**Institute for Information Technology**
Post address: Postboks 4 St. Olavs plass, 0130 Oslo
Visiting address: Holbergs plass, Oslo
Phone: 22 45 32 00

| PROJECT LETTER |
| --- |
| Group I |

| AVAILABILITY |
| --- |
| |

# European Project Semester

| MAIN PROJECT TITLE: | DATE |
| --- | --- |
| Predicting wind power from wind speed data | 03-12-2019 |
| | NUMBER OF PAGES/APPENDIXES<br>68 pages<br>3 Appendixes |
| PROJECT TEAM MEMBERS<br>Victor Guerrero, Lukas Spannenberger<br>Patrick Kuiper, Maxime Trinquard<br>Ellen Ophuis | INTERNAL SUPERVISOR(S)<br>Pedro Lind<br>Anis Yazidi |

| EXTERNAL SUPERVISOR | CONTACT PERSON |
| --- | --- |
| | |

ABSTRACT

The search for renewable and sustainable energy sources is a global concern. One way to produce renewable energy is by making use of wind turbines. However, wind speed is a stochastic process which fluctuates considerably. This means that energy cannot be produced at a constant rate. For many years, the most used probability density function has been the Weibull distribution. Experimental research is made in this report, in which Nakagami, Normal, and Rician distributions are compared to Weibull distribution using the Kullback-Leibler divergence for different time frames. The results reveal that the four distributions are all numerically acceptable for analysing wind speed data for different time frames. Since Weibull distribution is the commonly used distribution, wind generation companies can continue using this distribution as the standard one. However, almost every time frame manifest multimodal distributed wind speed data. Future research can focus on studying multimodal distributions instead of unimodal.

# PREDICTING WIND POWER FROM WIND SPEED DATA

**Authors**
Victor Guerrero
Patrick Kuiper
Ellen Ophuis
Lukas Spannenberger
Maxime Trinquard

**Supervisors**
Pedro Lind
Anis Yazidi

# List of Contents

# Summary

The search for renewable and sustainable energy sources is a global concern. One way to produce renewable energy is by making use of wind turbines. However, wind speed is stochastic process which fluctuates considerably. This means that wind energy cannot be produced at a constant rate. To overcome this problem, more research needs to be done in wind statistics. Nowadays, the most used probability density function is the Weibull distribution, and it is accepted as a standard for modeling wind speed. In this report, Nakagami, Rician, and Normal distribution will be compared to Weibull distribution using the Kullback-Leibler divergence for different time frames of wind speed data. The fifteen months of wind speed data is donated by FINO, a German company that operates at the Baltic and North Sea.

Based on the results, all the four distributions are numerically acceptable for analysing wind speed data for different time frames. However, it seems that almost every time frame is showing multimodal distributed wind speed data. Since the Weibull distribution is the commonly used distribution in research, one can continue using this distribution as the standard, even if Nakagami, Normal and Rician distribution are also comparable to Weibull when analysing wind speed data.

This report is structured as follows: The introduction exposes the background of the problem and motivates a solution. After that, the theoretical background describes the basic statistical features employed in this report. The research strategy gives information about the methodology employed to get the results from the data. The results chapter shows all the results for different time frames that were generated after executing the research strategy, followed by a conclusion and discussion chapter about the results found. Finally, the learning process of the authors and a group reflection about the European Project Semester-Program is written down in last chapter.

# List of Figures

# List of Tables

# 1. Introduction: the importance of wind

The search for renewable and sustainable energy sources is a global concern. Using fossil fuels to produce energy will contribute to global warming, and therefore renewable energy is getting more important every day. One way to produce energy free of carbon dioxide pollution is to make use of wind turbines. Therefore, making use of wind turbines delivers an important contribution to the UNDP goals (Schwan, 2019) (see appendix B for an explanation about how this project contributes to this goals). There is however a big problem with the energy that is produced by these turbines: wind is a stochastic process which fluctuates considerably. This feature makes it difficult to produce energy at a constant rate and causes unstable scenarios into the power grid. For example, too high energy outputs caused by wind speed fluctuations might lead to an overload of the grid. Because of this non-constant wind speed, it is difficult to predict what the power output will be. Therefore, it is important to better understand the fundamental features underlying the fluctuating wind power of a wind turbine and its statistical connection with the speed of the wind blowing across it.

Probability density functions (PDF's) are used to characterize wind speed observations (Ouarda et al., 2015). There are several PDF's and each one uses its own parameters. The choice of which PDF will be used for these observations are important because wind power is formulated as an explicit function of wind speed distribution parameters. A more accurately fitted PDF for wind speed data will reduce the uncertainties in wind power output estimates. The most used PDF's for wind speed data analysis are the 2-parameter Weibull distribution and the Rayleigh distribution (Hennessey, 1977; Justus, Hargraves, & Yalcin, 1976; Nfaoui, Buret, & Sayigh, 1998; Şahin & Aksakal, 1998). The Rayleigh distribution is a special case of the 2-parameter Weibull distribution, which will be explained later. The reason that these distributions are the most used for characterizing wind speed is because they are generally a good fit to the observed wind speeds; it only requires an estimation of 2 parameters and the estimation of the parameters is simple (Tuller & Brett, 1985).

Despite that the Weibull distribution is well accepted and gives some advantages, it is unable to fit all wind regimes (Carta, Ramírez, & Velázquez, 2009). The use of other PDF's has been proposed in the literature before. For example, the 3-parameter Weibull distribution has been used and it has been concluded to be a better fit than the ordinary 2-parameter Weibull distribution (Stewart & Essenwanger, 1978). Some other distributions, the Nakagami and the Rician distribution, are also used to observe wind speed characteristics (Yu et al., 2019). Yu et al. showed that Nakagami and Rician performed well in predicting wind speed distributions. However, this has only been done for one-, three- and twelve-month(s) time frames. In this project the Nakagami, Rician, Weibull and Normal distribution will be reviewed for different time frames. With help of the Kullback Leibler divergence, Nakagami, Rician and Normal distributions will be compared to Weibull distribution in order to know which distribution fits better for a certain time frame.

# 2. Theoretical background

This chapter describes the different probability distributions used in this report to explain about the statistics that will be used for analyzing wind speed data. First, the basics about wind is described. Next, the basic statistics used in this report are explained. After that, the different distributions that were used are explained.

## 2.1 What is wind?

Wind is atmospheric motion and one of the basic physical elements of our environment (Kalmikov, 2017). The strength of the wind is defined by the velocity of the air motion; this is directly related to the amount of energy in the wind: its kinetic energy. The source of this energy is solar radiation. The earth's surface is unevenly heated by the electromagnetic radiation from the sun. Differential absorption of sunlight by soil, rock, water and vegetation causes the air to warm up at different rate in these different regions. This uneven heating is being converted to air motion through convective processes and then becomes adjusted by rotation of the earth.

Since there is only data available about the wind speed for this project, it is not possible to look at the different wind directions. Therefore, this will be left out of consideration.

## 2.2 Basic statistics

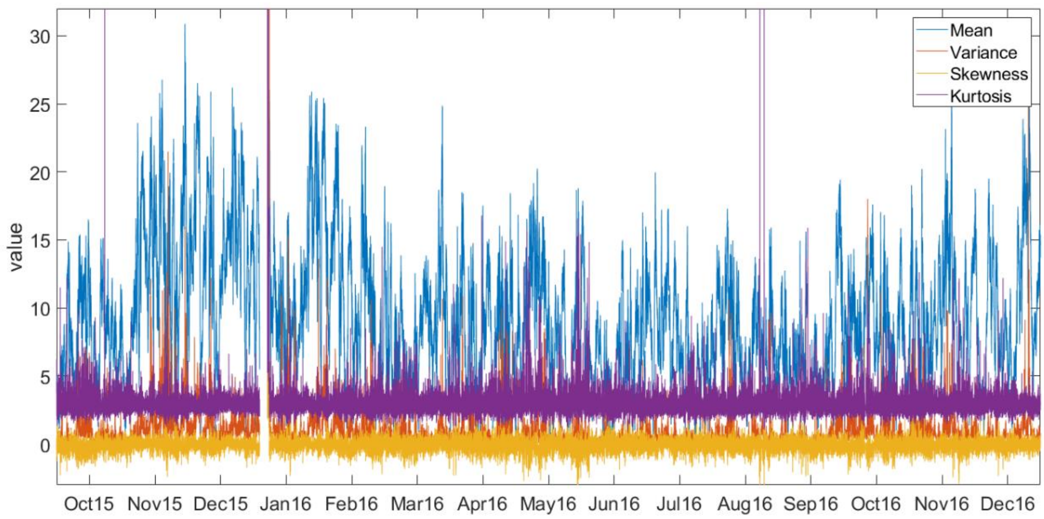


*Figure 1: The four statistical moments.*

**Probability density function**

A probability density function (PDF) is a mathematical distribution that provides the probabilities of occurrence of different possible outcomes in an experiment. Probability distributions are divided into two classes: discrete probability distributions and continuous probability distributions. A discrete probability distribution is applicable where the set of possible outcomes is discrete, for example: when rolling a dice the possible outcomes are 1, 2, 3, 4, 5 or 6. A continuous probability distribution is applicable to the scenarios where the set of possible outcomes can take on values in a continuous range, such as wind speed during the day. Plotting the PDF's of different time frames of wind speed will be done using discrete probability distributions.

Every PDF is shaped in a specific way. How they are shaped depends on the four mathematical moments. These moments are a set of constants that represent important properties of the distributions (Pilling, 2017). In this report the following moments will be used: the mean, variance, skewness, and kurtosis. The mean gives information about the value of the wind speed at that time. As an example, the average wind speed from 19:30 to 19:40 of a cold and rainy day of January is not the same as from 12:10 to 12:20 of a hot and sunny day of June. The variance gives information about how much difference there is between the values within the time frame. To show how the data varies over time, the four moments are calculated for the fifteen months for ten-minute time frames. This is shown in Figure 1.

The first moment is the mean which gives information about the average wind speed. For a continuous distribution the mean has the following formula:

$\mu = \int_{-\infty}^{\infty} x\, f(x) dx$; where *f(x)* is the probability density function. (1)

The second moment is the variance. The variance is a measure of the variability of all possible values of a variable. In this report the variance gives information about the fluctuation of the wind speed. The positive square root of the variance is the standard deviation. The formula for the standard deviation is (Spatz, 2007):

$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}}$; where $\sigma$ is the standard deviation, *N* is the number of scores, and $\mu$ is the mean. (2)

The third moment is the skewness. The skewness gives information about if there is more often a lower or a higher wind speed. A negative skewness means that the left tail of the distribution is longer, a positive skewness means that the right tail of the distribution is longer. The skewness can be calculated by making use of the following formula:

$\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}$; where $\mu$ is the mean and $\sigma$ is the standard deviation. (3)

The fourth moment is the kurtosis. The coefficient of kurtosis measures the degree of flatness of the distribution (Pilling, 2017). In other words, it tells if the tails are steep or not. The formula that calculates the kurtosis is:

$\gamma_2 = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4};$ where $\mu$ is the mean and $\sigma$ is the standard deviation. (4)

**Kullback-Leibler divergence**

To compare how well certain PDF's fit the discrete PDF's of the data, the Kullback-Leibler divergence (KL-divergence) is used. The KL-divergence, also known as relative entropy, is a mathematical tool to measure the difference between two PDF's *P* and *Q* on the same statistical manifold (Calin & Constantin, 2014). This tool creates a non-commutative measurement which is defined by:

$$D_{KL}(P||Q) = \sum_i P(i) \log\left(\frac{P(i)}{Q(i)}\right) \quad (5)$$

Since discrete probably distributions have been used, the sum of all KL 'distances' needs to be used instead of an integral. The relative entropy between the two PDF's P(*i)* and *Q(i)* will be calculated and give a 'distance'. The KL-divergence is always non-negative due to Gibbs' inequality (Hershey & Olsen, 2007). The divergence is only zero when the PDF *P(i)* exactly matches the expected distribution *Q(i)*. Furthermore, the value of the KL tells how much information is lost when approximating *P(i)* into *Q(i)*. A lower value for the KL-divergence means the PDF fits the PDF of the data better than when a higher value for divergence is computed using another PDF. As mentioned, the KL-divergence is not computing a real distance. However, to make it more practicable, from now on the word distance will be used whenever KL-divergence will be compared.

## 2.3 Statistics and wind

This subparagraph describes the different distributions and the Kullback-Leibler divergence that are used to analyze wind speed data.

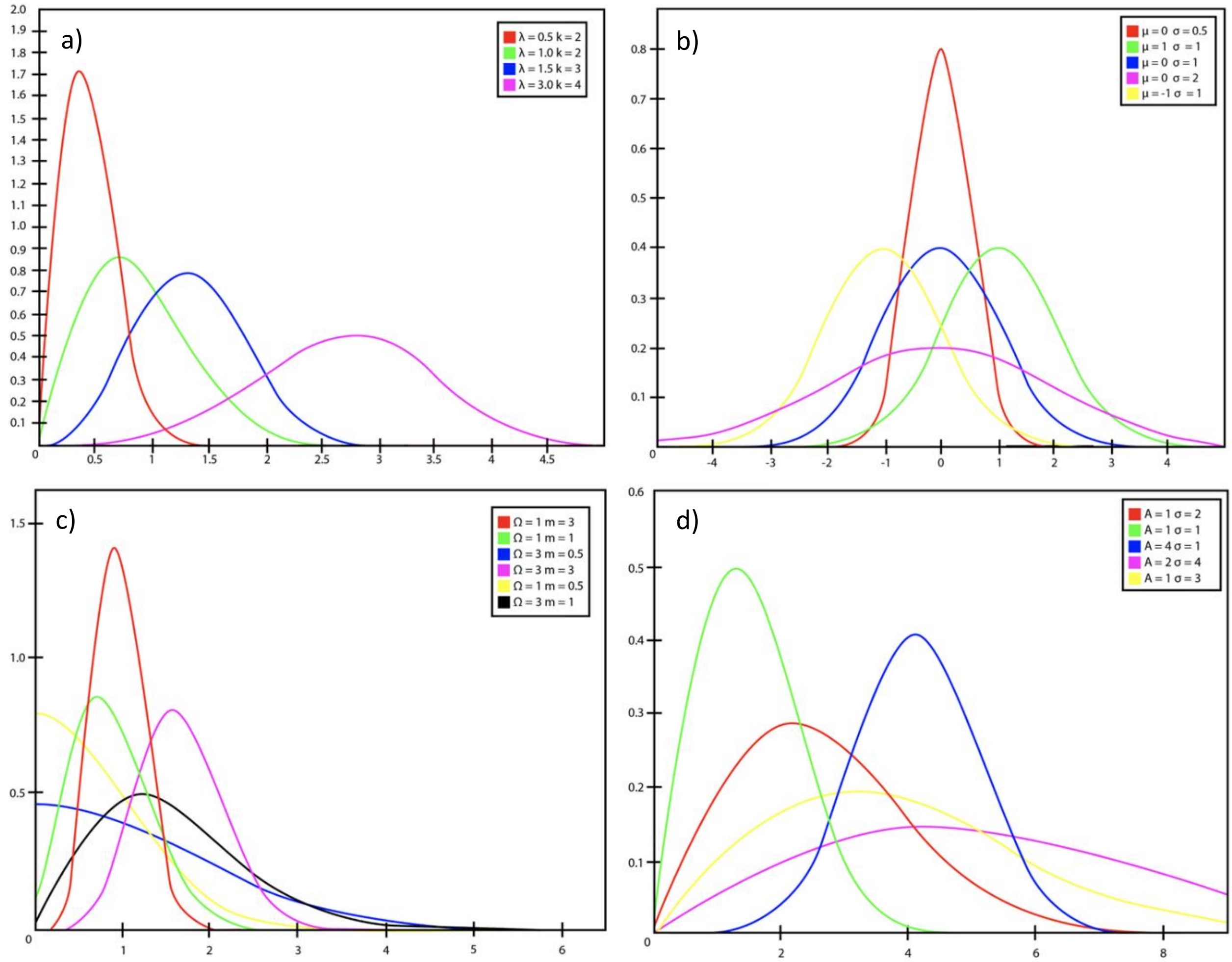


*Figure 2: Different shapes of a) Weibull distribution; b) Normal distribution; c) Nakagami distribution; d) Rician distribution.*

**Weibull distribution**

The Weibull distribution is one of the continuous probability distributions that will be used during this project. There are two types of Weibull distributions that are commonly used: the two-parameter Weibull distribution and the three-parameter Weibull distribution. The two-parameter has a scale and shape parameter. The three-parameter has these as well, but also includes a location parameter. The general form of Weibull PDF is as follows (Ali, Lee, & Jang, 2018) :

$$\mathrm{f}_{\mathrm{Wei}}(k,c,v) = \left(\frac{\mathrm{k}}{\lambda}\right)\left(\frac{\mathrm{v}}{\lambda}\right)^{\mathrm{k}-1} \mathrm{e}^{-\left(\frac{\mathrm{v}}{\lambda}\right)^{\mathrm{k}}} (v > 0,\ \lambda > 0) \tag{6}$$

In this function, $k$ is the shape parameter (dimensionless), $\lambda$ is the scale parameter (m/s), $v$ is the measured wind speed (m/s), and f(v) is the probability of occurrence of wind speed v. Figure 2a shows the different shapes of the Weibull distribution with the two parameters and their values. When k = 2 and λ = 1.0, the distribution is considered a Rayleigh distribution.

**Normal distribution**

The normal distribution (also known as the Gaussian distribution) is the most important continuous probability distribution in statistics, because it fits many natural phenomena. A normal distribution is a bell-shaped, symmetrical, theoretical distribution based on a mathematical formula rather than on empirical observations (Spatz, 2008). Different to other distributions, the mean, median and mode are the same with the normal distribution. They have the same number on the X-axis (the number where the peak is in the curve). Another special part of the normal distribution is if you draw a line from the place where the graph peaks to the mean (X-axis). The area under the curve on the left side of this line is the same as the area on the right side of this line. The tails of the curve are indefinite (they keep continuing for ever and the distance between the line and the X-axis keeps getting less). The inflection points (this means where the curve is the steepest and this is also the point from where the curve changes from bending upward to bending over (Spatz, 2008)) of a normal distribution are exactly -σ1 and +σ1. The PDF of the normal distribution is:

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{7}$$

In this function, $\mu$ is the mean (and median and mode), $\sigma$ is the standard deviation, and $\sigma^2$ is the variance. Figure 2b shows the different shapes of the Normal distribution with the parameters and their values.

