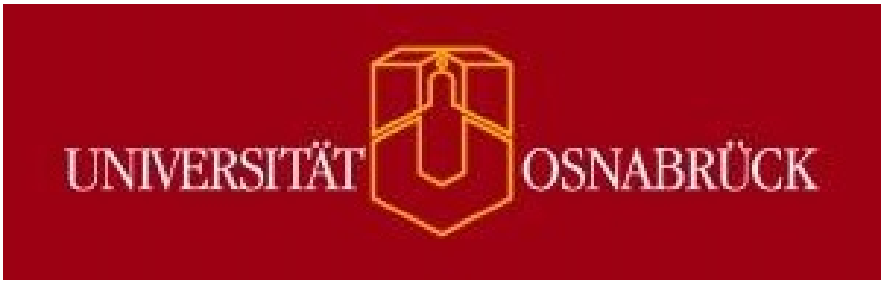

Fachbereich Physik

Masterarbeit

Stochastic analysis of off shore wind profiles

Kristoffer Schnieders

Betreuer: Dr. Pedro Gonçalves Lind

Zweitprüfer: Prof. Dr. Philipp Maass

Contents

1 Scope and motivation **1**

2 State of the art **3**

2.1 Modelling wind velocity profiles 3

2.2 Statistics of wind velocity measurements 4

2.3 Statistics of wind velocity increments 5

2.4 Markov processes and their application to wind data 6

2.5 Basics of stochastic data analysis 8

3 Data preprocessing **11**

4 Statistical analysis of single time series **15**

4.1 The wind speed 15

4.2 The wind direction 17

5 Time increment analysis **21**

5.1 Laplace distribution 21

5.2 Castaing distribution 25

6 Spatial increment analysis **31**

7 Stochastic modelling of wind profiles **37**

8 Discussion and conclusion **43**

Abstract

The aim of this thesis is twofold: Firstly, to describe the statistical features especially of short-time wind speed profiles offshore and secondly, to discuss a stochastic approach for modelling the short-time evolution of offshore wind profiles. To that end, we consider empirical data collected at the FINO1 platform between September 2015 and April 2017.

For the wind data, the Weibull distribution provides an adequate fit. The scale parameter strongly fluctuates depending on the wind direction. For the increments in time and in height, the Castaing distribution yields better results compared to the Laplace distribution for small lags, while for larger lags the errors are similar. The Laplace and Castaing coefficients are an exponential function of the time and distance, while for other parameters the dependence is more difficult to understand. For large increments it is possible to simulate adequate surrogate data using the Langevin equation, while for small increments distinct differences are observed in the distribution of the measured and simulated data.

All in all, we conclude that with short-time scales, there are large velocity fluctuations. Moreover, wind profiles deviate significantly from the standard (average) wind profile models. Our findings refer to the importance of incorporating stochastic models of wind profile fluctuations within short-time periods for better monitoring and predicting the fluctuation of wind.

Acknowledgements

First of all, I want to thank Prof. Dr. Phillip Maass for the permission to write my master thesis as a member of the Statistical Physics group at the University of Osnabrück. Furthermore, I want to express my special gratitude to the operators of the FINO1 platform for providing me with the wind speed data I used in my thesis. The FINO-project is financially supported by the Federal Ministry of Economics and Energy of the Federal Republic of Germany and the Jülich Research Centre.

For his patient and thoughtful help during all stages of my master thesis I want to sincerely thank my supervisor, Dr. Pedro Gonçalves Lind. Finally, I want to thank the members of the Statistical Physics group in Osnabrück for several discussions of the thesis and my parents for correcting the spelling.

1 Scope and motivation

Stochastic dynamics behind wind profiles and their short-time fluctuations are still not entirely understood, especially off shore.

In the past 20 years an increasing number of wind turbines has been integrated in electrical grids all over the world, and thus, the importance of wind energy has considerably increased. Nowadays, apart from the electrical outputs of these power plants, the load is a central property. The reason for this is that the load sets a limit to the lifetime of the power plant. The short-time fluctuations of wind profiles up to 100 m height play a fundamental role in cumulative loads on turbines. For small wind turbines up to a dozen meters, the wind can be considered as laminar. However, for large turbines extending over more than 100 m height, non-laminar profiles are observed. (See figure 1.1).

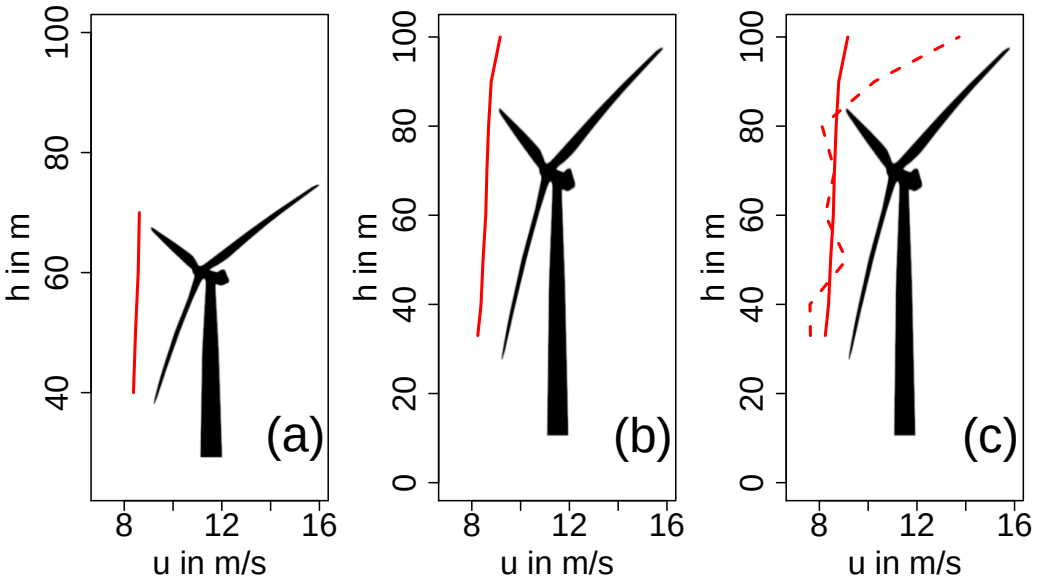

Figure 1.1: Graphic illustrating different scopes on wind turbine. (a) Small wind turbine with a diameter of about 30 m and quasi-laminar average wind speed. (b) Turbine with a diameter of about 70 m. (c) Turbine with a diameter of about 70 m. The mean wind speed (solid line) and an example for the measured wind speed in one second (dashed line) are shown.

The wind speed for small and large wind turbines is shown under different scopes in figure 1.1. In graphic (a), the long-term average wind speed for a small turbine with a diameter of about $30\,\mathrm{m}$ is shown. Here, a quasi-laminar flow can be assumed. In graphic (b), you see the long-time average for a wind turbine with a diameter of about $70\,\mathrm{m}$. In this case, assuming a quasi-laminar flow is no longer sufficient, even for the mean wind speed.

In graphic (c), the long-time average is shown along with the wind speed measured in one second. Here, a constant wind profile can no longer be found. The wind speed fluctuation has an effect on the load and output of the turbine.

To this end, we consider short-time scales and fluctuations between wind speeds at different heights. Short-time wind speed fluctuations cause fast fluctuating torsions on the turbine blades and indirectly on the tower, influencing the lifetime of the power plant. Therefore, a proper understanding of the statistical features of these fluctuations is important.

An adequate stochastic model of wind profiles may enable us to derive time series of the wind speed for the full range of the rotors of wind turbines, and provide the opportunity to improve power production forecasts (see [15]) and estimates of fatigue loads(see [10]).

For this purpose, we will give an overview over the necessary theoretical framework used in this thesis in chapter 2. In chapter 3, the location of the platform, the positioning and the type of gauges on the FINO1 tower and the pre-processing of the data are presented. In chapter 4, we are going to analyse single time measurements. Firstly, in section 4.1, we will consider the wind speed measurements with no further constraints. Secondly, in section 4.2, we will analyse wind speed constraints referring to the direction. In chapter 5 and 6, the time and spatial increments will be in the centre of interest. Initially, it is tried to describe the distribution by fitting it to the Laplace distribution. Afterwards, the Castaing model is used. In chapter 7, surrogate time series are simulated by the use of the Langevin equation. Then the distributions of the simulated and measured time series are compared. In chapter 8, a conclusion of the results will be presented and discussed.

2 State of the art

In this section, the theoretical background of the statistical analysis of the wind speed profile is given. The main models of wind speed data and wind speed increments are described. For the analysis of the increments, the Castaing and Laplace distributions are presented. The stochastic modelling of the wind profile with respect to the height is done with the Langevin equation. This is theoretically discussed in section 2.5.

2.1 Modelling wind velocity profiles

Here, the most common profiles for the modelling of wind speed as a function of the height will be presented. These functions are known as the power law and log wind profile. Those types of wind profiles describe the wind profiles on shore at heights of between $33\,\mathrm{m}$ and $100\,\mathrm{m}$ equally well. Even though the performance of the wind profile on shore is well known, in the case of offshore wind profiles, a lack of investigation still exists. This topic is of interest, because of the influence of wind on the evolution of waves as well as the force affecting ships and other man-made obstacles in the sea, and especially the increasing importance of offshore power generation by wind turbines.

First of all, we are presenting the log wind profile. This profile is based upon the idea that the wind above a height at which the wind velocity is nearly damped to zero due to objects on the ground approximately evolves in a logarithmic way. The top height, at which the wind velocity is assumed to be linearly independent of the upper layers of wind flux, is called the zero-plane displacement d. With respect to the surface structure, d varies between about $30\,\mathrm{m}$ in the Amazon rain forest [13] to about $1\,\mathrm{mm}$ for flat, not overgrown terrain like water or snowscape. Apart from this, the structure on the ground has an effect on the upper atmospheric layers due to different degrees of turbulences. This can be estimated by the surface roughness z_0. The log wind profile is given by

$$u(z) = \frac{u_*}{\kappa}\left(\log\left(\frac{z-d}{z_0}\right)\right). \tag{2.1}$$

Here u_* is the friction velocity, κ is the Von Kárnán constant and z is the height above

the ground. The wind speed at height z can be calculated by

$$u(z) = \frac{u(z_{\text{ref}}) \log(\frac{z-d}{z_0})}{\log(\frac{z_{\text{ref}}-d}{z_0})}. \tag{2.2}$$

Here, z_{ref} indicates a reference height for which the wind speed is known.

In the power law wind profile, the wind speed at height z is defined by

$$u(z) = u(z_{\text{ref}})(\frac{z}{z_{\text{ref}}})^{\alpha}. \tag{2.3}$$

Here, α is an empirical coefficient which is assumed to reproduce the mean wind speed at the other heights in the best way. On shore, this coefficient is regularly set to $\alpha = \frac{1}{7}$. As Hsu *et al.* [8] have shown in 1994, an α value of about $\frac{1}{10}$ is more adequate when talking about offshore wind profiles. Apart from these many more or less complicated approaches, wind profile models were introduced for special circumstances. These are beyond the scope of this investigation because of the comparably low roughness offshore.

2.2 Statistics of wind velocity measurements

In general, it is assumed that the wind speed data for longer periods of time converge towards the Weibull distribution which is given by

$$f_{k,\lambda}(u) = \frac{k}{\lambda}\left(\frac{u}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{u}{\lambda}\right)^{k}\right). \tag{2.4}$$

By choosing this notation, the coefficient λ can be seen as a scale coefficient and k as a shape parameter.

As discussed by C.Carrillo *et al.* [5], there are several ways to find the best fitting parameters for the given dataset. In this thesis the already implemented method `nls` in R is mainly used. This method works with the least square method to find the parameters that fit best to the distribution. On the one hand, we characterize the goodness of the fit by the Kolmogorov-Smirnov test associated with the metric

$$d_{KS} = \sup(F_0 - F_n), \tag{2.5}$$

where F_0 is the measured probability distribution and F_n is the fitted probability

distribution. On the other hand, we use the χ^2 defined by

$$\chi^2 = \sum_{i=0}^{N} \frac{(p_o(\nu_i) - p_t(\nu_i))^2}{p_o(\nu_i) + p_t(\nu_i)}. \tag{2.6}$$

Here, N is the number of bins, ν a variable on which the distribution depends, p_o is the probability observed and p_t is the probability that is theoretically assumed according to the model and parameters chosen. Both methods are often used and thus, are well known in statistics. The difference between them is that the Kolmogorov-Smirnov test is more sensitive to the magnitude of single deviations between measured and fitted probability, while the χ^2 test gives an idea of the scale of deviation all over the area considered.

It is also possible to simply use the turbulent part of the wind by subtracting the mean value of the wind speed over a defined time from the measured values, namely

$$u'_T(t) = u(t) - \langle u(t) \rangle_T. \tag{2.7}$$

The mean wind speed $\langle u(t) \rangle_T$ is chosen to be the mean wind speed over the whole time series of 20 months.

2.3 Statistics of wind velocity increments

Apart from the one-point statistics, a closer look is taken at the two-point statistics. The time increments are computed by

$$\Delta_\tau u(t) = u(t+\tau) - u(t) = u'_\tau(t+\tau) - u'_\tau(t). \tag{2.8}$$

For the description of small increments, no matter if in time or space, nowadays Castaing's model, which was presented in [6], is most commonly used. Here, we follow the procedure that was presented by R. Wagner et. al in [15]. The probability depending on $\Delta_\tau u$ is given by

$$p(\Delta_\tau u(T)) = \frac{1}{2\pi\Lambda(\tau)} \int_0^\infty \frac{d\sigma}{\sigma^2} \exp\left(-\frac{\Delta_\tau u(T)^2}{2\sigma^2} - \frac{\left(\ln\left(\frac{\sigma}{\Sigma}\right)\right)^2}{2\Lambda(\tau)^2} \right). \tag{2.9}$$

The parameters determining the shape of the resulting distribution are $\Lambda(\tau)$ and $\Sigma(\tau)$ [4]. These can be calculated by

$$\Lambda(\tau)^2 = \frac{\ln(F(\Delta_\tau u))}{4}, \tag{2.10}$$

and

$$\Sigma^2(\tau) = \langle \Delta_\tau u^2 \rangle \exp\left(-2\left(\Lambda(\tau)\right)^2\right), \tag{2.11}$$

where $F(\Delta_\tau u(T))$ is the flatness which is the fourth central moment normalized to the standard deviation (See [1]. The flatness is defined by

$$F(\Delta_\tau u) = \frac{\langle (\Delta_\tau u(T) - \bar{\Delta_\tau u})^4 \rangle}{\sigma^4_{\Delta_\tau u}} - 3. \tag{2.12}$$

The flatness can be used to quantify the deviation of a dataset from the Gaussian distribution, because the fourth moment of the Gaussian distribution equals 3. For this reason, the flatness of the Gaussian distribution equals 0.

