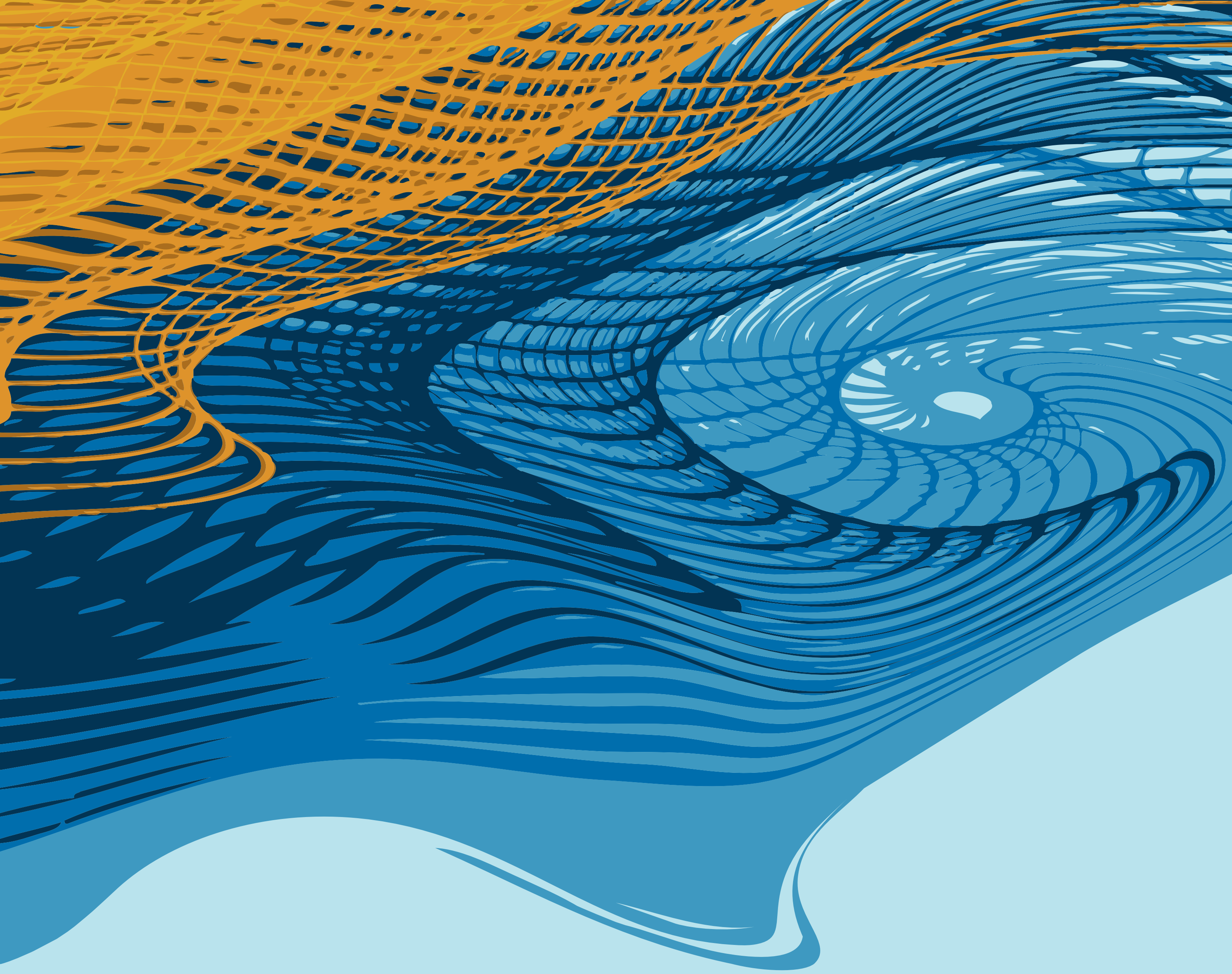

Aspects of Statistical Trends in High-Rate Wind Measurements

Christian Schou Oxvig | Patrick Steffen Pedersen

AALBORG UNIVERSITY
STUDENT REPORT

Title	Aspects of Statistical Trends in High-Rate Wind Measurements
Authors	Christian Schou Oxvig Patrick Steffen Pedersen
Supervisors	Assistant Professor Thomas Arildsen Professor Torben Larsen
Affiliation	Department of Electronic Systems Elite programme in Wireless Communication Aalborg University DK - 9220 Aalborg Denmark
Completed	June 2013
Page Count	104
Front Cover Artwork	*Turbulence* by Anne Mette Thy Tygesen Reproduced with permission from the artist

Resume

Vindhastigheden målt over tid på en given lokation er traditionelt set beskrevet ved dens gennemsnitlige fart og retning samt den statistiske spredning af disse. Middelværdien og spredningen er typisk opsamlet som gennemsnitlige værdier over 10 minutter. I dette projekt tages der udgangspunkt i målinger af vindhastigheder opsamlet med en samplingfrekvens i størrelsesordenen 1 Hz til 10 Hz. Ud fra disse målinger undersøges forskellige aspekter af såkaldte »trends« i vindhastighedssignalet - disse »trends« er lavfrekvente komponenter med periodetider større end 1 til 2 minutter.

De undersøgte aspekter relaterer sig til brugen af målte vindhastigheder til estimering af laster på vindmøller. Derfor gives først en beskrivelse af det setup, der gøres brug af ved beregning af laster - herunder den veletablerede model der bruges til simuleringer af vindhastigheder. Formaterne af de tilgængelige vindhastighedsmålinger er yderst forskelligartede og inkonsekvente, hvorfor de først behandles med henblik på at skabe en database af homogene målinger.

Det foreslås at skelne mellem »ubetydelige«, »lineære« og »periodiske« trends og en klassificeringsmetode til dette udvikles. Herefter foreslås en udvidelse af den etablerede model til simulering af vindhastigheder, således at modellen inkluderer en trend. Det vises, at et ideelt lavpasfilter kan bruges til at fjerne trenden, hvorefter resultatet kan beskrives af den oprindelige model. I praksis anbefales anvendelsen af et »Infinite Impulse Response« (IIR) filter på vindhastighedens tre komposanter hver for sig.

Det designede filter anvendes på de tilgængelige vindhastighedsmålinger, således at målingerne før og efter fjernelsen af trend kan sammenlignes. Ud fra en omfattende sammenligning af klassificeringsresultater vises det, at filtreringen sikrer, at alle betydelige trends fjernes fra vindhastighedsmålingerne. Yderligere vises det, at middelværdien af vindhastigheden over 10 minutter er uforandret efter anvendelsen af filteret, hvorimod spredningen af vindhastigheden over 10 minutter er faldet.

En større softwarepakke, navngivet *Wind Analysis Framework (WAF)*, er udviklet i forbindelse med projektet. Pakken er skrevet i programmeringssproget Python og giver mulighed for at genskabe alle resultater præsenteret i denne rapport. En kort introduktion til pakken gives mod rapportens slutning.

Preface

This Master's Thesis describes the findings of a study on *aspects of statistical trends in high-rate wind measurements*. The study was proposed by representatives from the wind turbine manufacturer Vestas[1], Frede Aakmann Tøgersen and Kim Emil Andersen, as a so-called long master's thesis project at Aalborg University. Consequently, the project has been carried out at Aalborg University in collaboration with Vestas.

Due to the project being proposed by representatives from Vestas, some of the background and views expressed in this report represent the beliefs of specialists employed at Vestas. Such beliefs have been marked with a footnote. All other beliefs and findings expressed in the report that are not those of the authors have been marked with appropriate references. A list of peer-reviewed references used in the report can be found on page 95. All other references are given in footnotes.

The authors have made their best attempt at keeping the research, presented in the report, reproducible. Thus, the source code of the *Wind Analysis Framework (WAF)* Python package developed as part of the project is freely available under the BSD 2-Clause License[2]. Furthermore, Python scripts that may be used to reproduce the figures and tables that present the research results in this report are freely available. The scripts are named according to the figures and tables they reproduce; for example, »fig_1_2.py« reproduces figure 1.2 whereas »tab_2_4.py« reproduces table 2.4. Both the package *WAF* and the scripts (collected in a Python package named *report*) can be found on the enclosed CD. The wind measurements used in the project are considered company confidential by Vestas and are thus, unfortunately, not freely available.

A list of symbols used in the report is found on page 99 whereas lists of figures and tables in the report are found on page 103. Finally, a reading guide to the report can be found in Section 1.4 on page 6.

Christian Schou Oxvig

Patrick Steffen Pedersen

[1]Plant Siting & Forecasting, Vestas Global Solutions & Services, Hedeager 42, DK - 8200 Aarhus N
[2]For further information about this license see `http://opensource.org/licenses/BSD-2-Clause`.

Contents

1 Introduction

This chapter serves as an introduction to the background of this report and its topic: information extraction from high-rate sampled wind velocities. The task of estimating loads on wind turbines is used to motivate the problem. Key parameters are presented and the subject to be addressed is further delimited. This leads to the problem thesis at the end of this chapter that formally states the specific tasks which are addressed in this report.

1.1 The Background

Electrical power has become a necessity in modern life style due to our dependence on electrical and electronic devices. We therefore rely on a constant power production to satisfy our omnipresent needs for electrical energy. In recent years, power production emerging from the renewable energy resource of wind has become more and more popular[1]: the global installed wind capacity, in terms of accumulated maximum power production of installed wind turbines measured in W, has increased each and every year for the last 16 years[2]. As of the end of 2012, the global installed wind capacity has reached over 282 GW[2].

The wind energy is harvested using wind turbines. Modern wind turbines are designed as so-called upwind horizontal axis wind turbines [1]. These are characterised by having an axis of rotation that is parallel to the ground with a rotor facing towards the direction of the wind. The mechanical rotation of the rotor is then transformed into electrical power in a generator [2].

The wind resource arises from uneven solar heating of the earth causing differences in pressure across the earth's surface [1]. Due to several meteorological phenomena, the resulting wind flow is often turbulent and characterised by seasonal components of different recurrence times [3]. Not only does this fluctuation in the wind entail a corresponding fluctuation in the output power of wind turbines; it also results in an uneven load of the wind turbine structure [2]. Hence, the components of a wind turbine are subject to complicated tear and wear effects.

In the early stages of the planning of a new wind turbine site, a measurement campaign is typically commissioned in order to analyse the wind regime at the location. Due to the high cost of such measurement campaigns they are often of short duration and only include few point measurements of the wind regime[3]. Naturally, wind turbines must be designed to sustain the uneven load effects imposed by the wind regime at the site.

[1]Based on Renewables 2012 Global Status Report, REN21, 2012. Available online: `http://www.ren21.net/Portals/0/documents/Resources/%20GSR_2012%20highres.pdf`

[2]Based on Global Wind Statistics 2012, Global Wind Energy Council (GWEC), 2013. Available online: `http://www.gwec.net/wp-content/uploads/2013/02/GWEC-PRstats-2012_english.pdf`

[3]Based on personal communication with senior specialist Kim Emil Andersen, Vestas, and specialist Frede Aakmann Tøgersen, Vestas, September 2012

Wind turbine manufactures are consequently faced with the demanding task of choosing components for wind turbines such that they are able to sustain the impact of a wind regime.

1.2 Selecting a Suitable Wind Turbine

The general task of selecting a suitable wind turbine for a given site is guided by the IEC 61400 standard *Wind turbines, Part 1: Design requirements* [4]. This standard describes the »external conditions« that should be taken into consideration in the process of designing a wind turbine. A guide to the usage of site specific information from measurement campaigns is included in the description. A short review of this description is given below.

The IEC 61400 standard defines three standard classes (I to III) of wind turbines; each with three subclasses (A to C). Additionally, a »special« class is defined in the standard. The special class is used when the site imposes special demands in terms of wind- or safety conditions that are not met by the three standard classes. The standard classes (I to III) differ in terms of the reference wind speed V_{ref} (Definition 1.1) that is aimed for. The subclasses (A to C) differ in terms of the expected turbulence intensity I_{ref} (Definition 1.2) that is aimed for. Hence, two important site specific parameters are [4]:

Definition 1.1
The *reference wind speed* V_{ref} is the upper bound on the extreme 10-minute averaged wind speed at hub height (the height of the centre of the rotor swept area) with a recurrence period of 50 years.