**Nakagami distribution**

One of the PDFs that will be used is the Nakagami distribution. This distribution is often used to model positive valued data with right skewness (Ozonur, Akdur, & Bayrak, 2018). The Nakagami (or Nakagami-m) distribution has found many applications in technical sciences. It has been applied for signal-to-noise ratio estimation (Lira de Queiroz, Brito Teixeira de Almeida, Madeiro, & Araújo Lopes, 2019), fading channel (Odeyemi & Owolawi, 2019), quantitative ultrasound imaging (Roy-Cardinal, Destrempes, Soulez, & Cloutier, 2019) and so on. The PDF of the Nakagami distribution is as follows (Huang, 2016):

$$f_{(n)} = \frac{2}{\Gamma(m)} \left(\frac{\mathrm{m}}{\Omega}\right)^{m} n^{2m-1} e^{-\frac{m}{\Omega}n^2} \tag{8}$$

The function has a shape parameter $m \geq 0.5$ and a scale parameter $\Omega > 0$ controlling spread. The shape parameter $m$ is also called the Nakagami parameter. Γ(m) is the Gamma-function, which is another distribution the Nakagami is related to. One must notice that illustrative variable $n > 0$. Figure 2c shows the different shapes of the Nakagami distribution with its associated values for the parameters.

**Rician distribution**

The Rician (or Rice, Nakagami-n) distribution is a probability distribution, which is describing the magnitude of a circular bivariate normal illustrative variable with potentially non-zero mean. The inventor is Stephan O. Rice and it was named after him. The first propose was in 1945 and the fields were telecommunication, information theory and communication theory (Le, 2019; Rice, 1945). The Rician distribution is similar to the Nakagami and Rayleigh distributions. In communications theory, Nakagami distributions, Rician distributions, and Rayleigh distributions are used to model scattered signals that reach a receiver by multiple paths. Depending on the density of the scatter, the signal will display different fading characteristics. Rayleigh and Nakagami distributions are used to model dense scatters, while Rician distributions model fading with a stronger line-of-sight.

Nakagami distributions can be reduced to Rayleigh distributions, but give more control over the extent of the fading ("Rician Distribution - MATLAB & Simulink - MathWorks Benelux," n.d.). The Rician distribution can be seen as the distribution that models the square root of the sum of squares of two independent and identically distributed Gaussian illustrative variable with the same *σ* (Aja-Fernández, Vegas-Sánchez-Ferrero, Aja-Fernández, & Vegas-Sánchez-Ferrero, 2016, p. 278). The probability density function of the Rician distribution defined as (Aja-Fernández et al., 2016, p. 278):

$$p(x|A,\sigma) = \frac{x}{\sigma^2} \times e^{-\frac{x^2+A^2}{2\sigma^2}} \times I_0\left(\frac{Ax}{\sigma^2}\right) \quad (9)$$

where $I_0$ is the zero-order modified Bessel function of the first kind, A is the noncentrality parameter, and σ is the scale parameter. Figure 2d shows the different shapes for the Rician distribution with the different values for the two parameters.

# 3. Research strategy: how to manage 40 million values?

In this chapter the steps that were taken to get a solution, the process and methodology used, and possible difficulties in the process are shown.

**Analysis of the data**

The wind speed is measured by an anemometer placed on a 100 m high tower from FINO1 research platform. FINO1 is located at the Alpha Ventus wind farm, at 45 km north from the Dutch-German border. The exact location of the platform is: N 54°00'53.5" E 6°35'15.5". The wind speed is measured at 100 m height in 1 Hz intervals from October 2015 until December 2016. This means that fifteen months of data has been collected containing around 40 million data points with five decimal accuracy (e.g., 3.14159 m/s). The data is analyzed using a computing program with a precision of 16 digits (e.g., 3.141592653589793 m/s).

For the data analysis, Matlab software with its statistical toolbox is used. The 15 months data is divided into shorter time frames for further analysis. To describe how the data is distributed, discrete PDF’s are used to fit the data into histograms. Matlab has a histogram function including an algorithm that automatically calculates the number of bins needed for the histograms of the different time frames. Therefore, higher accuracy for distributing the data is guaranteed. Unfortunately, there are some missing values in the data that appear as Not a Number (NaN) values. These NaN values will stay in the dataset and are also considered. However, when time frames only consist of NaN values, this will be left out of the results. In order to cover certain time ranges, only a certain selection of time frames will be used. These time frames are: one-month, two-week, one-week, two-day, twelve-hour, one-hour, and ten-minute.

**Conversion of the data into a distribution**

After distributing the data into time frames using discrete PDF’s, the data will be fitted into Weibull, Normal, Rician and Nakagami probability distributions. In order to define the different probability distributions, their own parameters need to be determined. These parameters determine the abscissa scale and width of a discrete PDF of wind speed, respectively. Therefore, it is important to determine the distribution parameters as accurately as possible.

In order to estimate these parameters one can use various mathematical approaches, such as least-squares, maximum likelihood, momentum, and empirical methods (Abdul Baseer & Meyer, 2016; Jiang, Wang, Wu, & Geng, 2017). This project is not about finding the best way to estimate the parameters. Thus, the choice has been made to work with maximum likelihood, since that is the easiest mathematical approach to work in Matlab.

**Compare distributions**

One illustrative PDF for each time frame will be shown in a graph. The four fitted probability distributions will be visualized in the same graph to get an idea how the distributions are fitted. To show how all the data of the 15 months is distributed within a certain time frame, the mean of every window from that specific time frame is used to fit a PDF. The four probability distributions will also be visualized in the same graph. For every window of the used time frame, the KL-divergence will be computed for every probability distribution. This can be visualized in a scatter plot using the Weibull distribution as reference. For every window of a certain time frame, the sum of all KL-divergences will be calculated.

To compare how well the probability distributions fit the PDF's of the data, the mean of the sum of the KL-divergences needs to be calculated. This needs to be done because there is a different number of windows for every time frame. All the means of the KL-divergences for every probability distribution will be compared to each other for every time frame. This can be visualized in a graph by showing the means of the KL-divergences vs the used time frame. The standard deviation will also be shown in this graph to make sure if there is a significant difference between the four probability distributions.

# 4. Results: which distribution has the best fit for the data?

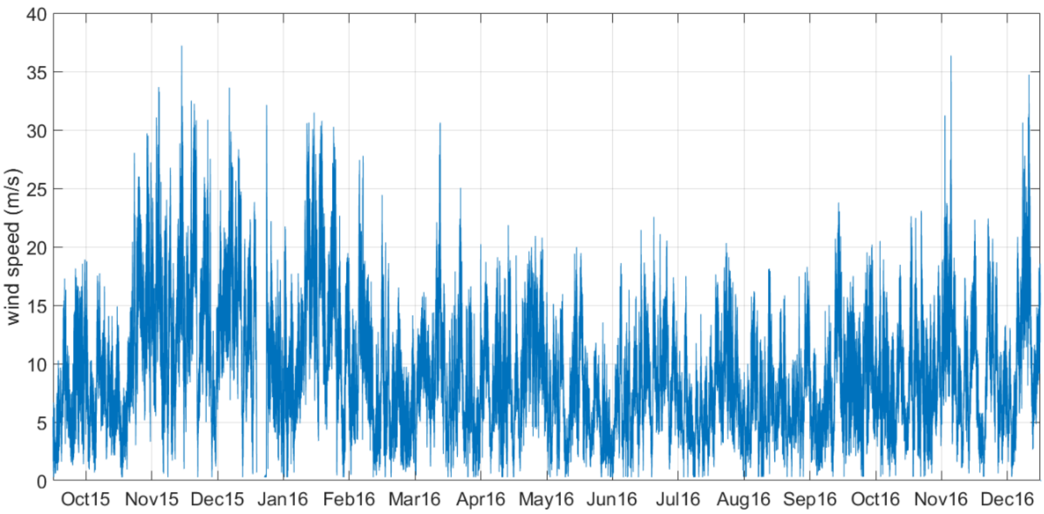

*Figure 3: Graphing fifteen months of 1 Hz wind speed data.*

As described before, the data used in this report is measured from 1 October 2015 up to 31 December 2016 at a rate of 1 Hz. Figure 3 shows all this data, which contains 40 million values of wind speed measurements in total. As can be seen, the wind speed fluctuates between 0 m/s (when there is no wind measured) and 38 m/s, which is the wind speed that occurs during a hurricane. This last wind speed only lasted for a few seconds. There is a gap in Figure 3 in January 2016, because there was only measured a value of 0.184 for about 4 days. This is considered as noise and changed into NaN values. As said before, the data also consists of missing values that already appeared as NaN values; these NaN values are spread within the months, mostly occurring at the end of every month. The noise and the NaN values could be caused by any malfunction of the anemometer. For example, when the anemometer is frozen it will not measure valid values of wind speed. The exact causes of the malfunctions are unknown and not investigated. Non-recorded data will not be estimated based on the existing data.

In the following subsections of this chapter, the four distributions that are explained in the second chapter are used to represent the PDF's of the data for different time frames. As mentioned, the time frames that will be used are: one-month, two-week, one-week, two-day, twelve-hour, one-hour and ten-minute time frames.

Every time frame has, in order of appearance:

- The histogram of an illustrative time frame, to visualize how the data is distributed fitting the four probability distributions in the histogram.
- The histogram for all the fifteen-month data using the mean value of every window for that time frame, fitting the four probability distributions in the histogram to represent the data.
- One table numerically showing the values of the KL-divergence for each of the four probability distributions for fifteen windows of that time frame. At the end of the table, the mean value of the sum of all KL-divergences is shown, considering all the windows of that certain time frame. The mean is computed because some probability distributions caused problems when computing the parameters for some windows of the time frames; which means that not every window is taken into account for some probability distributions. Thus, the mean is needed to make a fair comparison.
- A scatter plot of the KL-divergences for all the windows of a time frame, comparing Nakagami, Normal, or Rician distribution against Weibull distribution (the reference distribution). Only one scatter plot will be shown for every time frame, because the diagrams itself look very similar to each other and therefore no more scatter plots are needed to explain what happens in that time frame. For the first three time frames Normal distribution is used, for the two following time frames Rician distribution is used, and for the last two time frames Nakagami is used. The last three time frames present a large number of windows, so a zoom of these scatter plots is also added. The scatter plots that are not shown in this chapter, are shown in Appendix C.

The last chapter of this section compares all means of the KL-divergences for every time frame. To extract the big concept of the report, two tables and one diagram are shown to summarize all the time frames:

- The first table summarizes the mean of the KL-divergences considering all the windows of the time frames (mean values shown on the tables) and providing the mean of the means to see which distribution has the lowest divergence in average.
- The second table shows the standard deviation of the KL-divergences considering all the windows of the time frames and providing the mean of the standard deviations to see which distribution has the lowest divergence in average when compared with the other distributions.
- A diagram shows the means of the KL-divergences for every distribution of all the time frames including the standard deviations.

## 4.1 One-month time frame

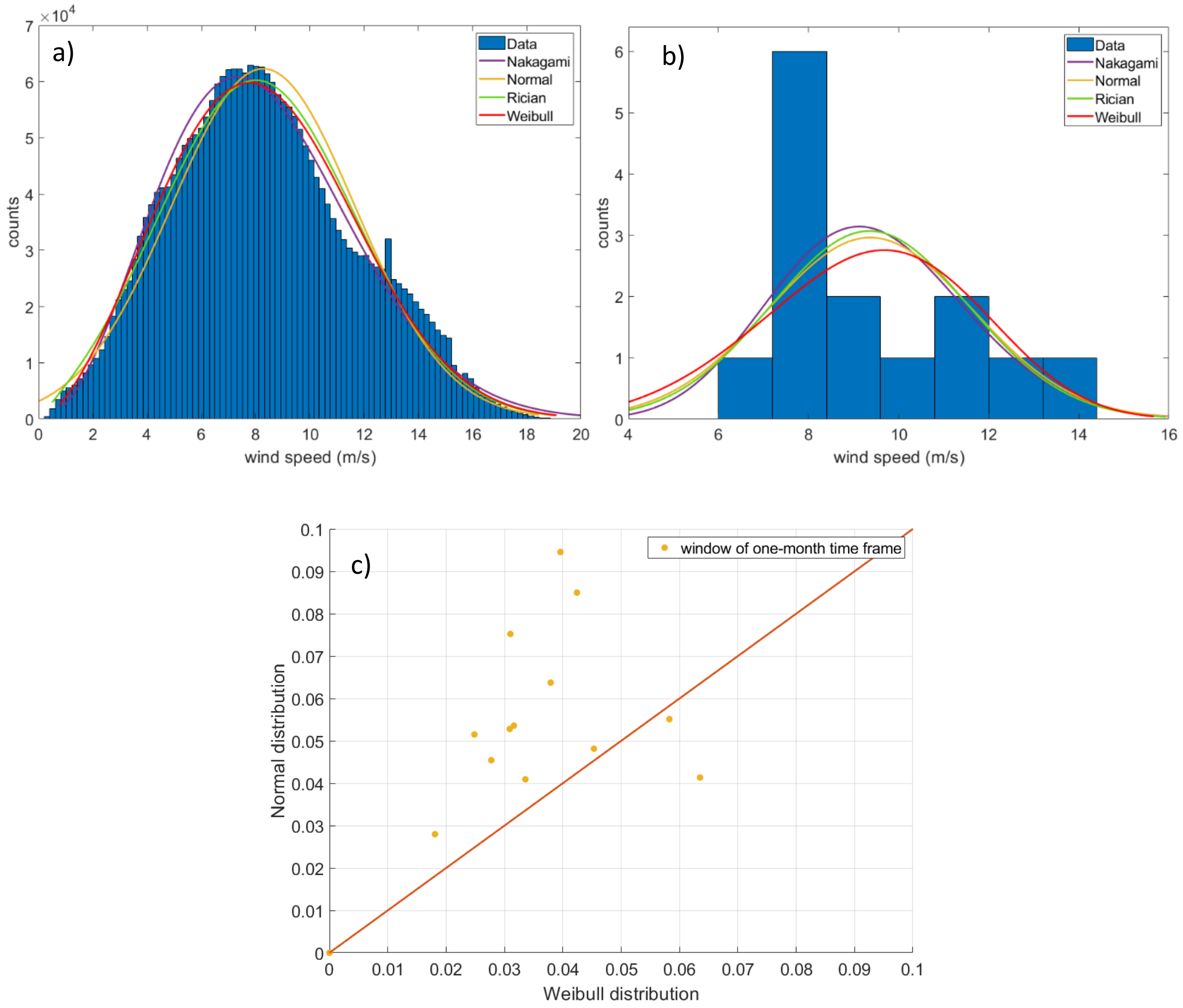


*Figure 4: a) Comparison of the different probability distributions for an illustrative window of one-month time frame of distributed wind speed data. b) Comparison of the different probability distributions for the means of one-month time frame for all the data. c) Scatter plot of all KL-divergences of Normal vs Weibull distribution for one-month time frame.*

In this subsection the comparison of the four distributions of the one-month time frame is showed. Figure 4a is an illustrative one-month time frame. The figure shows that Weibull, Normal, Nakagami, and Rician nearly have the same fit. But when looking at the KL-divergence in Table 1, it can be concluded that for one-month time frames Weibull and Rician distribution have the best fit. When looking at Figure 4a, it seems that the data is unimodal distributed. However, when looking at other windows of this time frame (see Appendix C; Figures 14a-b), it seems that the data is bimodal distributed. Figure 4b shows the comparison of the four probability distributions for the means of the one-month time frames for all the data. It shows all the data divided in the means of one-month time frames. As can be seen, this distribution of the data does not fit any of the probability distributions. Nakagami distribution has more divergence than the first two, while Normal distribution has the highest divergence. In other words, Normal distribution has the worst fit to the data.

