest

count at higher thresholds. Specifically, at a threshold of 0.5, the 100m elevation ranked 7th in terms of Mexican hat count; but at a higher threshold of 0.9, it ranked 1st. On the other hand, the 90m elevation had the highest Mexican hat count at lower thresholds but fell behind at higher thresholds. These findings are detailed in Figure 4.4.

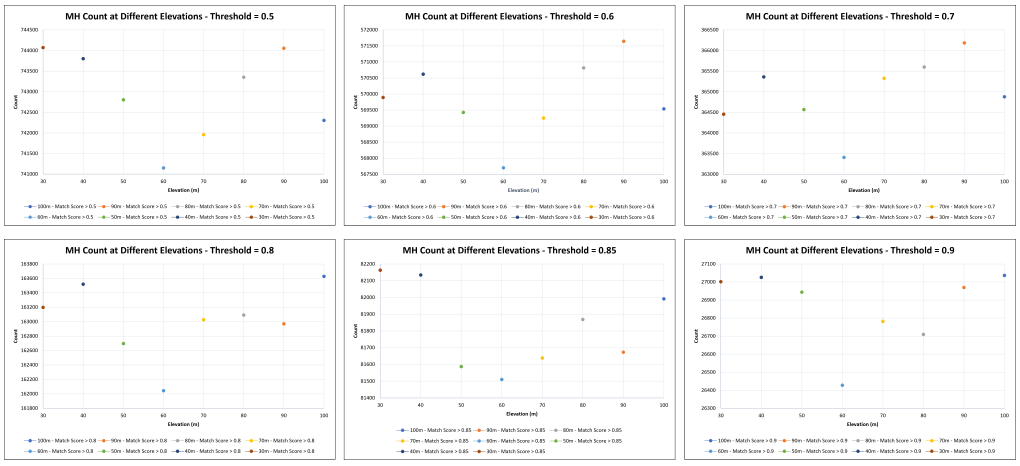


Figure 4.4: Mexican hat counts at different elevations and thresholds

- Meanwhile, Figure 4.5 shows that the decline in the number of detected Mexican hat events were similar across all elevations, indicating that height did not significantly affect the overall trend of the Mexican hat events.

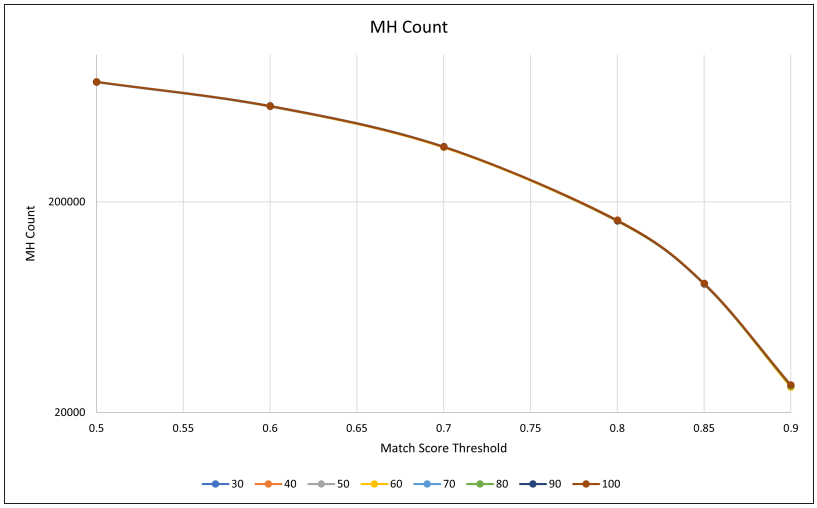


Figure 4.5: Trend of the Mexican hat count trend at different elevations and thresholds

- Figures 4.6 and 4.7 illustrate that the mean widths and amplitudes of the Mexican hat decreased as the match score threshold increased, and

this trend was consistent across all elevations. This means that as the threshold for detecting Mexican hat became stricter, the width and amplitude of the detected Mexican hat decreased, regardless of the height.

- Figure 4.8 shows that the maximum Mexican hat amplitude decreased with increasing the threshold at all elevations. However, the behavior was inconsistent, implying that the impact of elevation on maximum Mexican hat amplitude is not straightforward and may depend on other factors.

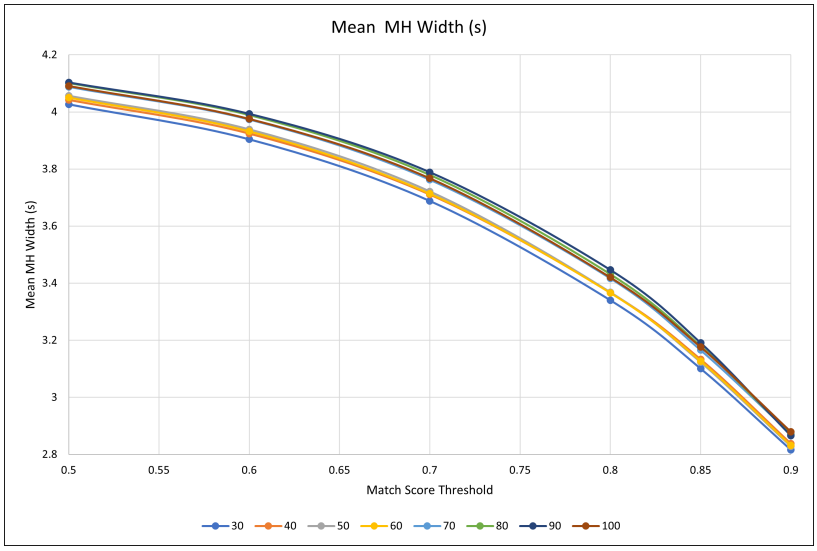


Figure 4.6: Trend of the mean Widths of the Mexican hats at different elevations and thresholds

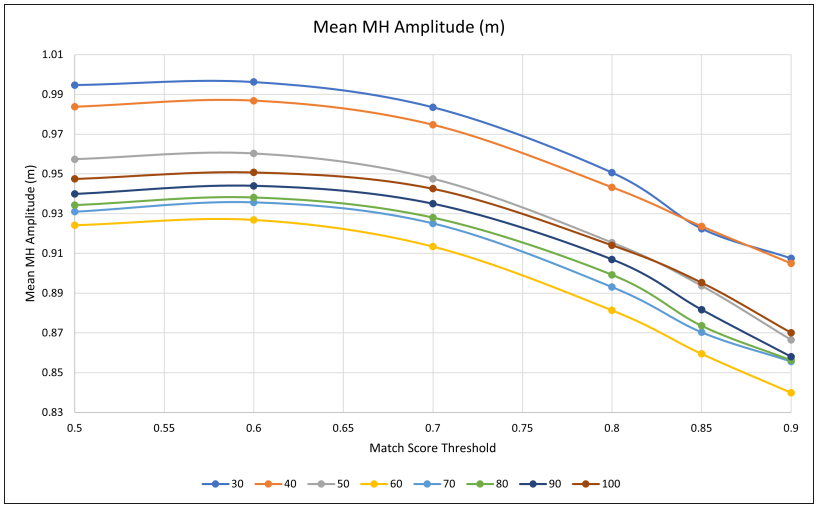


Figure 4.7: Trend of the mean Amplitudes of the Mexican hats at different elevations and thresholds

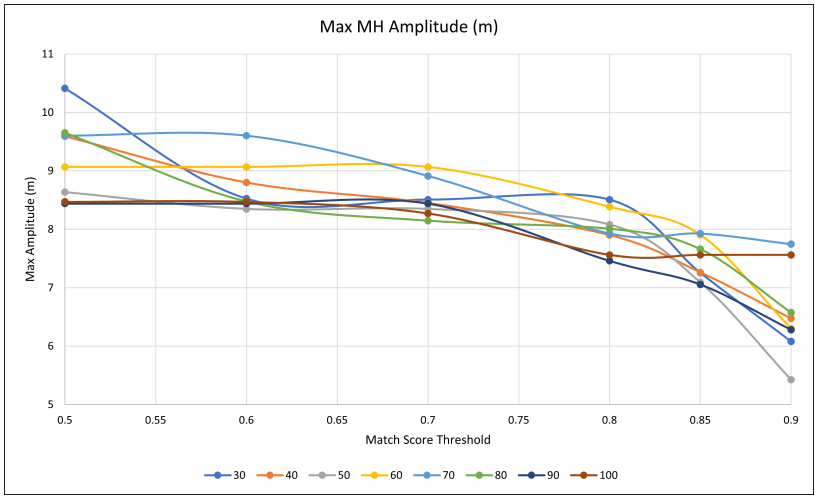


Figure 4.8: Trend of the max Amplitudes of the Mexican hats at different elevations and thresholds

- Figures 4.9 and 4.10 depict the relationship between elevation, match score threshold, and wind speed at Mexican hat peaks. As shown, there was a slight increase in both mean and max wind speed at Mexican hat peaks with elevation, regardless of the match score threshold. However, the mean wind speed at Mexican hat peaks displayed a small decline with an increase in the match score threshold for all elevations except for 90m, where the trend was reversed. On the other hand, the max wind speed at Mexican hat peaks remained constant for the 90m and 100m elevations, while all other heights experienced a decrease in max wind speed at higher thresholds, particularly after the 0.8 cutoff.

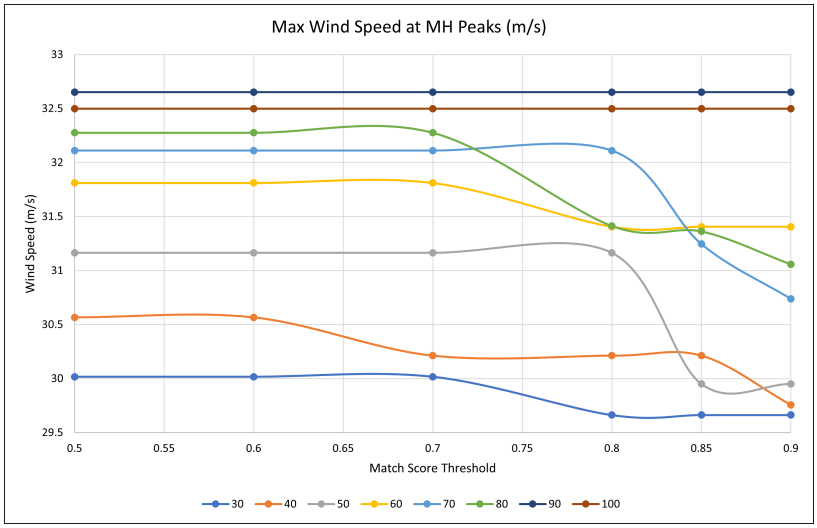


Figure 4.9: Trend of the max Wind Speed at Mexican hats Peaks at different elevations and thresholds

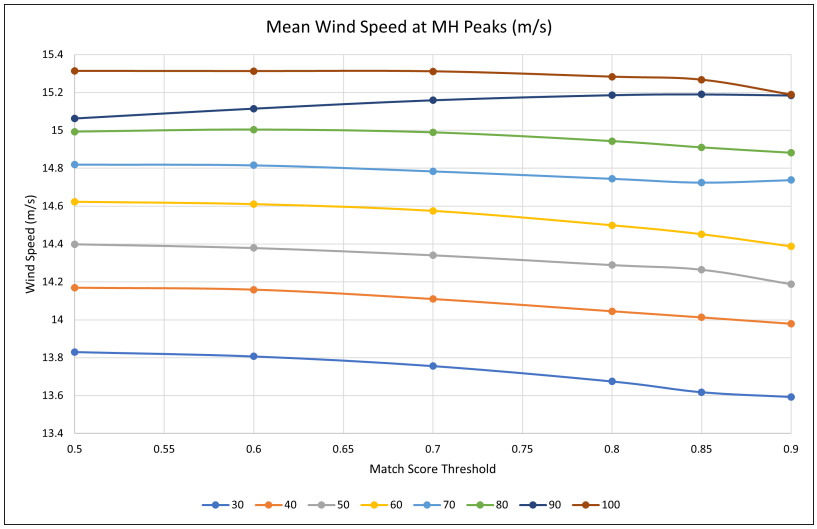
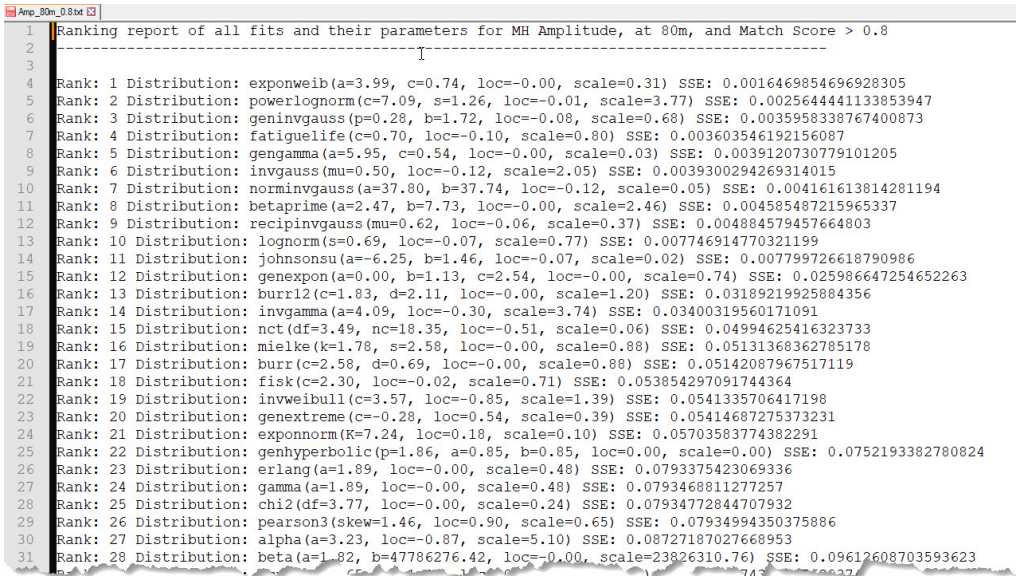


Figure 4.10: Trend of the mean Wind Speed at Mexican hats Peaks at different elevations and thresholds

To further refine our analysis, we endeavored to determine the optimal correlation that would effectively match the distribution of data across various elevations and each of the six unique thresholds. To accomplish this, we utilized the scipy.stats platform to apply 101 different statistical probability distribution functions and find the optimal fit for all 144 combinations of parameters (3), thresholds (6), and elevations (8).

To establish the most appropriate probability distribution functions, we developed a comprehensive ranking list that rated the degree of fit of each distribution function in all 144 variations. This list comprised the coefficients for each probability distribution function and the sum of squares error (SSE) of the fit, which enabled us to evaluate the accuracy and reliability of our results more effectively.