The time increments are also analysed by using the much simpler Laplace distribution. Both approaches will be compared to examine if, and if yes, for which scales of τ both descriptions are sufficiently equivalent.

The Laplace distribution is given by

$$F_{Laplace}(\Delta_\tau u) = \begin{cases} \frac{1}{2} \exp \frac{\Delta_\tau u - \mu}{\sigma_-} & \text{if } \Delta_\tau u < \mu \ , \\ 1 - \frac{1}{2} \exp \frac{\Delta_\tau u - \mu}{\sigma_+} & \text{if } \Delta_\tau u \geq \mu \ . \end{cases} \tag{2.13}$$

In this formula, μ is the median of the distribution. For all datasets analysed in this thesis, μ equals 0 and is the maximum of the distribution. σ_+ and σ_- are the scale parameters for values bigger and smaller than μ respectively.

2.4 Markov processes and their application to wind data

In this section, an overview is given on the properties which a time series has to fulfil, so that the Langevin equation can be applied on the time series.

Firstly, it is to be tested whether the time series has Markov properties. For this reason, a short stochastic view is taken at the concept of conditional probabilities

$$p(x|y) = \frac{p(x \cap y)}{p(y)}, \tag{2.14}$$

which gives the probability of the event x under the condition that the event y took place. If you now look at a time series with, in this case, one realization every second, it is possible that the conditioned probability at one point of time depends on a vast set of previous wind conditions $(u_i(r_i))_{i<N}$ at the places in space $(r_i)_{i<N}$. In this

case

$$p(u_1, r_1|u_2, r_2; ...; u_N, r_n) = \frac{p(u_1, r_1; u_2, r_2; ...; u_N, r_n)}{p(u_2, r_2; ...; u_N, r_n)} \tag{2.15}$$

is the conditioned probability for one data point of the time series. In case the time series has Markov probabilities, the conditioned probability is only dependent on the previous time step. In this instance you have

$$p(u_1, r_1|u_2, r_2; ....; u_N, r_N) = p(u_1, r_1|u_2, r_2). \tag{2.16}$$

This directly indicates that the probability $p(u_1, r_1|u_2, r_2; ....; u_N, r_N)$ can be written as

$$p(u_1, r_1|u_2, r_2; ....; u_N, r_N) = p(u_1, r_1|u_2, r_2) = \prod_{i=1}^{N-1} p(u_i, r_i|u_{i+1}, r_{i+1}). \tag{2.17}$$

To check whether the time series is Markovian, one examines the condition

$$p(u_1, r_1|u_2, r_2; u_3, r_3) = p(u_1, r_1|u_2, r_2). \tag{2.18}$$

We assume that the effect of a single event decreases exponentially in time. Therefore, time steps previous to $t = 3$ are expected to have a weaker influence on the present states.

The second step is to check whether the time series is stationary. In this case, we do not assume strict stationarity, but the so-called second-order stationarity. This is given if the following three properties do not depend on the times t_1 and T:

- the mean value $\mu_{t_1,\tau} = \langle u_{t_1} \rangle_\tau$
- the variance $\delta_{t_1,\tau} = \langle u_{t_1} - \mu_{t_1,\tau} \rangle_\tau^2$
- the autocorrelation $\rho_{t_1,T,\tau} = \frac{\langle (u_{t_1} - \mu_{t_1,\tau})(u_{t_1+T} - \mu_{t_1+T,\tau}) \rangle_\tau^2}{\delta_{t_1,\tau}\delta_{t_1+T,\tau}}$.

Here, t_1 is the starting point of the part of the time series that is taken into consideration, τ is the time span over which the mean value is calculated and T a time lag. There are several tests that are able to show whether these central properties of the time series stay constant in time. This is commonly investigated by the Priestley-Subba Rao test [12] focussing on the Fourier spectrum of the time series and the Wavelet Spectrum test which is based on a paper by R. von Sachs *et al.* [14].

A short overview over these techniques of testing the stationarity of time series will be given here. The Priestley-Subba Rao test is based on the idea that stationary time

series are expected to show a constant spectrum in the Fourier space. For this reason, the evolutionary spectral density function at time t $f_t(\omega)$ is defined by giving a measure of the change of the Fourier spectrum according to ω. Thus, the quantity

$$\hat{Y}(t,\omega) = \log(f_t(\omega)) \tag{2.19}$$

is defined. Here, the advantage of looking at the logarithm is that the influence of the variance can be neglected. For a sufficiently long period of time the estimator $\hat{Y}(t,\omega)$ is assumed to converge against $Y(t,\omega)$.

The Wavelet test is centred around coefficients $\alpha_{j,k,j',k'}$ that are mainly determined by the two halves of a segment of the time-frequency plain of the time series which are compared to each other. If the spectrum of the time series does change in time, at least one coefficient significantly differs from 0. In this case, the hypothesis of a stationary time series is rejected.

The third step is to check whether the fourth conditional moment $D^{(4)}$ of the time series equals 0. The meaning of conditional moments is explained in detail in section 2.5. This is important, because the Pawula theorem says that the Kramers–Moyal expansion of the Fokker-Plank equation can be cut off after the first or the second term. In all other cases, the number of terms necessary for a proper description is infinite. For this reason, the fourth conditional moment is calculated. If this moment is not close to 0 for a time series, it is clear that the Fokker-Plank equation cannot be transformed into the Langevin equation. Here, the fourth moment is chosen, because only positive values are summed up, while for the third moment negative values are possible. For this reason, $D^{(3)}$ can go to zero without all $(X(t) - X(t+\tau))^3$ being close to zero.

2.5 Basics of stochastic data analysis

As long as the conditions mentioned in the previous section hold true, the time series of the derivative of the measured variable X can be split into two parts. The first part is assumed to be the deterministic behaviour of the system in the current state $X(t)$. The second part takes the influence of a Gaussian noise $\Gamma(t)$ function on the next step into account. This process is described by the Langevin equation

$$\frac{dX}{dt} = D^{(1)}(X) + \sqrt{D^{(2)}(X)}\Gamma(t). \tag{2.20}$$

After using the Euler integration, the next step of the time series is given by

$$X(t+\tau) = X(t) + D^{(1)}(X) + \sqrt{D^{(2)}(X)}\eta(t) \tag{2.21}$$

Here, τ is a sufficiently small time step and η is a normally distributed random variable. The coefficients of this stochastic differential equation, $D^{(1)}(X)$ and $D^{(2)}(X)$ are called the drift and diffusion coefficient.

Next, the two-dimensional Langevin equation is presented. Here, the drift coefficient, the Gaussian noise and the stochastic variables are two-dimensional vectors, and the diffusion coefficient is a 2 x 2 matrix. The differential form is then given by

$$\begin{bmatrix} d(X_1(t)) \\ d(X_2(t)) \end{bmatrix} = \begin{bmatrix} D_1(X_1, X_2) \\ D_2(X_1, X_2) \end{bmatrix} + \begin{bmatrix} g_{1,1}(X_1, X_2) & g_{1,2}(X_1, X_2) \\ g_{2,1}(X_1, X_2) & g_{2,2}(X_1, X_2) \end{bmatrix} \begin{bmatrix} \Gamma_1(t) \\ \Gamma_2(t) \end{bmatrix}. \tag{2.22}$$

It is generally expected that $g_{2,1}(X_1, X_2) = g_{1,2}(X_1, X_2)$ holds true. For this reason, five parameters, two drift coefficients and three diffusion coefficients, are to be calculated.

The drift and diffusion coefficients are the first and second Kramers-Moyal coefficients. These are defined by

$$D^{(n)}(X) = \lim_{\tau \to 0} \frac{1}{n!\tau} M^{(n)}(X, \tau). \tag{2.23}$$

Here, $M^{(n)}$ is the n-th normalized moment. This is given by

$$M^{(n)}(x, \tau) = \langle (X(t) - X(t+\tau))^n \rangle |_{X(t)=x}. \tag{2.24}$$

This definition of the coefficients is sensible, because for drift coefficient $D^{(1)}(X)$ the mean value of the Gaussian white noise is by definition zero, so that the mean value for sufficiently long periods of measurement is expected to converge against the intrinsic behaviour of the system at a given point. For the calculation of the variance of $X(t)$ it is expected that the intrinsic behaviour of the system does not vary in time. For this reason, the variance of the data is exclusively determined by the white noise present in the data. The variance of the general white noise $\Gamma(t)$ is rescaled to the noise present in the data by calculating the square root of the second moment, so that

$$\langle (X(t) - X(t+\tau))^2 \rangle = \langle (\sqrt{D^{(2)}(X)}\Gamma(t))^2 \rangle = D^{(2)}(X)$$

holds true.

A problem in the practical calculation of the drift and diffusion coefficient for mea-

sured data is that measured data is always given by discrete values with a sampling frequency $f_S > 0$. For this reason, the coefficients can only be approximated. Thus, the choice of τ is restricted by the frequency of measurement f_m. This leads to problems if the smallest possible lag in the data cannot be assumed to be small enough. The costly calculation according to the determinations of the conditioned moment is implemented in the Langevin R package presented in P. Rinn *et al.* [13]. A theoretical overview of the applications of the Langevin equation in physics is given in [9].

Even if one or more of the conditions mentioned in section 2.4 are not fulfilled, the analysis of the data using this type of approach may provide an indication of the statistical nature of the dataset, even though from a mathematical point of view, it is not fully described by the results.

3 Data preprocessing

In this section, the origin of the measured data and the procedure of data preprocessing will be presented. For this reason, obvious errors are mentioned. These errors will be removed from the dataset. At the end of the section, a short overview over the data will be given.

As mentioned in chapter 1, the data was measured at the FINO1 platform about 35 km north of Borkum. A picture of the platform along with its measuring instruments and their precise location is shown in figure 3.1. At the FINO1 platform, several properties are measured. Here, we will focus on the wind velocity measured by cup anemometers as well as the corresponding wind direction given by the vane and the ultra-sonic anemometer (USA). A short overview over the variables and the measurement frequencies that are used in this thesis can be seen in table 3.1. Here, u is the wind speed and α is the wind direction. The data set covers 20 months, starting in September 2015 and ending in April 2017.

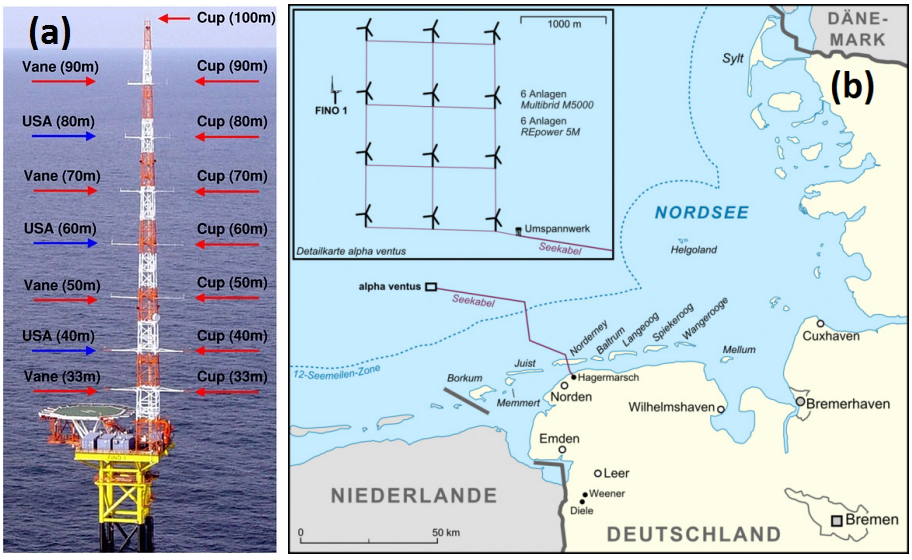

Figure 3.1: In graphic (a), the FINO1 platform with the heights at which the gauges relevant for this thesis are positioned is shown. Mostly, the data measured by the cup-anemometers on the left of the tower is used. In graphic (b), the location of the FINO1 tower is shown on a map. Source: © *Forschungs- und Entwicklungszentrum Fachhochschule Kiel GmbH.*

h in m	property	gauge	frequency of measurement in Hz
100	u	cup anemometer	1
90	u	cup anemometer	1
	α	vane	1
80	u	cup anemometer	1
	u and α	USA	20
70	u	cup anemometer	1
	α	vane	1
60	u	cup anemometer	1
	u and α	USA	20
50	u	cup anemometer	1
	α	vane	1
40	u	cup anemometer	1
	u and α	USA	20
30	u	cup anemometer	1
	α	vane	1

Table 3.1: Measured properties, gauges and frequencies at the FINO1 platform.

Moreover, the temperature, the air pressure, the relative humidity, the rain fall and the radiation at several heights are available. As these measurements are not used in this thesis, the specific positions and gauges are not listed here.

Next, a short explanation of the measuring devices and how they work will be given. A cup anemometer normally consists of four hollow hemispheres that are placed at the end of a short pole with an angle difference of 90 degrees to their next neighbours. The cup anemometer starts spinning in case of air fluctuations because of the force differences at the opposite sides of the anemometer resulting from the air drag difference of the outer and inner side of the hemispheres. On the basis of the spinning frequency, the wind velocity can be computed. The cup anemometer is a well-known and probably the most common device for the simple measurement of wind speeds.

A wind vane is a gauge that measures the wind direction. It is made of a flat metal plate of asymmetric shape that is attached to the end of a pole. A force difference results on the opposite sides of the plate if it is not parallel to the wind direction. Because of the fixture, the force difference is converted into a torque.

A USA calculates the wind velocity parallel to one direction by measuring the runtime of an ultrasonic signal over a known distance. For the measurement of wind velocity in a plain it is most common to determine this property in two orthogonal directions of measurement. This way the absolute value of wind velocity and the direction can be quantified.