Definition 1.2
The *expected turbulence intensity* is the ratio

$$I_{\mathrm{ref}} = \frac{\sigma_{\mathrm{ref}}}{V_{15}}$$

where σ_{ref} is the average standard deviation in case of an average wind speed $V_{15} = 15\,\frac{\mathrm{m}}{\mathrm{s}}$ at hub height.

A wind turbine aimed at one of the standard classes must be designed to safely withstand the wind conditions which it is subjected to throughout its lifetime of at least 20 years. In the assessment of this requirement, the wind regime is divided into »normal wind conditions« and »extreme wind conditions« [4]. Normal wind conditions describe behaviour of the wind that »*occurs frequently during normal operation*« whereas extreme wind conditions are wind conditions that have recurrence periods of 1 year or 50 years.

The reference wind speed, V_{ref}, and expected turbulence intensity, I_{ref}, are not the only parameters specified in the IEC 61400 standard. They are, however, the two basic parameters that determine the wind turbine class. In addition to this, they are essential in the process of estimating loads on wind turbines, as discussed later in this report.

1.2.1 The Load Estimation Problem

To comply with the IEC 61400 standard, a wind turbine must be able to safely withstand the wind conditions which it is subjected to; that is, a wind turbine must be able to withstand the loads imposed on its components by the wind regime, at the site where the wind turbine is erected. Thus the wind turbine designer needs to estimate the loads on the wind turbine, based on measurements of the wind regime.

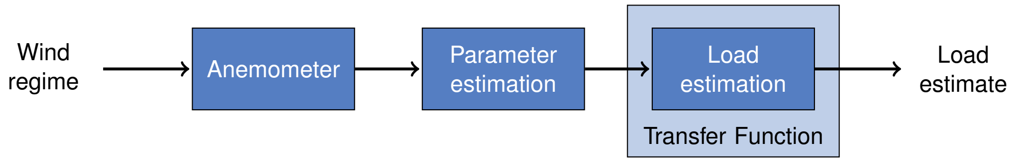

Figure 1.1. Illustration of the conceptual flow in estimating wind turbine loads based on the measurement of the wind regime at a given site. An anemometer is used to measure the wind velocity. A set of parameters is estimated from the resulting time series of the wind velocity. Through the use of a transfer function, this estimated set of parameters is used to estimate of the loads on the wind turbine.

The conceptual flow in estimating loads on a wind turbine[4] is illustrated in Figure 1.1. An anemometer is the device used to measure wind velocities. It should be noted, though, that some types of anemometers are only able to measure wind speeds; that is, the absolute value of the wind velocities [2]. As shown in Figure 1.1, the measured wind velocity time series are then used in the estimation of a set of parameters that characterise the wind regime. Finally, the loads are estimated using a transfer function from the parameters to the loads. Two of the estimated parameters[5] are the mean wind speed V (Definition 1.3) and the turbulence intensity I (Definition 1.4). Note that these parameters are generalisations of the reference wind speed V_{ref} and the expected turbulence intensity I_{ref}.

Definition 1.3
The *mean wind speed* V is the statistical mean of the instantaneous value of the wind speed averaged over a given time period which can vary from a few seconds to many years [4]

Definition 1.4
The *turbulence intensity*

$$I = \frac{\sigma_{\text{V}}}{V}$$

is the ratio of the wind speed standard deviation to the mean wind speed, determined from the same set of measured data samples of wind speed, and taken over a specified period of time [4]

[4] Based on personal communication with senior specialist, Kim Emil Andersen, Vestas, specialist Frede Aakmann Tøgersen, Vestas, and specialist Jesper Graugaard, Vestas, January 2013.

[5] Based on personal communication with senior specialist, Kim Emil Andersen, Vestas, specialist Frede Aakmann Tøgersen, Vestas, and specialist Jesper Graugaard, Vestas, January 2013.

The primary relation to realise from Definitions 1.3 on the previous page and 1.4 on the preceding page is that the mean and standard deviation of the wind speed are averaged over a specified time period. The measurements available to wind turbine manufactures for assessing a potential wind turbine site are usually only so-called 10-minute statistics[6]; that is, estimates of V and σ_{v} averaged over period of 10 minutes.

Although the IEC 61400 standard uses the turbulence intensity I as a basic parameter it is clear, from Definitions 1.2 on page 2 and 1.4 on the preceding page, that the turbulence standard deviation σ_{v} is an equivalent measure. As detailed later in this report, σ_{v} is the measure used in the current state-of-the-art model of the wind velocity observed at a given location. Hence, σ_{v} is generally used instead of I in the rest of this report. The conversion between I and σ_{v} is done using Definition 1.4 on the previous page which requires V to be known.

An aggregation (averaging over 10 minutes) cannot be directly compared to a sample rate. However, it is interesting to consider the frequency content of the wind speed that can be represented by use of a sample rate of 10 minutes. Applying the Nyquist-Shannon Sampling Theorem (see e.g. [5]), only frequency components with periods larger than 20 minutes are recoverable from data sampled at a rate of 10 minutes. Thus, using 10-minute statistics, only information about the fluctuations in the wind with periods shorter than 20 minutes is captured by the averaged standard deviation - quite a reduction of available information. The reason for using 10-minute statistics rather than statistics from shorter or longer aggregation periods is a bit unclear: looking towards the literature, the explanation is generally something like: "*Turbulence refers to fluctuations in the wind speed on a relatively fast time-scale, typically less than about 10 minutes*" [3]. Similar explanations stating that 10 minutes is *the* aggregation period to use can be found in [1], [2], [6], [7], [8].

Intuitively (see Figure 1.2 on the facing page), the fluctuations in the wind velocity, at a given location, are not fully captured by such 10-minute statistics. This observation has led to the desire[7] to be able to exploit high-rate wind measurements directly in the parameter estimation part of the system shown in Figure 1.1 on the previous page. In this context, high-rate measurements refers to data sampled at rates in the range of 1 Hz to 20 Hz (which are high rates relative to the 10-minute rate of $\frac{1}{600}$ Hz). Thus, the task that motivates the study of high-rate sampled wind velocities is the desire to establish a way to estimate the parameters input to the wind turbine load estimation transfer function from high-rate measurements.

[6] Based on personal communication with senior specialist Kim Emil Andersen, Vestas, and specialist Frede Aakmann Tøgersen, Vestas, September 2012.

[7] A desire expressed by senior specialist Kim Emil Andersen, Vestas, and specialist Frede Aakmann Tøgersen, Vestas, in personal communication, September 2012.

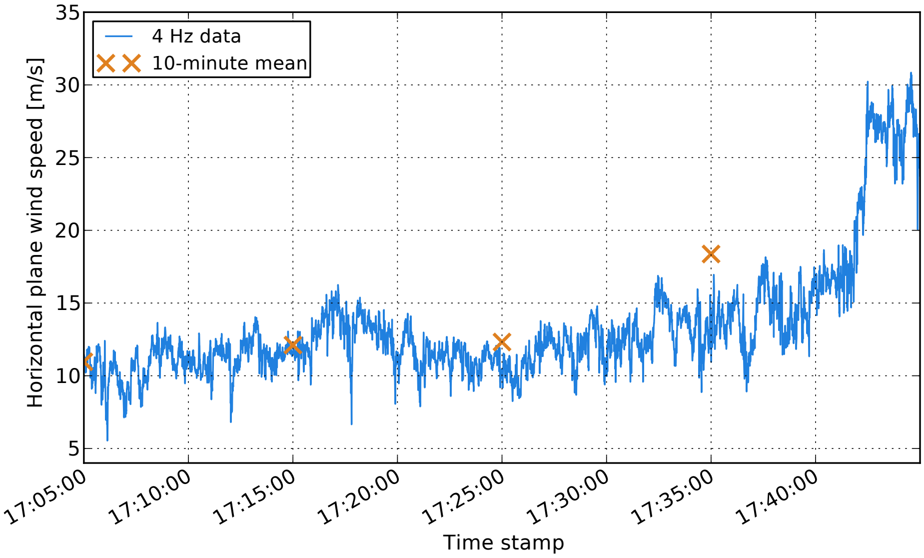

Figure 1.2. Example of wind speed measurements sampled at a rate of 4 Hz. Also shown are examples of 10-minute mean statistics calculated from the data. The time stamp used for the 10-minute mean is the start of the interval. The data is from Teufelsberge (29-10-2006).

1.3 Examining the Information in Wind Data

An important step towards being able to exploit high-rate wind measurements in load estimation is to gain insight into the mean wind speed, V, and turbulence standard deviation, σ_{v}, (or equivalently turbulence intensity, I) calculated from high-rate wind measurements. One element, which is of special interest, is the influence of the frequency components with periods larger than around 1 minutes to 2 minutes on the calculated values of V and σ_{v}. These components are of high interest since it is believed[8] that such slowly varying »trends« contribute insignificantly to the loads of wind turbines. Thus they may be the cause of over- or underestimation of loads that are estimated from 10-minute statistics of V and σ_{v}, since the trends, inevitably, contribute to the 10-minute statistics. Hence, the interesting aspect of this trend problem is not only the idea of exploiting high-rate measurements in the calculation of V and σ_{v}; it is also to establish links to the currently used 10-minute statistics in order to identify the changes in the estimates of V and σ_{v} when going from 10-minute statistics to high-rate measurements. This leads to the following Problem Thesis which is addressed in this report.

Problem Thesis

What information about »trends« in the mean wind speed V and turbulence standard deviation σ_v can be extracted from high-rate wind measurements? - and what is its relation to current 10-minute statistics used in wind turbine load estimation?

[8] A belief expressed by specialist Frede Aakmann Tøgersen, Vestas, and specialist Jesper Graugaard, Vestas, in personal communication, December 2012.

The following specific tasks have been identified to be connected to the problem thesis:

- Establish a consistent, high-performance database from in-homogeneous and frequently inconsistent high-rate wind measurements.
- Assess the impact of the aggregation period, or averaging period, on statistics calculated from the high-rate wind measurements.
- Compare the high-rate sampled wind velocities to current state-of-the-art models of such wind velocities.
- Relate the concept of trends to the current state-of-the-art models of wind velocities.
- Investigate and group the high-rate sampled wind velocities in terms of classes of trends.
- Review state-of-the-art methods used to remove a trend from wind velocities measurements.
- Evaluate the effect of removing trends from high-rate wind measurements.

1.4 Reading Guide

The remainder of this report consists of six chapters that address different parts of problem thesis and the tasks listed above. An overview of the contents of these chapters is given below.

Chapter 2 describes the very comprehensive conditioning of the wind velocities measurements, which is needed in order to obtain a consistent high-performance database of such measurements.
Chapter 3 provides an overview of the current state-of-the-art wind velocity model in addition to a comparison of the measured wind velocities to the model. Furthermore, the impact of the choice of aggregation period on calculated statistics is assessed.
Chapter 4 presents a machine learning based classification of the measured wind velocities into a number of classes of trends.
Chapter 5 relates the concept of a trend to the current state-of-the-art model of wind velocities and gives a review of methods which can be used to remove the trend. Finally, an evaluation of the effects of the removal of trends.
Chapter 6 gives a brief overview of the comprehensive software package that has been developed together with this report.
Chapter 7 states the collective set of conclusions which can be drawn from the findings presented in all the chapters preceding it.

2 Data Conditioning

A huge set of files containing wind velocity measurements is available for the work with the Problem Thesis stated in Section 1.3 on page 5. This data needs considerable conditioning before a consistent database of usable homogeneous high-rate measurements is in place. This chapter provides an overview of the available data files. A list of requirements to a consistent database is set forth and the format of the database that is needed in order to handle datasets with sizes of several hundred gibibytes is discussed. Necessary processing of the data such as error handling, time stamp handling, type conversions, coordinate tranformations, and outlier filtering is discussed. Finally, the performance of the resulting database is evaluated and an overview of the available datasets after the post processing is given.