In Figure 4.11, a segment of the ranking list is presented, depicting the distribution model fits for the Mexican hat amplitude at the elevation of 80 m for the data points with a match score of more than 0.8.



```
Amp_80m_0.8.txt
Ranking report of all fits and their parameters for MH Amplitude, at 80m, and Match Score > 0.8
-----------------------------------------------------------------------------

Rank: 1 Distribution: exponweib(a=3.99, c=0.74, loc=-0.00, scale=0.31) SSE: 0.0016469854696928305
Rank: 2 Distribution: powerlognorm(c=7.09, s=1.26, loc=-0.01, scale=3.77) SSE: 0.0025644441133853947
Rank: 3 Distribution: geninvgauss(p=0.28, b=1.72, loc=-0.08, scale=0.68) SSE: 0.0035958338767400873
Rank: 4 Distribution: fatiguelife(c=0.70, loc=-0.10, scale=0.80) SSE: 0.003603546192156087
Rank: 5 Distribution: gengamma(a=5.95, c=0.54, loc=-0.00, scale=0.03) SSE: 0.0039120730779101205
Rank: 6 Distribution: invgauss(mu=0.50, loc=-0.12, scale=2.05) SSE: 0.0039300294269314015
Rank: 7 Distribution: norminvgauss(a=37.80, b=37.74, loc=-0.12, scale=0.05) SSE: 0.004161613814281194
Rank: 8 Distribution: betaprime(a=2.47, b=7.73, loc=-0.00, scale=2.46) SSE: 0.004585487215965337
Rank: 9 Distribution: recipinvgauss(mu=0.62, loc=-0.06, scale=0.37) SSE: 0.004884579457664803
Rank: 10 Distribution: lognorm(s=0.69, loc=-0.07, scale=0.77) SSE: 0.007746914770321199
Rank: 11 Distribution: johnsonsu(a=-6.25, b=1.46, loc=-0.07, scale=0.02) SSE: 0.007799726618790986
Rank: 12 Distribution: genexpon(a=0.00, b=1.13, c=2.54, loc=-0.00, scale=0.74) SSE: 0.025986647254652263
Rank: 13 Distribution: burr12(c=1.83, d=2.11, loc=-0.00, scale=1.20) SSE: 0.03189219925884356
Rank: 14 Distribution: invgamma(a=4.09, loc=-0.30, scale=3.74) SSE: 0.03400319560171091
Rank: 15 Distribution: nct(df=3.49, nc=18.35, loc=-0.51, scale=0.06) SSE: 0.04994625416323733
Rank: 16 Distribution: mielke(k=1.78, s=2.58, loc=-0.00, scale=0.88) SSE: 0.05131368362785178
Rank: 17 Distribution: burr(c=2.58, d=0.69, loc=-0.00, scale=0.88) SSE: 0.05142087967517119
Rank: 18 Distribution: fisk(c=2.30, loc=-0.02, scale=0.71) SSE: 0.053854297091744364
Rank: 19 Distribution: invweibull(c=3.57, loc=-0.85, scale=1.39) SSE: 0.0541335706417198
Rank: 20 Distribution: genextreme(c=-0.28, loc=0.54, scale=0.39) SSE: 0.05414687275373231
Rank: 21 Distribution: exponnorm(K=7.24, loc=0.18, scale=0.10) SSE: 0.05703583774382291
Rank: 22 Distribution: genhyperbolic(p=1.86, a=0.85, b=0.85, loc=0.00, scale=0.00) SSE: 0.0752193382780824
Rank: 23 Distribution: erlang(a=1.89, loc=-0.00, scale=0.48) SSE: 0.0793375423069336
Rank: 24 Distribution: gamma(a=1.89, loc=-0.00, scale=0.48) SSE: 0.0793468811277257
Rank: 25 Distribution: chi2(df=3.77, loc=-0.00, scale=0.24) SSE: 0.07934772844707932
Rank: 26 Distribution: pearson3(skew=1.46, loc=0.90, scale=0.65) SSE: 0.07934994350375886
Rank: 27 Distribution: alpha(a=3.23, loc=-0.87, scale=5.10) SSE: 0.08727187027668953
Rank: 28 Distribution: beta(a=1.82, b=47786276.42, loc=-0.00, scale=23826310.76) SSE: 0.09612608703593623
```

Figure 4.11: Ranking report of all fits and their parameters for Mexican hat Amplitude at 80m, and match score > 0.8

After conducting a comprehensive analysis of the results, we concluded that there is no single statistical probability distribution function that can perfectly capture the distribution of a parameter across all elevations and thresholds. However, we did observe that certain probability distribution functions consistently outperformed others and displayed very good fits to the data distribution, with a minimal sum of squared errors (SSEs) in most cases.

Therefore, to determine the most suitable distribution functions for representing our data distribution, we focused on the top-ranked correlations and complemented our analysis by performing KS-Test and Q-Q plot analyses on these distribution functions. These analyses allowed us to identify the best-fitting distribution functions. Figure 4.12 illustrates an example of the Q-Q plot analysis for the Mexican hat amplitude data distribution and Exponentiated Weibull distribution function.

Armed with this understanding, we successfully identified a set of probability distribution functions that emerged as the most dependable choices. These selections were made based on their consistent performance across different elevations and thresholds, as well as their ability to closely resemble the observed data distribution.

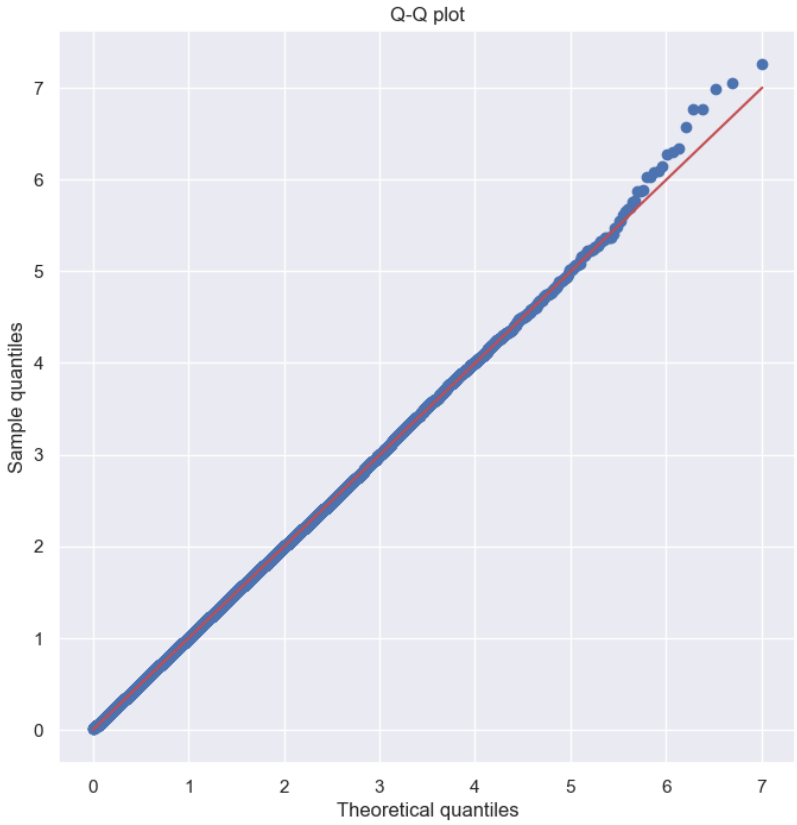


Figure 4.12: Sample Q-Q plot analysis: Mexican hat amplitude vs. Exponentiated Weibull function

- For the distribution of the wind speed at the Mexican hats peaks, the **Reciprocal Inverse Gaussian** distribution function (scipy.stats.recipinvgauss), was found to provide the best fit in most of the situations. This function is defined as:

$$f(x,\mu) = \frac{1}{\sqrt{2\pi x}} \exp\left(\frac{-(1-\mu x)^2}{2\mu^2 x}\right)$$

for $x \geq 0$.

where μ is the shape parameter.

Table 4.1 reports the SSE of this distribution function at different elevations and thresholds.

Elevation	Threshold					
	0.5	0.6	0.7	0.8	0.85	0.9
30	0.0053	0.00509	0.0047	0.00469	0.00466	0.00426
40	0.00481	0.0045	0.00466	0.00451	0.00424	0.00479
50	0.00511	0.00489	0.00466	0.0043	0.00419	0.00471
60	0.00528	0.00499	0.00476	0.00457	0.00448	0.00463
70	0.0053	0.0051	0.00491	0.00463	0.00392	0.00423
80	0.00503	0.00484	0.00466	0.00427	0.00413	0.00378
90	0.00495	0.00485	0.00456	0.0043	0.00401	0.00417
100	0.00645	0.00615	0.00586	0.00557	0.00571	0.00605

Table 4.1: SSE of the Reciprocal Inverse Gaussian function to fit the distribution of Wind Speed at Peaks at different elevations and thresholds

- For the distribution of the Mexican hats Widths, the **Pareto** distribution function (scipy.stats.pareto), was found to provide the best fit in most of the situations. This function is defined as:

$$f(x, b) = \frac{b}{x^{b+1}}$$

for $x \geq 1,\ b > 0$.

where **b** is the shape parameter.

Table 4.2 reports the SSE of this distribution function at different elevations and thresholds.

Elevation	Threshold					
	0.5	0.6	0.7	0.8	0.85	0.9
30	0.35357	0.28221	0.20518	0.15297	0.18706	0.53698
40	0.3547	0.28203	0.20176	0.14668	0.17794	0.53991
50	0.35745	0.28455	0.20358	0.15171	0.19417	0.58856
60	0.35819	0.28505	0.20446	0.15165	0.18683	0.55441
70	0.36392	0.29033	0.206	0.14249	0.17227	0.50313
80	0.35966	0.28411	0.20138	0.14562	0.18347	0.56822
90	0.36082	0.28692	0.20258	0.14195	0.17678	0.56787
100	0.36222	0.28838	0.20445	0.14351	0.17154	0.51158

Table 4.2: SSE of the Pareto function to fit the distribution of Mexican hats Widths at different elevations and thresholds

- For the distribution of the Mexican hats Amplitudes, the **Exponentiated**

Weibull distribution function (scipy.stats.exponweib), was found to provide the best fit in most of the situations. This function is defined as:

$$f(x, a, c) = ac\,[1 - \exp(-x^c)]^{a-1} \exp(-x^c)\, x^{c-1}$$

for $x > 0,\ \ a > 0,\ \ c > 0.$

where **a** and **c** are the shape parameters.

Table 4.3 reports the SSE of this distribution function at different elevations and thresholds. As evidenced by the data, the function exhibits a high degree of accuracy with negligible sums of squared errors (SSEs) across all elevations and thresholds, except for the threshold of 0.5 at 40m. In order to ensure a rigorous evaluation of the parameters and their variations with respect to both threshold and elevation, the fit function data for this specific combination was excluded from subsequent analysis.

Elevation	Threshold					
	0.5	0.6	0.7	0.8	0.85	0.9
30	0.018202	0.007915	0.003733	0.002917	0.00558	0.014264
40	0.312573	0.005077	0.002096	0.002747	0.004151	0.007386
50	0.025738	0.006809	0.001844	0.001391	0.003972	0.024973
60	0.021208	0.004549	0.000963	0.0012	0.001718	0.010898
70	0.019273	0.00425	0.001411	0.001156	0.002739	0.004677
80	0.021384	0.007472	0.002036	0.001647	0.004318	0.013016
90	0.028056	0.007214	0.00207	0.002502	0.003712	0.016192
100	0.028276	0.006384	0.001701	0.001614	0.002388	0.012054

Table 4.3: SSE of the Exponentiated Weibull function to fit the distribution of Mexican hat Amplitudes at different elevations and thresholds

Figures 4.13 to 4.15 showcase the distributions of wind speed at Mexican hats peaks, Mexican hats widths and amplitudes at the elevation of 30 meters, and different match score thresholds, along with the corresponding fitted distribution functions and their parameters. These plots provide a comprehensive overview of the distribution patterns and characteristics of the Mexican hats at this specific height.

For a more detailed analysis of these distributions at different elevations, please refer to Appendix B, which includes additional Figures and information. These visuals served as valuable tools for identifying Mexican hat events' distinct patterns and characteristics and how these patterns vary with altitude, helping us better understand their behavior and anticipate potential impacts.

The fitted distribution functions' characteristics and parameters are analyzed in greater depth in the following subsections, covering the changes in these parameters with elevation and thresholds.

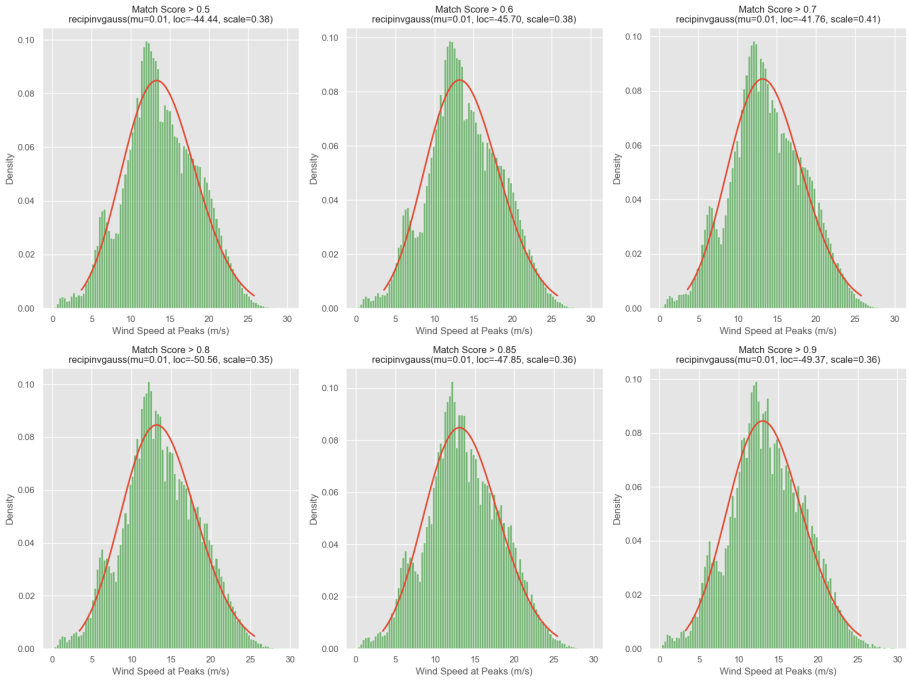


Figure 4.13: Distribution of wind speed at Mexican hat peaks and the fit function at different thresholds at 30m

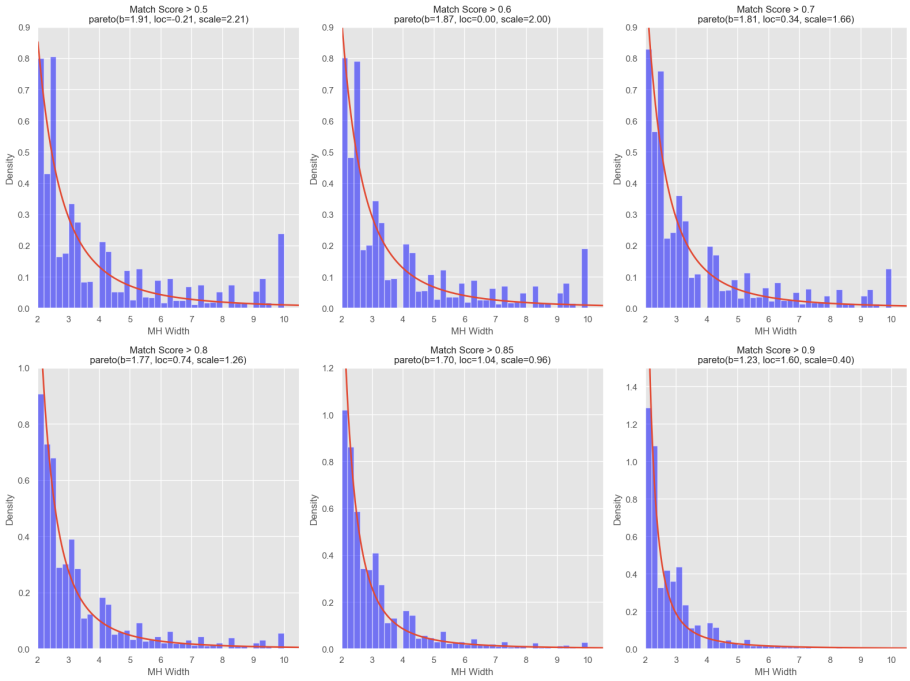


Figure 4.14: Distribution of Mexican hat widths and the fit function at different thresholds at 30m

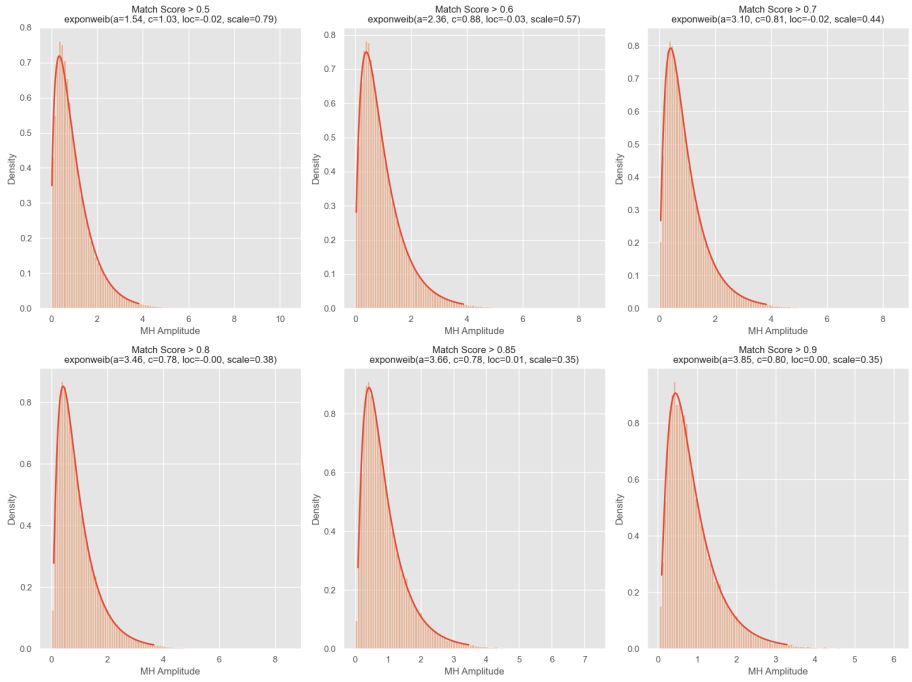


Figure 4.15: Distribution of Mexican hat amplitudes and the fit function at different thresholds at 30m

4.2.1 Reciprocal Inverse Gaussian distribution function parameters (fitted to Wind Speed at the Peaks distribution):

- The Shape parameter, μ, was found to remain constant at 0.01 at all elevations and thresholds, as shown in Figure 4.16.