The next step is to filter out all events that are obviously results of measurement

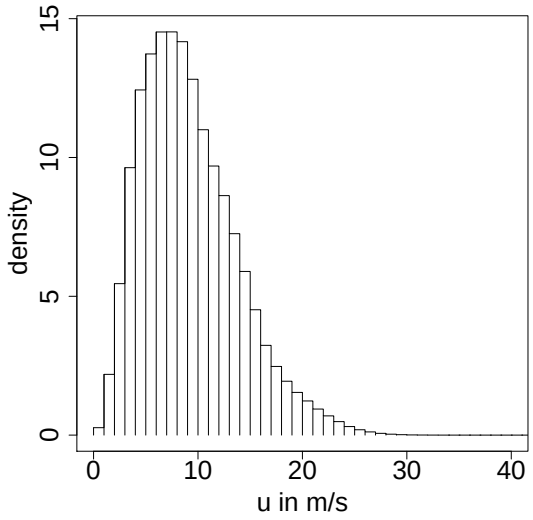

Figure 3.2: Histogram of wind speed data of all heights combined between September 2015 and April 2017.

errors or are not reliable for other reasons.

Firstly, it is chosen to take the data for the height of $100\,\mathrm{m}$ out of account. The data at this height show different properties in all contexts that are investigated. The reason for this might be that the data is measured at the top of the FINO1 tower and, thus, the gauge is not fixed in a similar way as the other cup anemometers. This can be seen in figure 3.1. Furthermore, no data set giving the wind direction is available for this height, so that part of the analysis carried out in this thesis is not possible. Secondly, a period of several days in January 2016 during which all cup anemometers were frozen are removed from the dataset. For this reason, all of the data points were measured to be $0.19\,\mathrm{m\,s^{-1}}$. In total, about 12 % of all data of this month shows that value. The values of these events where changed to NaNs, so that they do not affect the further analysis of the time series.

It is decided not to replace the NaNs in the time series by the mean value of the neighbouring data points. The reason for this decision is twofold: On the one hand, NaNs are results of an incorrect measurement during a period of time. It is generally not clear, whether the values measured before and after the NaN are affected by this fault either. On the other hand, the percentage of NaNs within the time series is about 0.8%. Thus, the rest of the time series is sufficiently large to neglect the NaNs.

At the end of this section, the distribution of the whole dataset is shown without the artefacts described in the upper part of this section. In figure 3.2, a histogram of the wind speed is presented.

4 Statistical analysis of single time series

The aim of this section is to find an accurate statistical description of the FINO1 data as well as its limits.

4.1 The wind speed

The first step is to find a fit for the total distribution of wind speed and afterwards to look at every height and month separately.

In figure 4.1 (a), the distributions of the wind speed data are shown in one graphic to illustrate the effect that the height has on the distributions. The dots show the average value of the wind velocity at particular heights. It can be seen that the mean value increases with the height.

The next step is to find a distribution that fits to the data. According to [5], this is best done by the Weibull distribution. Especially for long time scales, the data is assumed to fit well to the expectation of a λ correlating with the average wind speed and a k value between 1.5 and 2.5 (See [5]). In figure 4.1 (b), the distribution of the wind speed data for the whole dataset at 90 meters and the Weibull fit is shown.

In general, a good agreement of the data and the Weibull distribution can be observed for distributions over the whole period of measurement. The parameters of these fits along with their errors are seen in figure 4.2. The shape and scale parameters are seen to be functions of the height. These are given by

$$\lambda = 7.35 + 0.52 \cdot \log(h)$$

and

$$k = 2.18 - 2.6 \cdot 10^{-3} \cdot h.$$

In figure 4.2, it can also be observed that the values of k seem to decrease with the height while λ slowly increases. The behaviour of λ is in good agreement with figure 4.1 (a). Furthermore, the assumption that the mean wind speed can be modelled by a logarithmic wind profile is indirectly supported, because the scale parameter λ is closely

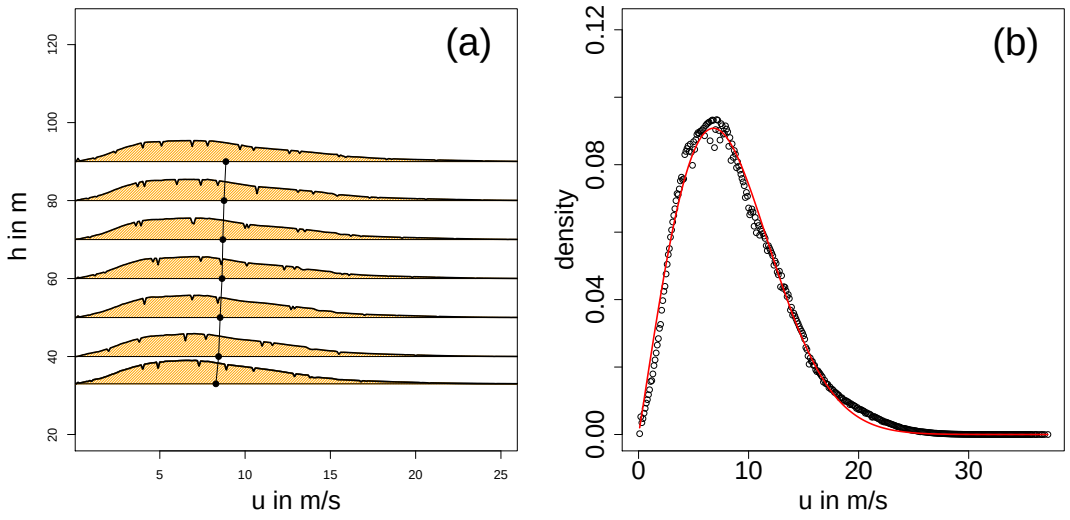

Figure 4.1: (a): Histogram of the wind speed measured by the cup anemometers at heights of 33 to 90 meters. (b): Wind speed histogram at a height of 90 m for the full data period used in this thesis and the Weibull distribution. As generally expected, a good agreement to the Weibull distribution is found for such a long period of time.

related to the mean wind speed. It has to be mentioned that the dependence of k on the height is of a small magnitude in absolute numbers. Taking into account that k is an exponent, this has an influence on the distribution that cannot be neglected.

The distributions of the data sets of one-month-periods do in some cases not show a perfect match with the Weibull distribution.

The shape parameter k and scale parameter λ are shown in figure 4.3. It was aimed to see a shape parameter close to 2 while the scale parameter was thought to show a regular behaviour according to the fact that, in general, higher wind speeds may be expected in winter. Apart from December 2015, the first assumption can be confirmed, while the second expectation cannot be answered clearly. It cannot be clearly confirmed or rejected, if an oscillation of the λ values occurs. For a more certain answer more values are needed. Apart from this, a recognizable movement of the scale coefficient around the mean value is seen. This is also expected for months with different weather conditions and, thus, a wider spread in the wind speed distribution.

For further decreasing time scales, a description of the data by the Weibull distribution becomes impossible. The problem describing the dataset of these short periods of time with the Weibull distribution is that data of single days or an even longer period of constant weather conditions cannot be averaged out. For this reason, in many data sets multi-modality is present. This cannot be modelled by using the Weibull distri-

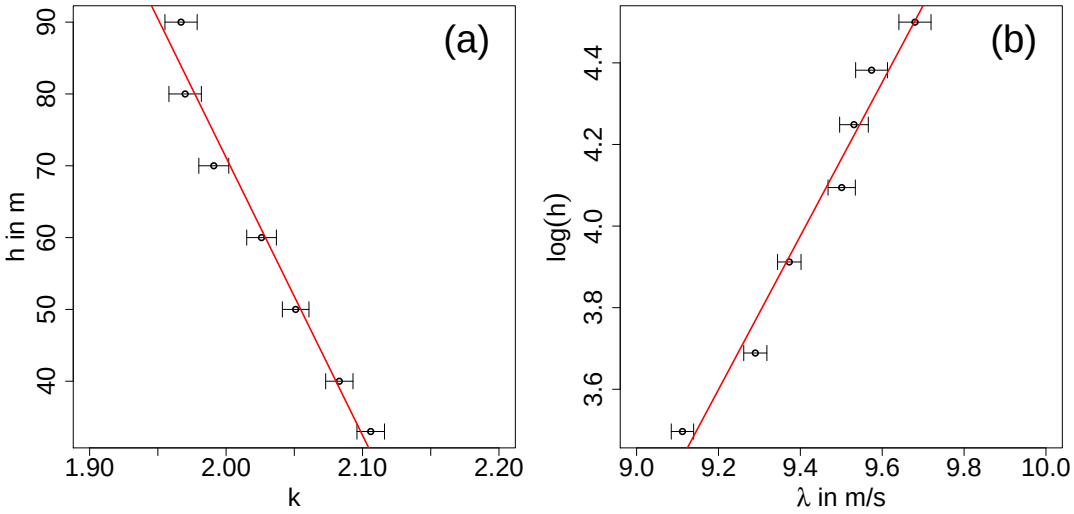

Figure 4.2: In graphic (a), the shape parameter k is shown as a function of the height. For k, a linearisation with respect to the height seems reasonable. In graphic (b), the scale parameter λ is plotted as a function of $\log(h)$. Equally well, the power law profile fits the data.

bution. The other considerable problem is that on these time scales conditions may show a behaviour that is not typical of wind data. This can be observed in the data of the second week in May 2016. Here, far more events were measured at $20\,\mathrm{m\,s^{-1}}$ than at $3\,\mathrm{m\,s^{-1}}$. Thus, the parameters of the Weibull fits have a low validity and, for this reason, will not be discussed. It may be an interesting field of further studies to find a fitting distribution for the measured data for these shorter time scales.

4.2 The wind direction

The idea behind filtering the wind data according to the wind direction is, firstly, to find the most common wind direction and, secondly, to filter out all possible effects the tower itself has on the distribution. In the data sets that were measured by the USA only the data points that were recorded parallel to the data of the cup anemometer and vane are taken into account, in order to get comparable data sets for every height. In general, it is chosen to split the wind directions into segments of $30°$. This guarantees that for all segments a sufficiently large dataset is at hand. For the analysis of the data in the preferred wind direction, the segment between $180°$ and $270°$ is divided into parts of $10°$. This is done, because in this segment enough events can be found to choose a more narrow division and because the effect of the tower on the measured data should be negligible.

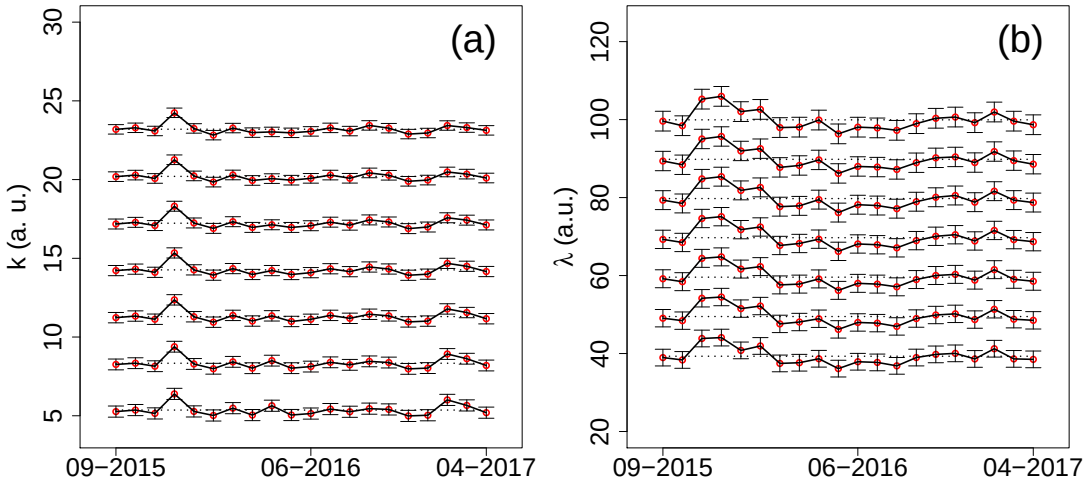

Figure 4.3: Shape (a) and scale (b) parameters of the Weibull distribution calculated separately for every month and height. For reasons of clarity, the data points are shifted arbitrarily upwards, so that they do not overlap each other. As a measure of the error, the standard deviations of the scale and shape parameter are taken separately for each height. The dotted lines mark the mean values of the parameters relative to the calculated data points for each month.

The mean wind direction and the mean wind speed are defined as

$$\bar{\alpha}(h) = \frac{1}{N} \sum \alpha_h, \tag{4.1}$$

and

$$\bar{v}_i = \frac{1}{N} \sum v_i, \tag{4.2}$$

where N is the number of data points in the according time series, α_h is the time series of the wind directions at one height and v_i is the wind speed time series in one segment of the wind directions.

Beside the mean wind direction, we consider the mean wind speed parallel to the x- and y-coordinate, defined as

$$\bar{v}_x(h) = \frac{1}{N} \sum v_h \sin(\alpha_h) \tag{4.3a}$$

and

$$\bar{v}_y(h) = \frac{1}{N} \sum v_h \cos(\alpha_h), \tag{4.3b}$$

where v_h is the wind speed time series at height h, α_h is time series of the wind

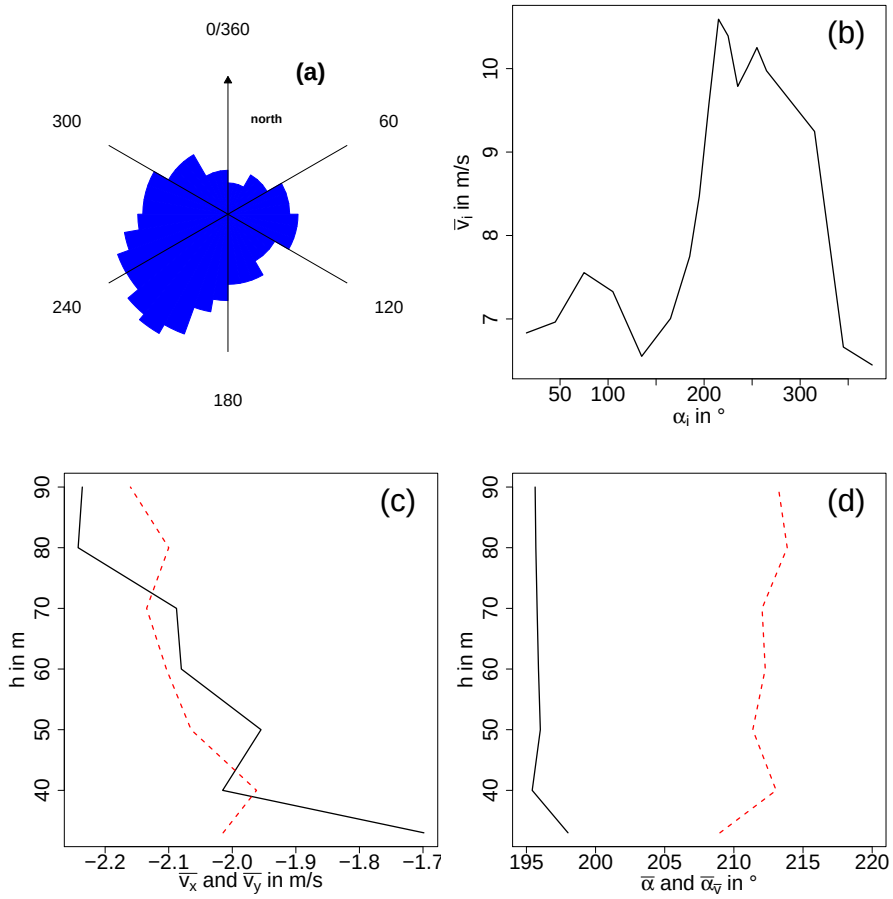

Figure 4.4: (a) Distribution of the wind speed according to the direction. (b) Mean wind speed in the different segments of the wind direction. (c) Mean wind speed parallel to the x- (solid line) and y-coordinate (dashed line) as a function of the height. (d) The mean wind direction (black solid line) and direction of mean wind speed (red dashed line) are shown as a function of the height.

direction at height h and N again is the length of the time series. The mean direction of the wind speed is defined as

$$\bar{\alpha}_{\bar{v}}(h) = \arctan\left(\frac{\bar{v}_x(h)}{\bar{v}_y(h)}\right). \tag{4.4}$$

By this calculation, not only the frequency of occurrence of the angle, but also the measured wind speed is taken into account. In figure 4.4, the distribution of the wind speed (a), the mean wind speed in the segments (b), the mean wind speed parallel to the x- and y-coordinate (c) and the mean wind direction and the direction of the mean wind speed (d) are presented.