2.1 The Available Data

The wind turbine manufacturer Vestas has provided the high-rate wind measurement datasets that are used to obtain the results presented in this report. Vestas has provided measurements of wind velocities from eight different locations around the world representing examples of both so-called flat terrain and complex terrain (see e.g [2] for a an explanation of the difference between flat and complex terrain). The measurements have been taken over periods of 3 to 13 months. No less than six different anemometers including lidars, ultra sonic, and cup anemometers have been used in the different measurement campaigns. These six different anemometers have been combined with four different data loggers to record the resulting time series of wind velocities.

The measured wind velocity time series are stored in over a thousand different comma separated value[1] (CSV) files. The files contain examples of both high-rate wind measurements and 10-minute statistics. The high number of combinations of measuring equipment and data loggers has resulted in datasets that are very in-homogeneous as can be seen from comparing the examples of CSV files in Tables 2.1 on the next page to 2.3 on the following page.

The datasets differ in terms of data types (e.g. strings, floats, and integers), sample rate, measurement coordinate system, time stamp style, and the kind of additional information provided along with the wind velocities to name just the most obvious differences. Furthermore, due to failures in the measurement equipment, the wind velocity time series are incomplete, sometimes leaving large time gabs between consecutive measurement.

[1]The term comma separated values is used to designate various different ways of representing tabular-like data in a text format. Generally, newlines are used to distinguish rows of the table whereas values within a row are separated by commas or tabs. See, e.g. the Python CSV module dialect documentation (`http://docs.python.org/3/library/csv.html`) for an overview of some of the possible differences between files, that are all termed CSV files

"TIMESTAMP" "TS"	"Spd3D" "m/s"	"Dir3D" "degrees"	"Vert" "m/s"	"Status" "Volts"
"2006-11-14 14:29:16.25"	27.26	296.9	0.701	-0.001
"2006-11-14 14:29:16.5"	27.24	296.9	0.682	0
"2006-11-14 14:29:16.75"	26.73	294.9	-1.022	0
"2006-11-14 14:29:17"	26.73	294.9	-0.9850001	-0.001

Table 2.1. Example of a subset of columns from a CSV file containing data from Teufelsberge. The column headers reflect a subset of the header information from the CSV file. The delimiter used in the CSV file is a comma.

"TIMESTAMP" "TS"	"Wsp1" ""	"Wdir1" ""	"Slope1" ""	"Temp1" ""	"Err1" ""
"2011-11-08 11:37:28.4"	7.08	293	15.46	21.66	0
"2011-11-08 11:37:28.5"	6.65	287	16.72	21.56	0
"2011-11-08 11:37:28.6"	6.59	294	9.3	21.58	0
"2011-11-08 11:37:28.7"	7.07	290	7.942	21.77	0

Table 2.2. Example of a subset of columns from a CSV file containing data from Snaresmoor. The column headers reflect a subset of the header information from the CSV file. The delimiter used in the CSV file is a tab.

Date	Vh-0	Azi (°)-0	u-0	v-0	w-0
08/01/2010 15:01:55.19	0.060981	314.456820	0.042709	-0.043527	-3.079550
08/01/2010 15:01:56.39	0.457359	83.855152	0.048957	0.454731	-3.232003
08/01/2010 15:01:57.49	0.457359	78.614205	0.090289	0.448358	-3.232003
08/01/2010 15:01:58.49	0.091472	342.408537	0.087194	-0.027645	-3.384456

Table 2.3. Example of a subset of columns from a CSV file containing data from another location than Teufelsberge and Snaresmoor. The column headers reflect a subset of the header information from the CSV file. The delimiter used in the CSV file is a tab.

Since it is a necessity to be able to compare different datasets in order to infer new relations from high-rate sampled wind velocities, such in-homogeneous data cannot be utilised directly. Hence, in order to be usable in the work on the tasks associated with the Problem Thesis presented in Section 1.3 on page 5, the data must be collected in a consistent format and post processed to yield a homogeneous and representative set of measured wind velocities.

A complete study of all the datasets provided by Vestas is out of the scope of this report. Consequently, two datasets have been chosen to be used as representative datasets for all the available datasets. The first dataset which represents a mountainous area (complex terrain) will be referred to as »Teufelsberge«. The other dataset represents a flat area and is referred to as »Snaresmoor«. The names »Teuflesberge« and »Snaresmoor« are aliases made up by the authors since the actual locations are considered company confidential by Vestas.

The measurement equipment used at Teufelsberge is a Gill WindMaster ultra sonic anemometer and a Campbell CR1000 data logger[2]. The Gill WindMaster has a specified resolution of $0.01\,\frac{\mathrm{m}}{\mathrm{s}}$ and an accuracy in terms of maximum deviation of $1.5\,\%$[3]. At Snaresmoor a combination of a Metek USA-1 ultrasonic anemometer and a Campbell CR1000 data logger has been used. The Metek USA-1[4] has a specified resolution of $0.01\,\frac{\mathrm{m}}{\mathrm{s}}$ and an accuracy in terms of maximum deviation of $2\,\%$. It is unclear whether or not the anemometers have been calibrated.

2.2 A Consistent Database

The discussion in Section 2.1 on page 7 of the available data makes it clear that a consistent database of wind velocities measurements must be established as a prerequisite to any other work related to the Problem Thesis in Section 1.3 on page 5. The requirements to such a consistent database is discussed in the following. Although only two datasets are used in the present report, all of the choices presented in the following have been made to allow for extensibility of the database such that the remaining datasets provided by Vestas may be included in the database in the future.

2.2.1 Data Storage Requirements

A set of requirements to a consistent way of storing the data may be outlined based on the Problem Thesis in Section 1.3 on page 5. The first three requirements relate to the actual wind velocity measurements. For each wind velocity measurement, the following must be available in the database:

1. A unique time stamp
2. The three dimensional wind velocity described in a consistent coordinate system
3. An indication of the validity of the data

The differences in information that can be extracted from wind velocities collected at different sample rates are central in the Problem Thesis. Thus, it is a necessity to have correct time stamps for each wind velocity measurement in order to keep track of the sample rates used in addition to the time of the measurement. To be able to compare measured wind velocities they must be described in a consistent coordinate system using a consistent set of units. Most of the measurement equipment is capable of delivering status flags or similar information that can be used to assess the validity of the measured wind velocities. Such information should be included in the database for later use in a post processing step such that all findings are based solely on valid data.

[2]The specifications of the Campbell CR1000 data logger are available at `http://www.campbellsci.com/cr1000`

[3]The specifications for the Gill WindMaster ultra sonic anemometer are available at: `http://www.gill.co.uk/products/anemometer/windmaster.htm` (Accessed: 15-10-2012).

[4]The specifications of the Metek USA-1 ultrasonic anemometer are available at: `http://www.metek.de/product-details/usonic-3-scientific.htm` (Accessed: 15-10-2012).

Two additional requirements that relate to the task of keeping track of wind velocity measurements from different locations and different measurement equipment are outlined. The database must support:

4. Storage of datasets with different sample rates, e.g. both 10 Hz measurements and 10-minute statistics.
5. Grouping of datasets based on locations and measurement equipment

Having such large datasets, it is necessary to have a strict storage structure to be able to keep track of the different datasets. However, such a structure must be flexible enough to allow for storage of datasets with different sample rates.

Several details related to a consistent description of the measured wind velocities and the mapping from high-rate measurements to 10-minute statistics must be established in order to fulfil the five requirements listed above. These details are described in the next two sections.

2.2.2 Reference Coordinate System

A wind velocity $\mathbf{v}(t)$ is a three dimensional time dependent quantity. However, some anemometers are only capable of measuring one or two of the components. The cup[5] anemometer is an example of a device that is only capable of measuring a single component, namely the horizontal wind speed. If a wind vane is used in addition to the cup anemometer, the wind direction in the horizontal plane may also be measured yielding a horizontal plane polar representation of the two dimensional wind velocity. If an ultra sonic anemometer is used, all three wind velocity components may be measured.

To be able to represent measurements from both cup anemometers and ultra sonic anemometers, a horizontal plane polar coordinate system with a vertical component that is perpendicular to the horizontal plane is chosen as the reference coordinate system. Specifically, the three wind velocity components that are dependent on the time t are:

- $v_{\mathrm{h}}(t)$ - The wind speed in the horizontal plane, i.e. the length of the horizontal wind velocity measured in $\frac{\mathrm{m}}{\mathrm{s}}$. Hence, $v_{\mathrm{h}}(t) \in [0; \infty)$
- $\theta(t)$ - The wind direction (direction of flow) in the horizontal plane measured in degrees from true North (with 90° being a flow towards east). Hence, $\theta(t) \in [0; 360)$.
- $v_{\mathrm{v}}(t)$ - The vertical wind speed (normal to the horizontal plane) measured in $\frac{\mathrm{m}}{\mathrm{s}}$. A positive $v_{\mathrm{v}}(t)$ indicates an upward flow whereas a negative $v_{\mathrm{v}}(t)$ indicates a downward flow. Hence, $v_{\mathrm{v}}(t) \in (-\infty; \infty)$.

The chosen $(v_{\mathrm{h}}(t), \theta(t), v_{\mathrm{v}}(t))$ reference coordinate system furthermore has the property that it allows for easy handling of measurements with no clear specification of the absolute horizontal wind direction. Such a specification of the absolute horizontal wind direction is

[5] See e.g. `http://www.windsensor.dk/products.htm` for details about the Risø P2546A cup anemometer that have been used in several of the datasets provided by Vestas.

not essential in the study of wind speeds and turbulence standard deviations since these quantities are calculated from the length of the wind velocity and not the direction. It is generally unclear if a measurement of $\theta(t_0) = 0$ at an arbitrary time t_0, in the datasets corresponds to a wind direction towards true North. Thus, a measurement of $\theta(t_0) = 0$ is assumed to be a wind velocity towards true North. The reference coordinate system and a spherical coordinate system are shown in Figure 2.1.

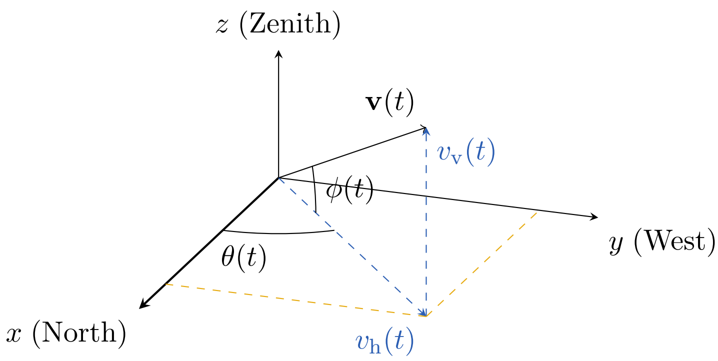

Figure 2.1. Illustration of the three dimensional wind velocity $\mathbf{v}(t)$ in the reference coordinate system $(v_\mathrm{h}(t), \theta(t), v_\mathrm{v}(t))$ together with the spherical coordinate system $(|\mathbf{v}(t)|, \theta(t), \phi(t))$.

2.2.3 Aggregation

Aggregations, which are also known as reductions, are operations that produce scalars from arrays [9]. Examples of aggregations include sums, means, minimums, and maximums. The connection between the high-rate measurements and 10-minute statistics is that the 10-minute statistics are formed as aggregations of the high-rate measurements. Let $a_1, \ldots, a_N \in \mathbb{R}$ be consecutive values from a time series of wind speed measurements with N being equal to the number of samples in an interval of 10-minutes, then the 10-minute mean wind speed statistic $\hat{a}$ is:

$$\hat{a} = \frac{1}{N} \sum_{i=1}^{N} a_i \tag{2.1}$$

Thus, the N wind speed measurements have been aggregated to a scalar value. An important thing to note at this point is that each of the a_i's are associated with a unique time stamp, since $a_1, \ldots, a_N$ are part of a time series. Thus, an »aggregation« of the time stamps must also be done in order to have a time stamp associated with the 10-minute statistic. To the authors knowledge there is no standard way to choose this time stamp. However, obvious choices for the 10-minute statistic time stamp are the first, the middle, or the last time stamp of the high-rate measurement series. The choice that seems to be the most frequent in the datasets provided by Vestas is the first time stamp in the series. For that reason, the first time stamp in the high-rate measurements series is used in this report for 10-minute statistic even though that leads to the (perhaps a bit unusual) choice of a non-causal system.