- The Location parameter, on the other hand, displayed a non-uniform behavior across all elevations. Nevertheless, the overall trend indicated an increase in the parameter value up to a threshold between 0.7 and 0.8, followed by a decrease, as illustrated in Figure 4.17.

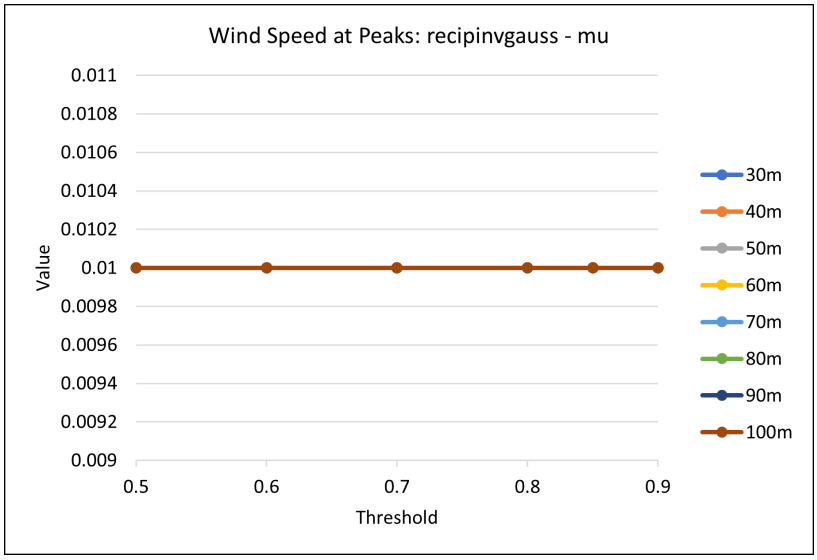


Figure 4.16: Reciprocal Inverse Gaussian distribution function Shape parameter (mu) at different elevations and thresholds

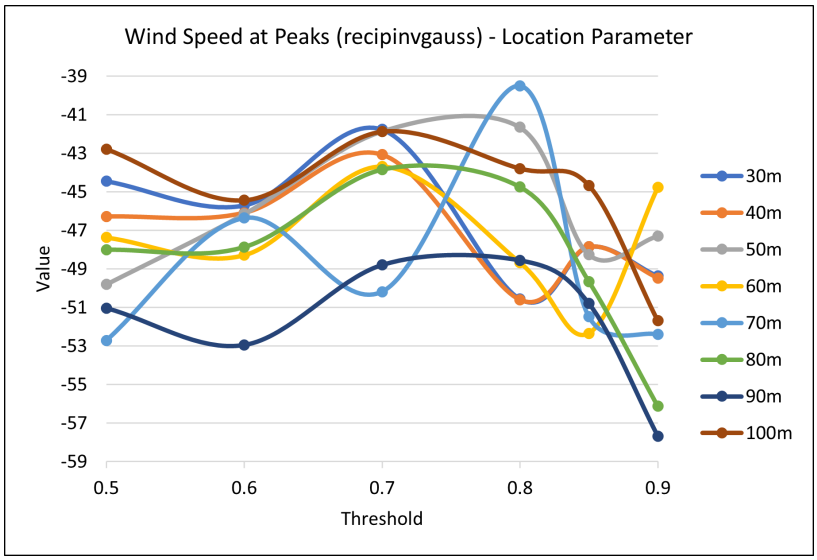


Figure 4.17: Reciprocal Inverse Gaussian distribution function Location parameter at different elevations and thresholds

- Similarly, the Scale parameter showed an inconsistent trend across all elevations. However, the overall pattern indicated an increase in the scale factor by increasing the thresholds up to somewhere between 0.7 and 0.8, followed by a decline, as shown in Figure 4.18.

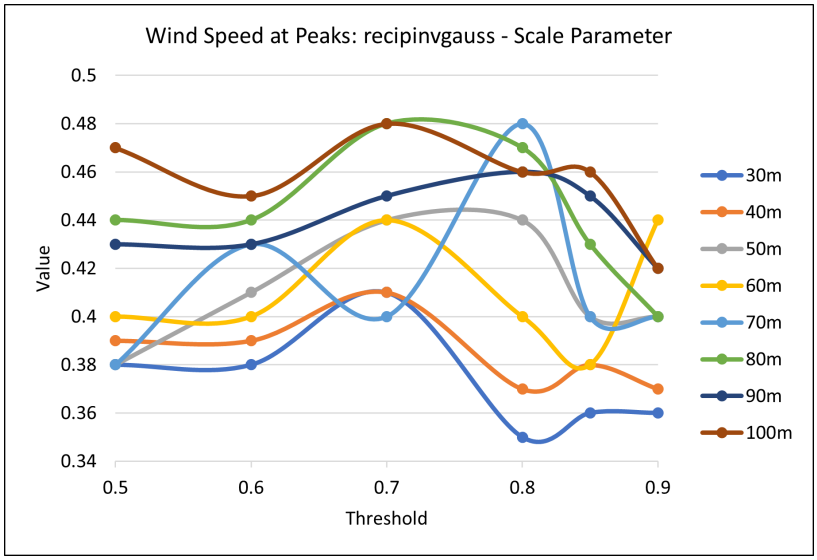


Figure 4.18: Reciprocal Inverse Gaussian distribution function Scale parameter at different elevations and thresholds

After analyzing the parameters of the Reciprocal Inverse Gaussian distribution function, it was determined that changes in elevations and match score thresholds did not impact the Shape parameter. This suggests that the shape of the wind speed distribution function at Mexican hats peaks remains consistent despite changes in other variables.

However, the behavior of the location and scale parameters was more intricate. Although the general trend indicated an increase in both parameters up to a specific threshold, followed by a decrease, this pattern was not consistent across all elevations. This nonuniformity indicates that changes in other variables, such as elevation and threshold, affect the location and scale of the distribution function.

These findings are significant, and additional research is required to fully comprehend their implications and investigate other factors that may impact the wind speed distribution at Mexican hats Peaks.

4.2.2 Pareto distribution function parameters (fitted to Mexican hats Widths distribution):

- As illustrated in the Figure 4.19, the shape parameter, **b**, exhibits a steady decline as the threshold increases until it reaches 0.85. Beyond this threshold, there is a sharp decrease in the **b** value, which is consistent across all elevations.

- Similarly, the location parameter displays a steady increase with the increase of the threshold until 0.85, followed by a sharp increase beyond this threshold. This behavior is consistent across all elevations, as shown

in Figure 4.20.

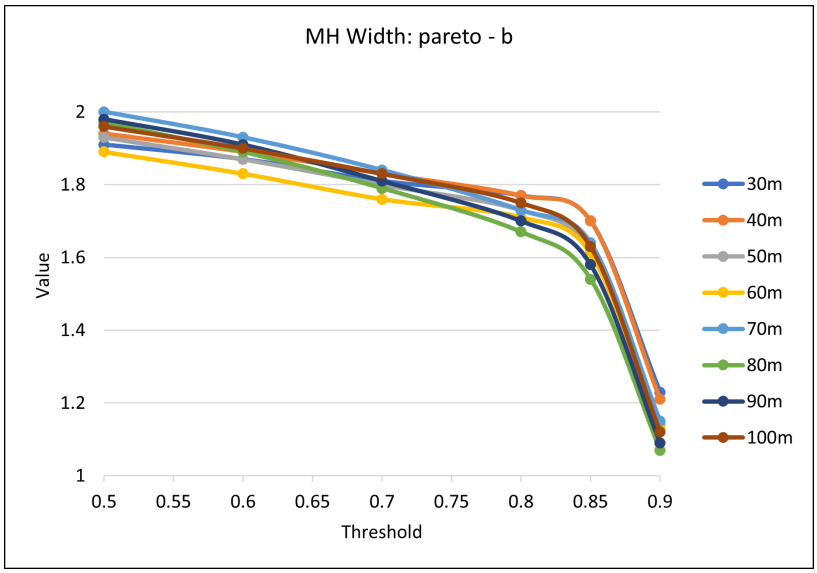


Figure 4.19: Pareto distribution function Shape parameter (b) at different elevations and thresholds

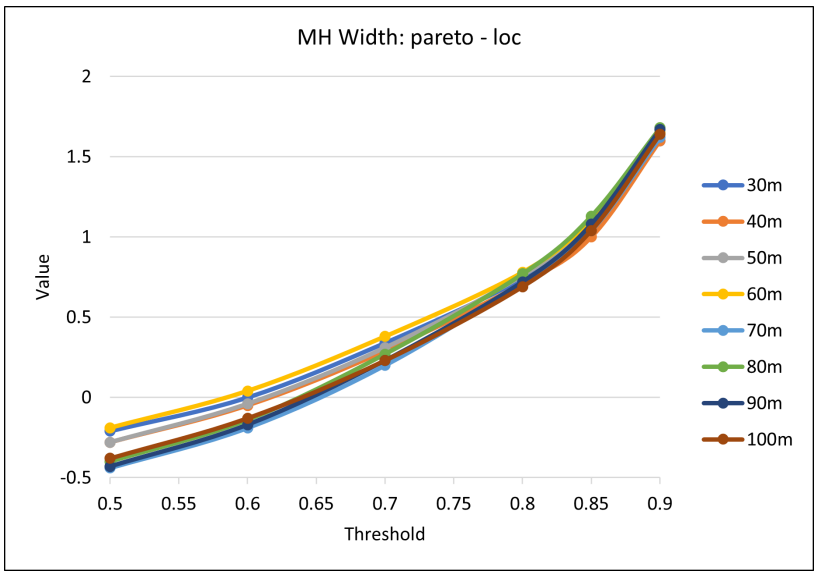


Figure 4.20: Pareto distribution function Location parameter at different elevations and thresholds

- Lastly, as demonstrated in the Figure 4.21, the scale parameter exhibits a steady decline with the increase of the threshold until 0.85, beyond

which there is a sharper decline in the scale parameter value. This consistent behavior is observed at all elevations.

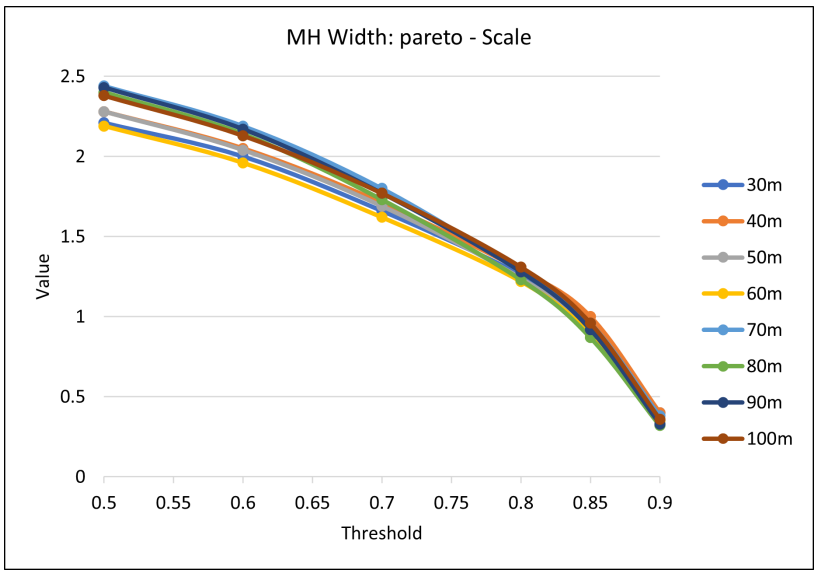


Figure 4.21: Pareto distribution function Scale parameter at different elevations and thresholds

Based on the analysis of the Pareto distribution function parameters, it can be concluded that there is a consistent behavior observed across all elevations. The shape parameter, location parameter, and scale parameter all display a steady change with increasing threshold up to 0.85, after which there is a sharp change in their values. These observations confirm that the Pareto distribution is a suitable model for analyzing the data under consideration and that a match score of 0.85 is the appropriate threshold to be considered for our Mexican hat detection model.

4.2.3 Exponentiated Weibull distribution function parameters (fitted to Mexican hats Amplitudes distribution):

As previously mentioned, the Exponentiated Weibull distribution function exhibited a relatively high SSE at 40m with a threshold of 0.5. As a result, it was determined that the fit function parameters for this condition should be excluded from future analyses. The findings of the analysis conducted on the function parameters for all other combinations of elevations and thresholds are presented below:

- α **(first Shape parameter):** The α shape parameter value increases steadily with the increase of threshold until 0.85.Beyond this point, the

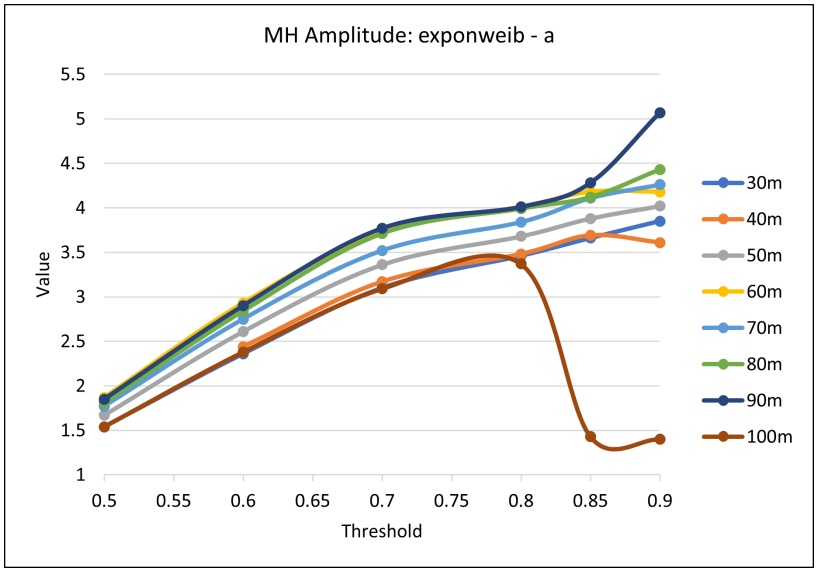


Figure 4.22: Exponentiated Weibull distribution function α Shape parameter at different elevations and thresholds

trend differs at different elevations, with 40m and 60m showing a slight decline in the α value while others continue to rise. It was observed that the α value declined significantly after the threshold of 0.8 at an elevation of 100m . These observations are illustrated in the Figure 4.22.

- **c (second Shape parameter):** As illustrated in the Figure 4.23, the value of the c shape parameter decreases sharply from the 0.5 to 0.6 threshold. The declining trend slows down after this point and reaches a plateau between the thresholds of 0.75 and 0.8. The c value showed a slight increase after the threshold of 0.8 at all elevations except for 100m where the c value declined significantly.

- **Location parameter:** The location parameter plot (Figure 4.24), showed two inflection points. The first one was around the threshold of 0.6, where the falling leg of the plot reached its minimum value and started a moderately fast rise. The second inflection point was between the thresholds of 0.8 and 0.85, where the location factor reached its maximum and showed a relatively slow decline afterward.