In 4.4 (a), it is seen that the distribution is asymmetric. For this reason, a preferred

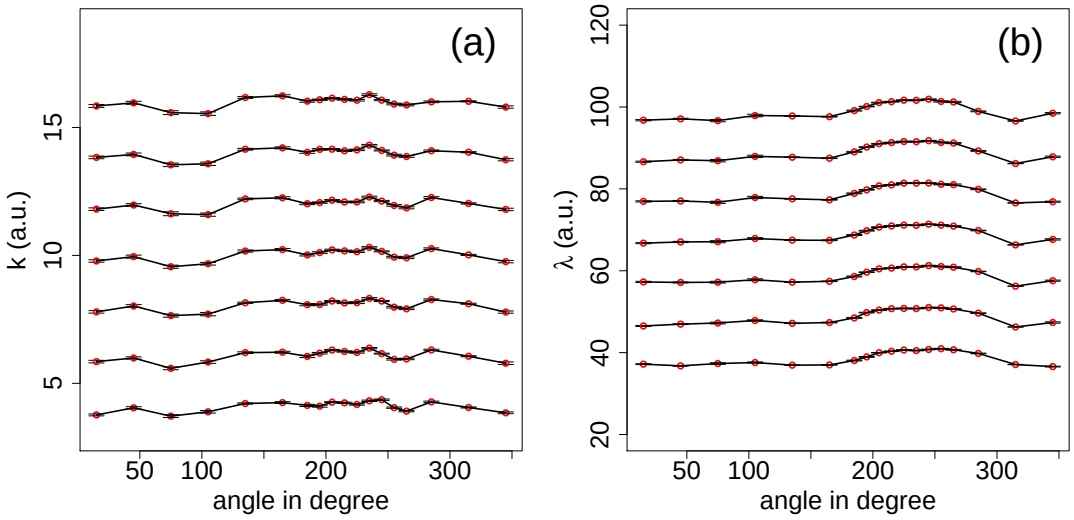

Figure 4.5: (a) The shape and (b) the scale parameters of the Weibull distribution for every height and direction segment separately. The values are arbitrarily shifted up, so that the curves do not overlap. The curves show the coefficients for a height of 33 m to a height of 90 m at the top.

wind direction exists. The average over the preferred wind direction for all heights is $\bar{\alpha} = 196^\circ \pm 0.9^\circ$. In figure 4.4 (b), it can be noticed that the mean wind speed seriously varies according to the wind direction. The maximum of $\bar{v}_i$ is slightly shifted compared to the the mean wind direction $\bar{\alpha}$. In figure 4.4 (c) it is shown that the magnitude of the mean wind speed generally increases. A slight shift of the direction of the mean wind speed can be observed. In figure 4.4 (d), it can be seen that the direction of the $\bar{\alpha}_{\bar{v}}(h)$ is directed towards the middle of the segment that was analysed more precisely. The mean value of the $\bar{\alpha}_{\bar{v}}(h)$ is given by $\bar{\alpha}_{\bar{v}} = 212^\circ \pm 2^\circ$. Thus, the mean wind direction and the direction of the mean wind speed differ from each other by about 15°. The angle difference is a result of the shift of the maximum of the mean wind speed.

Next, it is being verified whether the Weibull parameters are a function of the wind direction. This is done by fitting the Weibull distribution to the histogram of the wind speed at every height and segment separately. The results are shown in figure 4.5.

While the shape parameter stays more or less the same in every direction (figure 4.5 (a)), the scale parameter shows to be a function of the direction (figure 4.5 (b)). Approximately in the mean wind direction, at around 180° and 300°, the scale parameter reaches its maximum level with values around $10.5\,\mathrm{m\,s^{-1}}$, while in the segment between 0° and 90° values around $7\,\mathrm{m\,s^{-1}}$ can be observed. This may partly be an effect of the shielding of the gauges by the tower.

5 Time increment analysis

In this chapter, the time increments of the FINO1 data are analysed firstly using the Laplace distribution in section 5.1 and, afterwards, in section 5.2, the Castaing distribution which was presented by B. Castaing *et al.* [6]. This way, the rather simple and easy to interpret Laplace distribution is compared to the more sophisticated Castaing distribution that is generally thought to lead to a better fit of the observed increments, especially for short lags.

5.1 Laplace distribution

In this thesis, lags of 2^n s with $n = 0, ..., 12$ are chosen, because by using these lags, the short-time behaviour can be observed with sufficient accuracy, while also for larger time scales, an image of the principle behaviour can be gained. Firstly, the distribution of the time increments $\Delta_{1\,\mathrm{s}}u(t)$, $\Delta_{8\,\mathrm{s}}u(t)$, $\Delta_{64\,\mathrm{s}}u(t)$, $\Delta_{512\,\mathrm{s}}u(t)$ and $\Delta_{4096\,\mathrm{s}}u(t)$ as defined in (2.8) are shown in figure 5.1 (b). Hence, an idea of the general distribution of the increments at different time scales can be gained. Along with this, the increments with a lag of $1\,\mathrm{s}$ are shown in a normal scale for every height in figure 5.1 (a).

It should be noted that the distribution of the increments looks more or less symmetrical. Plotting σ_+ as a function of σ_-, a linear fit yields a slope of $m = 0.96 \pm 0.03$ and an intercept of $a = 0.05 \pm 0.03$. This indicates that generally σ_+ is a bit smaller than the absolute value of σ_-.

Furthermore, the mean value of the distribution is always bigger than 0 and grows with longer lags. For a lag of $4096\,\mathrm{s}$, the mean value can be up to $0.014\,\mathrm{m\,s^{-1}}$ and is, hence, distinctly smaller than the resolution of most cup anemometers and the accuracy of the binning used. For this reason, it is possible to assume that μ equals 0.

For a first approach, the bins with less than approximately 200 events are not taken into account. Two points about this procedure can be seen critically: It is not clear whether only taking bins above an arbitrary threshold into account is sufficient to fully understand the statistical nature of the distribution. For example, the heavy tails for short lags and the clear asymmetry in the distribution for longer lags due to a few extremely high increments are not fully taken into account.

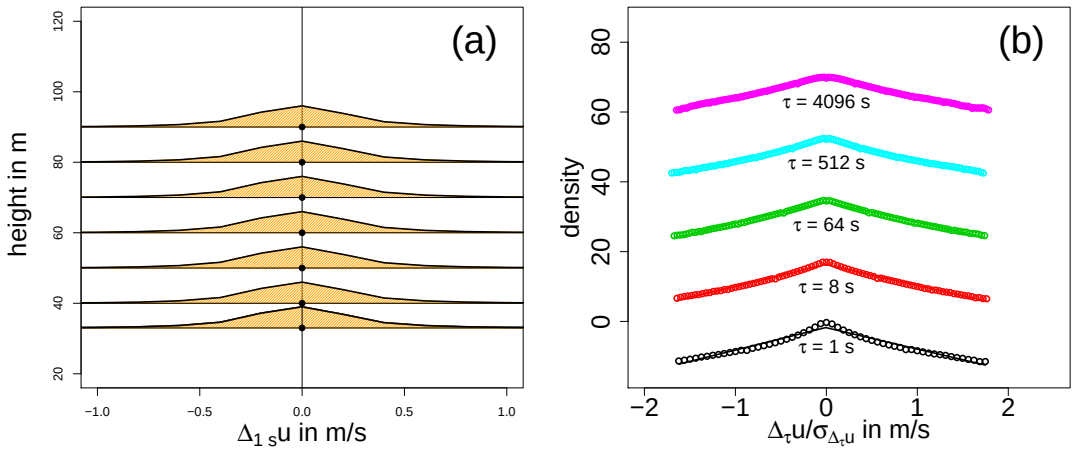

Figure 5.1: (a) Histograms of the increments for all data and a lag of 1 s with the mean value being marked by black dots. (b) Logarithmic density of increments at a height of 90 m and time lags of 1 s, 8 s, 64 s, 512 s and 4048 s.

Next, the properties of the scale parameters σ_+ and σ_- will be looked at. Therefore, the mean values of σ_+ and σ_- for all heights are plotted as a function of the logarithmic time lag in figure 5.2 (a). The average is taken, because the differences in magnitude of the Laplace coefficients according to the height are negligible. Here, a good agreement with the linear fits

$$\sigma_+ = 0.32 + 0.08 \cdot log(\tau)$$

and

$$\sigma_+ = -0.37 - 0.08 \cdot log(\tau)$$

can be found. All points apart from the first point are in good agreement with the fitted curve. In graphic (b) of figure 5.2, it can also be speculated about whether a seasonal dependence of the Laplace coefficients is present. It seems as if the Laplace coefficients have the tendency of being slightly higher in winter. In the end, the period of measurement is not long enough to confirm this proposition. In graphic (c) the Laplace coefficients calculated from the whole dataset at one height are plotted as a function of the height. A slight increase of the Laplace coefficients is seen. The tendency of a slight increase does not hold true for the top height of 90 m presented in this plot.

In the next step, it will be verified whether the Weibull coefficients and the Laplace coefficients show any dependence or correlation. The correlation coefficient is calculated

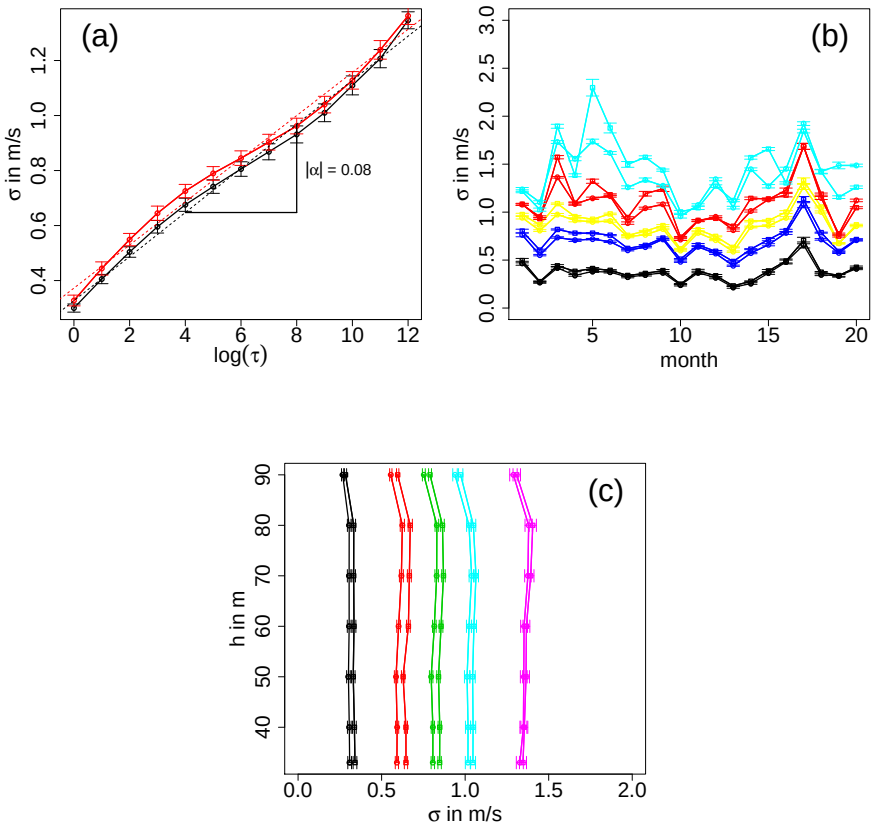

Figure 5.2: (a) Half-logarithmic plot of the Laplace coefficients σ_+ and σ_- as a function of the lag τ. The dotted lines indicate the linear fit of the Laplace coefficients. (b) Laplace coefficients σ_+ and σ_- for chosen lags as a function of the month of measurement and a height of 90 m. (c) Laplace coefficients as a function of the height. Here, the whole period of measurement was taken into account. In both, (b) and (c), lags of 1 s, 16 s, 256 s and 4096 s are shown.

by

$$\rho_{x,y} = \frac{(x - \langle x \rangle)\,(y - \langle y \rangle)}{\sigma_x \sigma_y}. \tag{5.1}$$

For this reason, the data measured in one month at one height is seen as one dataset. Then, for every lag used in this chapter, the correlation coefficients of the Laplace coefficients with the Weibull coefficients are calculated. In the case of a high correlation value, it can be assumed that for a high mean wind speed or a narrow spread of the data, bigger turbulences are present. The correlation coefficient is calculated separately for each lag. In figure 5.3 (a) and (b) the absolute values of the Laplace coefficients are plotted as a function of the Weibull coefficients. As the Laplace coefficients increase for growing lags, the points have a layer structure. In 5.3 (c) and (d) the correlation