The numerical values for Normal and Weibull distribution of Table 1 are plotted in figure 4c. It shows the KL-divergences for Normal distribution vs Weibull for the windows of one-month time frame. When looking at the graph, one can see that almost all the values are above the red line y=x. This means Weibull distribution has shorter distances than Normal distribution. It can be concluded that Weibull distribution fits the wind speed data the best for the one-month time frame.

| Weibull | Rician | Normal | Nakagami |
|---|---|---|---|
| **KL-divergence values** | | | |
| 0.041 | 0.034 | 0.028 | 0.037 |
| 0.051 | 0.025 | 0.035 | 0.022 |
| 0.053 | 0.031 | 0.035 | 0.029 |
| 0.028 | 0.018 | 0.021 | 0.022 |
| 0.048 | 0.045 | 0.039 | 0.053 |
| 0.075 | 0.031 | 0.031 | 0.031 |
| 0.045 | 0.028 | 0.027 | 0.029 |
| 0.095 | 0.040 | 0.041 | 0.040 |
| 0.041 | 0.064 | 0.047 | 0.100 |
| 0.054 | 0.032 | 0.039 | 0.033 |
| 0.055 | 0.058 | 0.050 | 0.062 |
| 0.121 | 0.112 | 0.113 | 0.113 |
| 0.085 | 0.042 | 0.055 | 0.035 |
| NaN | NaN | NaN | NaN |
| 0.064 | 0.038 | 0.037 | 0.039 |
| **KL-divergence mean** | | | |
| **0.043** | **0.043** | **0.061** | **0.046** |

*Table 1: KL-distances of the four distributions for one-month time frame.*

## 4.2 Two-week time frame

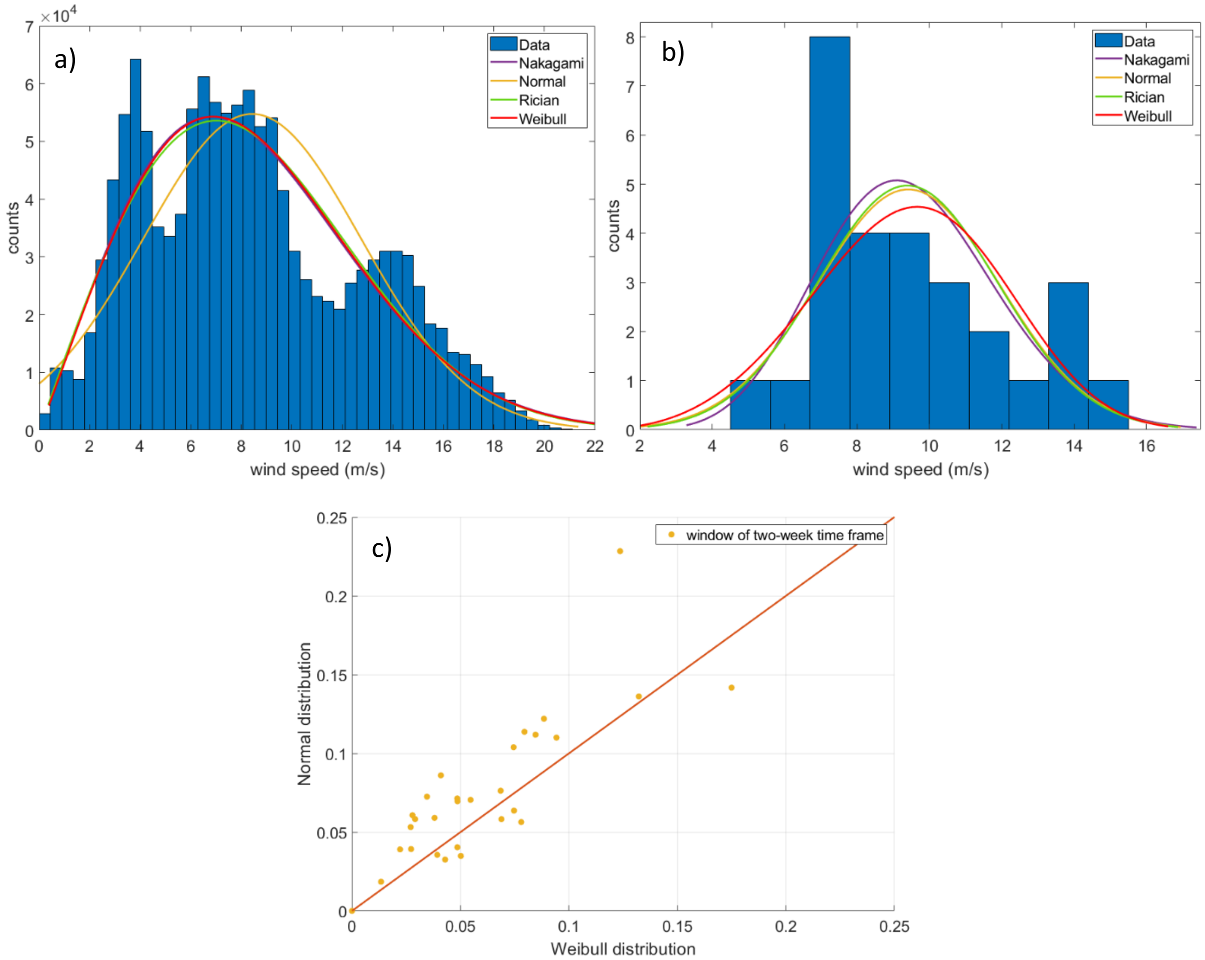


*Figure 5: a) Comparison of the different probability distributions for an illustrative window of two-week time frame of distributed wind speed data. b) Comparison of the probability distributions for the means of two-week time frame for all the data. c) Scatter plot of all KL-divergences of Normal vs Weibull distribution for two-week time frame.*

In this subsection the comparison of the four distributions of the two-week time frame is showed. Figure 5a is an illustrative two-week time frame. This figure shows that Weibull, Nakagami and Rician nearly have the same fit. Only the Normal distribution fits differently. However, the wind speed distribution does not fit any of the probability distributions well. In this report, the focus is only about the four unimodal distributions. When looking at Figure 5a and other windows of this time frame (see Appendix C; Figures 15a-b), it seems that the data is multimodal distributed. When looking at Table 2, it can be concluded that for this time frame Weibull and Rician distribution have the best fit. Figure 5b shows the comparison of the four probability distributions for the means of two-week time frames for all the data. It shows Normal, Nakagami and Rician nearly have the same fit. Only Weibull distribution is a bit different distributed than the other distributions.

Table 2 shows that Weibull and Rician have the same and the best fit for a two-week time frame. The Normal distribution fits the data less good than the others. Figure 5c is the visual representation of Table 2. It shows the KL-divergences for Normal vs Weibull distribution for the windows of the two-week time frame. When looking at the graph, almost all values are above the red line y=x. This means that Weibull distribution has shorter distances than Normal distribution. Based on this can be concluded that Weibull distribution fits the wind speed data the best for the two-week time frame.

| Weibull | Rician | Normal | Nakagami |
|---|---|---|---|
| **KL-divergence values** | | | |
| 0.056 | 0.078 | 0.061 | 0.132 |
| 0.070 | 0.049 | 0.052 | 0.047 |
| 0.040 | 0.049 | 0.038 | 0.058 |
| 0.018 | 0.013 | 0.014 | 0.026 |
| 0.114 | 0.079 | 0.086 | 0.077 |
| 0.112 | 0.085 | 0.085 | 0.086 |
| 0.228 | 0.124 | 0.141 | 0.133 |
| 0.076 | 0.068 | 0.066 | 0.072 |
| 0.071 | 0.048 | 0.062 | 0.046 |
| 0.122 | 0.088 | 0.098 | 0.082 |
| 0.058 | 0.029 | 0.038 | 0.025 |
| 0.142 | 0.175 | 0.167 | 0.178 |
| 0.033 | 0.043 | 0.035 | 0.070 |
| 0.104 | 0.074 | 0.074 | 0.074 |
| 0.035 | 0.050 | 0.037 | 0.071 |
| … | … | … | … |
| **KL-divergence mean** | | | |
| **0.062** | **0.062** | **0.077** | **0.068** |

*Table 2: KL-distances of the four distributions for two-week time frame.*

## 4.3 One-week time frame

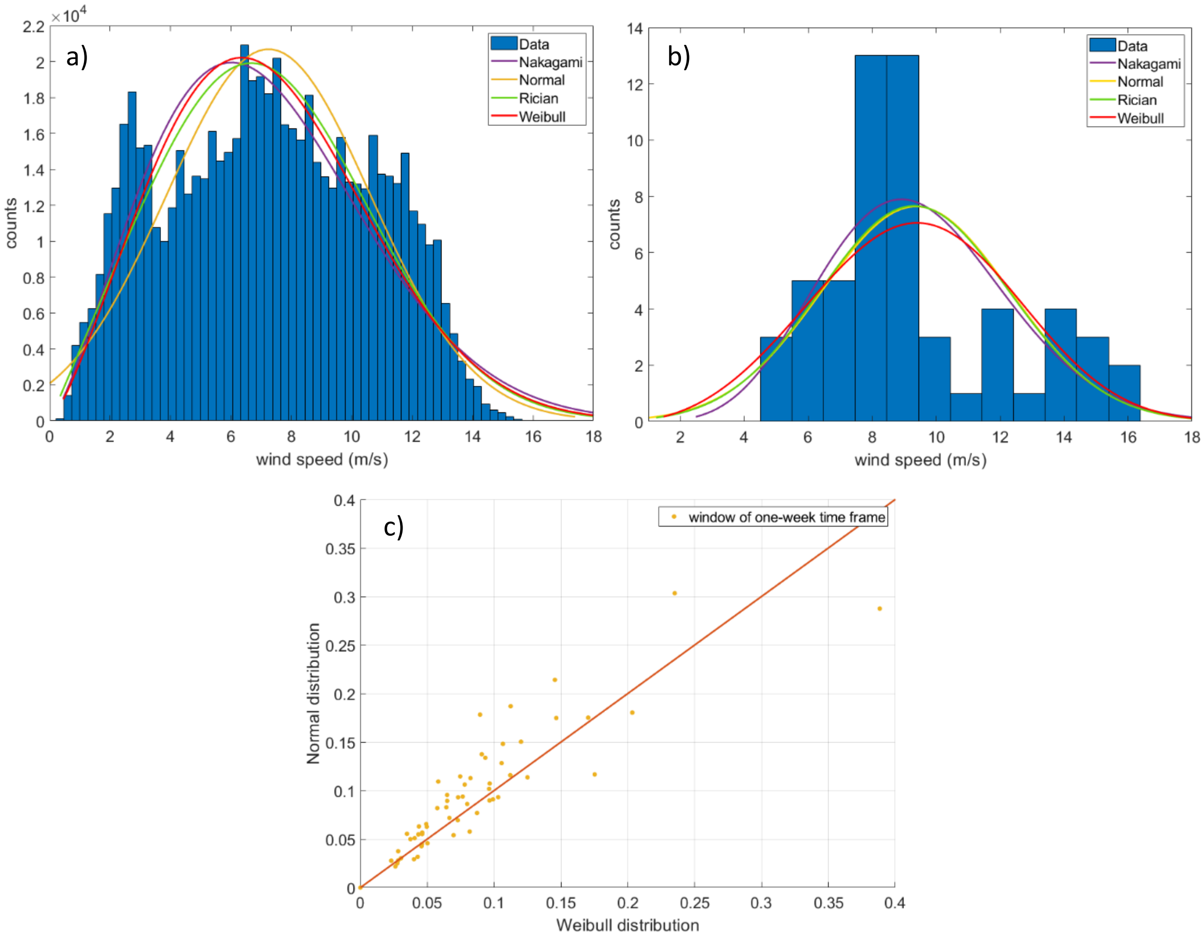


*Figure 6: a) Comparison of the different probability distributions for an illustrative window of one-week time frame of distributed wind speed data. b) Comparison of the probability distributions for the means of one-week time frame for all the data. c)* Scatter plot of all KL-divergences of Rician vs Weibull distribution for one-week time frame.

In this subsection the comparison of the four distributions of a one-week time frame is showed. Figure 6a is an illustrative one-week time frame. This figure shows that Weibull, Nakagami and Rician nearly have the same fit. The Normal distribution fits differently. However, just as the two-week time frame, it seems like this distribution is multimodal distributed. This can be concluded when looking at different windows of this time frame (see Appendix C; Figures 16a-b) and Figure 6a. This also means that the four probability distributions are not fitting the distributed data that well. When looking at Table 3, it can be concluded that for this time frame Weibull and Rician distribution have the best fit. The same conclusion is drawn for the one-month and two-week time frames. Figure 6b shows the comparison of the four probability distributions for the means of one-week time frames for all the data. It shows that Normal, Nakagami and Rician nearly have the same fit.

Only Weibull is a bit different distributed than the other distributions. Table 3 shows that Weibull and Rician have the best fit for a one-week time frame, while Normal fits the data less good. The same conclusion is drawn for the two-week time frame. Figure 6c is the visual representation of Table 3. It shows the KL-divergences for Normal vs Weibull distribution for the windows of one-week time frame. When looking at the graph, almost all values are above the red line y=x. This means that Weibull distribution has shorter distances than Normal distribution. Based on this can be concluded that the Weibull distribution fits the wind speed data the best for the one-week time frame.

| **Weibull** | **Rician** | **Normal** | **Nakagami** |
|---|---|---|---|
| **KL-divergence values** | | | |
| 0.214 | 0.145 | 0.143 | 0.145 |
| 0.106 | 0.078 | 0.095 | 0.071 |
| 0.090 | 0.097 | 0.085 | 0.103 |
| 0.175 | 0.170 | 0.170 | 0.190 |
| 0.093 | 0.103 | 0.092 | 0.118 |
| 0.148 | 0.107 | 0.121 | 0.095 |
| 0.054 | 0.070 | 0.057 | 0.091 |
| 0.055 | 0.046 | 0.048 | 0.046 |
| 0.055 | 0.043 | 0.040 | 0.047 |
| 0.083 | 0.064 | 0.062 | 0.065 |
| 0.072 | 0.067 | 0.070 | 0.079 |
| 0.102 | 0.096 | 0.093 | 0.100 |
| 0.028 | 0.023 | 0.027 | 0.036 |
| 0.037 | 0.028 | 0.031 | 0.038 |
| 0.063 | 0.044 | 0.046 | 0.047 |
| … | … | … | … |
| **KL-divergence mean** | | | |
| **0.085** | **0.087** | **0.097** | **0.089** |

*Table 3: KL-distances of the four distributions for one-week time frame.*

## 4.4 Two-day time frame

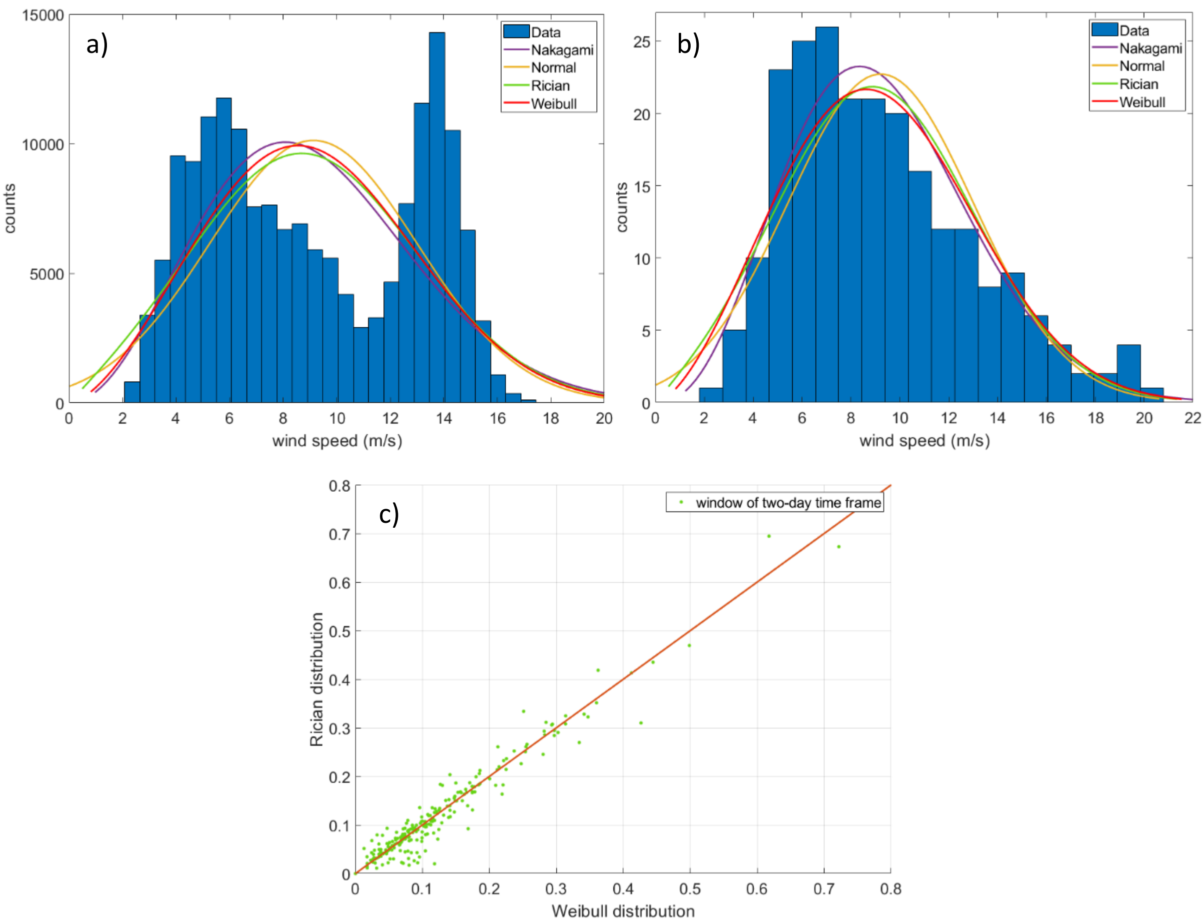


*Figure 7: a) Comparison of the different probability distributions for an illustrative window of two-day time frame of distributed wind speed data. b) Comparison of the probability distributions for the means of two-day time frame for all the data. c) Scatter plot of all KL-divergences of Rician vs Weibull distribution for two-day time frame.*

In this subsection the comparison of the four distributions of a two-day time frame is showed. Figure 7a is an illustrative two-day time frame. This figure shows that Weibull, Normal, Nakagami and Rician nearly have the same fit. However, when looking at other windows of this time frame (see Appendix C; Figures 17a-b) and Figure 7a, it looks bimodal distributed. The four unimodal distributions do not fit this data well. When looking at Table 4, it can be concluded that for this time frame Weibull and Rician distribution have the best fit. Figure 7b shows the comparison of the four probability distributions for the means of two-day time frames for all the data. It shows that the Weibull and Rician nearly have the same fit. Only Normal and Nakagami are a bit different distributed than the other distributions. Table 4 shows Weibull and Rician have the same and the best fit for a two-day time frame; while Normal fits the data less good.