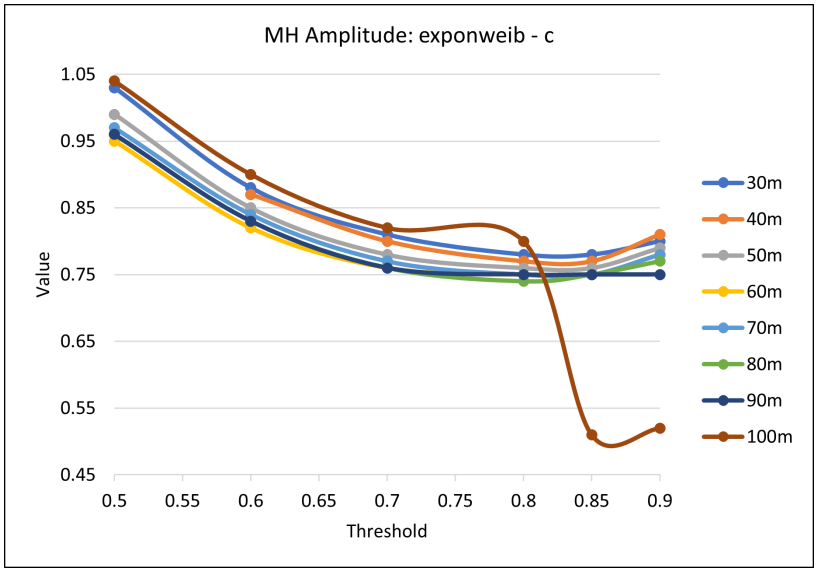


Figure 4.23: Exponentiated Weibull distribution function c Shape parameter at different elevations and thresholds

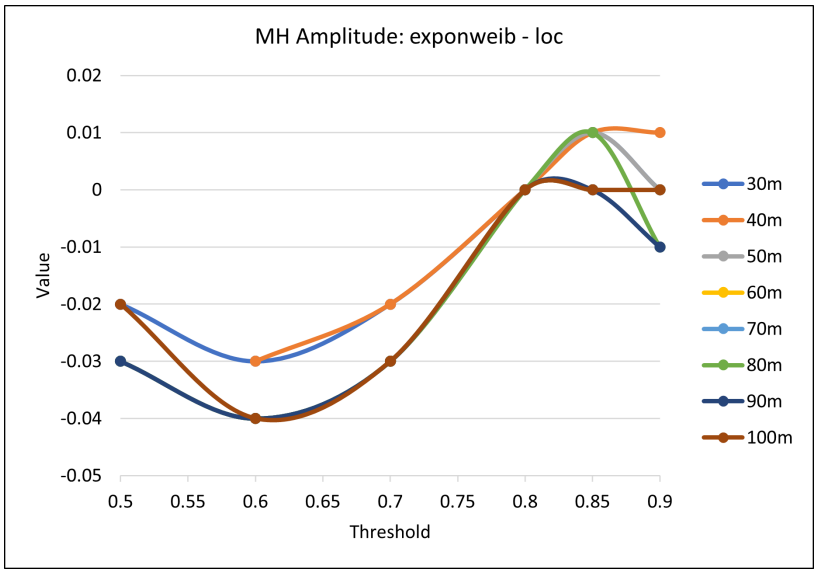


Figure 4.24: Exponentiated Weibull distribution function Location Shape parameter at different elevations and thresholds

- **Scale parameter:** The behavior and trend of the Scale parameter (Figure 4.25) were similar to that of the c parameter. It showed a sharp trend from the 0.5 to 0.6 threshold and slowed down after this point, reaching a plateau between the thresholds of 0.75 and 0.8. Similar to the c value, the Scale parameter showed a slight increase after the threshold of 0.8 at all elevations.

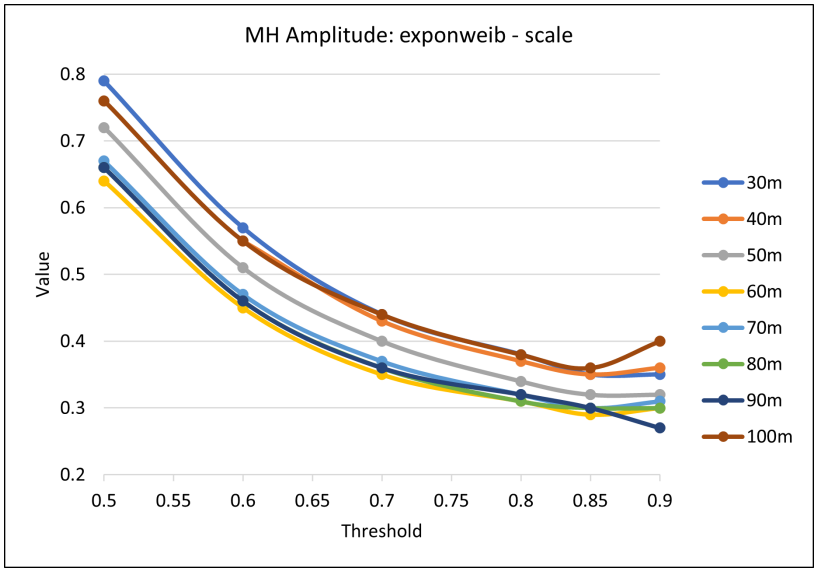


Figure 4.25: Exponentiated Weibull distribution function Scale parameter at different elevations and thresholds

After conducting an analysis of the Exponentiated Weibull distribution function parameters across different elevations and thresholds, it was observed that the parameters exhibited unique trends and tendencies. These findings suggest that this distribution function is a suitable model to indicate the behavior of the Mexican hat amplitudes under different conditions. Moreover, the parameters of the Exponentiated Weibull distribution function provide valuable insights into the underlying relationships and trends, which can aid in the development of an accurate detection model for the Mexican hat wavelet. Interestingly, similar to the Pareto distribution function, the Exponentiated Weibull distribution function parameters displayed a significant behavioral change at thresholds between 0.8 and 0.85. This observation supports the selection of a match score of 0.85 as the appropriate threshold for our Mexican hat detection model. By incorporating the insights gained from this analysis, we can further refine our detection model and improve its accuracy in identifying the Mexican hat wavelet.

Chapter 5

Discussion and Conclusions

5.1 Summary of the Findings

The results presented in Chapter 4 demonstrate that our innovative Mexican hat detection model is effective at identifying these events in our dataset while remaining unaffected by the noise level. Our model's ability to withstand the impact of noise is a significant advantage, as it ensures that false positives are minimized and the model's overall accuracy is high.

Furthermore, the statistical distribution fit functions utilized in the analysis identified a match score of 0.85 as the appropriate threshold that end-users should consider as the signature of a Mexican hat event. However, it is crucial to note that additional research is required to verify the threshold's effectiveness in other scenarios.

5.2 Limitations

Despite this research's significant contributions and yielded promising outcomes, it is crucial to recognize its limitations, which should be considered when interpreting the findings presented in this thesis.

One of the primary limitations was the short duration of the thesis, which constrained the scope and depth of the research. Consequently, the author was limited in the bandwidth to explore alternative machine learning models and additional variables or to examine different datasets and their implications in the predictive maintenance of wind turbines. Therefore, it is imperative to interpret the findings within the context of these time constraints.

Another significant limitation was the absence of a powerful GPU, which limited the scalability of the research. Machine learning algorithms require substantial computational power, and the lack of a robust GPU restricted the size and complexity of the datasets that could be processed and analyzed effectively. Therefore, the conclusions drawn from this research may be subject to the constraints imposed by the available computational resources.

Lastly, the limited sample size utilized in this study may pose a potential

limitation. While efforts were made to ensure the reliability and representativeness of the chosen dataset, the sample size may only partially capture the variability and nuances present in real-world wind turbine operations. A larger and more extensive dataset with additional parameters such as wind direction could have provided a more comprehensive understanding of the relationships between wind patterns, Mexican hat events frequencies and their characteristics, turbine performance anomalies, and maintenance optimization.

It is important to note that these limitations do not detract from the significance of the research presented in this thesis. Instead, they highlight the need for further exploration and advancements in wind turbine predictive maintenance. Future studies should consider allocating more time for research, securing access to powerful computational resources, and expanding the sample size to validate and enhance the findings presented here. By addressing these limitations, future research endeavors can provide a more comprehensive and robust understanding of wind turbine predictive maintenance and its potential for sustainable and efficient wind energy operations.

5.3 Recommendations for Future Research

Like any other study, this project has certain limitations that offer opportunities for future research to expand the knowledge of the influence of Mexican hat events on wind turbines. Several potential research directions can be pursued to advance this understanding further.

One possible track for future research is to investigate the mechanical response of wind turbine components to extreme wind events and Mexican hats through laboratory experiments and numerical simulations. Such studies can aid in the development of novel models that can predict the long-term structural behavior of wind turbines under different loading conditions.

Another avenue to pursue is to analyze the impact of Mexican hat events on different types of wind turbines under various operating conditions. Such research could include studying turbines with varying rotor sizes, hub heights, power ratings, wind regimes, and terrain conditions. The outcome of such studies would enable researchers to gain a more comprehensive understanding of how Mexican hat events affect wind turbines and their electrical power production.

Moreover, machine learning and AI-assisted detection algorithms can be utilized to identify Mexican hat events in the wind. Clustering or classification algorithms can identify and categorize these events based on their associated features. Developing innovative algorithms that can automatically identify patterns and anomalies in the data and integrate data from other sources, such as meteorological data and satellite imagery, can provide a more accurate understanding of these events.

Another potential area of research is the use of advanced sensors and edge IoT monitoring systems to detect and analyze Mexican hat events in real time. Developing novel sensing technologies and edge applications that can provide

high-resolution measurements of wind speed, direction, turbulence, and other relevant variables, as well as integrating data from different sensors, such as accelerometers and strain gauges, can enable more accurate detection and analysis of these events.

Overall, further research on the impact of Mexican hat events on wind turbines has the potential to lead to new insights into the behavior of wind turbines under extreme wind events. It can also provide new strategies for mitigating the impact of these events on wind turbines and advance the science and engineering of wind energy systems.

5.4 Final Thoughts

Incorporating the insights gained from our analyses and the practical implications of this study could lead to significant improvements in the design and efficient operation of wind turbines, making wind power a more reliable, sustainable, and cost-effective renewable energy source for the future. Wind turbine operators can use our methodology to detect and analyze Mexican hat events, allowing them to better understand the causes and develop strategies to mitigate their impacts. This can lead to improved performance, reduced maintenance costs, and increased energy production.

Furthermore, studying the behavior of the wind during extreme events can provide insights into the physical mechanisms that drive them, enabling researchers to develop predictive models to forecast their occurrence. By proactively preparing for and responding to extreme events, wind turbine operators can enhance the safety and efficiency of wind power generation. Overall, our study highlights the importance of developing effective detection and analysis methods for extreme events in renewable energy systems and the potential benefits of doing so.

Appendices

Appendix A

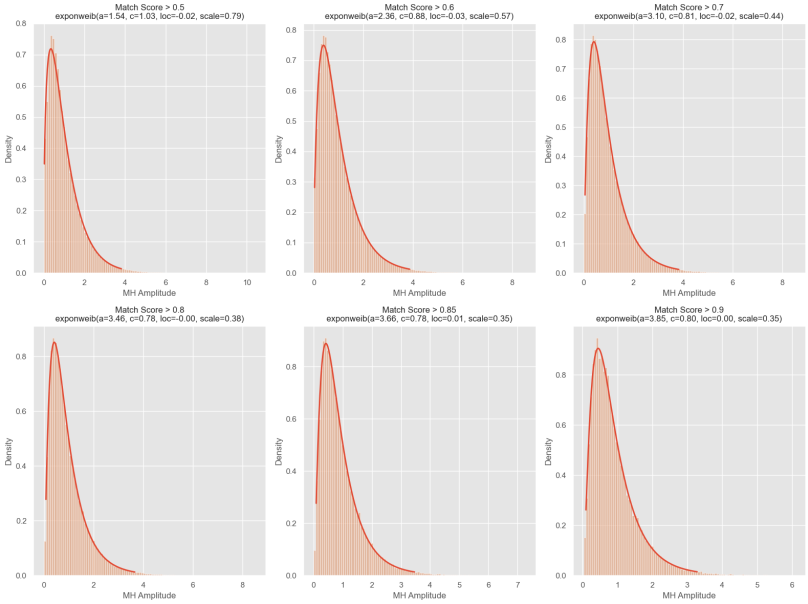
Code Repository

The code repository for the implementation of the algorithms and models discussed in this thesis is available on GitHub. This repository contains the source code, scripts, and relevant files necessary to replicate the experiments and analysis conducted in this research. The code files are written in Python format. To access the code repository and explore the Python file, please follow the link below:

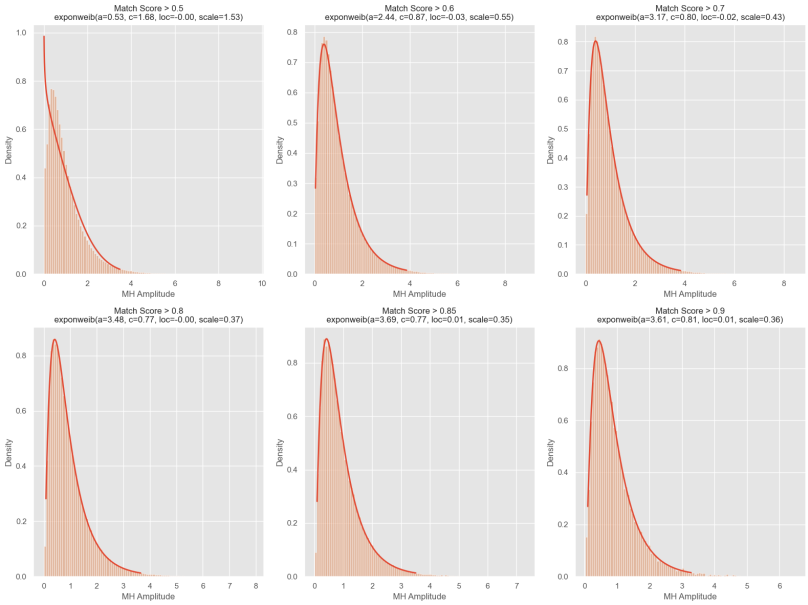
https://github.com/MMarasuli/Mexican-Hat-Detection

The provided Python code implements a cross-correlation technique between the wind speed data and the Mexican hat wavelet. It applies a rolling window approach to detect specific patterns in the wind speed time series. The code utilizes libraries such as numpy, pandas, and scipy.signal for data manipulation and signal processing operations. It defines the Mexican hat wavelet function and calculates the correlation between the wavelet filter and the wind speed data. The detected events are then sorted and organized into a table for further analysis.