The operation of calculating consecutive 10-minute mean statistics may be interpreted as an application of a finite impulse response (FIR) filter followed by a downsampling by a factor of N (see e.g. [5] for the details about a FIR filter). That is, if the high-rate measurements are filtered with a FIR filter of length N formed from a rectangular window scaled with $1/N$, it corresponds to calculating a moving average. Taking only every N'th sample of the filtered series results in the series of 10-minute statistics. Hence, the well established theory behind the design of FIR filters by use of windowing may be used to derive aggregations with properties slightly different from the sum aggregation. For instance, using a smoother window like the Hamming or Hann windows results in a filter frequency characteristic with smaller side lobes (at the cost of a wider main lobe), thus reducing the Gibbs phenomenon compared to a filter based on the rectangular window [5].

2.2.4 Database Format

Having identified the five data storage requirements listed in Section 2.2.1 on page 9, the task is to select a database format that can satisfy these requirements. The term »database format« is to be understood as a specification of the container used to store the datasets on disk. Examples of possible database format candidates are text file based databases such as CSV or XML[6] files, an SQL[7] database or an HDF[8] database.

The CSV files containing the Snaresmoor and Teufelsberge datasets have an accumulated size of 307 GiB. A serious bottleneck in processing such large datasets is getting the data to the CPU - a problem that is termed CPU data starvation [10]. Thus, a high performance database format that can handle datasets with sizes of several hundred gibibytes is needed. In a recent study, whose goal was to find the best suitable database format for storing time series data, it is concluded that datasets with gibibyte sizes should be stored in a binary format to be efficiently handled [11]. The recommended database format in that study is HDF5, since it is an open, standardised format with bindings for major programming languages such as C/C++, Matlab, and Python. Furthermore, the test cases used in the study showed that datasets larger than 200 GiB were successfully handled using the HDF5 format. The HDF5 format is also recommended in [9] for storage of large datasets that are only written once and read many times (which is the case for the wind velocity measurements). In another study, that is a comparison of the HDF5 format and a Microsoft SQL database for storage of large datasets, it is concluded that HDF5 performs far better than the Microsoft SQL database in terms of write and read speeds [12]. Thus, it is recommended that an SQL database is used for tasks that require a relational database whereas the HDF5 format should be used for efficient storage of large

[6]Extensible Markup Language (XML) files is a standardise flexible text format used to data. See `http://www.w3.org/XML/` for further details.

[7]Structured Query Language (SQL) is the language used in relational databases management systems like MySQL. See `www.mysql.com` for further details.

[8]Hierarchical Data Format (HDF) is an open file based database format. See `http://www.hdfgroup.org/` for further details

datasets. Based on the above discussion, the HDF5 database format has been chosen for storage of the measured wind velocities due to its flexibility and capability in terms of handling large datasets. Furthermore, its hierarchical structure ensures that datasets can be grouped as required – see Section 2.5.2 on page 18 for details on how this is done.

2.3 Data Preprocessing

Preprocessing of the data in the CSV files is needed before the wind velocity measurements can be stored in the HDF database in a consistent way that meets the requirements listed in Section 2.2.1 on page 9. The description of the applied data preprocessing is split into a discussion of handling of errors, type conversions and coordinate transformations.

2.3.1 Handling Errors

The CSV files contain different types of errors. These are:

1. Fragmented lines, i.e. lines in which some entries are missing
2. Missing values, i.e. entries with blank values
3. Corrupted values, i.e. entries with values with unknown encoding
4. Invalid value format, e.g. floats that are specified as 11.21.1
5. Values outside expected ranges

An example of a line from the Snaresmoor dataset that contains corrupted values is:

```
'2012-02-29 15:5\xe9\xfdb','m\x19\x87sk','\x8ew\xbfrZ#\xe9\xfdb',...,
'm\x19\x87sk','\x8ew\xbfrZ#\xe9\xfdb','16.969999','0.0','0.0','0.0','0.0'
```

Error types 1 to 4 all results in ambiguous lines. Since there is no way to retrieve the missing information, such errors are handled by skipping the lines in which they occur.

Error type 5 relates to wind velocity coordinate values that are outside their expected ranges. Two types of range violations are present in the datasets and must consequently be handled. The first one relates to coordinate values that, by the definition, does not exist, e.g. a negative $v_{\mathrm{h}}(t_0)$ at time t_0. The other range violation relates to values that exceed the limitations of the anemometers used. The Gill WindMaster[9] is capable of measuring wind speeds up to 45 $\frac{\mathrm{m}}{\mathrm{s}}$ whereas the Metek USA-1 is capable of measuring wind speeds up to 60 $\frac{\mathrm{m}}{\mathrm{s}}$[10]. Both measurement values that by definition cannot exist and values that exceed the limits of the measurement equipment used are ignored, i.e. lines in which they occur are skipped.

[9] The specifications for the Gill WindMaster ultra sonic anemometer are available at: `http://www.gill.co.uk/products/anemometer/windmaster.htm` (Accessed: 15-10-2012).

[10] The specifications of the Metek USA-1 ultrasonic anemometer are available at: `http://www.metek.de/product-details/usonic-3-scientific.htm` (Accessed: 15-10-2012).

2.3.2 Conversions

The data storage requirements listed in Section 2.2.1 on page 9 states that a unique time stamp and an indication of validity of the data (a validity flag) must be used in the database. Hence, the different specifications used across different datasets necessitates conversions of:

- Time stamp
- Valid flag

The time stamp is saved in the HDF database as a 64-bit floating point (float64) value with the format »seconds.milliseconds«, i.e. the place before the decimal operator is used to store the seconds since the Epoch[11] whereas the place after the decimal operator is used to augment the seconds since the Epoch to milliseconds precision. Since no time zone information or information about daylight savings is available for the datasets, all time stamps are assumed to be UTC[12] time. Thus, the string representation of the time stamps must be converted to this float64 format.

The valid flag saved in the HDF database is a bool value that is True if the measurement is valid and False otherwise. The Gill WindMaster ultra sonic anemometer used at Teufelsberge provides a voltage level to indicate if a measurement is valid. The Metek USA-1 used at Snaresmoor has an error flag, i.e. a negated valid flag. Conversions of such quantities to a bool valid flag are thus used.

2.3.3 Coordinate Transformations

The chosen reference coordinate system is described in Section 2.2.2 on page 10 and depicted in Figure 2.1 on page 11. The datasets are described in different coordinate systems. Hence, different transformations are needed in order to describe the wind measurements in the reference coordinate system.

The wind velocities measured at Teufelsberge are already described in the reference coordinate system with the single exception that the direction of rotation of $\theta(t)$ is reversed. Hence, the sign of the horizontal plane wind direction needs to be switched when the dataset is loaded.

The measurements from Snaresmoor, on the other hand, are described in a spherical coordinate system $(|\mathbf{v}(t)|, \theta(t), \phi(t))$ shown in Figure 2.1 on page 11. Here, $|\mathbf{v}(t)| \in [0; \infty)$ is measured in $\frac{\text{m}}{\text{s}}$, $\theta(t) \in [0, 360)$ is measured in degrees, and $\phi(t) \in [-90, 90]$ is measured in degrees. The direction of rotation of $\theta(t)$ is unknown for the Snaresmoor data. However, since the absolute direction is not essential, $\theta(t)$ is assumed to be defined as shown in Figure 2.1 on page 11. Thus, $\theta(t)$ is shared between the two coordinate systems. The

[11] Epoch is on January 1st 00:00 hours 1970 for Unix systems (see e.g. `http://cm.bell-labs.com/cm/cs/who/dmr/1stEdman.html` (Accessed: 05-06-2012)). This is the definition used in the present report.

[12] Coordinated Universal Time (UTC) is a way to specify time that is closely related to Greenwich Mean Time (GMT). See e.g. `http://www.timeanddate.com/time/aboututc.html` for further details.

transformation of the remaining coordinates from the spherical coordinates to the reference coordinates at the arbitrary time t_0 is then given by:

$$v_h(t_0) = |\mathbf{v}(t_o)| \cdot \cos(\phi(t_o)) \tag{2.2}$$
$$v_v(t_0) = |\mathbf{v}(t_0)| \cdot \sin(\phi(t_0)) \tag{2.3}$$

Hence, (2.2) and (2.3) are used to transform the Snaresmoor data to the reference coordinate system depicted in Figure 2.1 on page 11.

2.4 Data Postprocessing

Despite the preprocessing of the data to comply with the data storage requirement listed in Section 2.2.1 on page 9, »strange events« still occur in the data. Thus, despite the wind velocity measurements being stored in a consistent way, a further postprocessing of the data is necessary. The description of the applied data postprocessing is split into a discussion of irregular time stamps sequences, outliers, and quantization problems.

2.4.1 Time Stamp Related Problems

The time stamps associated with a time series of wind velocity measurements taken at successive points in time is expected to constitute a monotonically increasing sequence. That is, if the time stamp (the float64 value defined in Section 2.3.2 on the facing page) is plotted versus the measurement index (an enumeration of the sequence of measurements) the result is a straight line if no measurements are missing. If some measurements are missing, discontinuous jumps to higher time stamps for higher measurements indexes results. In no way should it be possible for the time stamp to suddenly drop to a lower value for higher values of the measurement index. However, such drops are present in the datasets. An example is given in Figure 2.2 on the next page. This problem is handled by ignoring measurements that are not part of the (overall) monotonically increasing time sequence, i.e. the problematic lines of a CSV file are skipped.

2.4.2 Outliers

An outlier is taken to be a measurement that is »unlikely to be valid« given its adjacent measurements. An example of an outlier is given in Figure 2.3 on the following page. Before the outliers can be handled, they must first be identified in the datasets. The outlier shown in Figure 2.3 on the next page may be identified by its extreme change in a very short time span. However, not all outliers are that clearly separated from the steady level surrounding it. An attempt at identifying outliers based on the mean value and variance of its surrounding measurements revealed that a much more advanced criteria is needed if the outliers are to be separated from turbulent measurements. Since the turbulent measurements are an important part of the study described in this report,

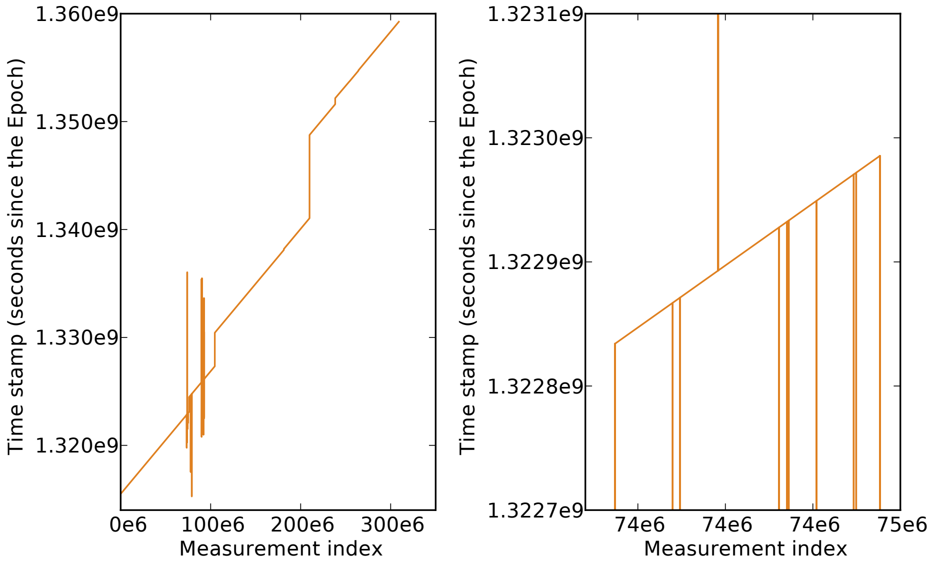

Figure 2.2. Example of data extract from Snaresmoor M2 60 which has time stamps that do constitute a monotonically increasing sequence. Left figure: The data time stamp since the Epoch (based on a sampling frequency of 10 Hz) versus data index. Right figure: Excerpt of left figure highlighting the problematic span of measurement indexes.

they cannot simply be removed without an extensive assessment of the consequences of removing them. Such an assessment is however an entire project in itself. Thus, since a single outlier only has little impact on aggregations like mean values (which is the kind of processing applied to the data as described in the rest of this report), the outliers are not removed.