Figure 7c is a visual representation of Table 4. It shows the KL-divergences for the Rician vs Weibull distribution for the windows of the two-day time frame. When looking at the graph, it cannot be seen if there are more values above or below the red line y=x. It looks like the same amount of values are above the line as below the line. As concluded out of Table 4, Weibull fits as well as Rician distribution the wind speed data for the two-day time frame.

| Weibull | Rician | Normal | Nakagami |
|---|---|---|---|
| **KL-divergence values** | | | |
| 0.080 | 0.017 | 0.018 | 0.012 |
| 0.069 | 0.023 | 0.023 | 0.021 |
| 0.142 | 0.117 | 0.107 | 0.179 |
| 0.085 | 0.093 | 0.107 | 0.087 |
| 0.073 | 0.071 | 0.095 | 0.073 |
| 0.132 | 0.120 | 0.125 | 0.140 |
| 0.153 | 0.167 | 0.168 | 0.174 |
| 0.722 | 0.673 | 0.652 | 0.750 |
| 0.186 | 0.185 | 0.173 | 0.340 |
| 0.149 | 0.149 | 0.155 | 0.123 |
| 0.115 | 0.114 | 0.121 | 0.126 |
| 0.139 | 0.115 | 0.117 | 0.093 |
| 0.314 | 0.325 | 0.346 | 0.311 |
| 0.060 | 0.073 | 0.072 | 0.082 |
| 0.091 | 0.086 | 0.088 | 0.075 |
| … | … | … | … |
| **KL-divergence mean** | | | |
| **0.129** | **0.129** | **0.133** | **0.131** |

*Table 4: KL-distances of the four distributions for two-day time frame.*

## 4.5 Twelve-hour time frame

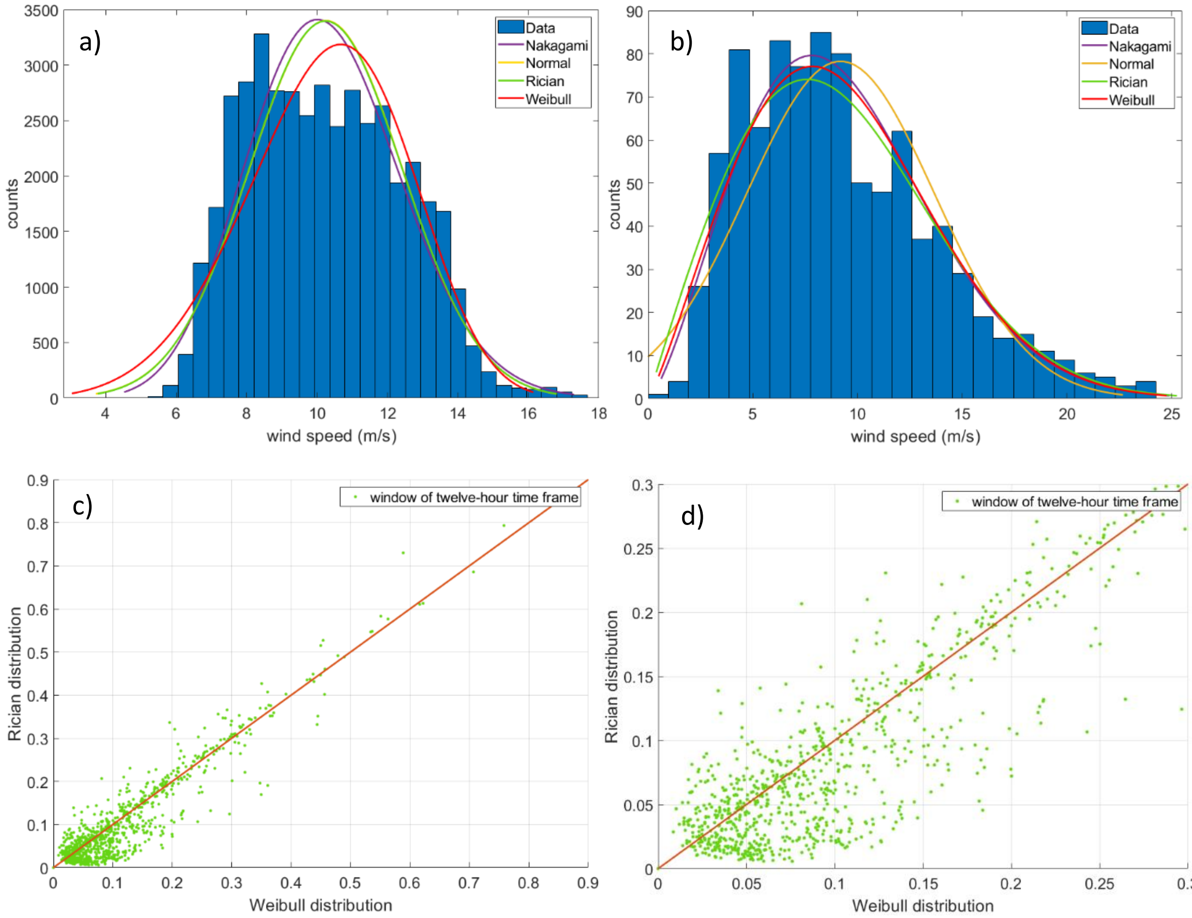


*Figure 8: a) Comparison of the different probability distributions for an illustrative window of twelve-hour time frame of distributed wind speed data. b) Comparison of the probability distributions for the means for all twelve-hour time frame. c) Scatter plot of all KL-divergences of Rician vs Weibull distribution for twelve-hour time frame, showing the outliers. d) A zoomed scatter plot of all KL-divergences of Rician vs Weibull distribution for twelve-hour time frame.*

In this subsection the comparison of the four distributions of a twelve-hour time frame is showed. Figure 8a is an illustrative twelve-hour time frame. This figure shows that Normal, Nakagami and Rician nearly have the same fit. Weibull distribution has a worse fit for this time frame. In contrast to the previously discussed time frames, the wind speed distribution of the twelve-hour time frame starts to fit the probability distributions better. This is because the distributed data seems to be unimodal. However, when looking at other windows of this time frame (see Appendix C; Figures 18a-b), it seems that the data can also be multimodal distributed. When looking at Table 5, it can be concluded that for this time frame Nakagami distribution has the best fit. Figure 8b shows the comparison of the four probability distributions for the means of twelve-hour time frames for all the data. It shows that Weibull, Rician and Nakagami nearly have the same fit. Only Normal is a bit different distributed than the other distributions. Table 5 shows that Nakagami has the best fit for a twelve-hour time frame; while Weibull fits the data the least good.

Figure 8c shows the visual representation of Table 5, which probably contains some outliers. Figure 8d shows a zoomed version of this scatter plot. When looking at Figure 8c, most values are below the red line y=x. This means that Rician distribution has shorter distances than Weibull distribution. Based on this can be concluded that Rician distribution has the best fit for the wind speed data for the twelve-hour time frame.

| Weibull | Rician | Normal | Nakagami |
|---|---|---|---|
| **KL-divergence values** | | | |
| 0.216 | 0.126 | 0.128 | 0.099 |
| 0.023 | 0.078 | 0.077 | 0.095 |
| 0.083 | 0.072 | 0.073 | 0.066 |
| 0.304 | 0.310 | 0.309 | 0.324 |
| 0.040 | 0.039 | 0.039 | 0.052 |
| 0.037 | 0.014 | 0.014 | 0.022 |
| 0.052 | 0.010 | 0.010 | 0.009 |
| 0.088 | 0.023 | 0.024 | 0.014 |
| 0.051 | 0.010 | 0.010 | 0.012 |
| 0.090 | 0.096 | 0.096 | 0.095 |
| 0.301 | 0.303 | 0.304 | 0.294 |
| 0.143 | 0.139 | 0.160 | 0.144 |
| 0.108 | 0.116 | 0.116 | 0.130 |
| 0.185 | 0.079 | 0.079 | 0.072 |
| 0.128 | 0.051 | 0.051 | 0.044 |
| … | … | … | … |
| **KL-divergence mean** | | | |
| **0.122** | **0.113** | **0.114** | **0.110** |

*Table 5: KL-distances of the four distributions for twelve-hour time frame.*

## 4.6 One-hour time frame

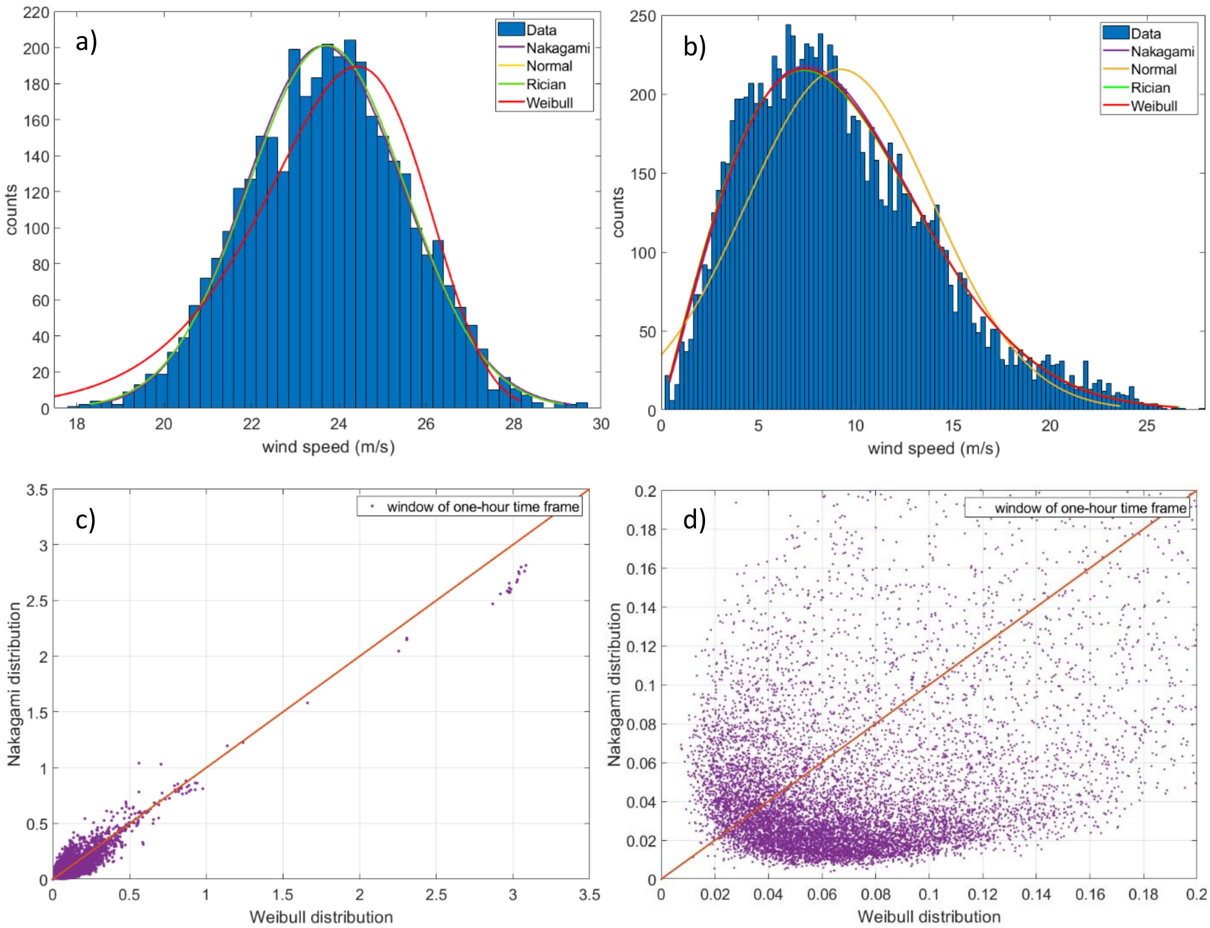


*Figure 9: a) Comparison of the different probability distributions for an illustrative window of one-hour time frame. b) Comparison of the probability distributions for the means for all one-hour time frame. c) Scatter plot of all KL-divergences of Nakagami vs Weibull distribution for one-hour time frame, including the outliers. d) A zoomed scatter plot of all KL-divergences of Nakagami vs Weibull distribution for one-hour time frame.*

In this subsection the comparison of the four distributions of a one-hour time frame is showed. Figure 9a is an illustrative one-hour time frame. This figure shows that Normal, Nakagami, and Rician nearly have the same fit and Weibull fits completely different. When looking at other windows of this time frame (see Appendix C; Figures 19a-b), it seems that the data is multimodal distributed. However, because there are approximately 11.000 windows, one cannot draw the conclusion that the data is multimodal distributed while only looking at 3 windows. Figure 9b shows the comparison of the four probability distributions for the means of one-hour for all the data. The figure shows that the Weibull, Rician and Nakagami nearly have the same fit. Only the Normal is a bit different distributed to the other distributions. Table 6 shows that the Normal and Rician distribution have the same best fit for one-hour time frame, because they have a lower divergence. Weibull fits the data less good, because it has a higher divergence.

Figure 9c shows the visual representation of Table 6, which probably contains some outliers. Figure 9d shows a zoomed version of this scatter plot. When looking at Figure 9c, can be seen that most values are below the red line y=x. This means that Nakagami distribution has shorter distances than Weibull distribution. Based on this can be concluded that Nakagami distribution fits the wind speed data better for the one-hour time frame.

| Weibull | Rician | Normal | Nakagami |
|---|---|---|---|
| **KL-divergence values** | | | |
| 0.066 | 0.021 | 0.021 | 0.022 |
| 0.040 | 0.013 | 0.013 | 0.015 |
| 0.036 | 0.018 | 0.018 | 0.023 |
| 0.071 | 0.011 | 0.011 | 0.012 |
| 0.087 | 0.019 | 0.019 | 0.018 |
| 0.075 | 0.141 | 0.140 | 0.159 |
| 0.082 | 0.018 | 0.018 | 0.017 |
| 0.095 | 0.027 | 0.027 | 0.023 |
| 0.053 | 0.040 | 0.040 | 0.044 |
| 0.050 | 0.042 | 0.042 | 0.045 |
| 0.119 | 0.037 | 0.037 | 0.029 |
| 0.093 | 0.172 | 0.172 | 0.190 |
| 0.048 | 0.016 | 0.016 | 0.016 |
| 0.071 | 0.015 | 0.015 | 0.013 |
| 0.021 | 0.036 | 0.036 | 0.046 |
| … | … | … | … |
| **KL-divergence mean** | | | |
| **0.091** | **0.066** | **0.066** | **0.067** |

*Table 6: KL-distances of the four distributions for one-hour time frame.*

## 4.7 Ten-minute time frame

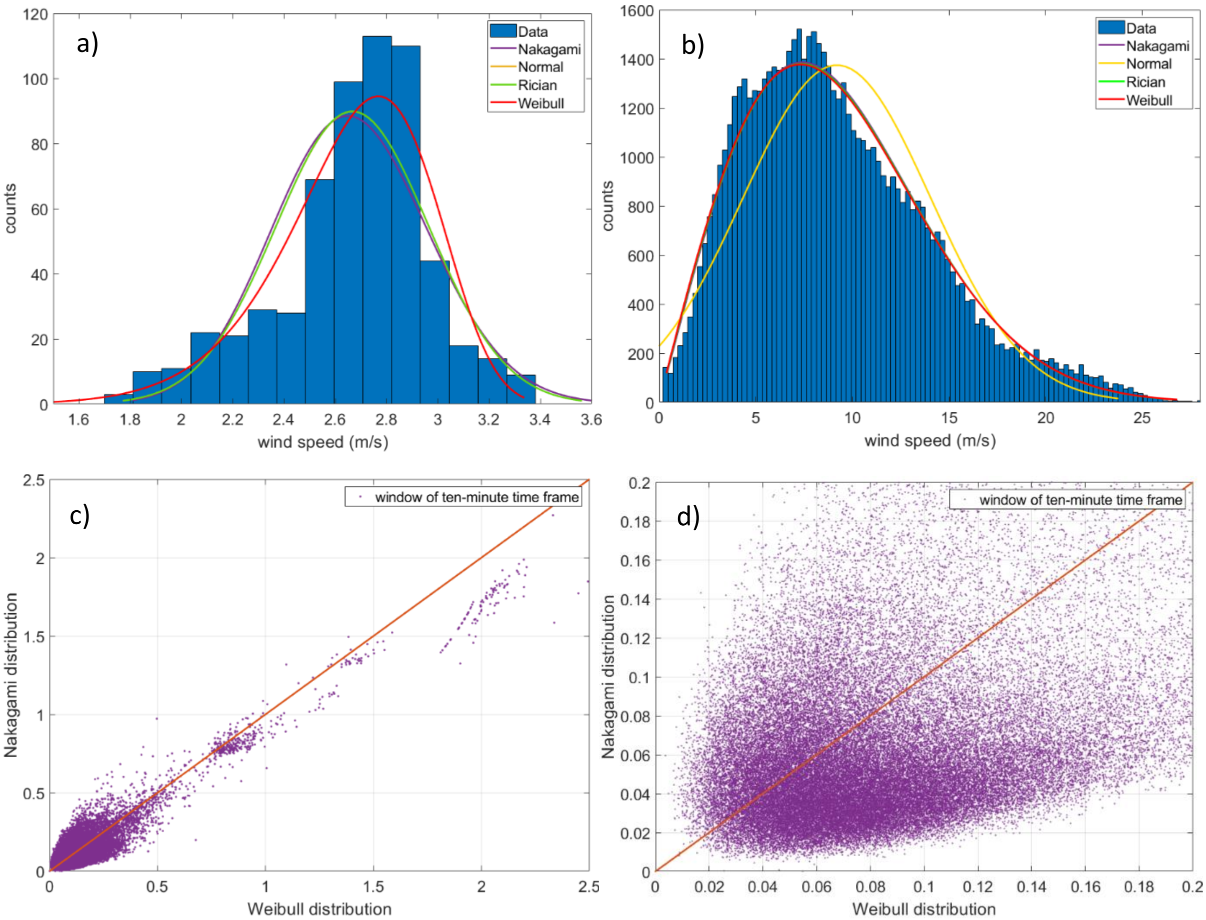


*Figure 10: a) Comparison of the different probability distributions for an illustrative window of ten-minute time frame. b) Comparison of the probability distributions for the means of windows for all ten-minute time frame. c) Scatter plot of all KL-divergences of Nakagami vs Weibull distribution for ten-minute time frame, including the outliers. d) A zoomed scatter plot of all KL-divergences of Nakagami vs Weibull distribution for ten-minute time frame.*

In this subsection the comparison of the four distributions of a ten-minute time frame is showed. Figure 10a is an illustrative ten-minute time frame. This figure shows that Normal, Nakagami, and Rician nearly have the same fit. Only Weibull distribution fits completely different. When looking at other windows of this time frame (see Appendix C; Figures 20a-b), it seems that the data is unimodal distributed. However, because there are approximately 60.000 windows, the conclusion that it is unimodel distributed is not representative while only looking at three windows. Figure 10b shows the comparison of the four probability distributions for the means of ten-minute time frames for all the data. This figure shows that Weibull, Rician and Nakagami nearly have the same fit. Only Normal is different distributed than the other distributions. When looking at Table 7, it can be concluded that Normal and Rician have the same and the best fit for a ten-minute time frame; while Weibull fits the data the least good.