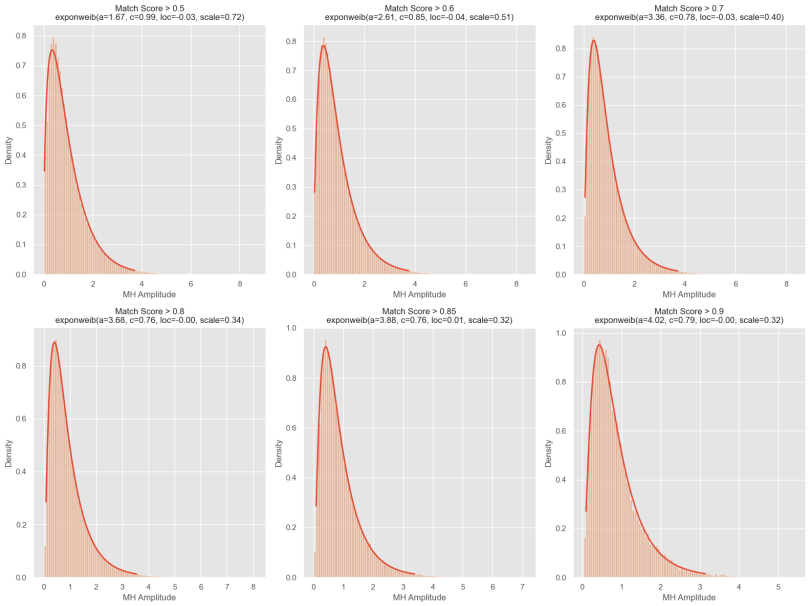
Appendix B



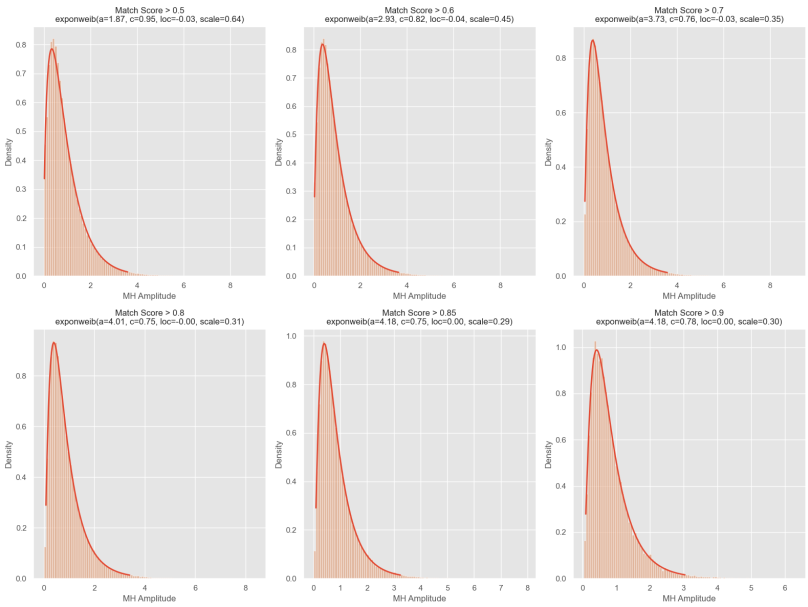
Mexican hat Amplitude distribution, fitted function, and its parameters at different thresholds at 30m elevation



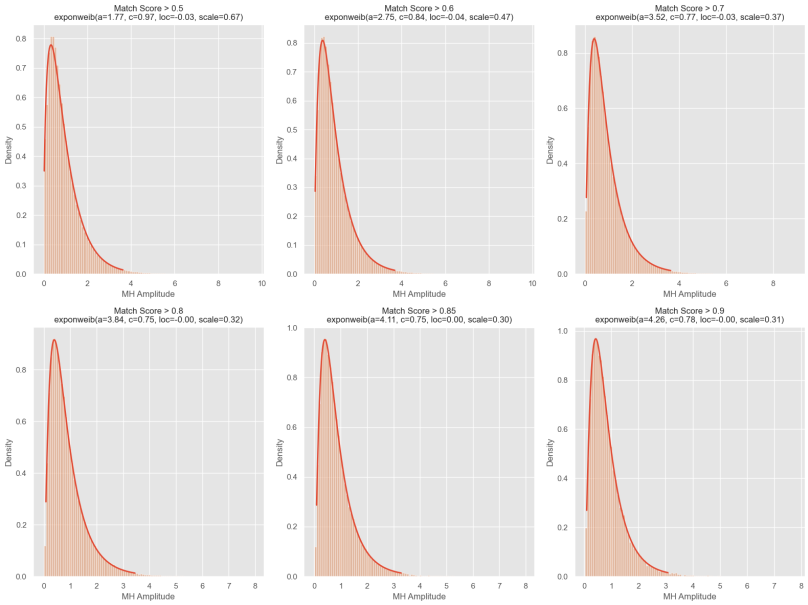
Mexican hat Amplitude distribution, fitted function, and its parameters at different thresholds at 40m elevation



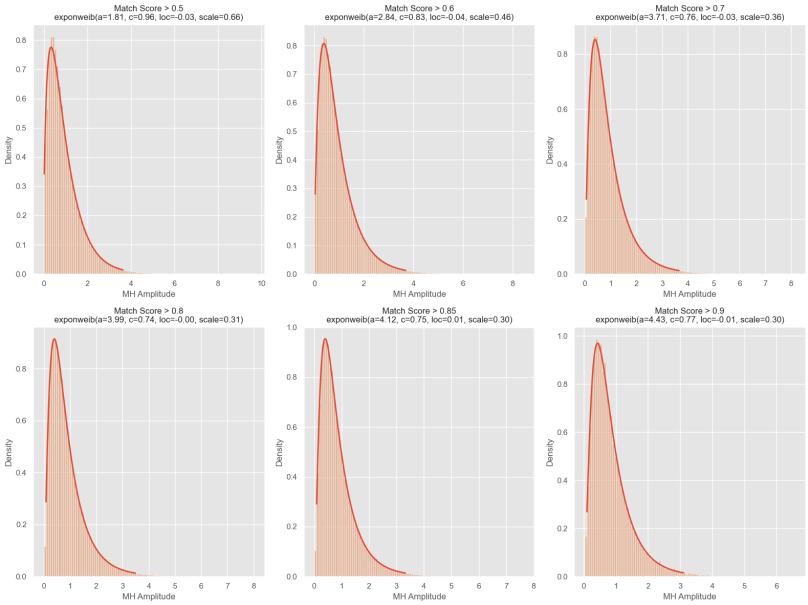
Mexican hat Amplitude distribution, fitted function, and its parameters at different thresholds at 50m elevation



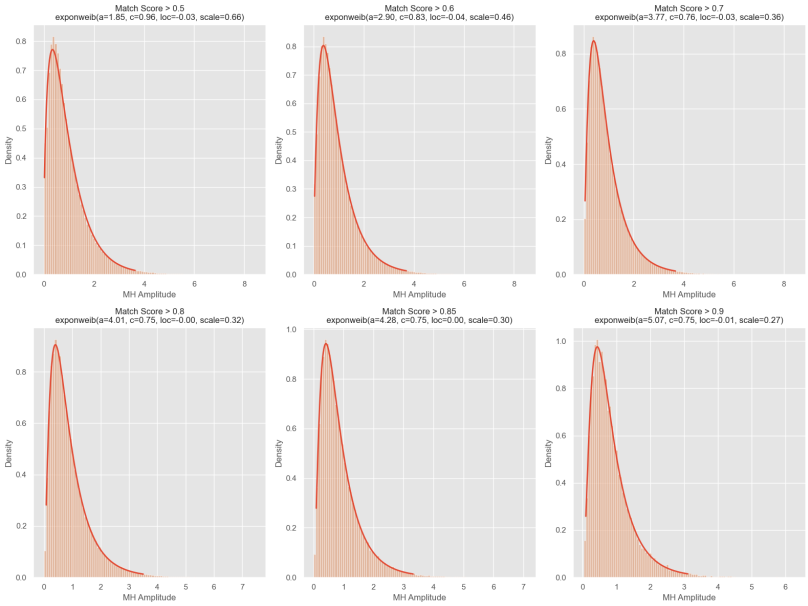
Mexican hat Amplitude distribution, fitted function, and its parameters at different thresholds at 60m elevation



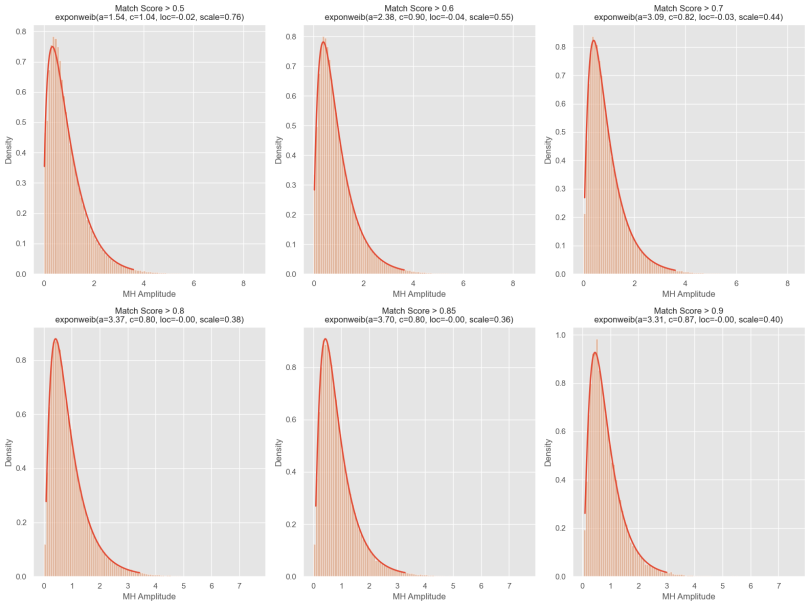
Mexican hat Amplitude distribution, fitted function, and its parameters at different thresholds at 70m elevation



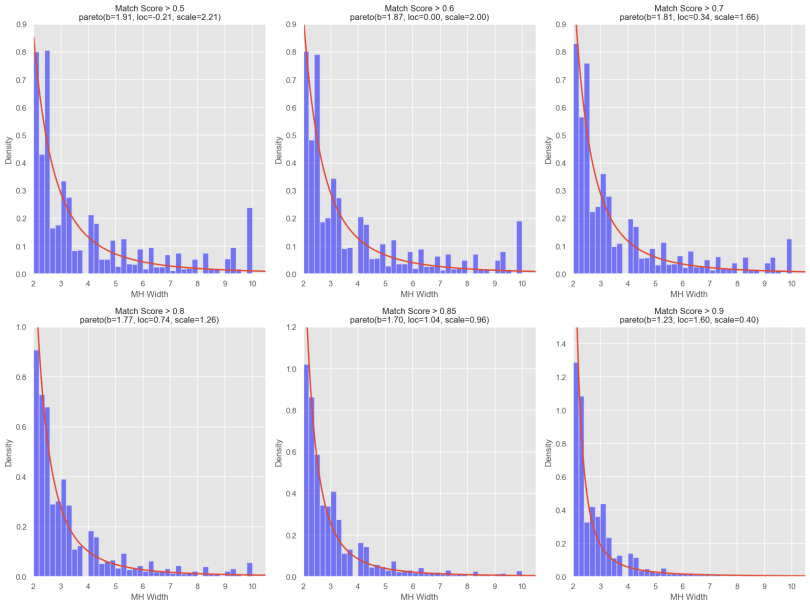
Mexican hat Amplitude distribution, fitted function, and its parameters at different thresholds at 80m elevation



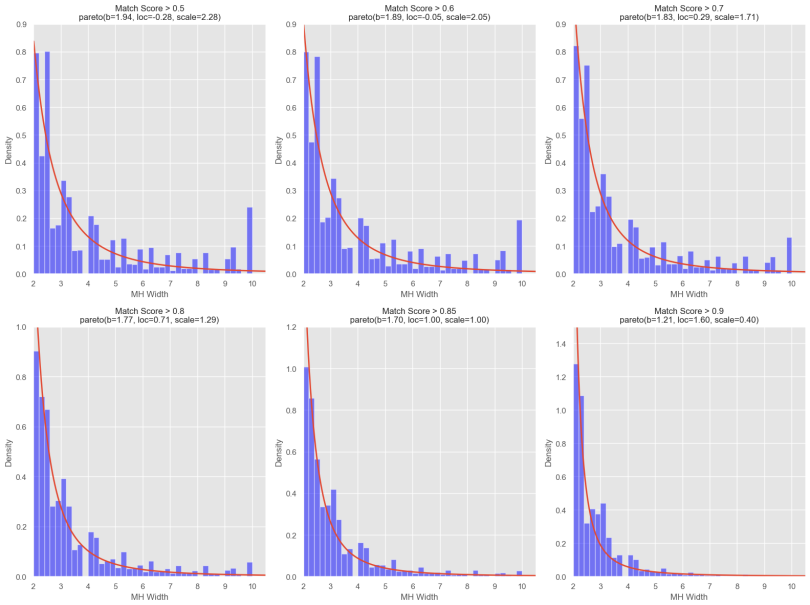
Mexican hat Amplitude distribution, fitted function, and its parameters at different thresholds at 90m elevation



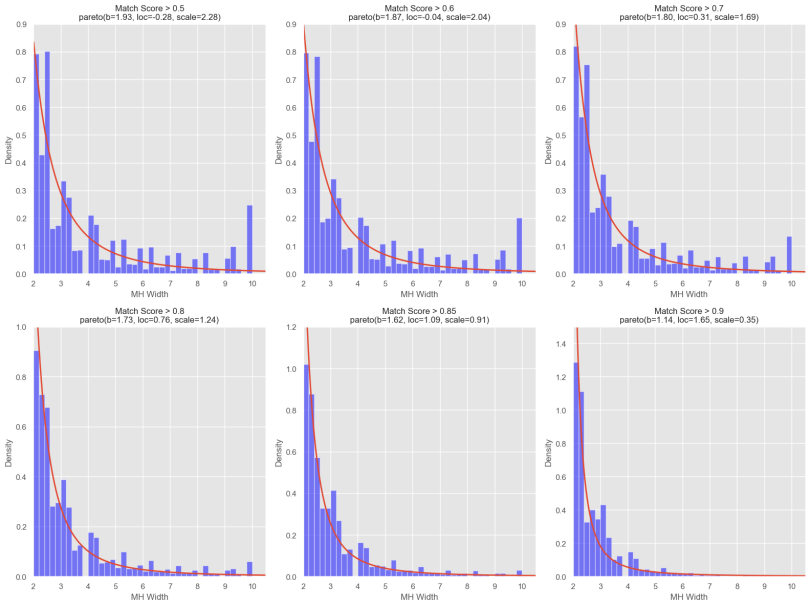
Mexican hat Amplitude distribution, fitted function, and its parameters at different thresholds at 100m elevation



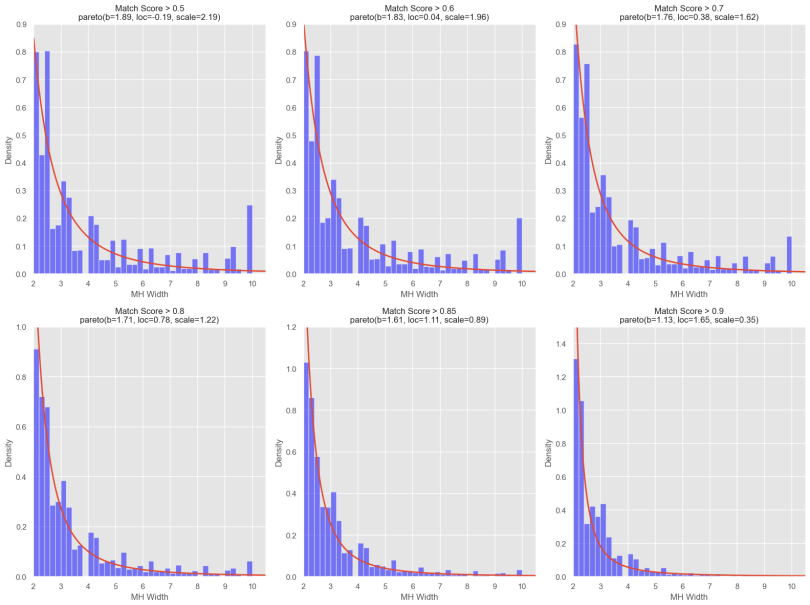
Mexican hat Width distribution, fitted function, and its parameters at different thresholds at 30m elevation



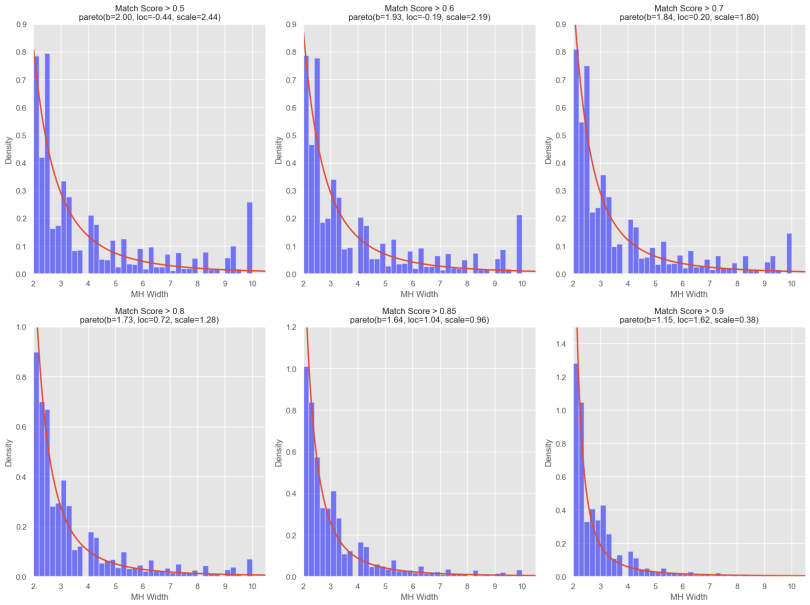
Mexican hat Width distribution, fitted function, and its parameters at different thresholds at 40m elevation