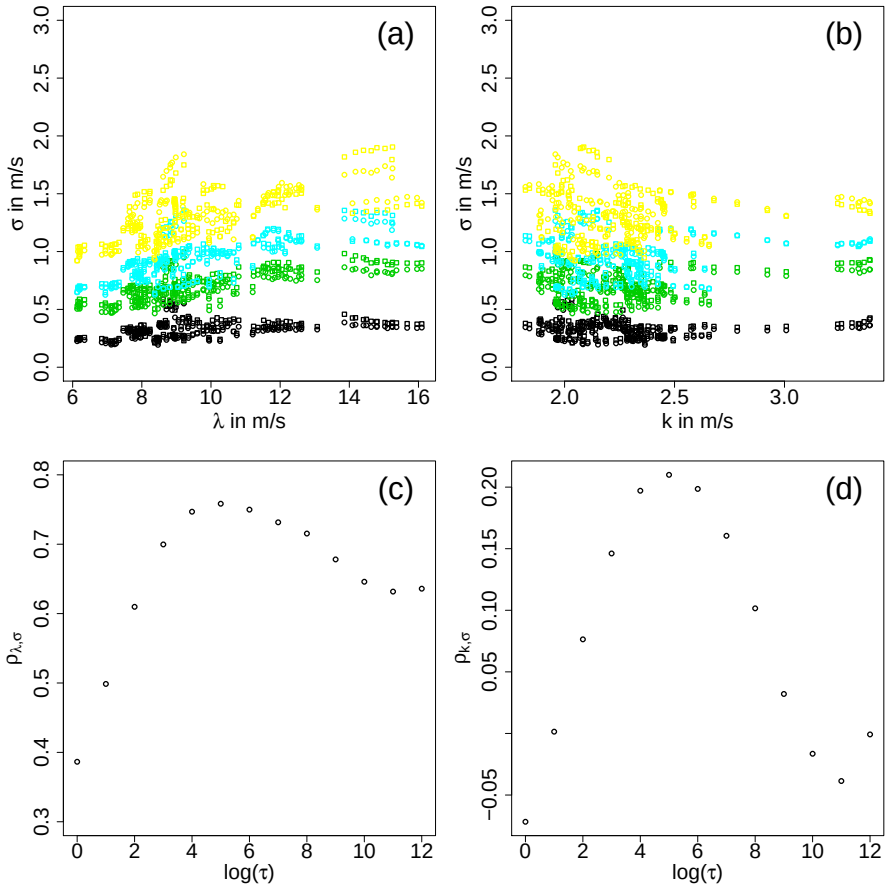

Figure 5.3: The Laplace coefficient $\sigma_{\pm}$ is in (a) plotted as a function of λ and (b) as a function of k. Below, the correlation coefficients $\rho_{\lambda,\sigma}$ are shown in (c) and $\rho_{k,\sigma}$ are shown in (d). While $\rho_{k,\sigma}$ stays at a relatively low level, $\rho_{\lambda,\sigma}$ here reaches a high level with a value of 0.8. $\rho_{\lambda,\sigma}$ is by far larger than $\rho_{k,\sigma}$ for lags taken into account.

coefficients of the Laplace coefficients and the Weibull coefficients are shown.

The most interesting point about figure 5.3 (c) is the level of the correlation coefficient $\rho_{\lambda,\sigma}$. The high level of correlation indicates that a high level of turbulence is more probable in the case of a high average wind velocity. Furthermore, also a small correlation of the Laplace coefficients with the shape parameter k is visible in figure 5.3 (d). As this correlation only exists on a very low scale and even the algebraic sign changes, it is not possible to deduce a general dependence between the scale parameter and the level of turbulence seen in a specific period of time. This is interesting, because it indicates that the shape of the distribution of wind speed has a smaller impact on the turbulences measured than the mean wind speed, even though the scale parameter determines the width of the distribution. Both correlation coefficients reach their maximum for a lag of about a half to a full minute

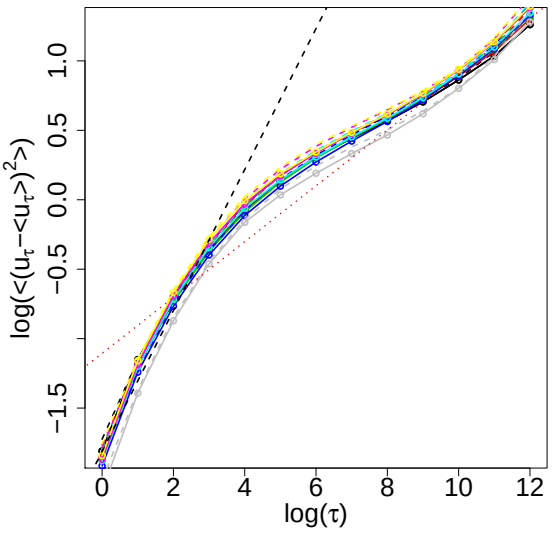

Figure 5.4: Double-logarithmic plot of the second centred moment of the time increments $\Delta_\tau u$ to τ is shown in this graphic. The data is linearised for smaller lags (up to about 8 s) and longer lags beginning from 9 min up to the maximum lag of about 68 min.

This statistic approach can be criticised, because only the measuring period of one month is looked at. The reason is that one month is the shortest period of time for which the data show a Weibull-like distribution.

Yet, the weather during a one-month period can be assumed to be rather constant compared to periods of half a year. Furthermore, the statistical significance of the correlation coefficient gets lower if few data points exist. Thus, the statistical significance of the correlation coefficient decreases if longer periods of time are taken into account.

After this, we calculate the variance of the increment time series. This should be connected to the turbulence, and thus, to the Laplace coefficients that quantify the magnitude of turbulence. To see whether this assumption is true, the logarithmic variance is plotted in a logarithmic time scale in figure 5.4.

The slope of the linearisation on short time scales is $a_{\text{short}} = 0.48 \pm 0.5$ and $a_{\text{long}} = 0.18 \pm 0.01$ on long time scale.

5.2 Castaing distribution

In this section, the Castaing distribution is used to describe the time increments. The Castaing distribution was presented in section 2.3. The goodness of the Castaing

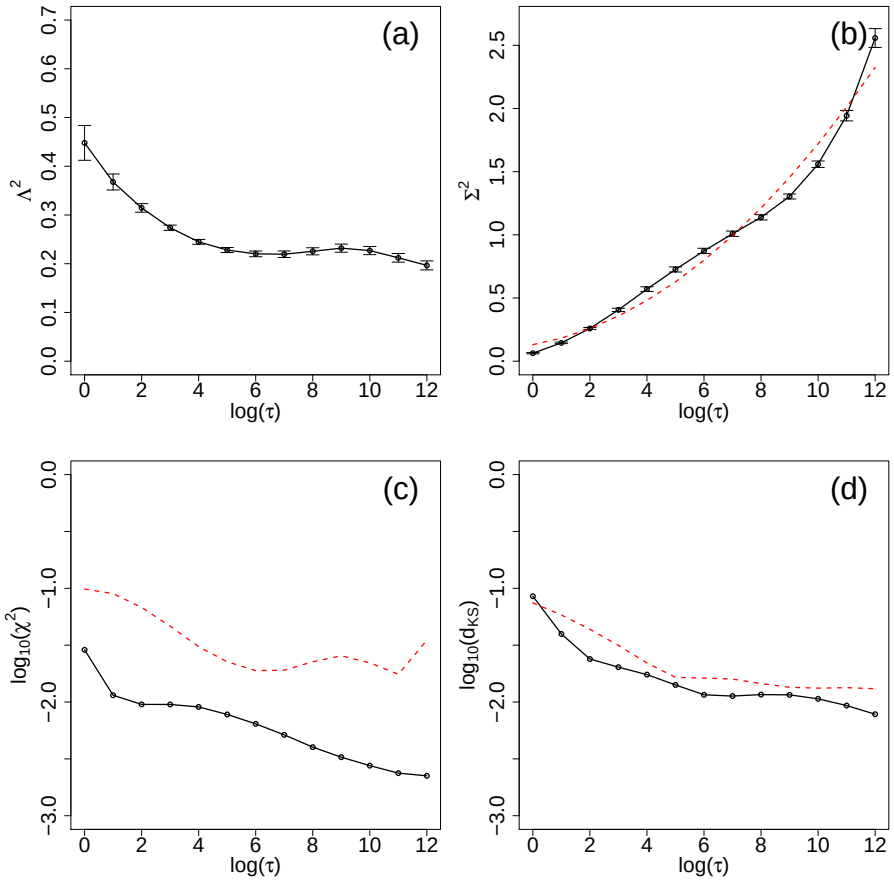

Figure 5.5: Castaing coefficients (a) Λ and (b) Σ in a logarithmic time. In (a) a decreasing tendency can be observed, which is connected to a less significant cusp, while in (b) an approximately linear increase can be seen. This is due to the growing spread of the distribution. In (c) and (d) the magnitude of χ^2 and d_{KS} can be seen. Here, the solid black line shows the errors for the Castaing distribution and the scattered red line the errors for the Laplace distribution.

distribution and the Laplace distribution are compared. For the evaluation of the goodness of the fit, the χ^2 and Kolmogorov-Smirnov test are used. They were presented in the equations (2.6) and (2.5). For the calculation of these errors, the tenfold standard deviation is considered. In this area, at least 98.6 % of the events can be found. The fact that a relevant percentage of the events is not located in this vast region, is a clear indicator for the presence of heavy tails in the distribution. Nevertheless, the vast majority of more than 99 % of the data is inside this region.

Firstly, the Castaing distribution is fitted to the data of the whole measurement period at every height separately. The average of the values given back for the coefficients Σ^2 and Λ^2 from 33 m to 90 m in dependence on the logarithmic time can be seen in

figure 5.5 (a) and (b). The average is taken, because the coefficients are very close to each other. Along with this, the magnitude of the errors of the Castaing and Laplace distribution are shown in 5.5 (c) and (d).

In general, it can be seen that Λ^2 does not vary much and shows a decreasing tendency while Σ^2 constantly increases. A small value of Σ^2 is connected to the shape of the cusp. For low values of Σ, the cusp becomes significant, while for higher values of Σ, the cusp gets rounder. With a value of Σ given and, thus, the shape of the cusp, Λ^2 varies the spread of the distribution. The higher the value of Λ, the wider are the tails of the distribution. Nevertheless, it cannot be stated that for a high value of Λ the distribution is always high, because the value of Σ also affects the width significantly. For this reason, it can be concluded that the cusp of the distribution gets rounder for longer lags. This for itself has the effect that the distribution gets wider. For the cusp which induces Σ^2, the distribution fits best with Λ^2 values that do not vary much. This is not paradox, given that the distribution shows a larger spread for longer lags, because the meaning of the Castaing coefficients is not as clear as the meaning of the Laplace coefficients.

In both cases, the Castaing coefficients Λ^2 (5.5 (a)) and Σ^2 (5.5 (b)) seem as if they could be approximated by a polynomial of the third degree. Nevertheless, the deviation from a quadratic polynomial is so low in the case of Σ^2 that it seems sufficient to use a polynomial of the second degree to provide a clue of the value of Σ^2. Due to the quadratic fit shown in figure 5.5 (b), Σ^2 is given by

$$\Sigma^2 = 0.13(\log(\tau))^2 + 0.012\log(\tau) + 0.04.$$

It can be seen that the χ^2 errors of the Laplace distribution are in nearly all cases 1 to 2 magnitudes bigger than the errors of the Castaing distribution, so that the Castaing distribution is to be regarded as the better fit. For the Kolmogorov-Smirnov error, the Castaing distribution is still better in most cases. The difference between these two kinds of errors indicates that the maximum deviation from the measured data is comparable for both models. Especially for the tails of the distributions, the Castaing distribution fits better to the data than the Laplace distribution. Nevertheless, this does not mean that showing the Laplace distribution is of no use, because it is a good marker for the spread of the distribution. The spread is not provided by the coefficients of the Castaing distribution in a comparable clarity as the above discussion shows.

Next, the Castaing coefficients for every single month are looked at. These should show whether the behaviour of the coefficients is similar for all lags. Furthermore, special peaks might be observed. The errors can provide an idea whether Castaing is always the better fitting distribution or if the better agreement of the measured

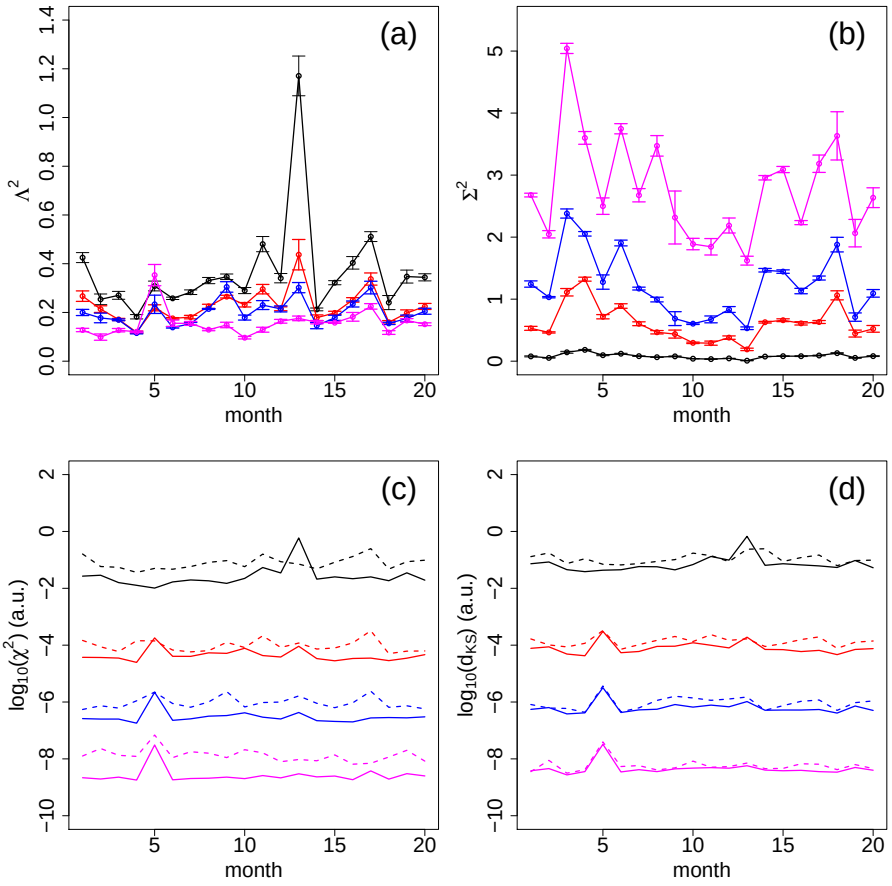

Figure 5.6: Castaing coefficients (a) Λ^2 and (b) Σ^2 in dependence on the month they were measured. The Castaing coefficients for the distribution of the increments with τ = 1s (black) , 16s (red) , 256s (blue) and 4048s (violet) were chosen to be shown. In (c) and (d), the magnitude of χ^2 and d_{KS} can be seen as well as the errors for the Castaing distribution (solid lines) and the Laplace distribution (dashed line).

distribution and fit only results from a single month with special conditions. The Casting coefficients are shown in figure 5.6 (a) for Λ^2 and (b) for Σ^2. In 5.6 (c) χ^2 and in (d) d_{KS} are shown in a half logarithmic plot.