Figure 10c is a visual representation of Table 7, which probably contains some outliers. Figure 10d shows a zoomed version of this scatter plot. When looking at Figure 10c, values are below the red line y=x. This means that Nakagami distribution has shorter distances than Weibull distribution. As concluded out of Table 7, Nakagami distribution fits the wind speed data better for the ten-minute time frame.

| Weibull | Rician | Normal | Nakagami |
|---|---|---|---|
| **KL-divergence values** | | | |
| 0.080 | 0.051 | 0.051 | 0.050 |
| 0.182 | 0.070 | 0.070 | 0.063 |
| 0.103 | 0.028 | 0.028 | 0.024 |
| 0.069 | 0.062 | 0.062 | 0.064 |
| 0.049 | 0.072 | 0.072 | 0.077 |
| 0.099 | 0.033 | 0.033 | 0.033 |
| 0.079 | 0.095 | 0.095 | 0.105 |
| 0.092 | 0.047 | 0.047 | 0.049 |
| 0.081 | 0.152 | 0.152 | 0.164 |
| 0.071 | 0.035 | 0.035 | 0.034 |
| 0.075 | 0.056 | 0.056 | 0.058 |
| 0.168 | 0.130 | 0.131 | 0.112 |
| 0.062 | 0.053 | 0.053 | 0.053 |
| 0.257 | 0.233 | 0.234 | 0.215 |
| 0.133 | 0.113 | 0.113 | 0.112 |
| … | … | … | … |
| **KL-divergence mean** | | | |
| **0.099** | **0.075** | **0.075** | **0.076** |

*Table 7: KL-distances of the four distributions for ten-minute time frame.*

## 4.8 Comparison between the time frames

This last subsection is an overview of the previous subsections. This subsection is about comparing the four distributions with each other for every time frame.

| | KL-divergence mean | | | | KL-divergence standard deviation | | | |
|---|---|---|---|---|---|---|---|---|
| | Weibull | Rician | Normal | Nakagami | Weibull | Rician | Normal | Nakagami |
| **One month** | 0.043 | 0.043 | 0.061 | 0.046 | 0.025 | 0.024 | 0.029 | 0.093 |
| **Two weeks** | 0.062 | 0.062 | 0.077 | 0.068 | 0.039 | 0.039 | 0.047 | 0.041 |
| **One week** | 0.085 | 0.087 | 0.097 | 0.089 | 0.062 | 0.069 | 0.062 | 0.060 |
| **Two days** | 0.129 | 0.129 | 0.133 | 0.131 | 0.107 | 0.127 | 0.115 | 0.110 |
| **Twelve hours** | 0.122 | 0.113 | 0.114 | 0.110 | 0.161 | 0.216 | 0.182 | 0.161 |
| **One Hour** | 0.091 | 0.066 | 0.066 | 0.067 | 0.143 | 0.135 | 0.136 | 0.135 |
| **Ten minutes** | 0.099 | 0.075 | 0.075 | 0.076 | 0.116 | 0.107 | 0.107 | 0.107 |
| | **Mean of KL-divergence means** | | | | **Mean of KL-divergence standard deviations** | | | |
| | 0.090 | 0.082 | 0.089 | 0.084 | 0.093 | 0.102 | 0.097 | 0.101 |

*Table 8: KL-divergence means and standard deviations of the four distributions for all time frames.*

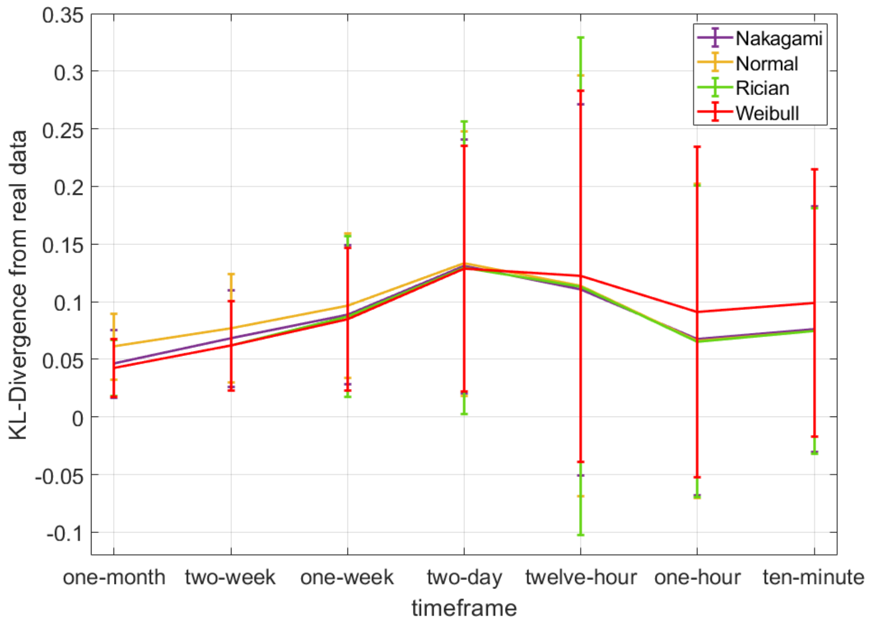


*Figure 11: Diagram of KL-divergence means with ± one standard deviation of the four distributions for all time frames.*

Table 8 is divided in four sections:

- **KL-divergence mean** section has the purpose to collect the means of each time frame without having to look at separate time frames.
- The **Mean of KL-divergence means** section has the purpose to provide which distribution has the lowest value, which means the best average fit for the data.
- **KL-divergence standard deviation** section has the purpose to highlight the standard deviations of each time frame. It provides information about which distribution has the highest deviation for each time frame.

- Finally, the **Mean of KL-divergence standard deviations** section has the purpose to provide which distribution has the lowest average deviation.

Figure 11 shows the means of the KL-divergences for all four distributions for every time frame including the standard deviation for all the time frames. From one month until around one day, Weibull and Rician distributions have the smallest mean of the KL-divergence. This means they have a better fit than the other two distributions. Around one day, there is a changing point where Weibull distribution becomes worse. From one day until ten-minute time frame, Normal, Rician, and Nakagami have the smallest mean (see Table 8). This means, for this period of time, the three distributions have a better fit than Weibull. However, all the error bars are large enough to overlap with each other. This means that there is no difference between the four distributions; they are either all considered to be equally good or equally bad for the investigated time frames. This is because the probability distribution can have a smaller mean, but also a larger deviation. Therefore, it can be concluded that all distributions are numerically acceptable. The large deviations can partly be explained by the fact that some time frames are multimodal for some windows (two weeks, one week, two days, twelve hours). These are models that are showing the largest deviations.

# 5. Conclusion and discussion

Based on the given data, there are some features that must be considered when reading the results:

- The results were registered from data produced by an anemometer located at FINO1. The measured data presents some NaN values, which can alter the feasibility of the data.
- As mentioned, the analysed data consists of 15 months of wind speed measurements; which are 40 million 1 Hz data values. This amount of values is considered as enough for reliable analysis of wind speed data; especially when analysing short time frames.
- The analysis of the wind speed data could be improved by using kernel-based estimators.
- The binning of the PDF is not perfect; there are some bins without representing values. This could also cause the error bars to be this large. In this report, there has been made use of discrete PDF's to fit the wind speed data. However, the binning is not representing all the data accurately/correctly, since it turned out some bins contained no values. That is why it can be considered to make use of continuous PDF's to fit the wind speed data.
- Maximum likelihood might not be the best approach to estimate the parameters of the probability distributions. Other studies can be done finding the best mathematical approach to estimate these parameters.

Based on the Table 8, the following conclusions can be drawn:

- From the Mean of KL-divergence means section of the table, it is extracted that Weibull distribution (the reference distribution) has the greatest mean considering all the time frames. This means that Weibull should be replaced by Rician distribution, which has the lowest mean considering all the time frames.
- As said in the introduction, when creating wind power from wind speed one must think that the high fluctuation of wind causes big fluctuations on the output power generated by wind turbines. Thus, finding which distribution has the lowest standard deviation will cause more reliable wind speed distribution predictions.
- The Mean of KL-divergence standard deviations section of Table 8 of the Results section shows that Weibull distribution has the lowest standard deviation considering all the time frames. Thus, it is concluded that Weibull is the best distribution of all four in terms of stability. However, when considering both mean and standard deviation together, one cannot conclude that Weibull distribution must be replaced by one of the other studied distributions.

Based on the Figure 11, the following conclusions can be drawn:

- For the one-month, two-week, one-week, and two-day time frames Weibull and Rician have the same and the best fit. For the twelve-hour time frame the best distribution is Nakagami distribution. Finally, for the one-hour and ten-minute time frames Normal and Rician distribution have the best fit.
- All standard deviation of the distributions, which are represented by the error bars of Figure 11, are large enough to overlap each other. This means that all distributions are numerically the same for all the time frames. As all the standard deviations of the KL-Divergences overlap each other, all the four distributions are acceptable to fit all time frames.
- Weibull does not exhibit good performance in short time frames, while Normal does not exhibit good performance in large time frames. Nakagami also loses performance for large time frames, but not as much as Normal. Rician manifests bad performance in time frames that present a large standard deviation.

For companies involved in wind energy, the recommendations are the following:

- On one hand, the recommendation for a new industry that has no previous experience using Weibull distribution and needs to make measurements for a certain time frame can look for that specific time frame and see which distribution fits better. This is because better performance might be achieved with other distributions for certain time frames.
- On the other hand, for companies that already use Weibull distribution as the standard one for modeling wind speed, the cost of changing from this distribution to another one might not deliver profitable changes, time, human resources and risk. Thus, the recommendation is not to change from distribution to fit wind speed data.

For further studies in wind speed analysis, the recommendations are the following:

- All probability distributions studied as candidates to replace Weibull distribution are unimodal distributions. However, as almost every time frame presents multimodality, multimodal models could be better to analyse the wind speed data.
- Unimodal distributions seem similar enough to not to have a significant gain to replace Weibull distribution when its already implemented.

# 6. Learning process

We can look back on a successful period. During this European Project Semester we learned a lot about ourselves, different European cultures and specially the Norwegian one. In the beginning of this semester we had problems with working together. The fact that we are five people from four different countries and five different educational backgrounds made it difficult to understand what the strengths and weaknesses of each group member were. It also made it difficult to know what way every group member was used to work, because in every country, at every university, students work in a different way. After the first mentor's meeting we decided that we needed an extra meeting with Petter to discuss the issues in our group. During this meeting we found out that some group members had the feeling that they were doing all the work, while others had the feeling that they were useless. Together with Petter we found out that there were also a lot of secondary tasks that needed to be done in order to deliver a successful project at the end. From this moment we started working together better as a team and there was more respect to each other.

In this paragraph we will reflect on the EPS-program. We like the idea of having a main project and supporting courses. However, we think that the supporting courses were not as supportive as they could be, especially at the beginning of the semester. On Mondays with the course "Working in Projects" we expected to learn more about project management or working in teams. It would have been great if in the beginning of the class, maybe only thirty minutes, Sinan would give a small lecture about one of these kinds of subjects. For the other supporting course "English and Academic Writing" some group members had the feeling that the information we got during this course was information they already knew. These gave us the feeling that we could have invest our time just working on the project without going to class. Gratefully, last English classes were very useful and instructive.

Another part of the program were the workshops. We think that it was nice to have some extra workshops and to learn more about specific subjects. The first workshop, the teambuilding workshop, was a fun experience, and we learned even more about ourselves when we became a team. The second workshop, the sustainability workshop, took too long and the things we learned, we could have learned by reading some articles. But on the other hand, the fact that we had this workshop provided us information about the UNDP goals in our report. So, it had its pros and cons. The last workshop was the cross-cultural collaboration workshop, where we became aware about the social problems that we face when meeting other nationalities.

When we saw this workshop in our schedule, we were not looking forward to it. This is because we, as Engineering students, don’t really like to act and express ourselves loudly. But it turned out this workshop was real fun. We enjoyed it way more than expected, even though we felt that it was not useful for our project (for the report). It was useful when interacting between us, know other EPS student more deeply and understand other cultures better.

In conclusion, the main thing we learned is that it is not easy to work together with people from different backgrounds. However, this challenge makes it even more fun to do. Besides that, all of us has become more independent during our stay abroad here in Oslo. Everyone got to know more about him or herself and this is an important thing for the future in our lives.

# Bibliography


Abdul Baseer, M., & Meyer, J. (2016). Wind power characteristics of seven data collection sites in Jubail, Saudi Arabia using Weibull parameters. *Renewable Energy*, *102*, 35–49. https://doi.org/10.1016/j.renene.2016.10.040

Aja-Fernández, S., Vegas-Sánchez-Ferrero, G., Aja-Fernández, S., & Vegas-Sánchez-Ferrero, G. (2016). The Problem of Noise in MRI. In *Statistical Analysis of Noise in MRI* (pp. 1–6). https://doi.org/10.1007/978-3-319-39934-8_1

Ali, S., Lee, S. M., & Jang, C. M. (2018). Statistical analysis of wind characteristics using Weibull and Rayleigh distributions in Deokjeok-do Island – Incheon, South Korea. *Renewable Energy*, *123*, 652–663. https://doi.org/10.1016/j.renene.2018.02.087

Calin, O., & Constantin, U. (2014). *Geometric modeling in probability and statistics* (1st ed.). https://doi.org/10.1007/978-3-319-07779-6

Carta, J. A., Ramírez, P., & Velázquez, S. (2009). A review of wind speed probability distributions used in wind energy analysis. Case studies in the Canary Islands. *Renewable and Sustainable Energy Reviews*, *13*(5), 933–955. https://doi.org/10.1016/j.rser.2008.05.005

Hennessey, J. P. (1977). Some Aspects of Wind Power Statistics. *Journal of Applied Meteorology*, *16*(2), 119–128. https://doi.org/10.1175/1520-0450(1977)016<0119:SAOWPS>2.0.CO;2

Hershey, J. R., & Olsen, P. A. (2007). Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models. *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, *4*. https://doi.org/10.1109/ICASSP.2007.366913

Huang, L. (2016). The Nakagami and its related distributions. *WSEAS TRANSACTIONS on MATHEMATICS*, *15*, 477–485.

Jiang, H., Wang, J., Wu, J., & Geng, W. (2017). Comparison of numerical methods and metaheuristic optimization algorithms for estimating parameters for wind energy potential assessment in low wind regions. *Renewable and Sustainable Energy Reviews*, *69*, 1199–1217. https://doi.org/https://doi.org/10.1016/j.rser.2016.11.241

Justus, C. G., Hargraves, W. R., & Yalcin, A. (1976). Nationwide Assessment of Potential Output from Wind-Powered Generators. *Journal of Applied Meteorology*, *15*(7), 673–678. https://doi.org/10.1175/1520-0450(1976)015<0673:NAOPOF>2.0.CO;2

Kalmikov, A. (2017). Wind Power Fundamentals. In T. M. B. T.-W. E. E. Letcher (Ed.), *Wind Energy Engineering* (1st ed., pp. 17–24). https://doi.org/10.1016/B978-0-12-809451-8.00002-3

Le, K. N. (2019). A review of selection combining receivers over correlated Rician fading. *Digital Signal Processing*, *88*, 1–22. https://doi.org/https://doi.org/10.1016/j.dsp.2019.01.015

Lira de Queiroz, W. J., Brito Teixeira de Almeida, D., Madeiro, F., & Araújo Lopes, W. T. (2019). Signal-to-noise ratio estimation for M-QAM signals in η−µand κ−µfading channels. *EURASIP Journal on Advances in Signal Processing*, *2019*(1), 20. https://doi.org/10.1186/s13634-019-0607-7

Nfaoui, H., Buret, J., & Sayigh, A. A. . (1998). Wind characteristics and wind energy potential in Morocco. *Solar Energy*, *63*(1), 51–60. https://doi.org/10.1016/S0038-092X(98)00040-1

Odeyemi, K. O., & Owolawi, P. A. (2019). Selection combining hybrid FSO/RF systems over generalized induced-fading channels. *Optics Communications*, *433*, 159–167. https://doi.org/10.1016/j.optcom.2018.10.009

Ouarda, T. B. M. J., Charron, C., Shin, J.-Y., Marpu, P. R., Al-Mandoos, A. H., Al-Tamimi, M. H., … Al Hosary, T. N. (2015). Probability distributions of wind speed in the UAE. *Energy Conversion and Management*, *93*, 414–434. https://doi.org/10.1016/j.enconman.2015.01.036

Ozonur, D., Akdur, H. T. K., & Bayrak, H. (2018). Optimal Asymptotic Tests for Nakagami Distribution. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *22*(Special), 487. https://doi.org/10.19113/sdufbed.32458

Pilling, M. (2017). Handbook of Statistical Distributions with Applications. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *180*(3), 941–941. https://doi.org/10.1111/rssa.12292

Rice, S. O. (1945). Mathematical Analysis of Random Noise. *Bell System Technical Journal*, *24*(1), 46–156. https://doi.org/10.1002/j.1538-7305.1945.tb00453.x

Rician Distribution - MATLAB & Simulink - MathWorks Benelux. (n.d.). Retrieved November 25, 2019, from https://nl.mathworks.com/help/stats/rician-distribution.html

Roy-Cardinal, M., Destrempes, F., Soulez, G., & Cloutier, G. (2019). Assessment of Carotid Artery Plaque Components With Machine Learning Classification Using Homodyned-K Parametric Maps and Elastograms. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, *66*(3), 493–504. https://doi.org/10.1109/TUFFC.2018.2851846

Şahin, A. Z., & Aksakal, A. (1998). Wind power energy potential at the northeastern region of Saudi Arabia. *Renewable Energy*, *14*(1–4), 435–440. https://doi.org/10.1016/S0960-1481(98)00100-1

Schwan, G. (2019). Sustainable development goals. *GAIA*, *28*(2), 73. https://doi.org/10.14512/gaia.28.2.1

Spatz, C. (2007). *Basic statistics Tales of Distributions* (9th ed.). Wadsworth Publishing.

Stewart, D. A., & Essenwanger, O. M. (1978). Frequency Distribution of Wind Speed Near the Surface. *Journal of Applied Meteorology*, *17*(11), 1633–1642. https://doi.org/10.1175/1520-0450(1978)017<1633:FDOWSN>2.0.CO;2

Tuller, S. E., & Brett, A. C. (1985). The goodness of fit of the weibull and rayleigh distributions to the distributions of observed wind speeds in a topographically diverse area. *Journal of Climatology*, *5*(1), 79–94. https://doi.org/10.1002/joc.3370050107

Yu, J., Fu, Y., Yu, Y., Wu, S., Wu, Y., You, M., … Li, M. (2019). Assessment of Offshore Wind Characteristics and Wind Energy Potential in Bohai Bay, China. *Energies*, *12*(15), 2879. https://doi.org/10.3390/en12152879

# Appendix A: Matlab code

This appendix describes the Matlab code used during this project to generate all the graphs and tables showed in the report.