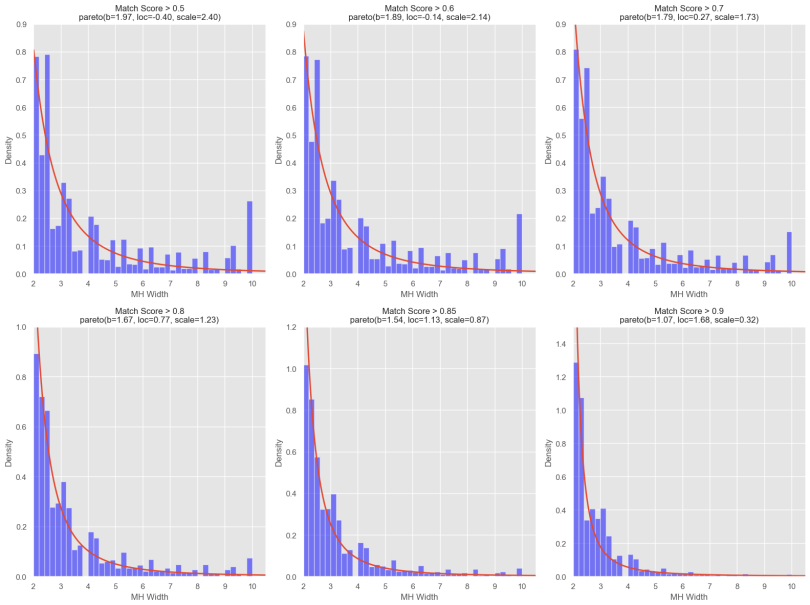
Mexican hat Width distribution, fitted function, and its parameters at different thresholds at 50m elevation



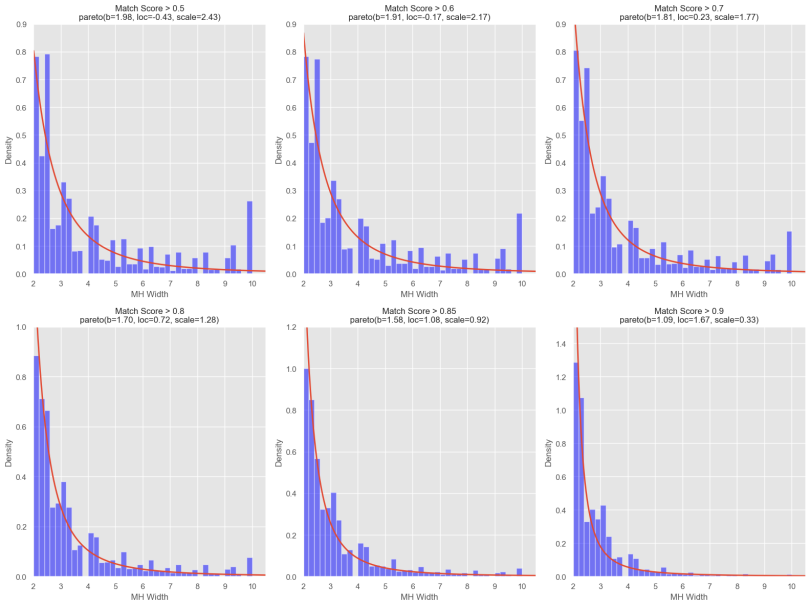
Mexican hat Width distribution, fitted function, and its parameters at different thresholds at 60m elevation



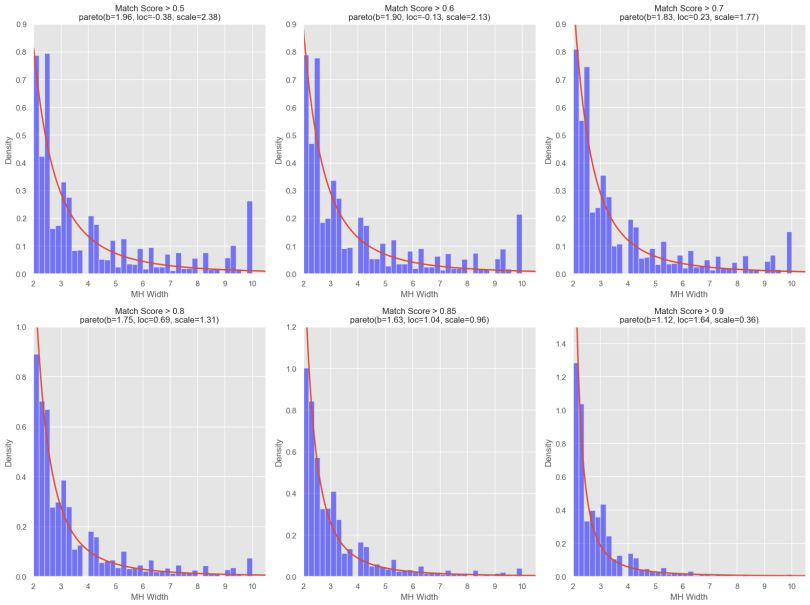
Mexican hat Width distribution, fitted function, and its parameters at different thresholds at 70m elevation



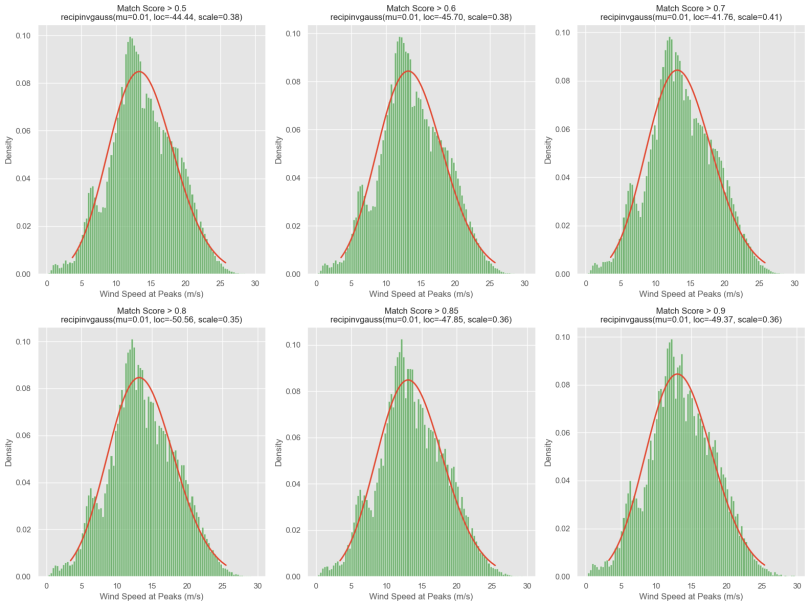
Mexican hat Width distribution, fitted function, and its parameters at different thresholds at 80m elevation



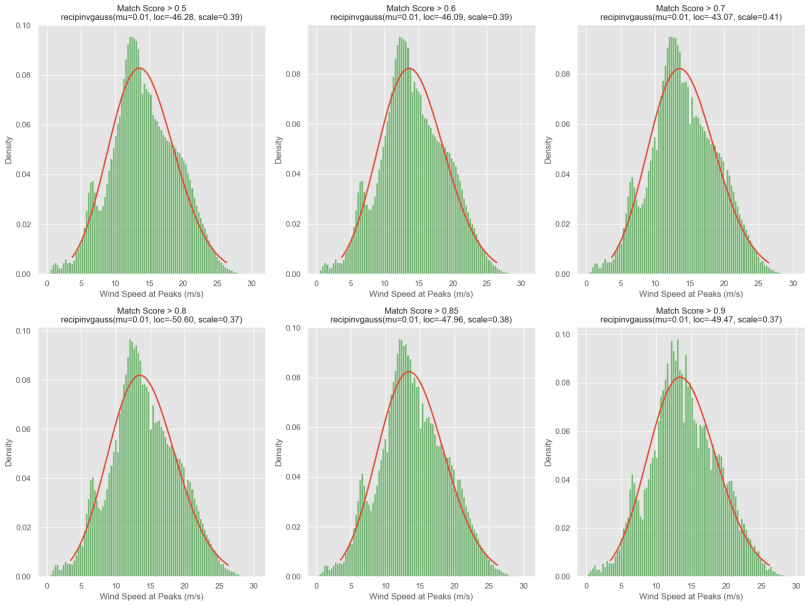
Mexican hat Width distribution, fitted function, and its parameters at different thresholds at 90m elevation



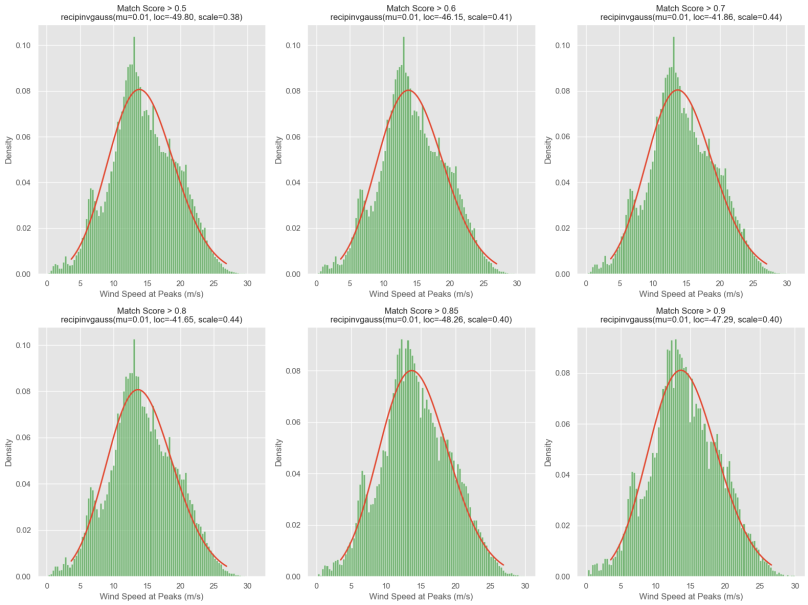
Mexican hat Width distribution, fitted function, and its parameters at different thresholds at 100m elevation



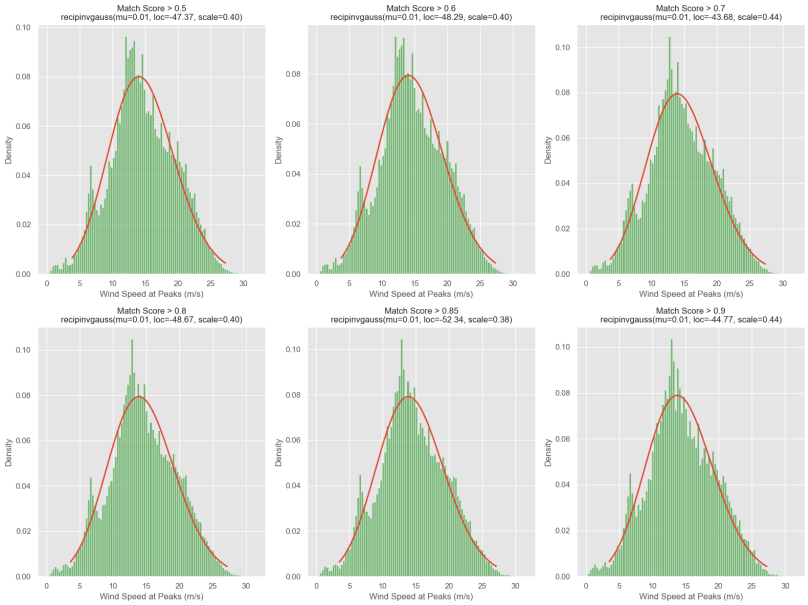
Mexican hat Speed distribution, fitted function, and its parameters at different thresholds at 30m elevation



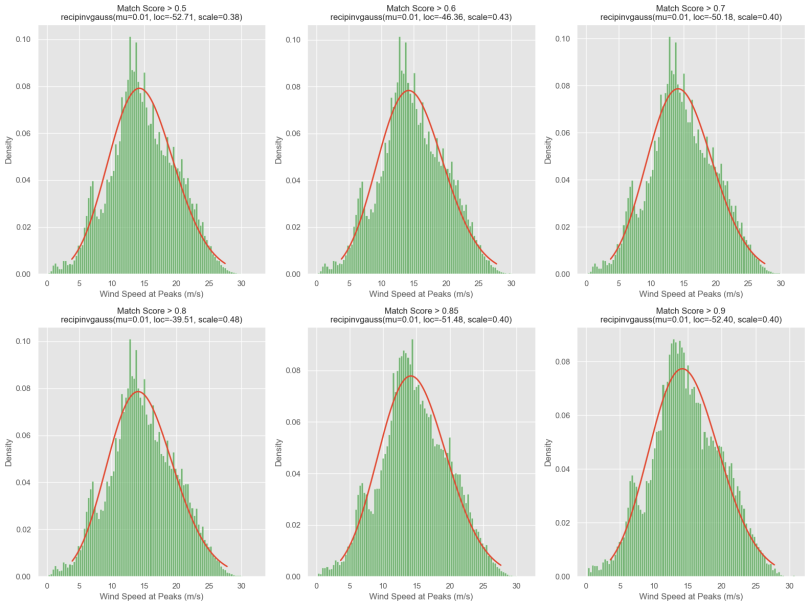
Mexican hat Speed distribution, fitted function, and its parameters at different thresholds at 40m elevation



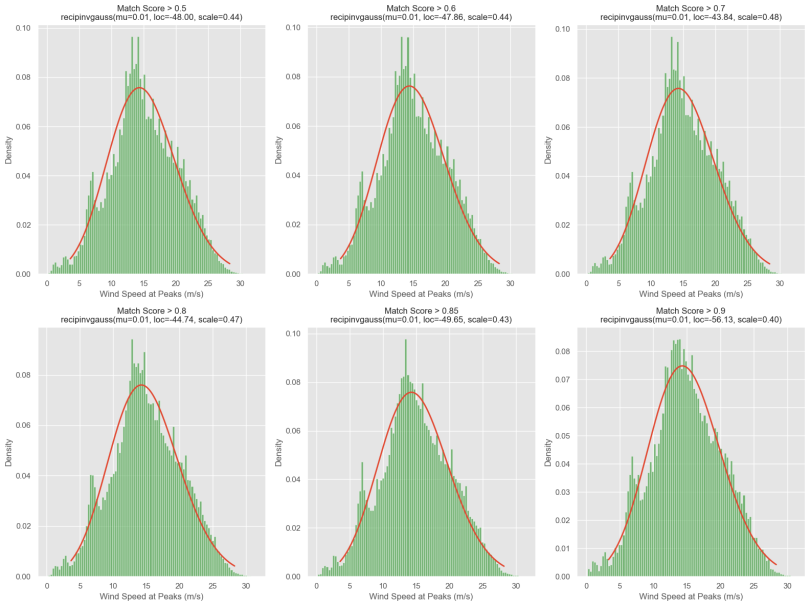
Mexican hat Speed distribution, fitted function, and its parameters at different thresholds at 50m elevation



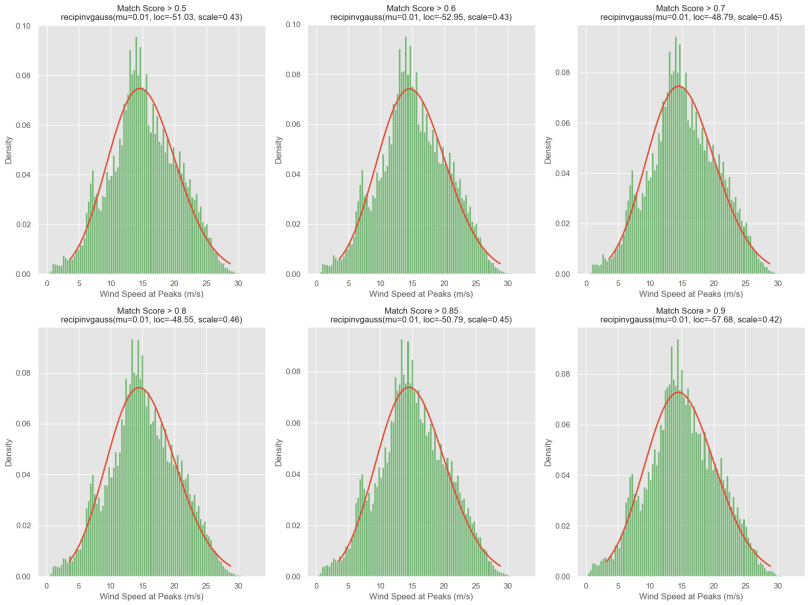
Mexican hat Speed distribution, fitted function, and its parameters at different thresholds at 60m elevation