Again, the coefficients are a function of the length of the time lags can especially be observed for Σ^2, while the Λ^2 are slightly decreasing. Just as it is seen in figure 5.5, the decreasing tendency of the Λ^2 is not strict. Furthermore, the coefficients show the same tendency for one month but different time lags. Here, the variation is increasing in height for larger time lags.

Apart from some exceptions, the errors of the Castaing distributions are about one order of magnitude smaller than the errors of the Laplace distributions regarding the χ^2 errors. Here a characteristic pattern can also be observed. As shown in figure 5.5, the

d_{KS} errors are only slightly better for most Castaing distributions and are in several points nearly exactly the same. This means that the centre region of the distribution is reproduced comparatively well by applying the Laplace distribution, while at the tails of the distribution the Castaing model fits better.

All in all, the Castaing distribution is a better fitting distribution for nearly all months. Only one example of a month and a lag can be found, where the Laplace distribution fits slightly better.

6 Spatial increment analysis

In this chapter, the spatial increments are being investigated. Thus, two approaches are used. The spatial increments for a reference height $\delta_{\text{ref}}u_h$ and the increments between two neighbouring heights δ_{next} are defined by

$$\delta_{\text{ref}}u_h = u_h - u_{h_{90\,\text{m}}} - (\bar{u}_h - \bar{u}_{90\,\text{m}}), \tag{6.1a}$$

$$\delta_{\text{next}}u_h = u_h - u_{h+10} - (\bar{u}_h - \bar{u}_{h+10}). \tag{6.1b}$$

As the reference height $h_{\text{ref}} = 90\,\text{m}$ is chosen. The half-logarithmic plot of the distributions of both kinds of spatial increments along with the Castaing distribution fitted to the data are shown in figure 6.1 (a) and (b).

Here, the Castaing distribution is chosen from the beginning, because the non-linearities even for larger spatial lags seem more distinct than they were observed for the time increments. To verify whether this choice is correct, the errors of the Castaing and Laplace distribution are quantified and compared. The Normal distribution

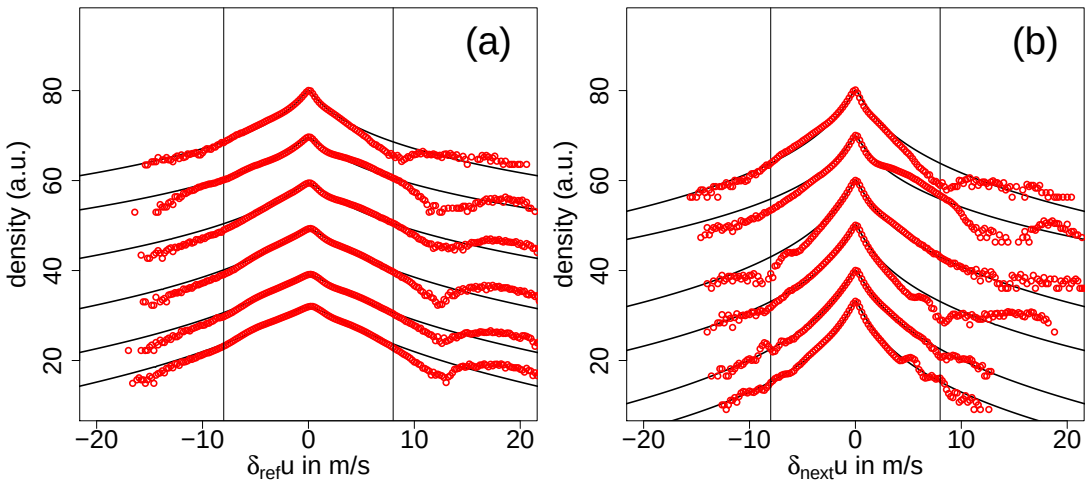

Figure 6.1: Histogram of height increments $\delta_{\text{ref}}u$ (a) and $\delta_{\text{next}}u$ (b) (red dots) in a half-logarithmic plot for every height along with a Castaing distribution (black line) fitted to the distribution of the measured data.

χ^2	h in m	Castaing	Laplace	Gauss	d_{KS}	h in m	Castaing	Laplace	Gauss
δ_{ref}	80	0.008	0.060	0.383	δ_{ref}	80	0.031	0.110	0.205
	70	0.018	0.092	0.398		70	0.044	0.132	0.213
	60	0.011	0.046	0.274		60	0.036	0.113	0.180
	50	0.013	0.018	0.192		50	0.038	0.039	0.147
	40	0.014	0.014	0.167		40	0.033	0.034	0.139
	33	0.016	0.018	0.141		33	0.050	0.083	0.120
δ_{next}	80	0.026	0.053	0.406	δ_{next}	80	0.025	0.055	0.197
	70	0.020	0.190	0.615		70	0.036	0.234	0.254
	60	0.004	0.039	0.328		60	0.029	0.093	0.186
	50	0.012	0.025	0.302		50	0.022	0.071	0.195
	40	0.006	0.022	0.268		40	0.036	0.043	0.163
	33	0.007	0.025	0.309		33	0.093	0.055	0.197

Table 6.1: The χ^2 and d_{KS} values are shown for the Castaing, Laplace and Normal distribution. The parameter is calculated by comparing the fit to the histogram of the spatial increments.

is also taken into the scope of the comparison of the goodness of the fit, because it is the most commonly used and the first approach for the description of many statistical phenomena.

For a comparison of the three models, the χ^2 and d_{KS} values for the distribution of the measured data and the three different fits are shown in table 6.1. The formulas for both methods for the quantification of errors of statistical models can be found in equations (2.6) and (2.5).

In table 6.1, it can be seen that the spatial increments are best described by the Castaing model. This result agrees with the results presented in C. Behnken's the master thesis [2] and is also quite similar to the observations made in 5.2. For this reason, the Castaing distribution will exclusively be discussed in the rest of this chapter. Nevertheless, the Laplace distribution has a comparable accuracy for large enough spatial increments.

As the next step, it is investigated whether the coefficients of the Castaing distribution are a function of the height for $\delta_{\text{next}}u$ and the magnitude of the distance for $\delta_{\text{ref}}u$. As mentioned in chapter 2, the Castaing model has two parameters that can be extracted out of the data analysed by simple statistical computations. Here, Σ^2 gives the median of the log-normal distribution and Λ^2 is connected to the intermittency of the distribution. For high values of λ^2 the Castaing distribution converges towards the Normal distribution. The Casting coefficients for $\delta_{\text{ref}}u$ and $\delta_{\text{next}}u$ are plotted as a function of the height in figure 6.2.

It can be seen that for $\delta_{\text{ref}}u$ a linearisation very well reproduces the values, while for

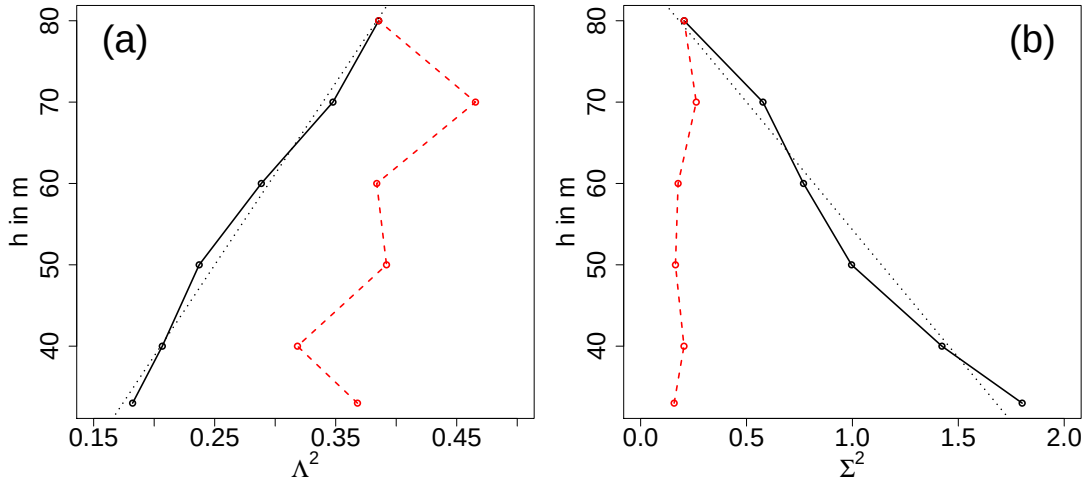

Figure 6.2: The parameters of the Castaing distribution Λ^2 (a) and Σ^2 (b) as a function of the height. Here, the lower of the two heights of δ_{next} is taken. The dashed red lines show the parameters of the distribution of δ_{next} and the solid black lines show the parameters of the δ_{ref}. The dotted black lines are linearisations of the parameters of δ_{ref}.

$\delta_{\text{next}}u$ no clear tendency according to the height can be observed. The linearisations for the Castaing coefficients as a function of the height are given by

$$\Lambda^2 = (2.6 + 0.4 \cdot h) \cdot 10^{-2}$$

and

$$\Sigma^2 = 2.73 - 0.03 \cdot h.$$

The next question is whether coefficients are vary in magnitude due to the season or whether special tendencies can be observed in certain months. For this reason, the spatial increment data is divided into months and is afterwards analysed again. The result is shown in figure 6.3.

Just like in figure 5.2 (b), it is not clear, whether a seasonal dependence can be extracted out of the data. For Λ^2, maxima can be observed in wintertime. It is interesting that the second maximum in November 2016 in graphic 6.3 (a) for the smallest increment cannot be observed for larger lags, while this is the case in January 2016. Furthermore, the Σ^2 coefficients for growing spatial difference show an increasing variance. In comparison to this trend, the coefficients for δu_{next} seem to stay relatively homogeneous.

Next, the Castaing and the Weibull coefficients are compared to each other. For this

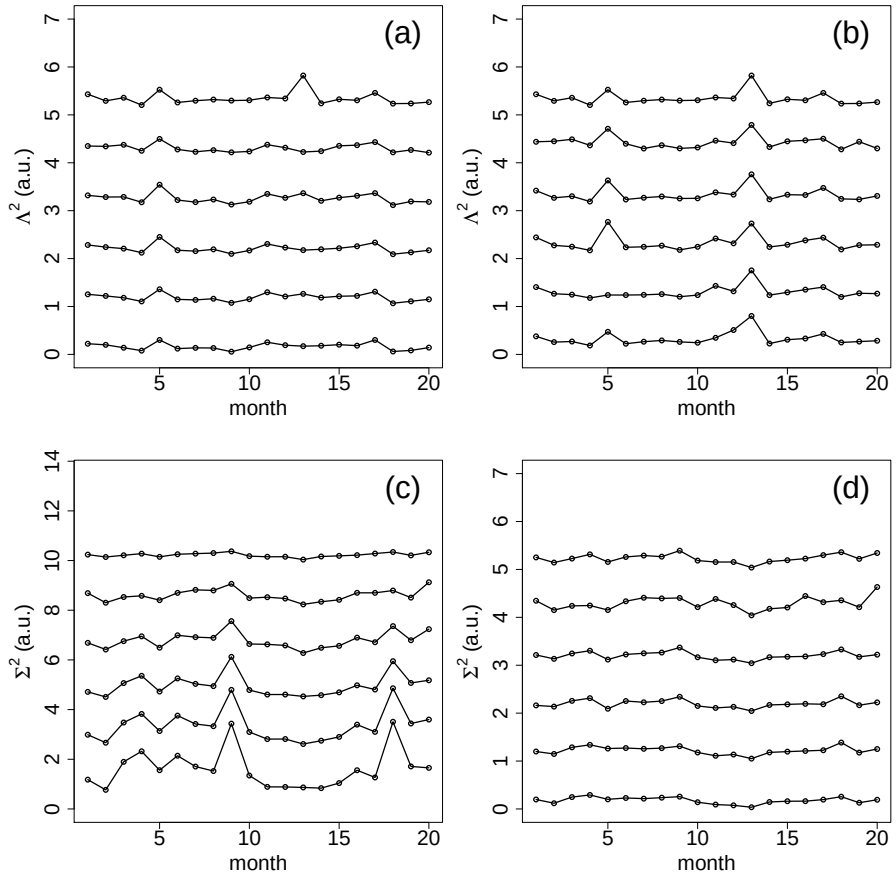

Figure 6.3: Coefficients Σ^2 and Λ^2 of the Castaing distribution for the spatial increments as a function of the month of measurement. Here, the figures (a) and (c) are the coefficients of δu_{ref} and (b) and (d) the coefficients of δu_{next}. Λ^2 and Σ^2 are displayed in arbitrary units for a clearer arrangement.

reason, firstly the Castaing coefficients for the whole dataset of 20 months are plotted as a function of the Weibull coefficients. This is shown in graphic 6.4.

In figure 6.4, it can be seen that Λ^2 is a linear function of the scale coefficient k (figure 6.4 (c)) and Σ^2 is a linear function of both Weibull coefficients (figure 6.4 (b) and (d)). It seems as if for figure 6.4 (a) a quadratic fit might reproduce the data better.