### Splitting the raw data into time frames

```
field1 = 'a102015';  value1 = wcup1;
field2 = 'a112015';  value2 = wcup2;
field3 = 'a122015';  value3 = wcup3;
field4 = 'a012016';  value4 = wcup4;
field5 = 'a022016';  value5 = wcup5;
field6 = 'a032016';  value6 = wcup6;
field7 = 'a042016';  value7 = wcup7;
field8 = 'a052016';  value8 = wcup8;
field9 = 'a062016';  value9 = wcup9;
field10 = 'a072016';  value10 = wcup10;
field11 = 'a082016';  value11 = wcup11;
field12 = 'a092016';  value12 = wcup12;
field13 = 'a102016';  value13 = wcup13;
field14 = 'a112016';  value14 = wcup14;
field15 = 'a122016';  value15 = wcup15;
Wcup =
struct(field1,value1,field2,value2,field3,value3,field4,value4,field5,value5,field6,value6,field7,value7,f
ield8,value8,field9,value9,field10,value10,field11,value11,field12,value12,field13,value13,field14,valu
e14,field15,value15)

Data=[wcup{1,1} wcup{1,2} wcup{1,3} wcup{1,4} wcup{1,5} wcup{1,6} wcup{1,7} wcup{1,8} wcup{1,9}
wcup{1,10} wcup{1,11} wcup{1,12} wcup{1,13} wcup{1,14} wcup{1,15}];

Data1=Data;
NaNValues=[];
for i=1:39571200
   Data1(i)
   if Data1(i)<0.27
     Data1(i)=NaN
     NaNValues=[NaNValues i];
   end
end

TenMin=mat2cell(Data1',diff([0:600:39571200]));
OneHour=mat2cell(Data1',diff([0:3600:39571200]));
TwelveHour=mat2cell(Data1',diff([0:43200:39571200]));
OneDay=mat2cell(Data1',diff([0:86400:39571200]));
TwoDay=mat2cell(Data1',diff([0:172800:39571200]));
ThreeDay=mat2cell(Data1',diff([0:263808:39571200]));
OneWeek=mat2cell(Data1',diff([0:659520:39571200]));
TwoWeek=mat2cell(Data1',diff([0:1319040:39571200]));
OneMonth=mat2cell(Data1',diff([0:2638080:39571200]));
```

**Computing the distribution parameters for one time frame using the raw splitted data**

```
windows =10992 ;
phat=cell(1,windows);
k=1;
falseparam=0;
 for k=1:windows
   try
      warning('') % Clear last warning message
     phat{1,k} = mle(OneHour{k,1},'distribution','Nakagami');

aNakagami(1,k) = phat{1,k}(1,1);
bNakagami(1,k) = phat{1,k}(1,2);

   phat2{1,k} = mle(OneHour{k,1},'distribution','Weibull');

aWeibull(1,k) = phat2{1,k}(1,1);
bWeibull(1,k) = phat2{1,k}(1,2);

phat3{1,k} = mle(OneHour{k,1},'distribution','Rician');

sRician(1,k) = phat3{1,k}(1,1);
bRician(1,k) = phat3{1,k}(1,2);

 phat4{1,k} = mle(OneHour{k,1},'distribution','Normal');

muNormal(1,k) = phat4{1,k}(1,1);
sigmaNormal(1,k) = phat4{1,k}(1,2);

   [warnMsg, warnId] = lastwarn;
      if ~isempty(warnMsg)
      k
falseparam=falseparam+1;
aNakagami(1,k) = NaN;
bNakagami(1,k) = NaN;
aWeibull(1,k) = NaN;
bWeibull(1,k) = NaN;
sRician(1,k) = NaN;
bRician(1,k) = NaN;
muNormal(1,k) = NaN;
sigmaNormal(1,k) = NaN;
      end
   catch
    k
falseparam=falseparam+1;
aNakagami(1,k) = NaN;
bNakagami(1,k) = NaN;
aWeibull(1,k) = NaN;
bWeibull(1,k) = NaN;
sRician(1,k) = NaN;
bRician(1,k) = NaN;
muNormal(1,k) = NaN;
sigmaNormal(1,k) = NaN;
      end
  end
```

**Computing the KL average and standard deviation**

```
Nbin=0;
windows = length(muNormal);
k=0;
 NaNWindow=[];

   for k=1:windows
  TimeWindow = TwoWeek{k,1};
      Pdata{1,k} = histcounts(TimeWindow,'Normalization','probability');
      NewPdata{1,k}=Pdata{1,k}/sum(Pdata{1,k});
      %k
      %sum(NewPdata{1,k})
      N = histcounts(TimeWindow,'Normalization','probability');
      Nbin=length(N);
     %length(N);
           [n,xout{1,k}] = hist(TimeWindow,Nbin);
   end

for k=1:windows
     Qtemp{1,k} = pdf('Normal',xout{1,k},muNormal(1,k),sigmaNormal(1,k));
     QpdfNormal{1,k} = Qtemp{1,k}/sum(Qtemp{1,k});
end
for k=1:windows
     Qtemp{1,k} = pdf('Weibull',xout{1,k},aWeibull(1,k),bWeibull(1,k));
     QpdfWeibull{1,k} = Qtemp{1,k}/sum(Qtemp{1,k});
end

 for k=1:windows
     Qtemp{1,k} = pdf('Nakagami',xout{1,k},aNakagami(1,k),bNakagami(1,k));
     QpdfNakagami{1,k} = Qtemp{1,k}/sum(Qtemp{1,k});
 end

 for k=1:windows
     Qtemp{1,k} = pdf('Rician',xout{1,k},sRician(1,k),bRician(1,k));
     QpdfRician{1,k} = Qtemp{1,k}/sum(Qtemp{1,k});
 end

AverageDistNormal=0;
AverageDistWeibull=0;
AverageDistRician=0;
AverageDistNakagami=0;
erreurKBL=0;

for i=1:windows
            P=NewPdata{1,i};
            Q=QpdfNormal{1,i};
            KLDNormal(1,i) = nansum( P .* log2( P./Q ) );
            if KLDNormal(1,i) <0 | KLDNormal(1,i) == NaN
                        erreurKBL=erreurKBL+1;
                        i
                        KLDNormal(1,i)=0;
            end
end
```

```
for i=1:windows

 P=NewPdata{1,i};
 Q=QpdfNakagami{1,i};
KLDNakagami(1,i) = nansum( P .* log2( P./Q ) );
if KLDNakagami(1,i) <0 | KLDNakagami(1,i) == NaN
   KLDNakagami(1,i)=0;
end

end

for i=1:windows

 P=NewPdata{1,i};
 Q=QpdfWeibull{1,i};
KLDWeibull(1,i) = nansum( P .* log2( P./Q ) );
if KLDWeibull(1,i) <0 | KLDWeibull(1,i) == NaN
   KLDWeibull(1,i)=0;
end
end

for i=1:windows

 P=NewPdata{1,i};
 Q=QpdfRician{1,i};
KLDRician(1,i) = nansum( P .* log2( P./Q ) );

if KLDRician(1,i) <0 | KLDRician(1,i) == NaN
   KLDRician(1,i)=0;
end
end

NaNNormal=[];
NaNWeibull=[];
NaNRician=[];
NaNNakagami=[];

for i=1:windows
if KLDNormal(1,i)== 0
   NaNNormal= [NaNNormal i];
   KLDNakagami(1,i)= 0;
   KLDWeibull(1,i)= 0;
   KLDRician(1,i)= 0;
   KLDNormal(1,i)= 0;
end
if KLDWeibull(1,i)<= 0
   NaNWeibull= [NaNWeibull i];
   KLDNakagami(1,i)= 0;
   KLDWeibull(1,i)= 0;
   KLDRician(1,i)= 0;
   KLDNormal(1,i)= 0;
end
if KLDRician(1,i)== 0
   NaNRician= [NaNRician i];
   KLDNakagami(1,i)= 0;
```

```
    KLDWeibull(1,i)= 0;
    KLDRician(1,i)= 0;
    KLDNormal(1,i)= 0;
end
if KLDNakagami(1,i)== 0
    NaNNakagami= [NaNNakagami i];
    KLDNakagami(1,i)= 0;
    KLDWeibull(1,i)= 0;
    KLDRician(1,i)= 0;
    KLDNormal(1,i)= 0;
end
AverageDistNormal = AverageDistNormal+KLDNormal(1,i);
AverageDistNakagami = AverageDistNakagami+KLDNakagami(1,i);
AverageDistWeibull = AverageDistWeibull+KLDWeibull(1,i);
AverageDistRician = AverageDistRician+KLDRician(1,i);
end
AverageDistNormal=AverageDistNormal/(windows-falseparam);
AverageDistNakagami = AverageDistNakagami/(windows-falseparam);
AverageDistWeibull = AverageDistWeibull/(windows-falseparam);
AverageDistRician = AverageDistRician/(windows-falseparam);

 SigmaDistNormal = std(KLDNormal);
SigmaDistWeibull = std(KLDWeibull);
SigmaDistNakagami = std(KLDNakagami);
SigmaDistRician = std(KLDRician);
```

**Obtaining the four mathematical moments (mean, variance, skewness and kurtosis) for ten-minute time frames**

```
for i=1:15
    for j=1:L(1,i)
        temp1(j,i) = kurtosis(wcup600{1,i}{1,j});
        temp2= nonzeros(temp1);
        krtosis10=temp2';
    end
end
krtosis15=mean(krtosis10,'omitnan')

for i=1:15
    for j=1:L(1,i)
        temp1(j,i) = skewness(wcup600{1,i}{1,j});
        temp2= nonzeros(temp1);
        skewnss10=temp2';
    end
end
skewnss15=mean(skewnss10,'omitnan')

for i=1:15
    for j=1:L(1,i)
        temp1(j,i) = var(wcup600{1,i}{1,j});
        temp2= nonzeros(temp1);
        variance10=temp2';
    end
end
variance15=mean(variance10,'omitnan')
```

```
for i=1:15
   for j=1:L(1,i)
      temp1(j,i) = mean(wcup600{1,i}{1,j});
      temp2= nonzeros(temp1);
      mu10=temp2';
   end
end
mu15=mean(mu10,'omitnan')
plot([1:65952],mu10)
hold on
plot([1:65952],variance10)
plot([1:65952],skewnss10)
plot([1:65952],krtosis10)
legend('Mean','Variance','Skewness','Kurtosis')
title('variation of the moments for 15 months')
ylabel('real value moments')
xlabel('time(10min)')
xlim([0 65952])
ylim([-3 32])
hold off
```

**Explanation of the KL-divergence using one window of a time frame**

```
x=[0,2];
y=[0,2];
hold on
plot (KLDWeibull(1,3970), KLDNormal(1,3970), 'o')
plot(x,y,'LineWidth',2);
xlabel('Distribution X')
ylabel('Distribution Y')
legend('window of a timeframe','y=x')
hold off
```

**Making a table of fifteen values for each distribution from the values of the KL-divergence**

```
MatrixTenMin=zeros(65952,4);
MatrixTenMin(:,1)=KLDWeibull;
MatrixTenMin(:,2)=KLDRician;
MatrixTenMin(:,3)=KLDNormal;
MatrixTenMin(:,4)=KLDNakagami;
RndCatch=randperm(65952);
MatrixTenMin15=MatrixTenMin(RndCatch(1:15),:);
```

**Plotting the time series of the data**

```
Data=[OneMonth{1,1} OneMonth{2,1} OneMonth{3,1} OneMonth{4,1} OneMonth{5,1}
OneMonth{6,1} OneMonth{7,1} OneMonth{8,1} OneMonth{9,1} OneMonth{10,1} OneMonth{11,1}
OneMonth{12,1} OneMonth{13,1} OneMonth{14,1} OneMonth{15,1}];
Data1=zeros(1,40000000);
Counter=1;
for i=1:15
   for j=1:2638080
      Data1(Counter)=Data(j,i);
      Counter=Counter+1;
   end
```

```
end
plot([1:40000000],Data1)
```

**Plotting a histogram of all the data for a specific timeframe**

```
histfit(muNormal,120,'nakagami')
hold on
histfit(muNormal,120,'normal')
histfit(muNormal,120,'rician')
histfit(muNormal,120,'weibull')
ylabel('counts')
xlabel('wind speed (m/s)')
xlim([0 28])
ylim([0 1600])
legend('Data','Nakagami','Data','Normal','Data','Rician','Data','Weibull')
hold off
```

**Plotting an illustrative histogram for a window of that time frame**

```
histfit(TenMin{18,1},20,'nakagami')
hold on
histfit(TenMin{18,1},20,'normal')
histfit(TenMin{18,1},20,'rician')
histfit(TenMin{18,1},20,'weibull')
xlabel('wind speed (m/s)')
ylabel('counts')
legend('Data','Nakagami','Data','Normal','Data','Rician','Data','Weibull')
hold off
```

**Plotting the KL-divergence scatterplots**

```
x=[0,2];
y=[0,2];
hold on
plot (KLDWeibull, KLDNormal, '.')
plot(x,y,'LineWidth', 2);
xlabel('Weibull distribution')
ylabel('Normal distribution')
legend('window of one-week time frame')
```

**Calculating the variance for a time frame**

```
VarTenMinNakagami=var(KLDNakagami);
VarTenMinNormal=var(KLDNormal);
VarTenMinRician=var(KLDRician);
VarTenMinWeibull=var(KLDWeibull);
```

**Plotting the variation of the mean ± standard deviation (μ ± σ) for all time frames**

```
x=1:7;

yNak=[MeanOneMonthNakagami,MeanTwoWeekNakagami,MeanOneWeekNakagami,MeanTwoDay
Nakagami,MeanTwelveHourNakagami,MeanOneHourNakagami,MeanTenMinNakagami];
errorNak=[DeviationOneMonthNakagami,DeviationTwoWeekNakagami,DeviationOneWeekNakagami,
DeviationTwoDayNakagami,DeviationTwelveHourNakagami,DeviationOneHourNakagami,DeviationTe
nMinNakagami];
errorbar(x,yNak,errorNak)
hold on

yNor=[MeanOneMonthNormal,MeanTwoWeekNormal,MeanOneWeekNormal,MeanTwoDayNormal,
MeanTwelveHourNormal,MeanOneHourNormal,MeanTenMinNormal];
errorNor=[DeviationOneMonthNormal,DeviationTwoWeekNormal,DeviationOneWeekNormal,Deviati
onTwoDayNormal,DeviationTwelveHourNormal,DeviationOneHourNormal,DeviationTenMinNormal];
errorbar(x,yNor,errorNor)

yRic=[MeanOneMonthRician,MeanTwoWeekRician,MeanOneWeekRician,MeanTwoDayRician,MeanT
welveHourRician,MeanOneHourRician,MeanTenMinRician];
errorRic=[DeviationOneMonthRician,DeviationTwoWeekRician,DeviationOneWeekRician,DeviationTw
oDayRician,DeviationTwelveHourRician,DeviationOneHourRician,DeviationTenMinRician];
errorbar(x,yRic,errorRic)

yWei=[MeanOneMonthWeibull,MeanTwoWeekWeibull,MeanOneWeekWeibull,MeanTwoDayWeibull
,MeanTwelveHourWeibull,MeanOneHourWeibull,MeanTenMinWeibull];
errorWei=[DeviationOneMonthWeibull,DeviationTwoWeekWeibull,DeviationOneWeekWeibull,Deviat
ionTwoDayWeibull,DeviationTwelveHourWeibull,DeviationOneHourWeibull,DeviationTenMinWeibull]
;
errorbar(x,yWei,errorWei)

xlabel('timeframe')
ylabel('KL-Divergence from real data')
legend('Nakagami','Normal','Rician','Weibull')
hold off
```

# Appendix B: UNDP-goals

In this appendix the report produced during the sustainability workshop about the UNDP-goals as well as the PowerPoint presentation (see Figure 12) is showed.

### Group I - Predicting wind power from wind speed data

In this text, first our project aim is described and after that there is a list of UN sustainable development goals where the project contributes to directly or indirectly.

**Description of our project:**

Wind is a very promising renewable energy source for a sustainable future. It is clean and is becoming cheaper than non-renewable sources of energy, such as coal or oil. However, there is an important disadvantage: wind fluctuates a lot. This is a problem for secure planning of the construction of wind farms, so it is needed to calculate previously the power that wind can be generating from a farm. To help with solving this issue, we are plotting several statistical distributions for wind speeds to better understand the fundamental features underlying the fluctuating wind power. Hereby, we try to find the best distribution to predict wind speeds with.

**We contribute to the following goals of the United Nations:**

**UN-7: Affordable and clean energy**

This goal is about ensuring access to affordable, reliable, sustainable and modern energy for all. This is the main goal we can contribute to with our project. The distributions we are going to look at with the given data of an offshore wind farm can be useful for a better understanding of the fundamental features underlying the fluctuating wind power of a wind turbine and its statistical connection with the speed of the wind blowing across it. This is needed because these features are essential for a secure planning of the construction of wind farms, for monitoring the functioning of wind turbines and for trustfully anticipate energy market quotations. Besides that, it will lead to a better energy efficiency, which is very important due to the amount of energy that is non-renewable and that needs to be reduced as quickly as possible.

7.1 and 7.2: Universal access to modern energy and Increase global percentage of renewable energy

Our project aims to make wind energy more reliable by trying to find a better probability distribution to fit wind speed data. If wind energy is more reliable, this means more efficient wind farms will be built, making renewable energy more present in our lives.

7.3: Double the improvement in energy efficiency

By finding a better statistical distribution for wind speed, electrical engineers are able to predict the wind power more accurately. Because of this, they will be able to be more energy efficient. Global electricity generation from wind increased 8 times from 2005 (105 TWh) to 2015 (832 TWh).