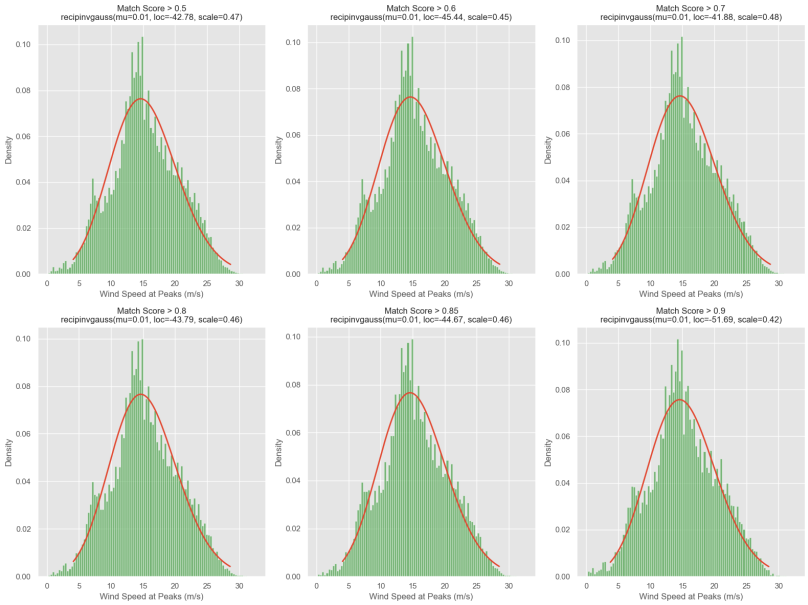
Mexican hat Speed distribution, fitted function, and its parameters at different thresholds at 70m elevation



Mexican hat Speed distribution, fitted function, and its parameters at different thresholds at 80m elevation



Mexican hat Speed distribution, fitted function, and its parameters at different thresholds at 90m elevation



Mexican hat Speed distribution, fitted function, and its parameters at different thresholds at 100m elevation

Appendix C

100m - Match Score > 0.5				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	742307	742307	742307	742307
mean	15.313501	0.699805	4.091142	0.947497
std	5.198141	0.113276	2.408233	0.78345
min	0.184	0.5	2	0
25%	12.05482	0.606619	2.4	0.398593
50%	14.94643	0.697079	3	0.741146
75%	18.9541	0.78805	5.2	1.275497
max	32.49901	0.997843	10	8.470295

100m - Match Score > 0.6				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	569541	569541	569541	569541
mean	15.312619	0.74461	3.975762	0.950779
std	5.205372	0.088591	2.316588	0.770461
min	0.184	0.6	2	0
25%	12.00409	0.670504	2.2	0.410097
50%	14.94643	0.738132	3	0.746456
75%	18.9541	0.811896	5	1.271756
max	32.49901	0.997843	10	8.470295

100m - Match Score > 0.7				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	364879	364879	364879	364879
mean	15.311345	0.797246	3.767797	0.942568
std	5.214915	0.063835	2.145491	0.747567
min	0.184	0.7	2	0
25%	12.00409	0.743438	2.2	0.418138
50%	14.94643	0.789769	2.8	0.742821
75%	18.9541	0.843592	4.4	1.251726
max	32.49901	0.997843	10	8.272987

100m - Match Score > 0.8				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	163628	163628	163628	163628
mean	15.283269	0.856952	3.420536	0.914125
std	5.225729	0.040182	1.821655	0.703066
min	0.184	0.8	2	0
25%	11.95336	0.823397	2.2	0.42275
50%	14.94643	0.850139	2.6	0.728323
75%	18.90337	0.884445	4	1.202188
max	32.49901	0.997843	10	7.564163

100m - Match Score > 0.85				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	81992	81992	81992	81992
mean	15.267489	0.889861	3.176768	0.895308
std	5.232683	0.029087	1.571772	0.671357
min	0.184	0.85	2	0.009059
25%	11.95336	0.865876	2.2	0.427581
50%	14.94643	0.884356	2.6	0.722299
75%	18.90337	0.908878	3.4	1.170968
max	32.49901	0.997843	10	7.564163

100m - Match Score > 0.9				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	27037	27037	27037	27037
mean	15.189387	0.924593	2.879669	0.870086
std	5.278975	0.018853	1.23553	0.622684
min	0.184	0.9	2	0.0308
25%	11.8519	0.909224	2	0.434829
50%	14.8957	0.920547	2.4	0.717467
75%	18.85264	0.93588	3.2	1.140217
max	32.49901	0.997843	10	7.564163

90m - Match Score > 0.5				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	744056	744056	744056	744056
mean	15.184258	0.699863	4.103063	0.939904
std	5.32588	0.113062	2.413959	0.786918
min	0.183	0.5	2	0
25%	11.78592	0.606976	2.4	0.39258
50%	14.89021	0.697262	3	0.727015
75%	18.96141	0.787672	5.4	1.255287
max	32.65082	0.99744	10	8.440221

90m - Match Score > 0.6				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	571649	571649	571649	571649
mean	15.189903	0.744411	3.993244	0.944019
std	5.334043	0.088459	2.328708	0.775215
min	0.183	0.6	2	0
25%	11.78592	0.67034	2.2	0.403485
50%	14.9411	0.738027	3	0.732923
75%	18.96141	0.811268	5	1.25346
max	32.65082	0.99744	10	8.440221

90m - Match Score > 0.7				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	366186	366186	366186	366186
mean	15.185839	0.796969	3.788859	0.935009
std	5.342922	0.063757	2.163911	0.751969
min	0.183	0.7	2	0
25%	11.73503	0.743472	2.2	0.410755
50%	14.9411	0.789214	2.8	0.729186
75%	18.96141	0.843027	4.6	1.232772
max	32.65082	0.99744	10	8.440221

90m - Match Score > 0.8				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	162971	162971	162971	162971
mean	15.158976	0.856979	3.446493	0.906971
std	5.364018	0.040221	1.855853	0.711009
min	0.23389	0.8	2	0
25%	11.68414	0.823448	2.2	0.41439
50%	14.9411	0.850137	2.6	0.71246
75%	18.96141	0.884437	4	1.186649
max	32.65082	0.99744	10	7.460264

90m - Match Score > 0.85				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	81673	81673	81673	81673
mean	15.114517	0.889916	3.191116	0.881712
std	5.393547	0.029136	1.598042	0.6701
min	0.23389	0.85	2	0.009087
25%	11.63325	0.865913	2.2	0.416373
50%	14.89021	0.884358	2.6	0.700343
75%	18.91052	0.908861	3.4	1.147338
max	32.65082	0.99744	10	7.055077

90m - Match Score > 0.9				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	26970	26970	26970	26970
mean	15.062274	0.924652	2.865962	0.858011
std	5.458462	0.018989	1.234656	0.629034
min	0.23389	0.9	2	0.031503
25%	11.53147	0.909121	2	0.424083
50%	14.89021	0.920568	2.4	0.694648
75%	18.91052	0.936325	3.2	1.104938
max	32.65082	0.99744	10	6.279755

80m - Match Score > 0.5				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	743355	743355	743355	743355
mean	14.992715	0.699833	4.101236	0.934244
std	5.211545	0.113088	2.413348	0.786343
min	0.19	0.5	2	0
25%	11.66402	0.606849	2.4	0.388028
50%	14.71022	0.697215	3	0.721233
75%	18.72105	0.787901	5.4	1.249447
max	32.27664	0.997698	10	9.651434

80m - Match Score > 0.6				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	570816	570816	570816	570816
mean	15.003625	0.744469	3.98918	0.938113
std	5.217101	0.088436	2.325648	0.773473
min	0.19	0.6	2	0
25%	11.66402	0.670418	2.2	0.400579
50%	14.71022	0.73817	3	0.727099
75%	18.72105	0.811436	5	1.247491
max	32.27664	0.997698	10	8.477561

80m - Match Score > 0.7				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	365599	365599	365599	365599
mean	14.988909	0.797059	3.779389	0.927985
std	5.224708	0.063665	2.156529	0.750304
min	0.19	0.7	2	0
25%	11.61325	0.743555	2.2	0.40616
50%	14.71022	0.789496	2.8	0.722569
75%	18.72105	0.843287	4.4	1.223806
max	32.27664	0.997698	10	8.148899

80m - Match Score > 0.8				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	163093	163093	163093	163093
mean	14.942912	0.856872	3.43142	0.899215
std	5.248442	0.040081	1.844527	0.707039
min	0.19	0.8	2	0
25%	11.56248	0.8234	2.2	0.408578
50%	14.71022	0.85022	2.6	0.705245
75%	18.67028	0.88409	4	1.175701
max	31.41355	0.997698	10	8.010706

80m - Match Score > 0.85				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	81869	81869	81869	81869
mean	14.910013	0.889603	3.179958	0.873574
std	5.27924	0.029072	1.593114	0.670275
min	0.19	0.85	2	0.009066
25%	11.51171	0.865528	2.2	0.408578
50%	14.65945	0.883924	2.6	0.690543
75%	18.61951	0.908484	3.4	1.135402
max	31.36278	0.997698	10	7.662083

80m - Match Score > 0.9				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	26710	26710	26710	26710
mean	14.881753	0.924593	2.868753	0.856243
std	5.307787	0.018845	1.245616	0.631425
min	0.19	0.9	2	0.030825
25%	11.46094	0.909135	2	0.418248
50%	14.65945	0.920646	2.4	0.690663
75%	18.61951	0.936327	3.2	1.110667
max	31.05816	0.997698	10	6.577601

70m - Match Score > 0.5				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	741958	741958	741958	741958
mean	14.819512	0.699862	4.087504	0.930983
std	5.06304	0.113214	2.40112	0.785783
min	0.189	0.5	2	0
25%	11.57492	0.606502	2.4	0.384856
50%	14.52306	0.69732	3	0.718881
75%	18.43697	0.787961	5.2	1.245335
max	32.11024	0.99633	10	9.602674

70m - Match Score > 0.6				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	569255	569255	569255	569255
mean	14.815616	0.744689	3.97333	0.935701
std	5.070468	0.088448	2.309156	0.774269
min	0.189	0.6	2	0
25%	11.57492	0.670683	2.2	0.396958
50%	14.52306	0.738556	3	0.725808
75%	18.43697	0.811573	5	1.24411
max	32.11024	0.99633	10	9.602674

70m - Match Score > 0.7				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	365329	365329	365329	365329
mean	14.783904	0.797102	3.762357	0.92513
std	5.080739	0.063703	2.134514	0.749691
min	0.189	0.7	2	0
25%	11.52409	0.743679	2.2	0.403009
50%	14.47223	0.789401	2.8	0.721217
75%	18.38614	0.843197	4.4	1.221735
max	32.11024	0.99633	10	8.913938

70m - Match Score > 0.8				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	163027	163027	163027	163027
mean	14.744743	0.856911	3.416859	0.89309
std	5.099783	0.040169	1.817202	0.701486
min	0.189	0.8	2	0
25%	11.42243	0.823407	2.2	0.406035
50%	14.4214	0.850107	2.6	0.703753
75%	18.33531	0.884401	4	1.166714
max	32.11024	0.99633	10	7.926628

70m - Match Score > 0.85				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	81639	81639	81639	81639
mean	14.724657	0.889825	3.165813	0.870351
std	5.113908	0.029096	1.552307	0.666405
min	0.189	0.85	2	0.009077
25%	11.42243	0.865807	2.2	0.408773
50%	14.4214	0.884353	2.6	0.693216
75%	18.33531	0.908789	3.4	1.128901
max	31.24613	0.99633	10	7.926628

70m - Match Score > 0.9				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	26782	26782	26782	26782
mean	14.737695	0.924708	2.87237	0.855683
std	5.159975	0.018834	1.208958	0.630555
min	0.189	0.9	2	0.036307
25%	11.42243	0.909295	2	0.418742
50%	14.4214	0.920695	2.4	0.697041
75%	18.38614	0.936314	3.2	1.103737
max	30.73783	0.99633	10	7.7457

60m - Match Score > 0.5				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	743802	741150	741150	741150
mean	14.623284	0.699344	4.050513	0.924206
std	4.981749	0.113229	2.381449	0.776149
min	0.187	0.5	2	0
25%	11.38728	0.606021	2.4	0.386133
50%	14.32672	0.696629	3	0.71676
75%	18.22908	0.7875	5	1.233213
max	31.81132	0.996859	10	9.067089

60m - Match Score > 0.6				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	567703	567703	567703	567703
mean	14.610795	0.744335	3.931946	0.926863
std	4.99405	0.088521	2.285791	0.764223
min	0.187	0.6	2	0
25%	11.38728	0.670201	2.2	0.39639
50%	14.32672	0.737872	3	0.720507
75%	18.1784	0.811315	5	1.229282
max	31.81132	0.996859	10	9.067089

60m - Match Score > 0.7				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	363409	363409	363409	363409
mean	14.575122	0.797007	3.7135	0.913448
std	5.005892	0.063757	2.101459	0.738222
min	0.23768	0.7	2	0
25%	11.3366	0.743389	2.2	0.400613
50%	14.27604	0.789407	2.8	0.713743
75%	18.12772	0.843434	4.4	1.203796
max	31.81132	0.996859	10	9.067089

60m - Match Score > 0.8				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	162043	162043	162043	162043
mean	14.498835	0.85693	3.365976	0.881376
std	5.009172	0.040111	1.769782	0.691503
min	0.23768	0.8	2	0
25%	11.23524	0.823519	2.2	0.403027
50%	14.17468	0.850336	2.6	0.692934
75%	18.07704	0.884227	3.8	1.151818
max	31.40588	0.996859	10	8.385169

60m - Match Score > 0.85				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	81510	81510	81510	81510
mean	14.451436	0.889587	3.127177	0.859502
std	5.024563	0.029133	1.517184	0.656452
min	0.23768	0.85	2	0.00905
25%	11.18456	0.865525	2.2	0.404837
50%	14.17468	0.883959	2.6	0.682798
75%	17.97568	0.908508	3.4	1.119787
max	31.40588	0.996859	10	7.904606

60m - Match Score > 0.9				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	26428	26428	26428	26428
mean	14.387004	0.924825	2.831588	0.839935
std	5.068612	0.018917	1.153515	0.617361
min	0.23768	0.9	2	0.03077
25%	11.03252	0.909363	2	0.410047
50%	14.02264	0.920817	2.4	0.680341
75%	17.97568	0.93672	3.2	1.09071
max	31.40588	0.996859	10	6.299299

50m - Match Score > 0.5				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	743802	742806	742806	742806
mean	14.399162	0.699552	4.056511	0.957392
std	4.928487	0.113232	2.381468	0.803229
min	0.204	0.5	2	0
25%	11.18203	0.606329	2.4	0.395084
50%	14.01507	0.696761	3	0.741384
75%	17.96109	0.787689	5	1.282818
max	31.16508	0.998081	10	8.635459

50m - Match Score > 0.6				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	569430	569430	569430	569430
mean	14.37927	0.744455	3.938865	0.960255
std	4.936036	0.088518	2.286094	0.789962
min	0.204	0.6	2	0
25%	11.18203	0.670479	2.2	0.406527
50%	14.01507	0.738008	3	0.746203
75%	17.96109	0.811385	5	1.279346
max	31.16508	0.998081	10	8.348109

50m - Match Score > 0.7				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	364569	364569	364569	364569
mean	14.340873	0.797073	3.721399	0.947539
std	4.947061	0.063845	2.103031	0.764817
min	0.204	0.7	2	0
25%	11.13144	0.743488	2.2	0.411947
50%	13.96448	0.789466	2.8	0.739392
75%	17.85991	0.843255	4.4	1.254512
max	31.16508	0.998081	10	8.348109

50m - Match Score > 0.8				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	162698	162698	162698	162698
mean	14.289257	0.857009	3.368328	0.915408
std	4.960057	0.040273	1.765368	0.719198
min	0.204	0.8	2	0
25%	11.08085	0.823414	2.2	0.412326
50%	13.96448	0.850164	2.6	0.718991
75%	17.80932	0.884622	3.8	1.201354
max	31.16508	0.998081	10	8.081491

50m - Match Score > 0.85				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	81587	81587	81587	81587
mean	14.264521	0.889962	3.123451	0.893675
std	4.959201	0.029173	1.503877	0.686622
min	0.204	0.85	2	0.009034
25%	11.03026	0.865862	2.2	0.414958
50%	13.96448	0.884484	2.6	0.706727
75%	17.75873	0.909159	3.4	1.165979
max	29.95092	0.998081	10	7.092093