To verify whether these findings hold true on other time scales, it is investigated if this linear dependence can also be found if shorter scales of time are considered. For this purpose, datasets of one month are chosen. On the one hand, this period of time is sufficiently long to assume the Weibull distribution as a proper fit. On the other hand, it is short enough to observe seasonal or other differences that result in a significant variation in the magnitude of the according parameters. Furthermore, the number of

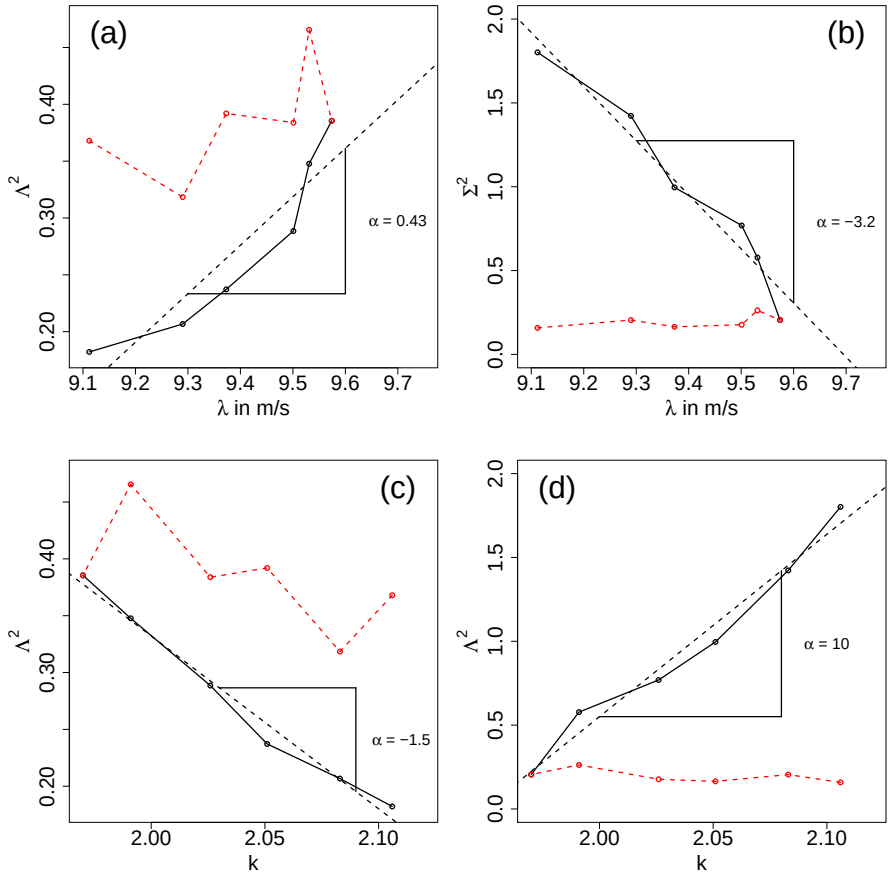

Figure 6.4: The Castaing coefficients are plotted as a function of the Weibull coefficients for the same height. The Casting coefficients for δ_{ref} are plotted with a solid black and for δ_{next} as a dashed red line. The linearisation of the coefficients for δ_{ref} can be seen as a dashed black line.

points is sufficiently high to assume statistical approaches as being sensible. The results are shown in figure 6.5.

A slight linear dependence of Σ^2 is found for both Weibull parameters. The slope is relatively constant for all datasets and does not seem to be a function of the Weibull coefficients. Regarding Λ^2, the slope of the linearisation has the tendency to increase for higher values of the scale and to decrease for higher shape coefficients. Linearisation is possible for almost all datasets taken into account.

For this reason, its might be possible to derive the Castaing coefficients by just knowing the value for one height and the slope of the linear fit, as long as you take a long enough space of time into account. Thus, a possibility can be provided to derive the dynamics of the wind speed by only looking at the one-point statistic. From a theoretical point of view, gaining a deeper understanding of the linear dependence

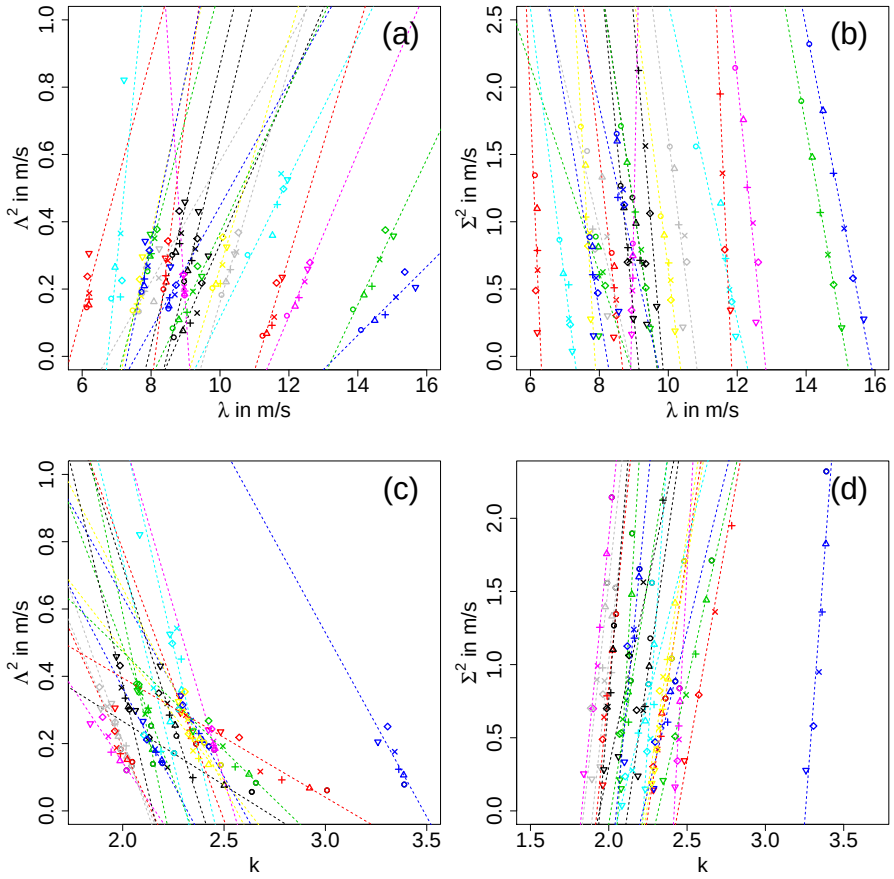

Figure 6.5: Correlation of the Castaing parameters Σ and Λ of the spatial increments δu_{ref} with the Weibull parameters λ and k. The points of one month are shown in the same colour and the points of one height each have one symbol. For all months, a linearisation is plotted as a thin dashed line.

between the Castaing and Weibull coefficients seems to be a field of studies that might be worth of further investigations.

Finding an equal dependence of the coefficients derived from the δ_{next} data is not possible. Here, only a slightly decreasing tendency of Σ^2 coefficients can be observed for growing Weibull coefficients. This is mainly caused by the fact that outliers are less frequently observed for bigger values of λ and k. For this reason, more than a range of values in which the Castaing coefficients can probably be found, cannot be given.

7 Stochastic modelling of wind profiles

The wind speed profile is modelled by using a system of coupled Langevin equations. We introduce a framework to derive the wind speed profile over the full height of a wind turbine with having only the wind speed at the nacelle of the wind turbine given. We use the wind speed increments $u'(t)$.

The first step to be taken for this purpose is to provide a possibility to easily convert wind speed data into wind speed increment data and vice versa. The transformation used is given by

$$u_h(t) = \langle u \rangle_T(h) + u'(t,h), \tag{7.1}$$

where $\langle u \rangle_T(h)$ is the average wind speed within a time span of length T as a function of the height, and $u'(t,h)$ is the short time fluctuation of the wind speed. It is chosen to take the mean wind speed over the whole measurement period of 20 months. The mean wind speed profile is modelled by the use of the power law. This is given by equation (2.3) with

$$\langle u \rangle_{h,T} = (\frac{h}{z_{ref}})^{\alpha}. \tag{7.2}$$

Here, $z_{ref} = 60$, $u_{ref} = 8.57\mathrm{m\,s^{-1}}$ and $\alpha = 0.060 \pm 0.003$ are to be used.

The parameter that has to be determined is the exponent α. This is done by a linearisation of the double logarithmic plot. (See figure 7.1).

The result is surprisingly low compared to the α suspected on shore of about $\frac{1}{7}$ and even more to a coefficient α of $\frac{1}{10}$ as it was estimated off shore. (See [8]). This is probably a result of the low degree of roughness off shore and the fact that the FINO1 platform is relatively far away from the nearest landscape. Using this method, the mean wind speeds are approximated with an error smaller than $\delta\langle u \rangle = 3.5 \cdot 10^{-2}\mathrm{m\,s^{-1}}$. This error is small compared to the fluctuation of the wind speed. The wind speed fluctuations can be modelled by the use of the Euler method. It is possible to derive the wind speed fluctuation iteratively by

$$u'(t + \Delta t, h) = u'(t,h) + d(u'(t,h)). \tag{7.3}$$

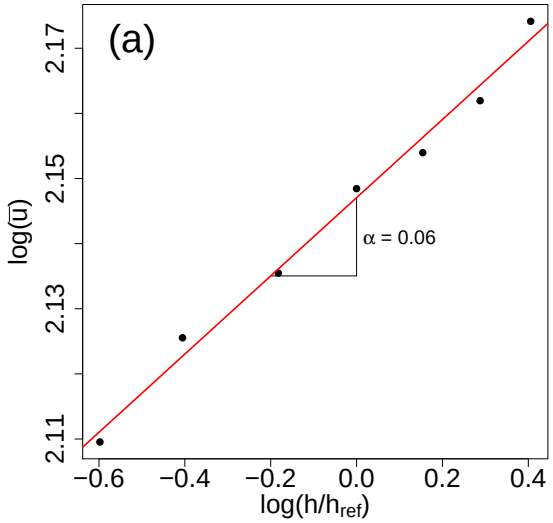

Figure 7.1: Determination of the power law exponent α by calculating the slope of the double logarithmic plot of the mean wind speed for the whole dataset against the relative height h/h_{ref} (black dots). The coefficient α is found to be $\alpha = 0.060 \pm 0.003$. The linearisation is plotted as a solid red line.

To simulate the fluctuations $d(u'(t,h))$, we use the Langevin approach mentioned in chapter 2. The wind speed is regularly measured at least at one height for every wind turbine. This yields a series of measured wind increments, $u'_h(t)$ and gives the opportunity to simulate the wind speed series at the neighbouring heights using

$$d(u'_{h_1}(t)) = D^{(1)}_{h_1}\left(u'_{h_1}(t), u'_{h_2}(t)\right) + g_{h_1,h_1}\left(u'_{h_1}(t), u'_{h_2}(t)\right)\Gamma_1(t) \\ g_{h_2,h_1}\left(u'_{h_1}(t), u'_{h_2}(t)\right)\Gamma_2(t), \tag{7.4a}$$

$$d(u'_{h_2}(t)) = D^{(1)}_{h_2}\left(u'_{h_1}(t), u'_{h_2}(t)\right) + g_{h_1,h_2}\left(u'_{h_1}(t), u'_{h_2}(t)\right)\Gamma_1(t) \\ g_{h_2,h_2}\left(u'_{h_1}(t), u'_{h_2}(t)\right)\Gamma_2(t), \tag{7.4b}$$

where we choose the initial condition $u'_{h_2}(0) = 0\ \mathrm{m\,s^{-1}}$.

The Kramers-Moyal coefficients can be derived by the `Langevin2D` function implemented in the Langevin package in R applied to the data set consisting of $5.2 \cdot 10^7$ data points. This function uses the equations (2.24) and (2.23) to calculate the functions $D^{(1)}_{h_i}\left(u'_{h_i}(t), u'_{h_j}(t)\right)$ and $D^{(2)}_{h_i,h_j}\left(u'_{h_i}(t), u'_{h_j}(t)\right)$ for $i,j = 1,2$, which are given back as matrices. These matrices are derived by dividing the range of values of u'_{h_1} and u'_{h_2} into a user-defined number of bins. If the number of events in a bin is below a defined threshold, the value of the bin is set to NaN. Here, the threshold is chosen to

be 100 events. We choose to use 20 x 20 bins.

For the derivation of the simulated time series, the functions $D^{(1)}_{h_1}\left(u'_{h_1}(t), u'_{h_2}(t)\right)$ and $g_{h_i,h_j}\left(u'_{h_1}(t), u'_{h_2}(t)\right)$ with $i, j = 1, 2$ are needed. The functions $g_{h_1,h_2}\left(u'_{h_1}(t), u'_{h_2}(t)\right)$ are given by

$$D^{(2)} = \begin{bmatrix} D^{(2)}_{h_1,h_1} & D^{(2)}_{h_1,h_2} \\ D^{(2)}_{h_2,h_1} & D^{(2)}_{h_2,h_2} \end{bmatrix} = \begin{bmatrix} g_{h_1,h_1} & g_{h_2,h_1} \\ g_{h_1,h_2} & g_{h_2,h_2} \end{bmatrix} \begin{bmatrix} g_{h_1,h_1} & g_{h_1,h_2} \\ g_{h_2,h_1} & g_{h_2,h_2} \end{bmatrix} = g^T g. \tag{7.5}$$

The equation holds true, because $D^{(2)}$ is symmetrical and the values on the main diagonal are non-negative. Furthermore, the eigenvalues must be non-negative. Due to the spectral theorem, it is possible to find an orthogonal matrix S consisting of the eigenvectors of $D^{(2)}$ and a diagonal matrix $\tilde{D}$. The main diagonal of $\tilde{D}$ consists of the eigenvalues of $D^{(2)}$ $\lambda_{D^{(2)},i}$ with $i = 1, 2$. This leads to

$$D^{(2)} = S\tilde{D}S^1,$$

As $D^{(2)}$ is positive definite, the eigenvalues of $D^{(2)}$ are positive, so that it is possible to take the square root, without having imaginary results. The eigenvalues of the matrix g are given by

$$\lambda_{g,i} = \sqrt{\lambda_{D^{(2)},i}}. \tag{7.6}$$

A diagonal matrix $\tilde{g}$ with these $\lambda_{g,i}$ on the main diagonal is defined, and the matrix g is given by

$$g = S^{-1}\tilde{g}S.$$

g is called the square root of $D^{(2)}$. For a deeper look into this topic, literature like [3] and [7] is recommended. [1]

The next step is the simulation of the wind speed fluctuation at neighbouring heights using equation (7.3). The simulated time series is to be seen as a surrogate of the time series.

For the illustration of the resulting matrices, $D^{(1)}_{h_i}$ and g_{h_i,h_j} with $i, j = 1, 2$ for the heights of 60 m and 70 m are shown in figure 7.2. These matrices are to be seen as discretized representations of steady functions describing the drift and diffusion coefficients.

[1] Notice that sometimes values smaller than zero can be found on the main diagonal of the matrix $D^{(2)}$ given by the `Langevin2D`. This is caused by a fault in a linearisation used for the determination of the values of the Kramers-Moyal coefficient. This error can only be observed if values very close to zero are reached. In this case, the resulting matrix is possibly no longer positive definite, so that the eigenvalues of g calculated in equation (7.6) are no longer guaranteed to be real. This problem is solved by trying to find the nearest positive definite matrix. This is done by the use of the R function `nearPD`, which gives back the nearest symmetrical, positive definite matrix.