7.A: Promote access to research, technology and investments in clean energy

Our project content/aim is to find a better probability distribution to model wind speed. So, we are directly doing research in clean energy, especially wind power. With our research we are promoting to R&D to look our report and investigate renewable energy in the same way we do.

7.B: Expand and upgrade energy services for developing countries

Our project aims to improve wind energy by making it more reliable and cost-efficient and therefore more affordable in developing countries for example.

**UN-9: Industry, innovation and infrastructure**

This goal is about building resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation. Our project can contribute to this goal because of the search for the best statistical distribution for wind speed that is useful for upgrading our industry. The three sub goals of this goal that fit best to our project are:

9.2: Promote inclusive and sustainable industrialization

Renewable energy, especially wind energy, is becoming more and more important. If we find a better statistical distribution fitting our wind speed data, it will be more accurate to predict wind power out of wind speed. So, wind energy would be more reliable. The wind farm companies will be able to build more wind farms and then the market will grow, lowering the costs due to the economy of scale. With this, we promote a sustainable industrialization.

9.4: Upgrade All Industries and Infrastructures for Sustainability

The industry must use clean and environmentally friendly technologies. But they also need clean energy. With our research to find the best statistical distribution, wind power would be better and more reliable. The wind farm companies will build more farms and the whole energy industry becomes cleaner.

9.5: Enhance Research and Upgrade Industrial Technologies

Our project content/aim is a kind of research. If the result of our project is good, we can enhance research to upgrade the technological capabilities of the wind farm companies. Then, they can build more wind farms providing the highest energy efficiency.

**Presentation about the UN goals:**


*Figure 12: Powerpoint slides of the presentation about the UNDP-goals. From the top to the bottom: slide 1, slide 2, and slide 3.*

# Appendix C: Additional Figures and tables

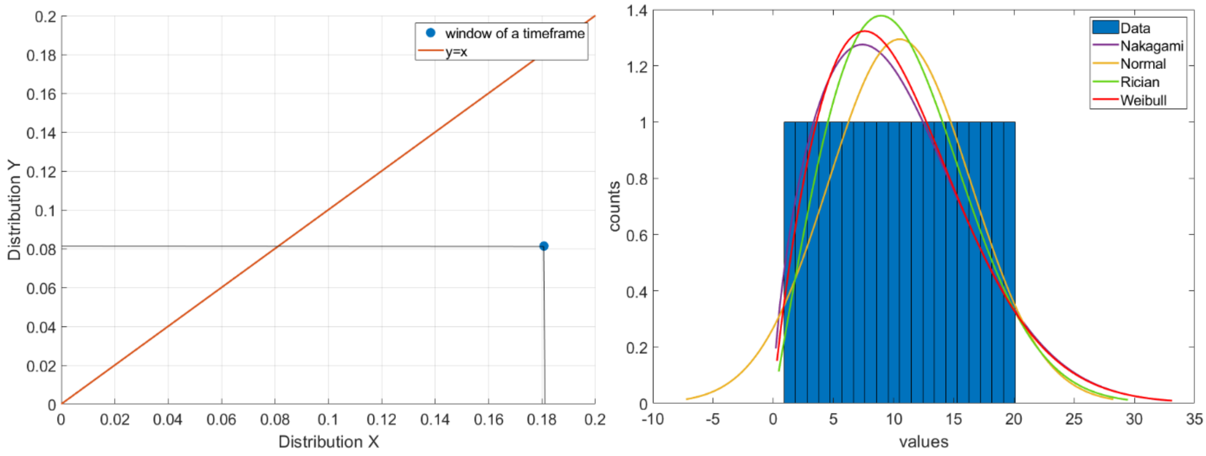


*Figure 13: a) How to read the KL-divergence with a simplification for a single window in a specific timeframe. b) Representation of the variation that the four distributions can experience when a uniform distributed data series is applied, as a simplification for the non-uniform data series of the wind speed.*

Figure 13a shows how to read the scatter plots of KL-divergence. Instead of showing all the windows for a certain time frame, this figure shows only one window. The red line is the function y=x. Any window located in the red line means that the Distribution X and the Distribution Y have the same divergence from the real data in that window. Above the red line the distribution X has better performance than Distribution Y and below the red line, as the window shown, is the opposite case. As can be seen, the window is located at the position (x=0.18, y=0.08) which means that the Distribution Y has less distance than distribution X so the first one represents better the data than the second one for that window of time.

The wind speed data analyzed consists in 40 million non-uniformly distributed values. Figure 13b is another simplification which has the purpose to highlight the different shapes of the four distributions when representing a uniformly distributed dataset sample of 20 values from 1 to 20 (1, 2, 3,… , 20).

### One-month time frame

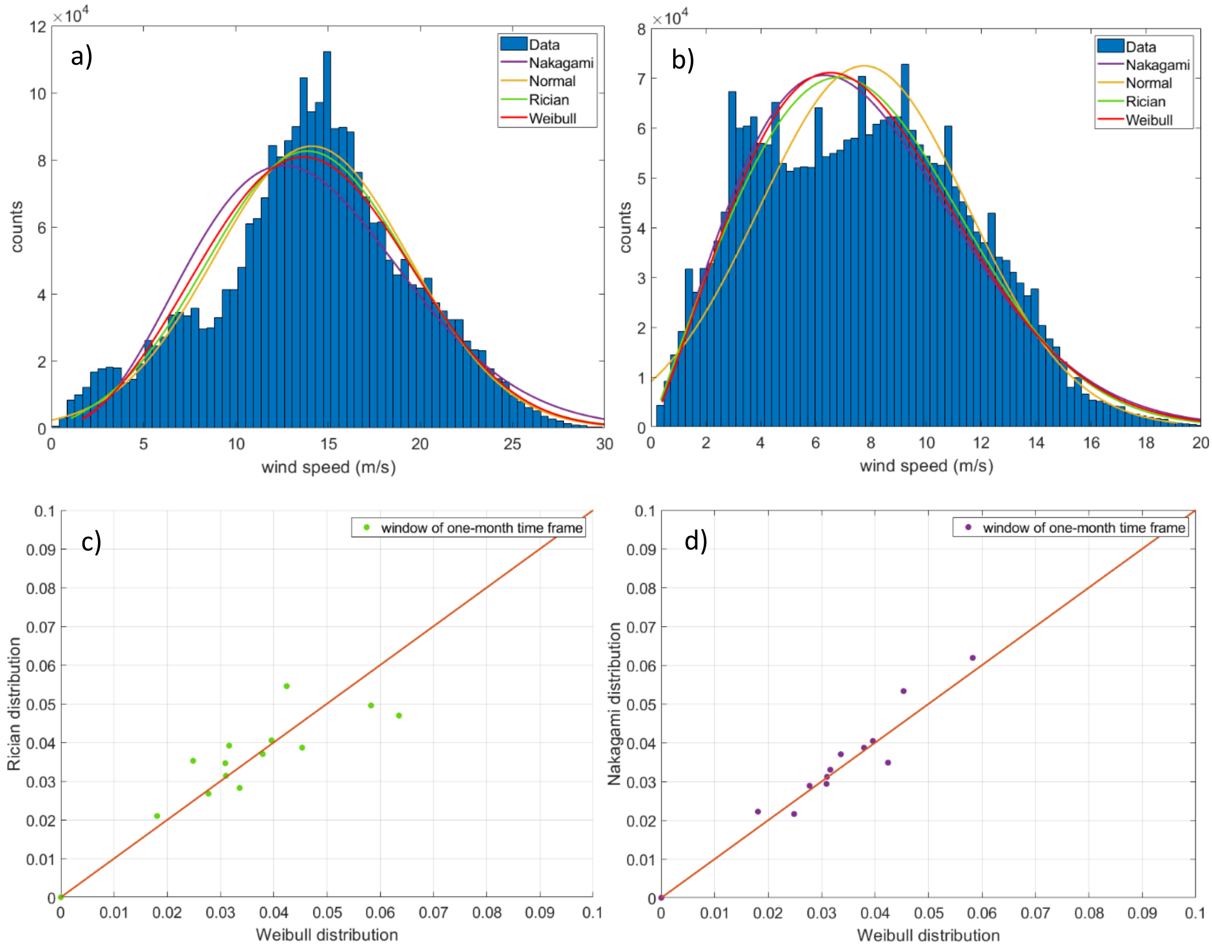


*Figure 14: a) Comparison of the different probability distributions for a second illustrative window of one-month time frame of distributed wind speed data. b) Comparison of the different probability distributions for a third illustrative window of one-month time frame of distributed wind speed data. c) Scatter plot of all KL-divergences of Rician vs Weibull distribution for one-month time frame. d) Scatter plot of all KL-divergences of Nakagami vs Weibull distribution for one-month time*

When looking at Figure 14c, most values are above the red line y=x. This means that Weibull distribution has shorter distances than Rician distribution. Based on this can be concluded that the Weibull distribution fits the wind speed data better for the one-month time frame. When looking at Figure 14d, it can be concluded that Weibull distribution also fits better than the Nakagami distribution for the one-month time frame.

**Two-week time frame**

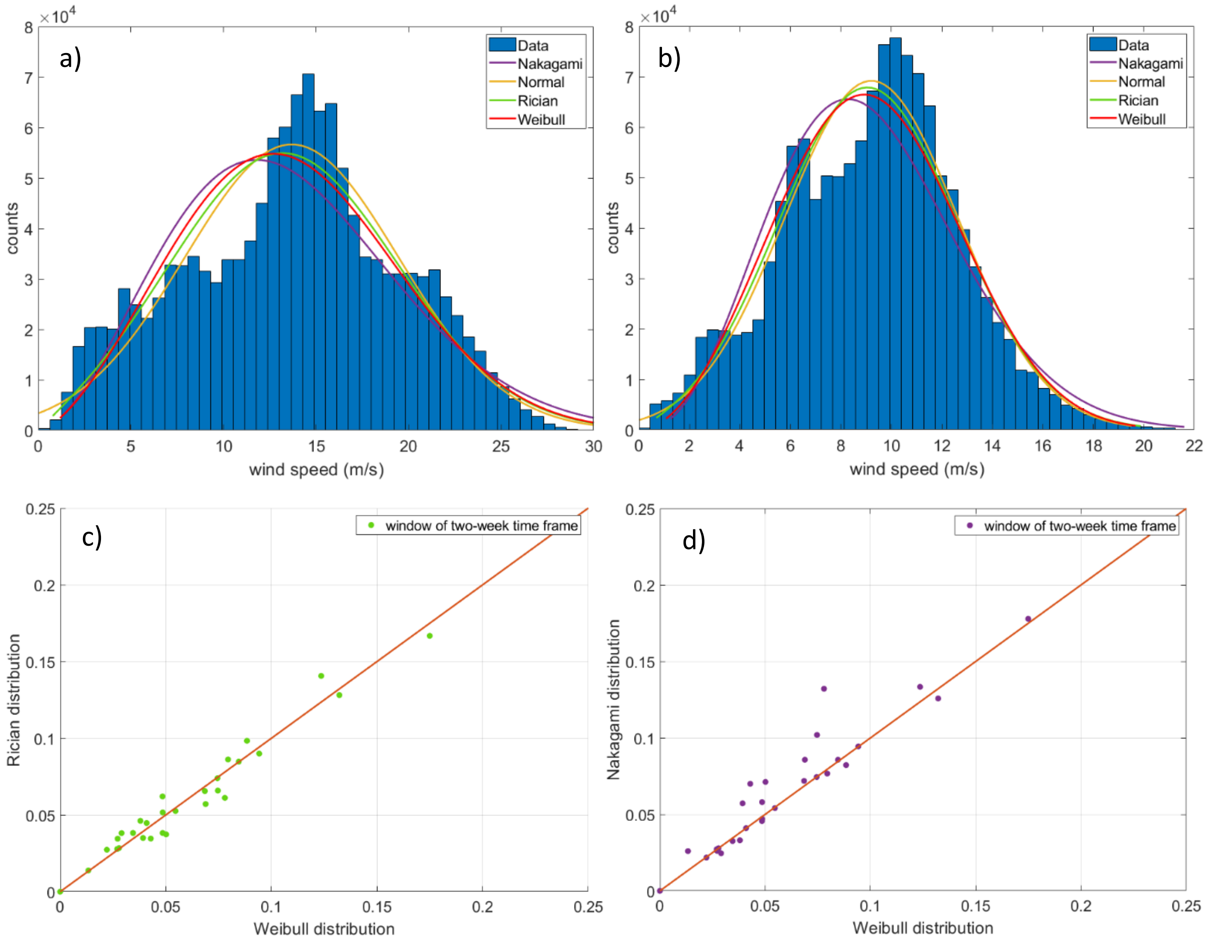


*Figure 15: a) Comparison of the different probability distributions for a second illustrative window of two-week time frame of distributed wind speed data. b) Comparison of the different probability distributions for a third illustrative window of two-week time frame of distributed wind speed data. c) Scatter plot of all KL-divergences of Rician vs Weibull distribution for two-week time frame. d) Scatter plot of all KL-divergences of Nakagami vs Weibull distribution for two-week time frame.*

When looking at Figure 15c and Figure 15d, most values are above the red line y=x. Based on this observation can be concluded that Weibull distribution fits the wind speed data better for two-week time frame compared to Rician distribution. The same can be concluded when comparing Weibull distribution with Nakagami distribution.

### One-week time frame

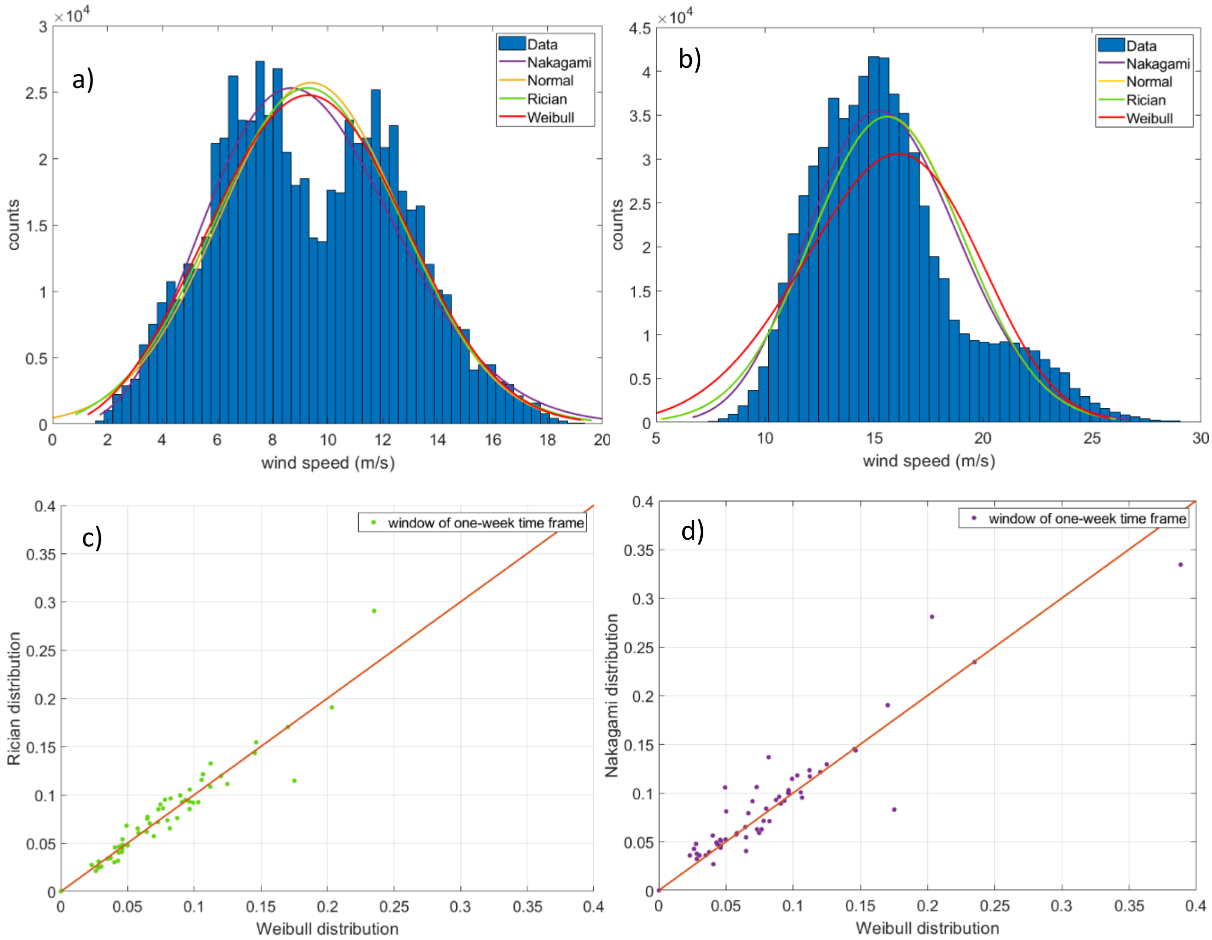


*Figure 16: a) Comparison of the different probability distributions for a second illustrative window of one-week time frame of distributed wind speed data. b) Comparison of the different probability distributions for a third illustrative window of one-week time frame of distributed wind speed data. c) Scatter plot of all KL-divergences of Rician vs Weibull distribution for one-week time frame. d) Scatter plot of all KL-divergences of Nakagami vs Weibull distribution for one-week time frame.*

When looking at Figure 16c and Figure 16d, most values are above the red line y=x. Based on this observation can be concluded that Weibull distribution fits the wind speed data better for one-week time frame compared to Rician distribution. The same can be concluded when comparing Weibull distribution to Nakagami distribution.