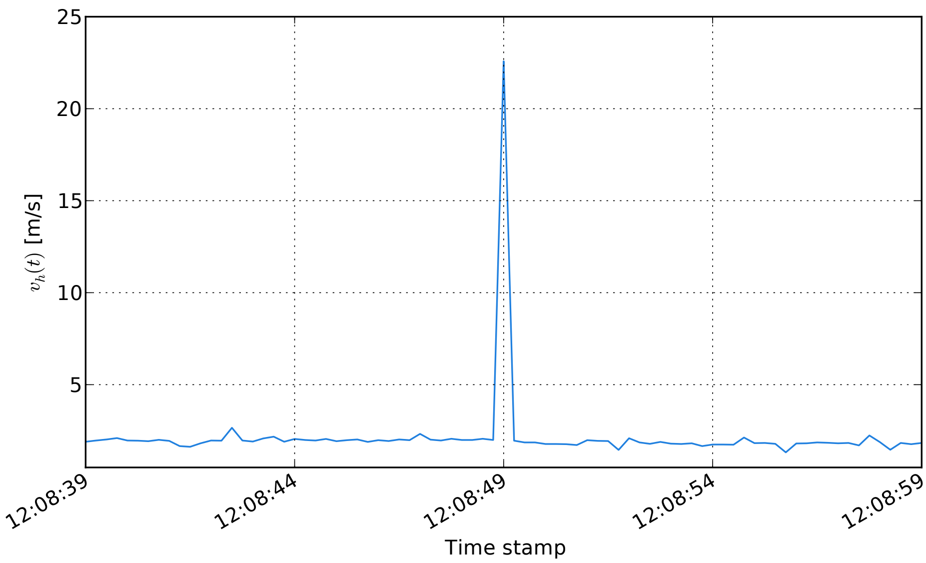

Figure 2.3. Example of wind speed outlier in the data from Teufelsberge (22-08-2006). Within a single second the horizontal wind speed changes from a steady level of around 2 $\frac{\mathrm{m}}{\mathrm{s}}$ to 22 $\frac{\mathrm{m}}{\mathrm{s}}$ and back to the steady 2 $\frac{\mathrm{m}}{\mathrm{s}}$.

2.4.3 Quantization Problems

In addition to wind velocity measurements from an ultra sonic anemometer, measurements from a set of cup anemometers are also part of the Snaresmoor dataset. However, the quantization used in the storage of the measurements is very coarse as can be seen from the example in Figure 2.4. Comparing the wind speed measurements in Figure 1.2 on page 5 and Figure 2.4 it is clear, that the notion of turbulence is vague for the cup anemometer data due to the coarse quantization. The measurements are still useable. However, due to the low accuracy, which must be considered in the interpretation of results based on such measurements, they are not used in the present study.

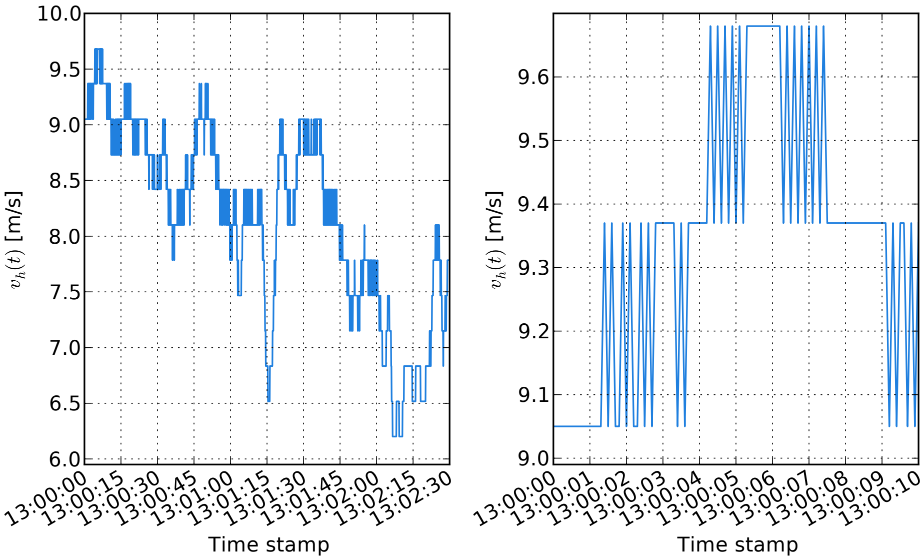

Figure 2.4. Example of very coarse quantization used in the cup anemometer measurements from Snaresmoor M1 Cup3 (30-07-2011). Right figure: An extract of horizontal wind speed measurement over a period of two-and-a-half minute. Right figure: Excerpt of left figure with highlights of the (low) precision of around 0.3 $\frac{\text{m}}{\text{s}}$.

2.5 The Resulting Database

The requirements to the database of wind velocity measurements are described in Section 2.2 on page 9. Sections 2.3 on page 13 and 2.4 on page 15 outlines the processing applied to the available data described in Section 2.1 on page 7. This section provides an overview of the resulting database after all of the conditioning has been applied. A simple evaluation of the performance gain in using an HDF database over CSV files is presented. Finally, an overview of the available post processed wind velocity measurements from the Teufelsberge and Snaresmoor datasets is given.

2.5.1 Performance

To assess the gain in performance (in terms of load time and size) of using an HDF5 based database over plain CSV files, a simple performance comparison has been carried out. The results of this comparison are shown in Table 2.4. The entire Teufelsberge dataset (63 215 050 rows including rows marked as invalid) has been stored in each of the database formats with the resulting file sizes given in the table. The load times have been tested on a 4-way Intel Xeon E7-4850 2.00 GHz (a total of 40 CPU cores) based 64-bit system with 512 GB memory, and SSD hard drives. The software used was the Enthought Python Distribution 7.3-2 (64-bit). The entire dataset was loaded from the SSD hard drive to memory (which is considered a worst case problem). The data in the CSV file was loaded using the `loadtxt` function from the high performance `numpy`[13] library whereas the data in the HDF file was loaded using the `pytables`[14] library. See the script (»tab_2_4.py« on the enclosed CD) used to generate the results in table for the specific details about the storing and loading the dataset.

Database	Load time [s]	Size [GiB]
CSV	2371.6	2.9
HDF5	2.3	1.7

Table 2.4. Measured performance of a CSV file and a HDF5 file as a database. The comparison is on the resulting file size and the time it takes to load the entire dataset from the file. The entire Teufelsberge dataset which contains 63 215 050 rows has been used in the comparison.

From Table 2.4, it is clear that the CSV file is larger than the HDF5 file. However, the most notable difference is with respect to the load time where the HDF5 database is superior.

2.5.2 Data

Measurements are available from four different masts at Snaresmoor (at the four different heights: 20, 40, 60, and 80 meters) and a single mast at Teufelsberge (at the height of 61 meter). In the HDF database the structure is as follows: The data measured using a given anemometer is stored in a table located under a group representing its location, e.g. »/Snaresmoor/Raw_M1_20« contains the measurements from the ultra sonic anemometer mounted on mast 1 at a height of 20 m at Snaresmoor. In the measurement processing discussed in this report, it is important to have large time spans of consecutive measurements. Table 2.5 on the next page provides an overview of the number of seconds of data that is available if requiring consecutive measurements of 10 minutes, 20 minutes, 30 minutes, and 60 minutes, respectively. It can be seen that only very few measurements are available from Snaresmoor if consecutive time spans of more than 20 minutes are required.

[13] Numpy is a state-of-the-art library for scientific computing with Python.
[14] Pytables is a state-of-the-art library for interfacing with HDF5 databases in Python.

Group/Table	10 minutes	20 minutes	30 minutes	60 minutes
/Snaresmoor/Raw_M1_20	18972000	16723200	0	0
/Snaresmoor/Raw_M1_40	16548600	13971600	0	0
/Snaresmoor/Raw_M1_60	20514600	16645200	1800	0
/Snaresmoor/Raw_M1_80	19685400	16612800	0	0
/Snaresmoor/Raw_M2_20	18460800	16221600	0	0
/Snaresmoor/Raw_M2_40	19015800	17788800	0	0
/Snaresmoor/Raw_M2_60	3918000	3279600	0	0
/Snaresmoor/Raw_M2_80	18951000	17630400	0	0
/Snaresmoor/Raw_M4_20	19210800	15264000	0	0
/Snaresmoor/Raw_M4_40	21730800	17848800	0	0
/Snaresmoor/Raw_M4_60	20328600	16192800	1800	0
/Snaresmoor/Raw_M4_80	21075000	17926800	0	0
/Snaresmoor/Raw_M5_20	18907200	16174800	0	0
/Snaresmoor/Raw_M5_40	19525800	17043600	1800	0
/Snaresmoor/Raw_M5_60	17705400	14844000	0	0
/Snaresmoor/Raw_M5_80	14916000	12582000	0	0
/Teufelsberge/T07_HF	15697200	15661200	15624000	15526800

Table 2.5. Overview of tables in the resulting HDF database. All the tables in the resulting HDF database are listed. For each table the number of seconds of available data when requiring consecutive time spans of 10 minutes, 20 minutes, 30 minutes, and 60 minutes, respectively, are listed.

Throughout the report results from the two locations Teufelsberge and Snaresmoor are presented and compared. In order to have a reasonable number of figures and tables, the Raw_M1_60 has been chosen to represent the Snaresmoor location. Thus, in the rest of this report the »Teufelsberge dataset« refers to the data in the HDF table »Teufelsberge/T07_HF« whereas the »Snaresmoor M1 60 dataset« refers to the data in the HDF table »/Snaresmoor/Raw_M1_60«. Note that both datasets are measured at (approximately) the same height.

A set of descriptive statistics of the wind velocity measurements sampled at 4 Hz at Teufelsberge is presented in Table 2.6 on the following page. A set of descriptive statistics of the wind velocity measurements sampled at 10 Hz at Snaresmoor M1 60 is presented in Table 2.7 on the next page. The Teufelsberge data is characterised by having both higher mean wind speed, maximum wind speed and turbulence standard deviation than the Snaresmoor M1 60 data. There also seems to be a slight average upward flow at the location of the anemometer used at Teufelsberge whereas the average flow is downward at the location of the anemometer used at Snaresmoor M1 60.

Statistic	v_{h} $[\frac{\mathrm{m}}{\mathrm{s}}]$	θ [°]	v_{v} $[\frac{\mathrm{m}}{\mathrm{s}}]$
mean	10.56	254.05	0.23
std	5.33	77.49	1.19
min	0.00	0.00	-42.25
max	45.00	360.00	44.97
5%	2.33	98.90	-1.42
10%	3.45	126.90	-0.95
15%	4.54	148.00	-0.68
25%	6.48	229.60	-0.34
50%	10.42	285.90	0.11
75%	14.13	303.70	0.70
85%	16.18	311.80	1.15
90%	17.63	317.80	1.55
95%	19.71	327.70	2.27

Table 2.6. Descriptive statistics of the post processed data. The statistics are the mean value, the standard deviation, the minimum, the maximum, and nine different percentiles. The dataset consists of a total of 63 147 098 measurements (at 4 Hz) **[Teufelsberge]**.

Statistic	v_{h} $[\frac{\mathrm{m}}{\mathrm{s}}]$	θ [°]	v_{v} $[\frac{\mathrm{m}}{\mathrm{s}}]$
mean	6.84	160.27	-0.14
std	2.99	101.01	0.53
min	0.00	0.00	-8.76
max	33.26	360.00	7.76
5%	2.34	12.00	-1.00
10%	3.17	25.00	-0.74
15%	3.80	42.00	-0.58
25%	4.75	76.00	-0.38
50%	6.61	157.00	-0.13
75%	8.65	225.00	0.06
85%	9.87	282.00	0.25
90%	10.74	320.00	0.42
95%	12.12	344.00	0.74

Table 2.7. Descriptive statistics of the post processed data. The statistics are the mean value, the standard deviation, the minimum, the maximum, and nine different percentiles. The dataset consists of a total of 325 414 056 measurements (at 10 Hz) **[Snaresmoor M1 60]**.