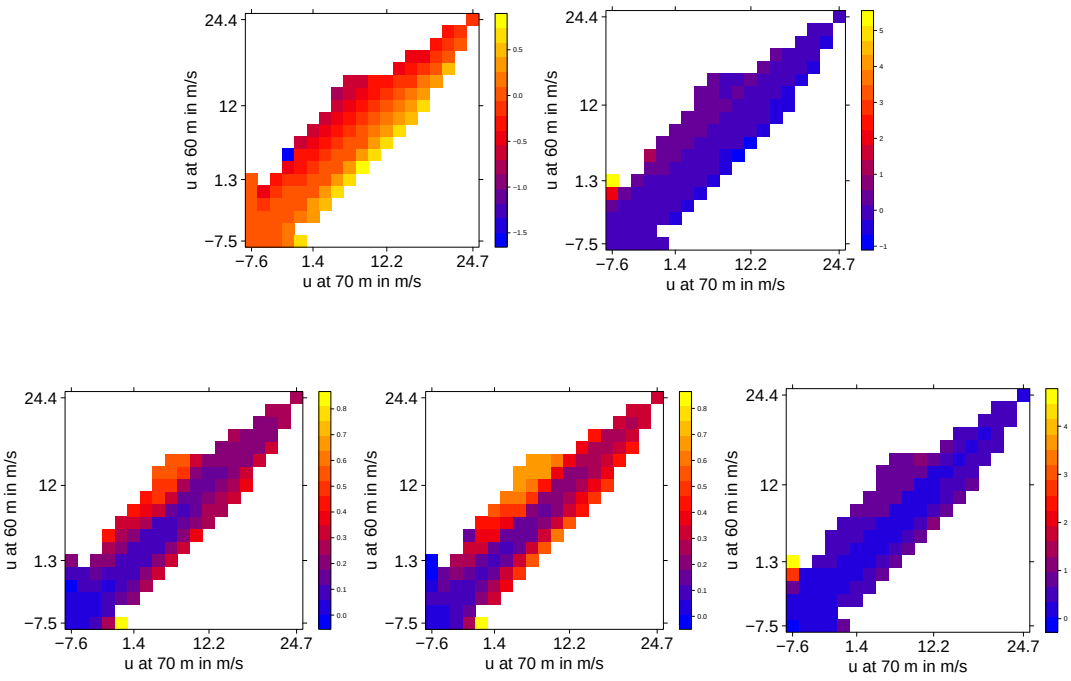

Figure 7.2: The drift coefficients $D^{(1)}_{60}$ (left) and $D^{(1)}_{70}$ (right) are shown in the upper row. In the lower row, the diffusion coefficients $g_{60,60}$ (left), $g_{60,70}$ (middle) and $g_{70,70}$ (right) are shown.

The drift coefficients are expected to fit a linear or a cubic two-dimensional polynomial. The diffusion coefficients could be described well by a quadratic two-dimensional polynomial. As the matrices given by the `Langevin2D` function are directly used for further calculations, they are not fitted in the way mentioned above.

The time series of the wind speed fluctuations are calculated using equations (7.4b), and (7.3).

Thus, repeating this procedure for the other heights using a previously generated surrogate series of the neighbouring height, the full wind profile is reconstructed and the simulated time series of the wind speed is calculated using equation (7.1).

The distributions of the simulated wind speed data are plotted in 7.3 (a). For heights below 60 m, the deviations are slightly higher than for the other heights. In general, the distributions of the simulated and measured data are in good agreement.

After this, we compare the statistics of the increments in time and space of the simulated time series to the statistics of the measured increment data. For this purpose, the height and time increments are calculated following the methods described in chapters 5 and 6. The time increments of 1 s, 64 s and 4096 s are chosen, because they cover the

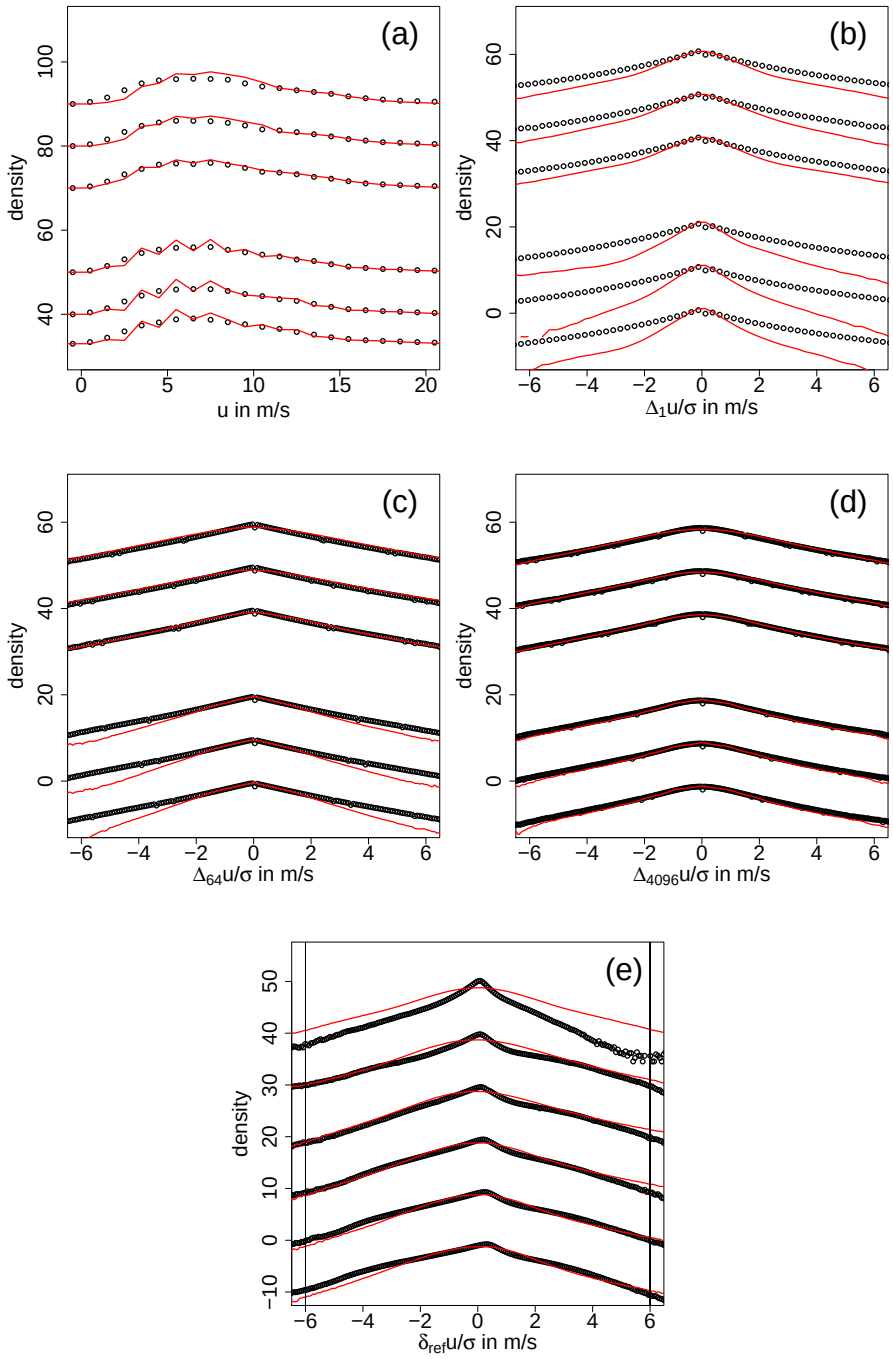

Figure 7.3: Distributions of the measured and simulated wind speed time series (a), the simulated and measured wind speed time increments for lags of (b) 1 s,(c) 64 s and (d) 4096 s. Figure (e) shows the distribution of the measured and simulated spatial increments $\delta_{ref} h_h$. The distribution of the measured data is shown by the black dots and the distribution of the simulated data by the red line. It can be seen that for small increments in time the width of the distribution is underestimated by the simulated data. This might be a result of the simulation procedure.

significant ranges of time discussed in this thesis.

The distributions of the time increments are shown in figure 7.3 (b), (c) and (d). For short time lags extreme events are clearly underestimated. This might be caused by the fact that every time $(u'_{h_1}(t), u'_{h_2}(t))$ fell into a bin in which less than 100 events are found in the data set. In this case the simulated value $u_{h_2}(t+\Delta t)$ is set to NaN. It seems difficult to solve this problem, because without this restriction in the case of an unrealistic combination of $(u'_{h_1}(t), u'_{h_2}(t))$ unrealistic values for the drift and diffusion coefficients might be used. This is due to the fact that for those combinations of $u'_{h_1}(t)$ and $u'_{h_2}(t)$, the fit to the matrices giving the drift and diffusion coefficients are of no physical relevance. For longer time lags the distributions are in good agreement.

It is chosen to calculate the height increments following equation (6.1a). The height increments are shown in 7.3 (e). The simulated and measured distributions are in poor agreement for comparably small spatial increments like $10\,\mathrm{m}$ and get better for growing spatial increments. Furthermore, it seems as if the skewness of the measured distribution is not well reproduced for longer spatial increments.

In general, it is possible to reproduce the distribution of the wind speed relatively well. Regarding the wind speed increments, this is only possible for larger lags and spatial increments. The main problem is the fact that the heavy tails of the distributions of the increments for small lags are not well reproduced.

8 Discussion and conclusion

The aim of this thesis was twofold: Firstly, we aimed at characterizing the wind profile. This was done for a single time series recorded at the FINO1 tower, the time increments and the spatial increments. Secondly, we aimed at reconstructing the wind speed profile using one-point measurements. To this end, we used a system of coupled Langevin equations.

The distributions of the wind speed can be well reproduced by the Weibull distribution for time periods of at least one month. The shape coefficients k found to be best fitting for the measured distribution, were close to two. The scale parameters were nearly proportional to the mean wind speed. These results corroborate the findings of C. Carrillo *et al.* [5]. For shorter periods with the duration of about a week, it might be possible to find a fitting distribution. This distribution would necessarily be able to reproduce data with distinct bimodularities.

The increment distributions were described by the Laplace and the Castaing distribution. It was found that for almost all datasets the Castaing distribution fits the data better. The Laplace coefficients and the Castaing coefficient Σ follow approximately an exponential function of the time lag. This is caused by the increasing variance of the distribution for growing lags. The coefficients were not clearly depending on the month or season in which they were measured. A larger correlation was found between the scale coefficient of the Weibull distribution and the coefficients of the Laplace distribution. This was not seen for the shape coefficients. This indicates that the mean wind speed of the single time distribution has a high impact on the width of the distribution of the time increments.

The height increments were calculated by the use of a reference height δ_{ref} and for neighbouring heights δ_{next}. It was seen that it is necessary to describe the distribution of height increments by the use of the Castaing distribution. The Castaing coefficients of δ_{ref} are seen to be a linear function of the distance. The parameters of the Weibull coefficients and the Castaing distribution are seen to be approximately linear for the data sets the parameters were calculated for. This provides the opportunity to derive properties of the two-point time series from the properties of the single time series. Understanding this linearity might be a good target of further investigations.

Surrogate data was generated with the Langevin model. The distributions of the simulated wind speed data were in good agreement with the measured data. The simulated increment time series were also in good agreement with the measured data for larger time lags and height differences. For shorter increments the heavy tails were not well reproduced. This problem might be solved by using a polynomial fit of the drift and diffusion coefficient for the generation of the data under other restrictions. Perhaps also further preprocessing of the data might give the opportunity for a more precise calculation of the drift and diffusion coefficients.

Bibliography

[1] C. Beck. Superstatistics in hydrodynamic turbulence. *Physica D: Nonlinear Phenomena*, 193(1):195 – 207, 2004.

[2] C. Behnken. Stochastische Charkterisierung des Offshore Windprofils auf kurzen Zeitskalen. Master's thesis, Carl von Ossietzky University Oldenburg, 2014.

[3] A. Beutelspacher. *Lineare Algebra*. Gabler, Betriebswirt.-Vlg, 2013.

[4] F. Boettcher, C. Renner, H.-P. Waldl, and J. Peinke. On the statistics of wind gusts. *Boundary-Layer Meteorology*, 108(1):163–173, 2003.

[5] C. Carrillo, J. Cidrás, E. Díaz-Dorado, and A. F. Obando-Montaño. An Approach to Determine the Weibull Parameters for Wind Energy Analysis: The Case of Galicia (Spain). *Energies*, 7(4):2676–2700, 2014.

[6] B. Castaing, Y. Gagne, and E.J. Hopfinger. Velocity probability density functions of high Reynolds number turbulence. *Physica D: Nonlinear Phenomena*, 46(2):177 – 200, 1990.

[7] G. Fischer. *Lineare Algebra*. Springer-Verlag GmbH, 2013.

[8] S. A. Hsu, E. A. Meindl, and David B. Gilhousen. Determining the Power-Law Wind-Profile Exponent under Near-Neutral Stability Conditions at Sea. *Journal of Applied Meteorology*, 33(6):757–765, 1994.

[9] D. S. Lemons and V. Ambegaokar. *An Introduction to Stochastic Processes in Physics*, volume 71. American Association of Physics Teachers (AAPT), 2003.

[10] P. G. Lind, I. Herráez, M. Wächter, and J. Peinke. Fatigue Load Estimation through a Simple Stochastic Model. *Energies*, 7(12):8279–8293, 2014.

[11] A. Morales, M. Wächter, and J. Peinke. Characterization of wind turbulence by higher-order statistics. *Wind Energy*, 15(3):391–406, 2011.

[12] M. B. Priestley and T. Subba Rao. A Test for Non-Stationarity of Time-Series. *Journal of the Royal Statistical Society. Series B*, 31(1):140–149, 1969.

[13] R. A. Silva de Santana, C.Q Dias-Junior, R. Sounza do Vale, J. Tota, and D. R. Fetzjarrald. Observing and Modeling the Vertical Wind Profile at Multiple Sites

in and above the Amazon Rain Forest Canopy. *Hindawi Advances in Meteorology*, page 8, 2017.

[14] R. von Sachs and M. H. Neumann. A Wavelet-Based Test for Stationarity. *Journal of Time Series Analysis*, 21(5):597–613, 2000.

[15] R. Wagner, M. Courtney, J. Gottschall, and P. Lindelöw-Marsden. Accounting for the speed shear in wind turbine power performance measurement. *Wind Energy*, 14(8):993–1004, 2011.

Eidesstattliche Versicherung

Ich versichere, dass ich die eingereichte Bachelorarbeit selbstständig und ohne unerlaubte Hilfe verfasst habe. Anderer als der von mir angegebenen Hilfsmittel und Schriften habe ich mich nicht bedient. Alle wörtlich oder sinngemäß den Schriften anderer Autoren entnommenen Stellen habe ich kenntlich gemacht.

Datum, Ort

Unterschrift