**Two-day time frame**

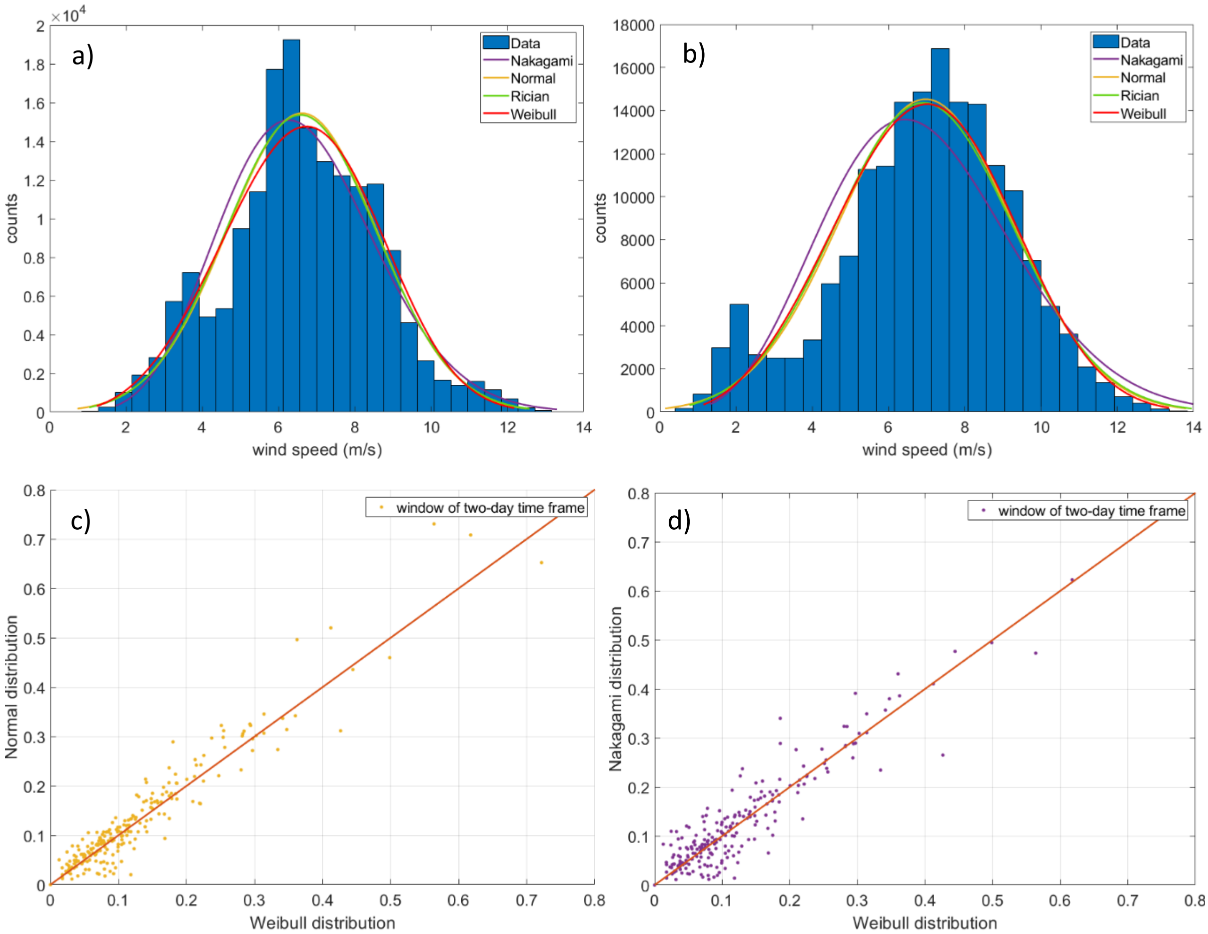


*Figure 17: a) Comparison of the different probability distributions for a second illustrative window of two-day time frame of distributed wind speed data. b) Comparison of the different probability distributions for a third illustrative window of two-day time frame of distributed wind speed data. c) Scatter plot of all KL-divergences of Normal vs Weibull distribution for two-day time frame. d) Scatter plot of all KL-divergences of Nakagami vs Weibull distribution for two-day time frame.*

When looking at Figure 17c and Figure 17d, most values are above the red line y=x. Based on this observation can be concluded that Weibull distribution fits the wind speed data better for two-day time frame compared to Normal distribution. the same can be concluded when comparing Weibull distribution to Nakagami distribution.

## Twelve-hour time frame

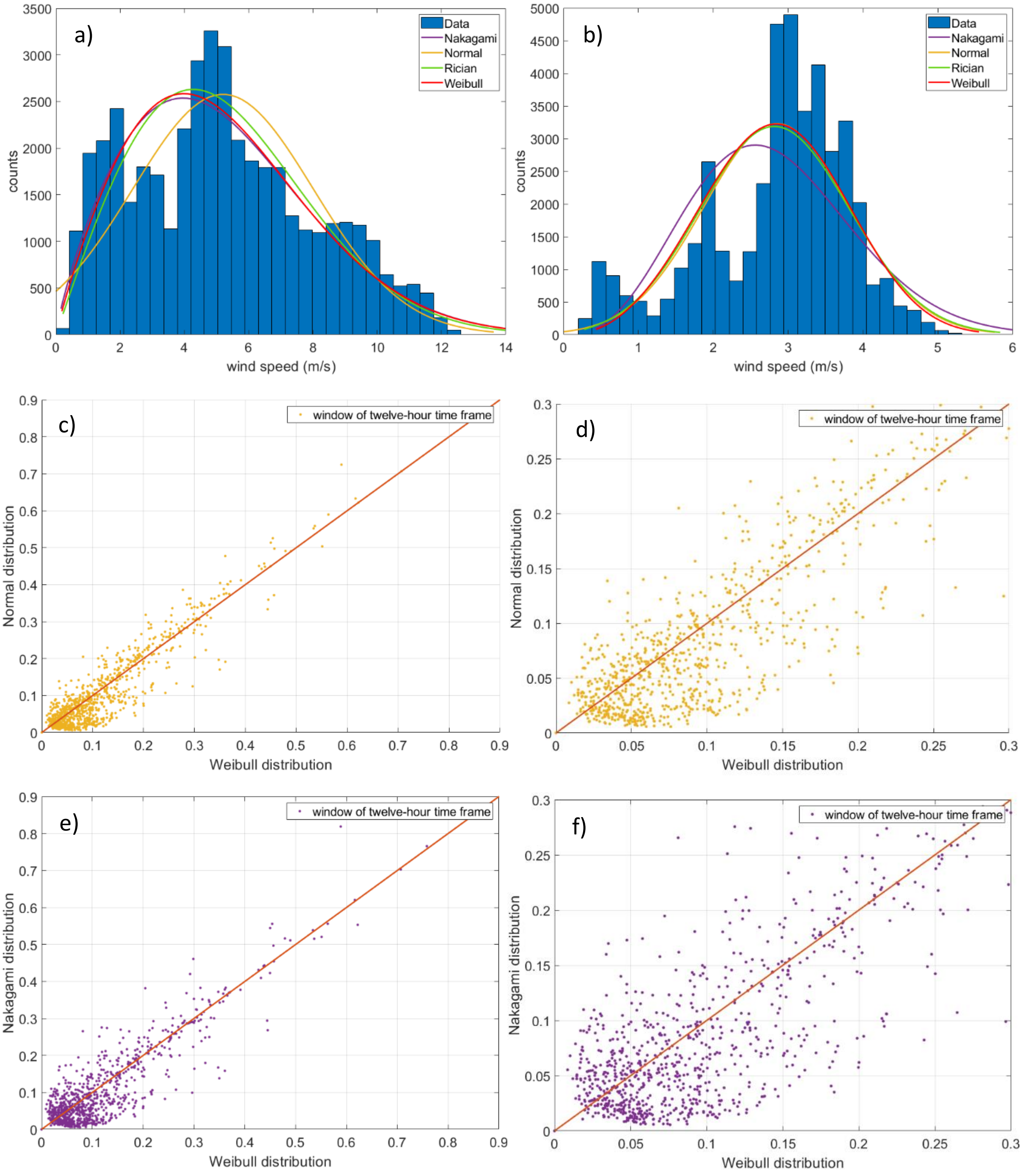


*Figure 18: a) Comparison of the different probability distributions for a second illustrative window of twelve-hour time frame of distributed wind speed data. b) Comparison of the different probability distributions for a third illustrative window of twelve-hour time frame of distributed wind speed data. c) Scatter plot of all KL-divergences of Normal vs Weibull distribution for twelve-hour time frame. d) Zoomed scatter plot of all KL-divergences of Normal vs Weibull distribution for twelve-hour time frame. e) Scatter plot of all KL-divergences of Nakagami vs Weibull distribution for twelve-hour time frame. f) Zoomed scatter plot of all KL-divergences of Nakagami vs Weibull distribution for twelve-hour time frame.*

When looking at Figure 18c, most values are below the red line y=x. This means that Normal distribution has shorter distances than Weibull distribution. Based on this can be concluded that Normal distribution fits the wind speed data better for the twelve-hour time frame. When looking at Figure 18e, most values are below the red line y=x. This means that Nakagami distribution has shorter distances than Weibull distribution. Based on this can be concluded that Nakagami distribution fits the wind speed data better for the twelve-hour time frame.

## One-hour time frame

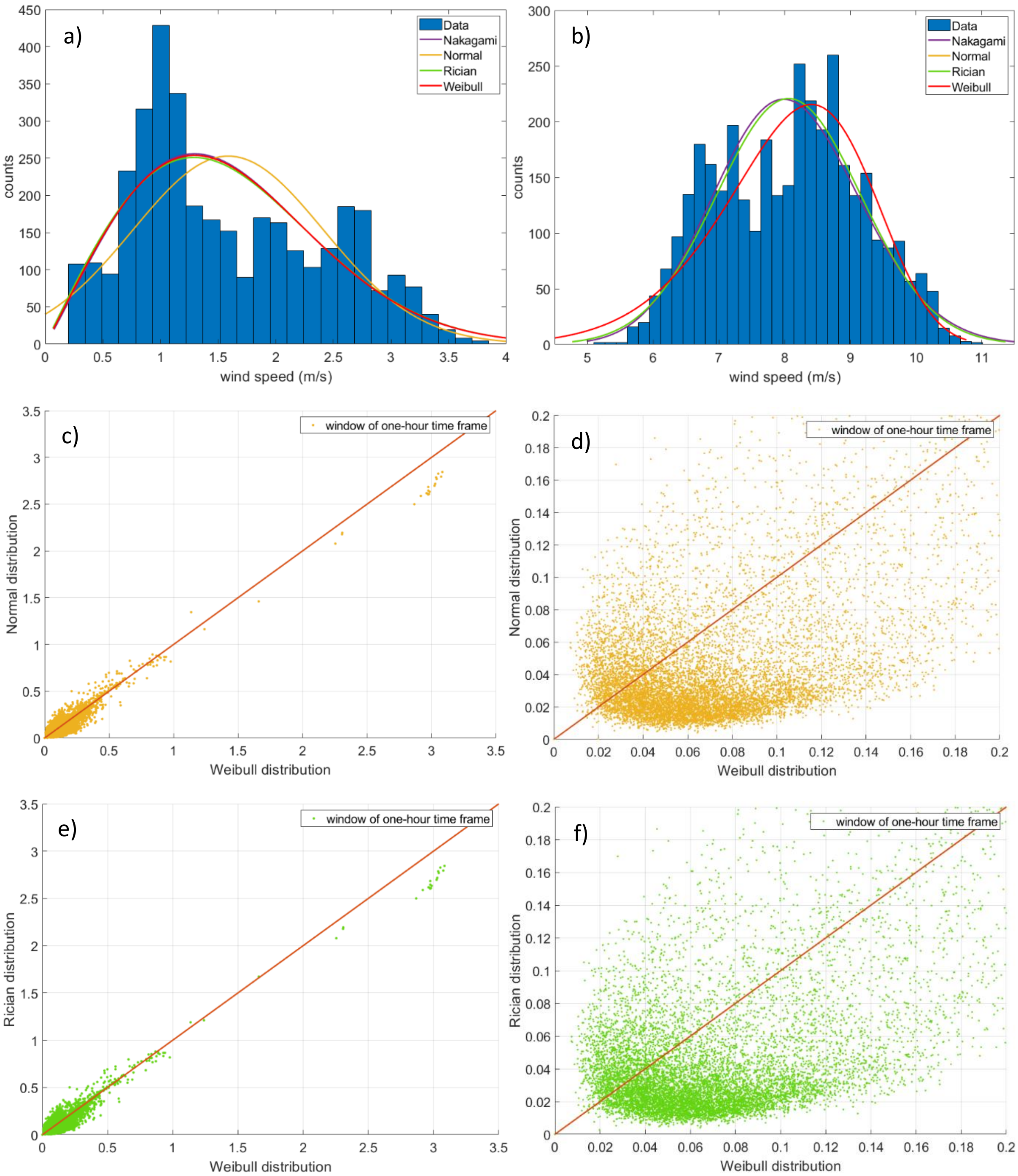


*Figure 19: a) Comparison of the different probability distributions for a second illustrative window of one-hour time frame of distributed wind speed data. b) Comparison of the different probability distributions for a third illustrative window of one-hour time frame of distributed wind speed data. c) Scatter plot of all KL-divergences of Normal vs Weibull distribution for one-hour time frame. d) Zoomed scatter plot of all KL-divergences of Normal vs Weibull distribution for one-hour time frame. e) Scatter plot of all KL-divergences of Rician vs Weibull distribution for one-hour time frame. f) Zoomed scatter plot of all KL-divergences of Rician vs Weibull distribution for one-hour time frame.*

When looking at Figure 19c, most values are below the red line y=x. This means that Normal distribution has shorter distances than Weibull distribution. Based on this can be concluded that Normal distribution fits the wind speed data better for the one-hour time frame. When looking at Figure 19e, most values are below the red line y=x. This means that Nakagami distribution has shorter distances than Weibull distribution. Based on this can be concluded that Nakagami distribution fits the wind speed data better for the one-hour time frame.

## Ten-minute time frame

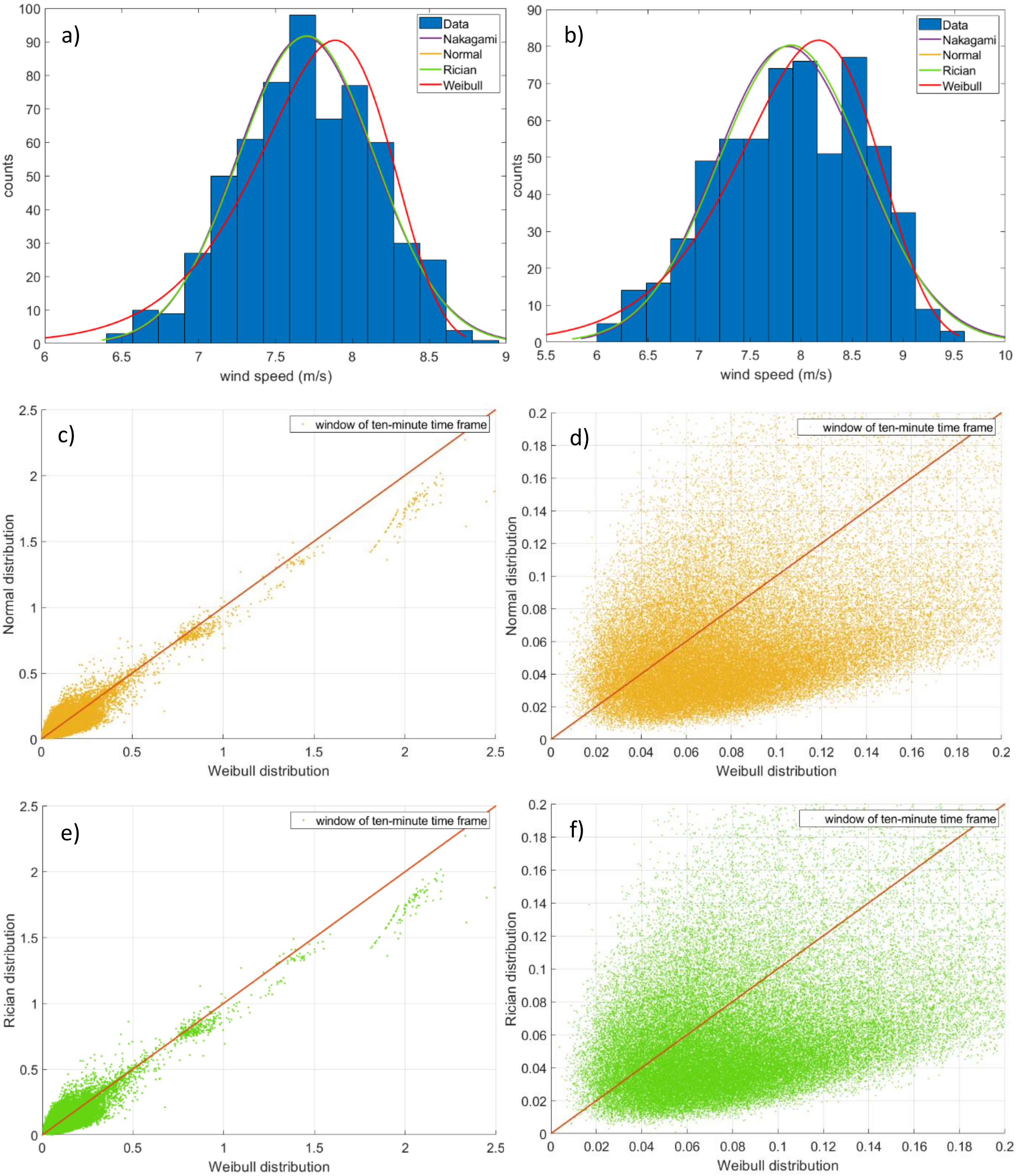


*Figure 20: a) Comparison of the different probability distributions for a second illustrative window of ten-minute time frame of distributed wind speed data. b) Comparison of the different probability distributions for a third illustrative window of ten-minute time frame of distributed wind speed data. c) Scatter plot of all KL-divergences of Normal vs Weibull distribution for ten-minute time frame. d) Zoomed scatter plot of all KL-divergences of Normal vs Weibull distribution for ten-minute time frame. e) Scatter plot of all KL-divergences of Rician vs Weibull distribution for ten-minute time frame. f) Zoomed scatter plot of all KL-divergences of Rician vs Weibull distribution for ten-minute time frame.*

When looking at Figure 20c, most values are below the red line y=x. This means that Normal distribution has shorter distances than Weibull distribution. Based on this can be concluded that Normal distribution fits the wind speed data better for the ten-minute time frame. When looking at Figure 20e, most values are below the red line y=x. This means that Nakagami distribution has shorter distances than Weibull distribution. Based on this can be concluded that Nakagami distribution fits the wind speed data better for the ten-minute time frame.