3 Assumptions

The mean wind speed V and turbulence standard deviation σ_{V} are central in the Problem Thesis stated in Section 1.3 on page 5. This chapter provides an overview of the current state-of-the-art model of wind velocities, its parameters which are closely related to V and σ_{V}, and its underlying assumptions. Key elements such as choice of coordinate system and the concept of using representative values of the mean wind speed and turbulence standard deviation are discussed. The correspondence between the available measured wind velocities described in Chapter 2 on page 7 and the model is assessed.

3.1 The Load Estimation Setup

The current state-of-the-art load estimation setup is the system shown in Figure 3.1. Figure 3.1 is an elaboration of the concept shown in Figure 1.1 on page 3. Each of the »rows«: Data Acquisition, Data Characterisation, and Load Estimation are discussed separately in reverse order.

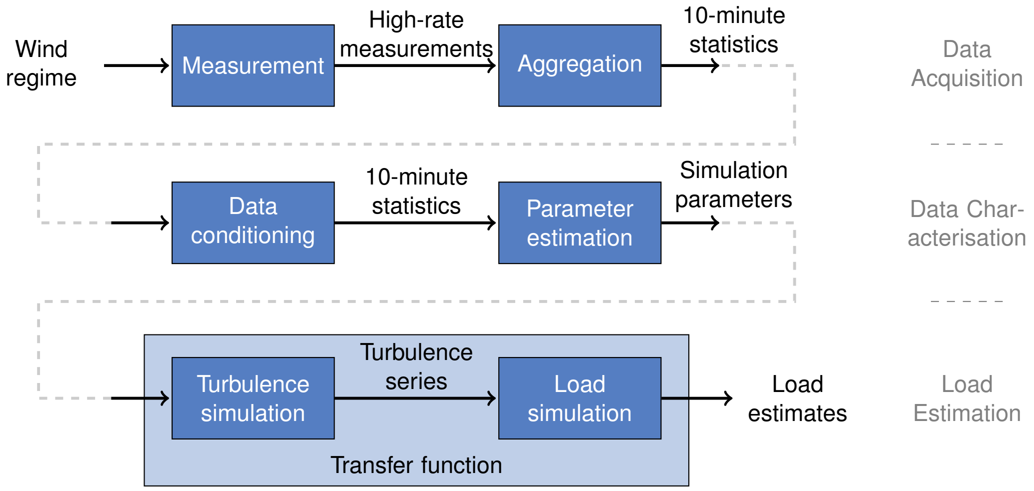

Figure 3.1. Illustration of the state-of-the-art setup used in estimating loads on wind turbines.

3.1.1 Load Estimation

In order to asses loads on a wind turbine, one could, ideally, collect a representative set of measurements of the wind regime at the location of interest and use these measurements as-is for load calculations. In that case only the »Load simulation« in Figure 3.1 is needed. However, the high cost of collecting measurements and the high number of load cases that must be considered (typically several hundred [13]) makes this method less attractive. Instead, the setup in Figure 3.1 is used[1]. The key part of this setup is the use of a transfer function from a set of parameters used in a model of the wind regime to the load estimates.

[1]Based on personal communication with senior specialist, Kim Emil Andersen, Vestas, specialist Frede Aakmann Tøgersen, Vestas, and specialist Jesper Graugaard, Vestas, January 2013

The reasoning behind the use of a transfer function is as follows. Once a model of the wind regime has been introduced, one could use the model to generate a set of time series of wind velocities that are statistically equivalent with the measurements of the wind regime and then use these in load simulations. In that case the input to »Turbulence simulation« in Figure 3.1 on the previous page is a set of parameters used in a model of the wind regime. The output is the generated series which is used in »Load simulations« that in turn output a set of load estimates resulting from the simulations. However, the computational burden of doing the simulations is still be present. On the other hand, if a transfer function from parameters to load estimates is present, the simulations are excess, i.e. no simulations are needed once such a transfer function has been established. Hence, the use of a transfer function eases the load estimation task.

Existing models of the wind regime are based on observing the wind velocity in a point in space in a given time span $t \in [t_0, t_1)$, such that a decomposition in mean wind velocity and turbulence yields [1] ,[7]:

$$\mathbf{v}(t) = \mathbf{V}(t) + \tilde{\mathbf{v}}(t) \tag{3.1}$$

where $\mathbf{V}(t) \in \mathbb{R}^{3\times1}$ is the mean wind velocity, $\tilde{\mathbf{v}}(t) \in \mathbb{R}^{3\times1}$ is a zero mean fluctuating wind velocity (the turbulence), and $\mathbf{v}(t) \in \mathbb{R}^{3\times1}$ is the resulting wind velocity. Typically, a time span of 10 minutes is chosen [1]. The process is considered wide sense stationary although it is known that this assumption is questionable in »longer« time intervals due to the change in the weather [8]. Thus, using longer averaging periods may yield mean wind speeds that are not representative for the entire interval. To have a reasonable length of the interval $[t_0, t_1)$, the trade-off of 10-minute time spans has been adopted [8]. Hence, in 10-minute intervals, the mean wind velocity $\mathbf{V}(t)$ is constant. Generally, $\tilde{\mathbf{v}}(t)$ is considered a wide sense stationary vector valued stochastic process with components being marginally zero mean Gaussian distributed although attempts at using other distributions have been done, see e.g. [14].

Simulations of wind velocities based on the model in (3.1) is done using so-called Fourier simulation [14] - see e.g. [15] and [16] for an introduction to the method. The key in using this method is to know all the relevant auto- and cross-correlations in the turbulence component $\tilde{\mathbf{v}}(\mathbf{t})$.

The state-of-the-art model used to capture the auto- and cross-correlations in the turbulence component $\tilde{\mathbf{v}}(\mathbf{t})$ is Manns spectral tensor [17] that is a linearised solution to the Navier-Stokes equation. Manns spectral tensor is also the recommended model in the IEC61400 standard [4]. The spectral tensor allows for a generalisation of $\tilde{\mathbf{v}}(t)$ to a time varying, three dimensional vector field (in space) of three components since all the relevant auto- and cross-correlations are described by the spectral tensor. The main advantage of the spectral tensor is its ability to model anisotropic wind velocities, i.e. wind velocities with different variances among its components (the isotropic case assumes equal variances among the components). The assumptions taken to make the spectral tensor tractable are stationarity in time, homogenity (stationarity in space), and the correctness of Taylor's frozen turbulence hypothesis [18]. Taylors frozen turbulence hypothesis [19] essentially states that observing a given translation of the field along the direction of the mean wind velocity in space is equal to observing a translation in time such that the distance travelled equals the mean wind speed times the time translation. This mapping between space and time allows for expressing the temporal dimension in terms of the spacial dimension along the mean wind. The assumptions of homogenity is generally a good approximation in the horizontal plane, but not in the vertical plane [18]. The change of wind velocity in the vertical direction is called wind shear and is attempted modelled in the spectral tensor. The assumption of stationarity is, as already mentioned, questionable due to changes in the weather. It is worth noting that the use of discrete Fourier simulations enforce periodicity in the solution. A pragmatic solution to this problem is to simply simulate a large enough field such that the effect of periodicity may be neglected.

3.1.2 Data Characterisation

The middle »row« in Figure 3.1 on page 21 consists of »Data conditioning« and »Parameter estimation«. As discussed in Section 3.1.1 on page 21, the parameters used in the chosen model of the wind regime must be estimated from measured data, which is exactly what is done in »Parameter estimation«. However, since the measurements may include faulty data due to failures in the data acquisition equipment it is necessary to first »condition« it in order to ensure its validity. That is, operations such as outlier filtering, time stamp correction, or sorting based on temperature or humidity may be needed. »Data conditioning« takes care of this.

The model in (3.1) with $\tilde{\mathbf{v}}(t)$ described by Manns spectral tensor is the wind regime model of interest in this report due to its position as the »standard model« in wind engineering and load estimation. In general three parameters are needed to describe the spectral tensor. Different sets of parameters result depending on the exact assumptions forced onto the spectral tensor. The IEC61400 standard assumes a Kaimal spectrum [20] of the wind component along the direction of the mean wind. In that case the resulting parameters are $\gamma \in \mathbb{R}$, $\sigma_\zeta \in \mathbb{R}^+$, and $\ell \in \mathbb{R}$. The shear distortion parameter γ determines the anisotropi in the simulated field. The scale parameter ℓ relates to the spectrum of the component along the direction of the mean wind (see [4] for further details). Finally, σ_ζ is the turbulence standard deviation along the direction of the mean wind.

In (3.1) no explicit coordinate system is stated. However, it is essential to observe that a rectangular right handed coordinate system that is aligned with the mean wind velocity is used. The three components of this coordinate system are denoted $v_\zeta \in \mathbb{R}$, $v_\xi \in \mathbb{R}$, and $v_\psi \in \mathbb{R}$, i.e. $\mathbf{v}(t) = [v_\zeta(t), v_\xi(t), v_\psi(t)]^{\mathrm{T}}$. Thus, v_ζ is aligned with the direction of the mean wind in the time span $t \in [t_0, t_1)$, v_ξ is horizontal and perpendicular to v_ζ, and v_ψ is perpendicular to both v_ζ and v_ξ. The dependence on the time t is generally based on the 10-minute interval concept presented in Section 3.1.1 on page 21. Bearing in mind that the quantities that are described in the rest of the report are, unless otherwise noted, considered in such a 10-minute interval, the notation that explicitly marks the dependence on t is dropped from now on.

By assumption, the mean wind velocity is zero in the vertical direction [18] and hence v_ψ ends up being the vertical component whereas v_ζ and v_ξ span the horizontal plane. This assumption also means that $V_\xi \in \mathbb{R}$ and $V_\psi \in \mathbb{R}$ are zero, i.e the mean wind velocity $\mathbf{V}$ is zero except for the $V_\zeta \in \mathbb{R}^+$ component. An important thing to note is that having made this choice means that the coordinate system (v_ζ, v_ξ, v_ψ) rotates with the mean wind direction in the horizontal plane when looking across different time spans. Hence, it is important to state the exact time span over which the wind velocity is observed in order to calculate the orientation of (v_ζ, v_ξ, v_ψ), if such a relative coordinate system is to be meaningful. Also noteworthy is that when using Mann's spectral tensor only σ_ζ is chosen whereas $\sigma_\xi \in \mathbb{R}^+$ and $\sigma_\zeta \in \mathbb{R}^+$ are controlled by Mann's spectral tensor such that $\sigma_\xi \geq 0.7\sigma_\zeta$ and $\sigma_\psi \geq 0.5\sigma_\zeta$ when the Kaimal spectrum is used [4].

To summarise, the parameters that are used in the model based on (3.1) are the mean wind velocity $\mathbf{V}$ and the three parameters describing Manns spectral tensor: γ, σ_ζ, and ℓ. However, to avoid having to deal with the underlying physics and meteorology of Mann's spectral tensor, γ and ℓ are not discussed further in this report. Thus, the focus is on gaining knowledge about $\mathbf{V}$ and σ_ζ which are the components of the model in (3.1) that relates to the mean wind speed V and turbulence standard deviation σ_{v} introduced in Chapter 1 on page 1.