50m - Match Score > 0.9				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	26944	26944	26944	26944
mean	14.188183	0.924791	2.82863	0.866461
std	4.956565	0.01882	1.148965	0.637224
min	0.204	0.9	2	0.031318
25%	10.97967	0.909453	2	0.42339
50%	13.91389	0.920921	2.4	0.696259
75%	17.65755	0.936414	3.2	1.117798
max	29.95092	0.998081	10	5.42714

40m - Match Score > 0.5				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	743802	743802	743802	743802
mean	14.168576	0.699739	4.041674	0.98376
std	4.831116	0.113306	2.366447	0.827247
min	0.18	0.5	2	0
25%	11.0504	0.60649	2.4	0.402924
50%	13.8312	0.696919	3	0.760168
75%	17.6232	0.787943	5	1.322485
max	30.56656	0.99917	10	9.591735

40m - Match Score > 0.6				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	570623	570623	570623	570623
mean	14.158547	0.744559	3.923819	0.986835
std	4.83504	0.088629	2.269818	0.814944
min	0.18	0.6	2	0
25%	10.99984	0.670312	2.2	0.414311
50%	13.8312	0.738133	3	0.765623
75%	17.6232	0.811803	5	1.31845
max	30.56656	0.99917	10	8.801291

40m - Match Score > 0.7				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	365361	365361	365361	365361
mean	14.109363	0.797281	3.710411	0.974681
std	4.843169	0.063827	2.088244	0.789713
min	0.18	0.7	2	0
25%	10.94928	0.743584	2.2	0.419259
50%	13.78064	0.789662	2.8	0.759067
75%	17.52208	0.843578	4.4	1.29476
max	30.21264	0.99917	10	8.451526

40m - Match Score > 0.8				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	163520	163520	163520	163520
mean	14.045148	0.857076	3.367227	0.943255
std	4.856862	0.040192	1.757954	0.742397
min	0.18	0.8	2	0
25%	10.89872	0.8236	2.2	0.421333
50%	13.73008	0.850259	2.6	0.741547
75%	17.47152	0.884476	3.8	1.244082
max	30.21264	0.99917	10	7.901307

40m - Match Score > 0.85				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	82134	82134	82134	82134
mean	14.012836	0.889879	3.133462	0.923621
std	4.850383	0.029149	1.50651	0.710132
min	0.18	0.85	2	0.009029
25%	10.84816	0.865815	2.2	0.425547
50%	13.67952	0.884267	2.6	0.730822
75%	17.42096	0.908882	3.4	1.211468
max	30.21264	0.99917	10	7.263637

40m - Match Score > 0.9				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	27026	27026	27026	27026
mean	13.979614	0.924747	2.838341	0.905009
std	4.846726	0.018895	1.16458	0.668751
min	0.18	0.9	2	0.030697
25%	10.7976	0.909259	2	0.438187
50%	13.67952	0.920725	2.4	0.731916
75%	17.3704	0.936481	3.2	1.179427
max	29.7576	0.99917	10	6.475918

30m - Match Score > 0.5				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	744072	744072	744072	744072
mean	13.829926	0.699403	4.0265	0.994702
std	4.698309	0.113376	2.356452	0.830139
min	0.19	0.5	2	0
25%	10.77376	0.606076	2.4	0.40934
50%	13.50832	0.696371	3	0.770451
75%	17.20504	0.787718	5	1.339549
max	30.01696	0.99919	10	10.412834

30m - Match Score > 0.6				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	569894	569894	569894	569894
mean	13.806807	0.744437	3.903976	0.996243
std	4.708452	0.088694	2.255362	0.816449
min	0.19	0.6	2	0
25%	10.72312	0.670189	2.2	0.419589
50%	13.45768	0.738021	3	0.774069
75%	17.20504	0.811679	4.8	1.334418
max	30.01696	0.99919	10	8.52627

30m - Match Score > 0.7				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	364456	364456	364456	364456
mean	13.755695	0.797301	3.688	0.983537
std	4.714251	0.06388	2.069626	0.789962
min	0.24064	0.7	2	0
25%	10.67248	0.743621	2.2	0.425069
50%	13.45768	0.789779	2.8	0.766834
75%	17.10376	0.843742	4.4	1.309644
max	30.01696	0.99919	10	8.50946

30m - Match Score > 0.8				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	163199	163199	163199	163199
mean	13.674294	0.857107	3.340134	0.950623
std	4.720575	0.040244	1.7318	0.743672
min	0.24064	0.8	2	0
25%	10.62184	0.823537	2.2	0.426823
50%	13.3564	0.850409	2.6	0.745131
75%	17.00248	0.884524	3.8	1.254574
max	29.66248	0.99919	10	8.50946

30m - Match Score > 0.85				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	82163	82163	82163	82163
mean	13.617301	0.88985	3.100766	0.92244
std	4.713094	0.029231	1.471237	0.702193
min	0.24064	0.85	2	0.009043
25%	10.52056	0.865555	2.2	0.428029
50%	13.3564	0.884257	2.6	0.733104
75%	16.95184	0.909081	3.4	1.208202
max	29.66248	0.99919	10	7.257387

30m - Match Score > 0.9				
	Wind Speed (m/s)	Abs. Match Score	D	Abs. Amp
count	27002	27002	27002	27002
mean	13.592061	0.924892	2.816028	0.907484
std	4.722423	0.01885	1.120602	0.667906
min	0.24064	0.9	2	0.031349
25%	10.52056	0.909487	2	0.437345
50%	13.3564	0.920976	2.4	0.733074
75%	16.9012	0.936562	3.2	1.181689
max	29.66248	0.99919	10	6.08172

Bibliography

[1] Bouckaert, S., Pales, A. F., McGlade, C., Remme, U., Wanner, B., Varro, L., D'Ambrosio, D., & Spencer, T. (2021). Net zero by 2050: A roadmap for the global energy sector.

[2] Tremeac, B., & Meunier, F. (2009). Life cycle analysis of 4.5 mw and 250 w wind turbines. *Renewable and sustainable energy reviews*, *13*(8), 2104–2110.

[3] Windeurope - wind energy. (2023-04-27).

[4] Association, E. W. E., et al. (2011). *Eu energy policy to 2050*. EWEA.

[5] Ren, Z., Verma, A. S., Li, Y., Teuwen, J. J., & Jiang, Z. (2021). Offshore wind turbine operations and maintenance: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, *144*, 110886.

[6] Albadi, M. H., & El-Saadany, E. (2010). Overview of wind power intermittency impacts on power systems. *Electric power systems research*, *80*(6), 627–632.

[7] Brasseur, O. (2001). Development and application of a physical approach to estimating wind gusts. *Monthly Weather Review*, *129*(1), 5–25.

[8] Suomi, I., & Vihma, T. (2018). Wind gust measurement techniques—from traditional anemometry to new possibilities. *Sensors*, *18*(4), 1300.

[9] Hewston, R., & Dorling, S. R. (2011). An analysis of observed daily maximum wind gusts in the uk. *Journal of Wind Engineering and Industrial Aerodynamics*, *99*(8), 845–856.

[10] Ágústsson, H., & Ólafsson, H. (2009). Forecasting wind gusts in complex terrain. *Meteorology and Atmospheric Physics*, *103*(1-4), 173–185.

[11] Lehmann-Matthaei, B. (2023). Research platform in the north and baltic seas [Accessed: 2023].

[12] Pasari, S., & Shah, A. (2020). Time series auto-regressive integrated moving average model for renewable energy forecasting. *Enhancing Future Skills and Entrepreneurship: 3rd Indo-German Conference on Sustainability in Engineering*, 71–77.

[13] Liu, X., Lin, Z., & Feng, Z. (2021). Short-term offshore wind speed forecast by seasonal arima-a comparison against gru and lstm. *Energy*, *227*, 120492.

[14] Singh, S., Mohapatra, A., et al. (2019). Repeated wavelet transform based arima model for very short-term wind speed forecasting. *Renewable energy*, *136*, 758–768.

[15] Woolf, P., Burge, C., Keating, A., & Yaffe, M. (2004). Statistics and probability primer for computational biologists. *Massachusetts Institute of Technology*.

[16] Association, E. W. E., et al. (2012). *Wind energy-the facts: A guide to the technology, economics and future of wind power*. Routledge.

[17] St. Amant, R., & Cohen, P. R. (1997). Interaction with a mixed-initiative system for exploratory data analysis. *Proceedings of the 2nd international conference on Intelligent user interfaces*, 15–22.

[18] Quan, J., Zhan, W., Chen, Y., Wang, M., & Wang, J. (2016). Time series decomposition of remotely sensed land surface temperature and investigation of trends and seasonal variations in surface urban heat islands. *Journal of Geophysical Research: Atmospheres*, *121*(6), 2638–2657.

[19] Castillo, E., Hadi, A. S., Balakrishnan, N., & Sarabia, J.-M. (2005). Extreme value and related models with applications in engineering and science.

[20] Milan, P., Wächter, M., & Peinke, J. (2013). Turbulent character of wind energy. *Physical review letters*, *110*(13), 138701.

[21] Garcia, C., Lind, P., Wächter, M., Otero, A., & Peinke, J. (2019). On the existence and characterization of extreme events in wind data. *7th European conference on renewable energy sytems, Madrid/Spain 10-12 June 2019*.

[22] TC88-MT, I. (2005). Iec 61400-3: Wind turbines–part 1: Design requirements. *International Electrotechnical Commission, Geneva*, *64*.

[23] Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology*, *49*(4), 764–766.

[24] Meng, Y., Sun, S., Wang, Y., & Wang, C. (2022). A new compound structure combining dawnn with modified water cycle algorithm-based synchronous optimization for wind speed forecasting. *Energy Reports*, *8*, 12255–12271.

[25] Carcangiu, C., Pujana-Arrese, A., Mendizabal, A., Pineda, I., & Landaluze, J. (2014). Wind gust detection and load mitigation using artificial neural networks assisted control. *Wind Energy*, *17*(7), 957–970.

[26] Ryan, H. (1994). Ricker, ormsby; klander, bntterwo-a choice of wavelets. *CSEG Recorder*, *19*(07).

[27] Taghavifar, H., & Mardani, A. (2014). Wavelet neural network applied for prognostication of contact pressure between soil and driving wheel. *Information Processing in Agriculture*, *1*(1), 51–56.

[28] van den Burg, G. J. J., & Williams, C. K. I. (2020). An evaluation of change point detection algorithms. *ArXiv*, *abs/2003.06222*.

[29] Gandhi, S., Foschini, L., & Suri, S. (2010). Space-efficient online approximation of time series data: Streams, amnesia, and out-of-order. *2010 IEEE 26th International Conference on Data Engineering (ICDE 2010)*, 924–935.

[30] Cannon, D. J., Brayshaw, D. J., Methven, J., Coker, P. J., & Lenaghan, D. (2015). Using reanalysis data to quantify extreme wind power generation

statistics: A 33 year case study in great britain. *Renewable Energy*, *75*, 767–778.

[31] Kanev, S., & van Engelen, T. (2010). Wind turbine extreme gust control. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, *13*(1), 18–35.

[32] Fitzwater, L. M., Cornell, C. A., & Veers, P. S. (2003). Using environmental contours to predict extreme events on wind turbines. *Wind Energy Symposium*, *75944*, 244–258.

[33] Chkeir, S., Anesiadou, A., Mascitelli, A., & Biondi, R. (2023). Nowcasting extreme rain and extreme wind speed with machine learning techniques applied to different input datasets. *Atmospheric Research*, *282*, 106548.

[34] Larsen, G. C., Hansen, K. S., & Pedersen, B. J. (2004). 4.2 constrained simulation of critical wind speed gusts by means of wavelets. *Database on Wind Characteristics*, 40.

[35] Haselsteiner, A. F., & Thoben, K.-D. (2020). Predicting wave heights for marine design by prioritizing extreme events in a global model. *Renewable Energy*, *156*, 1146–1157.

[36] Kumar, S., & Sengupta, S. (2022). Multi peak detection algorithm of fiber bragg grating using mexican hat wavelets and hilbert transform. *2022 IEEE India Council International Subsections Conference (INDISCON)*, 1–5.

[37] Hu, W., Letson, F., Barthelmie, R., & Pryor, S. (2018). Wind gust characterization at wind turbine relevant heights in moderately complex terrain. *Journal of Applied Meteorology and Climatology*, *57*(7), 1459–1476.

[38] Soman, S. S., Zareipour, H., Malik, O., & Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. *North American Power Symposium 2010*, 1–8.

[39] D Chandra, R., Sailaja Kumari, M., Sydulu, M., Grimaccia, F., Mussetta, M., et al. (2014). Adaptive wavelet neural network based wind speed forecasting studies. *Journal of Electrical Engineering & Technology*, *9*(6), 1812–1821.

[40] Singh, A., Rawat, A., & Raghuthaman, N. (2022). Mexican hat wavelet transform and its applications. In J. Singh, H. Dutta, D. Kumar, D. Baleanu & J. Hristov (Eds.), *Methods of mathematical modelling and computation for complex systems* (pp. 299–317). Springer International Publishing.

[41] Hu, X., Fang, G., Yang, J., Zhao, L., & Ge, Y. (2023). Simplified models for uncertainty quantification of extreme events using monte carlo technique. *Reliability Engineering & System Safety*, *230*, 108935.

[42] Branlard, E. (2009). Wind energy: On the statistics of gusts and their propagation through a wind farm. *ECN Wind Memo*, *7*, 5–9.

[43] Hannesdóttir, Á., Kelly, M., & Dimitrov, N. (2019). Extreme wind fluctuations: Joint statistics, extreme turbulence, and impact on wind turbine loads. *Wind Energy Science*, *4*(2), 325–342.

[44] Rapella, L., Faranda, D., Gaetani, M., Drobinski, P., & Ginesta, M. (2023). Climate change on extreme winds already affects off-shore wind power availability in europe. *Environmental Research Letters*, *18*(3), 034040.

[45] Mi, X., Ren, H., Ouyang, Z., Wei, W., & Ma, K. (2005). The use of the mexican hat and the morlet wavelets for detection of ecological patterns. *Plant Ecology*, *179*, 1–19.

[46] Hou, T., & Qin, H. (2012). Continuous and discrete mexican hat wavelet transforms on manifolds. *Graphical Models*, *74*(4), 221–232.

[47] Liu, W., Gao, Q., & Zhang, Y. (2016). A novel wind turbine de-noising method based on the genetic algorithm optimal mexican hat wavelet. *2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, 1003–1006.

[48] Liu, W., & Han, J. (2013). The optimal mexican hat wavelet filter de-noising method based on cross-validation method. *Neurocomputing*, *108*, 31–35.

[49] Ren, H., Liu, W., Jiang, Y., & Su, X. (2017). A novel wind turbine weak feature extraction method based on cross genetic algorithm optimal mhw. *Measurement*, *109*, 242–246.

[50] Tian, Z. (2023). Analysis and research on chaotic dynamics behaviour of wind power time series at different time scales. *Journal of Ambient Intelligence and Humanized Computing*, *14*(2), 897–921.

[51] Zhou, P., & Yin, P. (2019). An opportunistic condition-based maintenance strategy for offshore wind farm based on predictive analytics. *Renewable and Sustainable Energy Reviews*, *109*, 1–9.

[52] Tchakoua, P., Wamkeue, R., Ouhrouche, M., Slaoui-Hasnaoui, F., Tameghe, T. A., & Ekemb, G. (2014). Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges. *Energies*, *7*(4), 2595–2630.

[53] Li, Y., Zhang, J., Li, Z., Yang, P., & Wang, H. (2021). Design and verification of a novel double rotor without stator wind turbine generation system. *Energy Reports*, *7*, 161–172.

[54] Abubaker, A., Kostić, I., & Kostić, O. (2018). Numerical modelling of velocity profile parameters of the atmospheric boundary layer simulated in wind tunnels. *IOP Conference Series: Materials Science and Engineering*, *393*(1), 012025.

[55] Zivot, E., Wang, J., Zivot, E., & Wang, J. (2003). Rolling analysis of time series. *Modeling financial time series with S-Plus®*, 299–346.

[56] Puth, M.-T., Neuhäuser, M., & Ruxton, G. D. (2014). Effective use of pearson's product–moment correlation coefficient. *Animal behaviour*, *93*, 183–189.

[57] Hallinan Jr, A. J. (1993). A review of the weibull distribution. *Journal of Quality Technology*, *25*(2), 85–93.

[58] HOMER-Energy. (2021). Weibull distribution [Accessed: April 26, 2023].