3.1.3 Data Acquisition

The data acquisition, i.e. the first »row« in Figure 3.1 on page 21 is of course needed in order to obtain the measurements of the wind regime that are central in the discussion in Sections 3.1.1 on page 21 and 3.1.2 on the preceding page. The »Measurement« part of the data acquisition consists of the anemometer that measures the wind regime and outputs wind velocity time series. These high-rate measurements are then aggregated to yield 10-minute statistics. There are two major reasons why aggregation is used[2]: 1) It has been and still is somewhat too computationally demanding to directly use the high-rate measurements in the further processing. 2) Storing high-rate measurements results in dataset sizes far larger than the available storage space on the flash memory that is used in the data loggers. The often hard to reach locations at which the data loggers are positioned during measurement campaigns makes it very expensive to change or empty the data logger memory. Typical 10-minute statistics of a given wind velocity component, say v_a, are the sample mean $\hat{V}_a$, the sample standard deviation $\hat{\sigma}_a$, the sample minimum $v_{a,\min}$, and the sample maximum $v_{a,\max}$ within the 10-minute interval. These are calculated as:

[2]Based on personal communication with senior specialist, Kim Emil Andersen, Vestas, specialist Frede Aakmann Tøgersen, Vestas, and specialist Jesper Graugaard, Vestas, January 2013

$$\hat{V}_a = \frac{1}{N}\sum_{i=1}^{N} v_i, \qquad v_i \in \mathbb{R} \tag{3.2}$$

$$\hat{\sigma}_a = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(v_i - \hat{V}_a)^2}, \qquad v_i \in \mathbb{R}, \qquad \hat{\sigma}_a \geq 0n \tag{3.3}$$

$$v_{a,\min} = \min\{v_1, \ldots, v_N\}, \qquad v_i \in \mathbb{R} \tag{3.4}$$

$$v_{a,\max} = \max\{v_1, \ldots, v_N\}, \qquad v_i \in \mathbb{R} \tag{3.5}$$

where $v_1, \ldots, v_N$ are the N measurements of v_a in the 10-minute interval. In (3.2) to (3.6) any wind velocity component may be substituted for v_a. However, the 10-minute turbulence intensity statistic $\hat{I}$ is only defined for v_ζ [4]. Thus it can be derived from $\hat{V}_\zeta$ and $\hat{\sigma}_\zeta$ as:

$$\hat{I} = \frac{\hat{\sigma}_\zeta}{\hat{V}_\zeta} \tag{3.6}$$

Coordinate Systems

As mentioned in Section 2.2.2 on page 10 the reference coordinate system in which the data is stored in the HDF database is the $(v_\mathrm{h}, \theta, v_\mathrm{v})$ coordinate system. Thus, a change of coordinates to the (v_ζ, v_ξ, v_ψ) coordinate system is necessary. In this change of coordinates, the orientation of the (v_ζ, v_ξ, v_ψ) is based on the mean wind velocity $\mathbf{V}$ in 10-minute periods. The change of coordinates is done in two steps. First a non-linear transformation to the absolute rectangular coordinates $(v_\mathrm{x}, v_\mathrm{y}, v_\mathrm{z})$ is used:

$$v_\mathrm{x} = v_\mathrm{h} \cdot \cos(\theta) \tag{3.7}$$

$$v_\mathrm{y} = v_\mathrm{h} \cdot \sin(\theta) \tag{3.8}$$

$$v_\mathrm{z} = v_\mathrm{v} \tag{3.9}$$

Then the a (linear) rotation of the $(v_\mathrm{x}, v_\mathrm{y}, v_\mathrm{z})$ coordinate system is used:

$$\begin{bmatrix} v_\zeta \\ v_\xi \\ v_\psi \end{bmatrix} = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_\mathrm{x} \\ v_\mathrm{y} \\ v_\mathrm{z} \end{bmatrix} \tag{3.10}$$

where the rotation angle $\alpha \in [0, 360)$ is determined from:

$$\alpha = \tan^{-1}\left(\frac{V_\mathrm{y}}{V_\mathrm{x}}\right) \tag{3.11}$$

The horizontal wind velocity coordinates $v_\mathrm{x} \in \mathbb{R}$ and $v_\mathrm{y} \in \mathbb{R}$ are defined according to Figure 3.2 whereas $v_\mathrm{z} \in \mathbb{R}$ is the vertical wind velocity component. The corresponding mean wind velocities are $V_\mathrm{x} \in \mathbb{R}$, $V_\mathrm{y} \in \mathbb{R}$, and $V_\mathrm{z} \in \mathbb{R}$. The transformation from $(v_\mathrm{x}, v_\mathrm{y}, v_\mathrm{z})$ to (v_ζ, v_ξ, v_ψ) is a rotation by the angle α, in order to make v_ζ coincide with mean wind velocity direction in the horizontal plane as depicted in Figure 3.3.

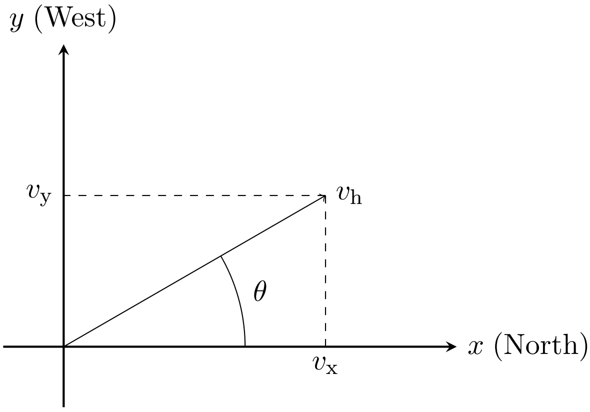

Figure 3.2. Illustration of the horizontal plane transformation from $(v_\mathrm{h}, \theta, v_\mathrm{v})$ coordinates to $(v_\mathrm{x}, v_\mathrm{y}, v_\mathrm{z})$ coordinates.

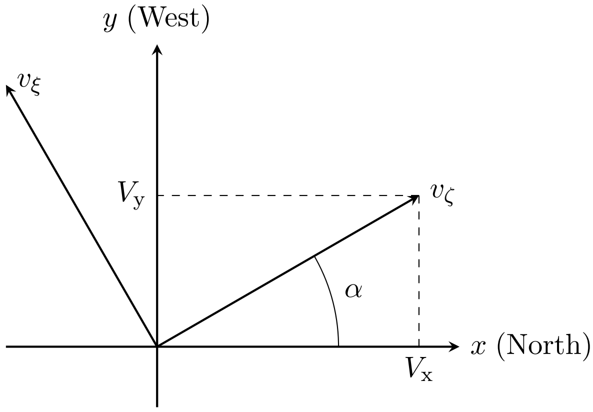

Figure 3.3. Illustration of the horizontal plane rotation from $(v_\mathrm{x}, v_\mathrm{y}, v_\mathrm{z})$ coordinates to (v_ζ, v_ξ, v_ψ) coordinates.

Taking an average of $\mathbf{V}$ over a time span of 10 minutes is too short a period to ensure that the resulting V_ξ and V_ψ are zero (if they even should be). However, V_ξ is forced to zero due to the rotation of the coordinate system in the horizontal plane. The consequence is that V_ξ is zero whereas V_ψ is not, since no vertical rotation is applied in the change of coordinates. Whether this behaviour is to be characterised as an error or not is essentially related to the rather philosophical question of what is the »best« time span to average over. There is no obvious answer to the question of how to »distribute« the horizontal mean wind velocity between V_ζ and V_ξ if such effects of finite averaging period is to be taken into account.

3.1.4 Representative Parameter Values

Restricting the time span used in (3.1) to 10 minutes implies the need for an assessment of the behaviour across different 10-minute intervals. Looking across different 10-minute intervals, different mean wind velocities $\mathbf{V}$ and turbulence standard deviations σ_ζ, σ_ξ, and σ_ψ are observed. Hence, 10-minute statistics of the important parameters V_ζ and σ_ζ follow some distributions. This in turn allows for choosing representative values of V_ζ and σ_ζ when the distributions are know. e.g. a representative value may be chosen as the mean value or a percentile. Using representative values simplifies load estimation since only scalar values and not distributions of V_ζ and σ_ζ are used. However, imposing the limitation of only using a single value, means that this value must be chosen conservatively.

The IEC 61400 standard prescribes that following distributions are assumed [4]:

10-minute mean wind speed distribution

The 10-minute mean wind speed at hub height, i.e. the mean wind speed at the height of the centre of the rotor swept area, is assumed to be Rayleigh distributed:

$$f_{V_\zeta}(\acute{v}_\zeta) = \frac{\acute{v}_\zeta}{\sigma_\mathrm{s}} \cdot \exp\left(-\frac{\acute{v}_\zeta^2}{2\sigma_\mathrm{s}}\right), \qquad \acute{v}_\zeta \geq 0 \tag{3.12}$$

where σ_s is related to V_ref by: $0.2V_\mathrm{ref} = \sqrt{\frac{\pi}{2}\sigma_\mathrm{s}^2}$ for the standard wind turbine classes.

10-minute turbulence standard deviation distribution

The 10-minute turbulence standard deviation conditioned on the mean wind speed at hub height is assumed to be log-normal distributed:

$$f_{\sigma_\zeta|V_\zeta}(\acute{\sigma_\zeta}|\acute{v}_\zeta) = \frac{1}{\acute{\sigma_\zeta}\sigma'\sqrt{2\pi}} \cdot \exp\left(-\frac{(\ln(\acute{\sigma_\zeta}) - \mu')^2}{2\sigma'^2}\right), \qquad \acute{\sigma_\zeta} > 0 \tag{3.13}$$

where μ' and σ'^2 are the mean and variance of the underlying normal distribution, respectively, such that: $\mathbb{E}[\sigma_\zeta|V_\zeta] = I_\mathrm{ref}(0.75V_\zeta + 3.8) = \exp\left(\mu' + \frac{\sigma'^2}{2}\right)$ and $\mathrm{Var}(\sigma_\zeta|V_\zeta) = (1.4I_\mathrm{ref})^2 = \exp\left(2\mu' + \sigma'^2\right)\left(\exp\left(\sigma'^2\right) - 1\right)$[3].

Note how (3.12) states the relationship between the distribution of V_ζ and the representative value V_ref and how (3.13) states the relationship between V_ζ and σ_ζ and the representative value I_ref. The representative values V_ref and I_ref are the two values used to represent a wind turbine site in the IEC 61400 standard as described in Chapter 1 on page 1.

To motivate the Rayleigh distribution of V_ζ note the following: If a vector $\mathbf{a} \in \mathbb{R}^2$ has components a_1, a_2 that are i.i.d. zero mean Gaussian with variance σ_a^2 then the length of the resulting vector, i.e. $\sqrt{a_1^2 + a_2^2}$, is Rayleigh distributed with the parameter σ_a [21]. Hence, if the two underlying horizontal wind speed components from which V_ζ in (3.12) emerges are i.i.d $\mathcal{N}(0, \sigma_\mathrm{s})$, then the parameter σ_s is indeed the standard deviation of this underlying Gaussian distribution.

3.2 Observations Across 10-Minute Intervals

The following sections provide an evaluation of the fit of the distributions in (3.12) and (3.13) to a subset of the wind measurements provided by Vestas, i.e. the Teufelsberge and Snaresmoor M1 60 datasets described in Chapter 2 on page 7. The total number of 10-minute statistics available are 26162 for Teufelsberge and 34191 for Snaresmoor M1 60.

[3]For further details about the relation between μ' and σ' and the mean and standard deviation of the log-normal distribution see e.g. `http://mathworld.wolfram.com/LogNormalDistribution.html`.

An evaluation of a fit to a given distribution may be based on a variety of different methods including statistical test, measures of goodness of fit or graphical methods such as the Q-Q plot or histograms. The quantities to be evaluated are statistics, i.e. they are quantities derived from the data. If methods such as statistical test or goodness of fit measures are used in the evaluation, it is very easy to end up with a measure that is derived from a measure with in turn is derived from the data. This nested hierarchy of measures makes it very difficult if not impossible to give a sensible interpretation of the results. Hence, in order to ease the interpretation of the results and obtain as clear a presentation of the results as possible, the evaluation is based on a graphical comparison of histograms of the quantities with maximum likelihood estimates of their assumed distributions.

3.2.1 Histograms

Histograms generally provide an easily comprehensible graphical presentation of the distribution of the data. However, it is of key importance to choose a reasonable width of the bins used, or equivalently, the number of bins. Having more bins gives a better resolution but also results in less samples falling in each bin. In the »language« of estimation this means a smaller bias at the cost of a higher variance of the estimator when more bins are used [22].

Different criteria for choosing an optimal bin width in different senses may be found in the literature, see e.g. [23] for an overview. If the mean squared error (MSE) criteria is used as a measure of optimality, then it can be shown that a bin width equal to $1/\sqrt{N}$ (where N is the number samples) minimises the MSE as $N \to \infty$ [22]. The quantities to be shown in histograms in the following have ranges from 2 to 30. The number of samples in the histograms range from 100 to nearly 34000. Hence, according to the MSE criteria, the number of bins to be used in the histograms range from $2 \cdot \sqrt{100} = 20$ to $30 \cdot \sqrt{34000} \approx 5532$. However, another very important criteria is the »external limitations«. These are the resolution that makes sense to use in a histogram to be printed on paper and the human ability to compare several different histograms. During the writing of this report, the authors have found that the use of a fixed number of bins across all histograms, makes the comparison and interpretation much easier. Note that the MSE criteria suggest a number of bins closer to 20 than 5532 for most of the histograms shown in the following sections. Combining the MSE criteria with the paper print requirement and the requirement of a fixed number of bins, 100 uniformly spaced bins has been decided on as the number of bins to be used.

3.2.2 Maximum Likelihood Estimation

Maximising the likelihood of observing a set of outcomes given an assumed model is probably the most common method for deriving estimators. It results in estimators that are optimal in the sense of maximising the chosen likelihood function which seems to be a natural approach when the choice of a given distribution must be evaluated since only the available data is used in the assessment and consequently not any prior belief about the shape of the distribution. Given a set of outcomes $x_1, \ldots, x_N$ of a random variable X, the likelihood function, i.e. the joint probability of observing the outcomes given a vector of parameters $\boldsymbol{\theta}$ in an assumed model, is:

$$\mathcal{L} = f_{X_1,\ldots,X_N}(x_1,\ldots,x_N|\boldsymbol{\theta}), \qquad \mathcal{L} \in \mathbb{R} \tag{3.14}$$

Below are derivations of the maximum likelihood (ML) estimators of the distributions of interest in the evaluation. All of the derivations rely on an assumption of observing i.i.d. outcomes $x_1, \ldots, x_N$. The assumption of independence is obviously not fulfilled when looking at statistics calculated from small set of adjacent wind measurements. The datasets used in the evaluation do, however, consist of a very large set of measurements that are separated into a lot of non-consecutive segments. The missing »bands« between the segments probably makes the assumption of independence more likely to be fulfilled although it is still questionable.

Gaussian Distribution

The likelihood function for a Gaussian random variable with mean μ and variance σ^2 is [22]:

$$\mathcal{L} = f_{X_1,\ldots,X_N}(x_1,\ldots,x_N|\mu,\sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \tag{3.15}$$

From this the log-likelihood function $\ln\text{-}\mathcal{L}$ and its partial derivatives may be found:

$$\ln\text{-}\mathcal{L} = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2 \tag{3.16}$$

$$\frac{\partial \ln\text{-}\mathcal{L}}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\mu) \tag{3.17}$$

$$\frac{\partial \ln\text{-}\mathcal{L}}{\partial\sigma^2} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{N}(x_i-\mu)^2 \tag{3.18}$$

Equating the partial derivatives to zero and solving the resulting set of equations yields the estimators:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3.19}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 \tag{3.20}$$

Thus, the ML estimators of the mean μ and variance σ^2 of a Gaussian random variable have been shown to be given by (3.19) and (3.20), respectively.

Log-normal Distribution

If a random variable X is log-normal distributed then $Y = \ln(X)$ is Gaussian with mean μ' and variance σ'^2. Using the pdf of the Log-normal distribution in (3.13) and same procedure as for the Gaussian distribution, the result is:

$$\mathcal{L} = f_{X_1,\ldots,X_N}(x_1,\ldots,x_N|\mu',\sigma') = \prod_{i=1}^{N} \frac{1}{x_i\sigma'\sqrt{2\pi}}\exp\left(-\frac{(\ln(x_i)-\mu')^2}{2\sigma'^2}\right) \quad (3.21)$$

$$\ln\text{-}\mathcal{L} = -\sum_{i=1}^{N} x_i - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\sigma'^2 - \frac{1}{2\sigma'^2}\sum_{i=1}^{N}(\ln(x_i)-\mu')^2 \quad (3.22)$$

$$\frac{\partial \ln\text{-}\mathcal{L}}{\partial \mu'} = \frac{1}{\sigma'^2}\sum_{i=1}^{N}(x_i-\mu') \quad (3.23)$$

$$\frac{\partial \ln\text{-}\mathcal{L}}{\partial \sigma'^2} = -\frac{N}{\sigma'} + \frac{1}{\sigma'^3}\sum_{i=1}^{N}(x_i-\mu')^2 \quad (3.24)$$

$$\hat{\mu}' = \frac{1}{N}\sum_{i=1}^{N}\ln(x_i) \quad (3.25)$$

$$\hat{\sigma}'^2 = \frac{1}{N}\sum_{i=1}^{N}(\ln(x_i)-\mu')^2 \quad (3.26)$$

Thus, the ML estimators of the mean μ' and variance σ'^2 of the underlying Gaussian random variable of a log-normal random variable have been shown to be given by (3.25) and (3.26), respectively

Rayleigh Distribution

The pdf of the Rayleigh distribution is given in (3.12). For the Rayleigh distribution the same procedure as above yields:

$$\mathcal{L} = f_{X_1,\ldots,X_N}(x_1,\ldots,x_N|\sigma) = \prod_{i=1}^{N} \frac{x_i}{\sigma^2}\exp\left(-\frac{x_i^2}{2\sigma^2}\right) \quad (3.27)$$

$$\ln\text{-}\mathcal{L} = \sum_{i=1}^{N}\ln(x_i) - N\cdot\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}x_i^2 \quad (3.28)$$

$$\frac{\partial \ln\text{-}\mathcal{L}}{\partial \sigma} = -\frac{2N}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{N}x_i^2 \quad (3.29)$$

$$\hat{\sigma} = \sqrt{\frac{1}{2N}\sum_{i=1}^{N}x_i^2} \quad (3.30)$$

Thus, the ML estimator of the parameter σ of a Rayleigh random variable have been shown to be given by (3.30).

3.2.3 Rayleigh Mean Wind Speed

Figures 3.4 and 3.5 on the next page show that the Rayleigh distribution is somewhat capable of describing $\hat{V}_\zeta$ although the fit is not perfect. It seems that $\hat{V}_\zeta$ might as well be modelled by a Gaussian distribution, especially at Snaresmoor M1 60. It therefore seems less likely that the underlying components of $\hat{V}_\zeta$ are i.i.d. $\mathcal{N}(0, \sigma_s)$ in 10-minute intervals.

As stated in Section 3.1.2 on page 23, V_ξ and V_ψ are assumed to be zero. Although not shown in the figures, it is indeed the case that $\hat{V}_\xi$ is (forced) zero due to the orientation of the (v_ζ, v_ξ, v_ψ) coordinate system. However, it is clear from Figures 3.4 and 3.5 on the following page that $\hat{V}_\psi$ is not zero in 10-minute intervals. At Teufelsberge it seems that $\hat{V}_\psi$ is closely resembled by a Guassian distribution. At Snaresmoor M1 60 a Gaussian fit is less convincing. The most interesting thing to note though is that $\hat{V}_\psi$ does not seem to be zero mean, i.e. at Teufelsberge it seems that there is an average upward flow whereas an average downward flow appears to be present at Snaresmoor M1 60.

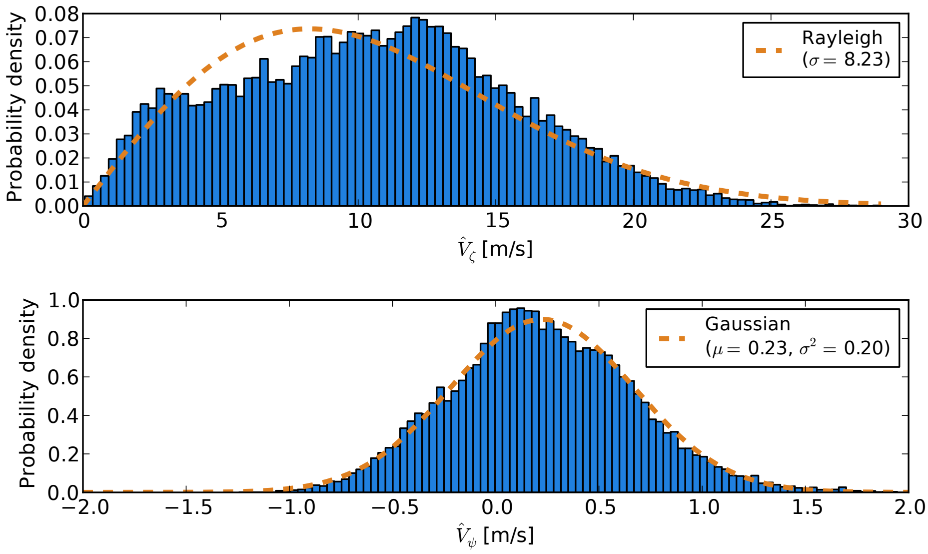

Figure 3.4. Probability density histograms of the mean wind speed in 10-minute intervals. The orange dashed curves show ML estimates of the assumed distributions **[Teufelsberge]**.

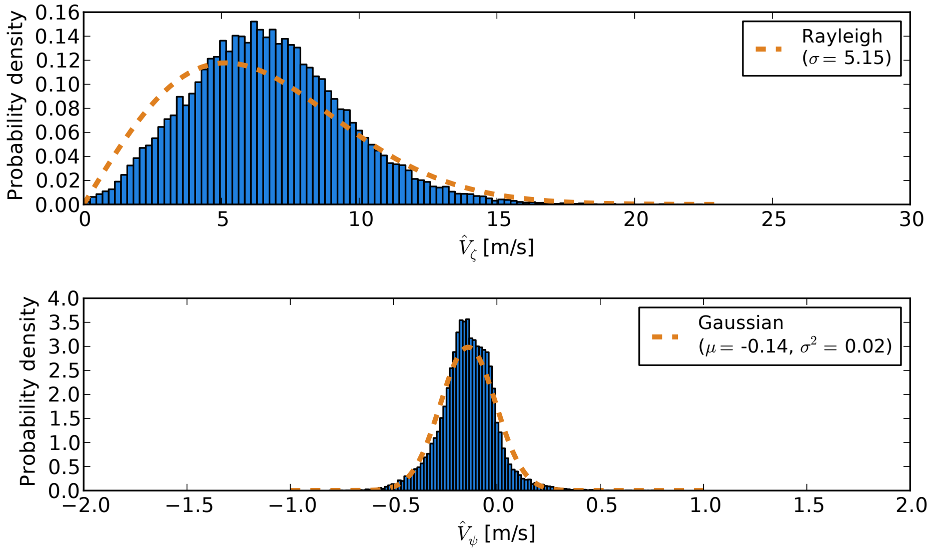

Figure 3.5. Probability density histograms of the mean wind speed in 10-minute intervals. The orange dashed curves show ML estimates of the assumed distributions **[Snaresmoor M1 60]**.

3.2.4 Log-Normal Turbulence Standard Deviation

As stated in (3.13) it is assumed that the 10-minutes turbulence standard deviation σ_ζ *conditioned on* V_ζ is log-normal distributed. The data available to assess this assumption consist of $(\hat{\sigma}_\zeta, \hat{V}_\zeta)$ pairs. Hence, in order to plot probability density histograms it is necessary to make a 2-dimensional discretisation, i.e. both $\hat{\sigma}_\zeta$ and $\hat{V}_\zeta$ must be binned. A resolution of 100 bins is used for $\hat{\sigma}_\zeta$ in accordance with the discussion in Section 3.2.1 on page 28. A fixed bin width of $1\,\frac{\text{m}}{\text{s}}$ is used for the binning of $\hat{V}_\zeta$ as suggested in [24].

Figure 3.6 on the facing page shows probability density histograms of $\hat{\sigma}_\zeta$ for three different bins of $\hat{V}_\zeta$ at Teufelsberge. The top figure shows the bin which holds the most samples. The log-normal fit generally seems to capture the distribution of σ_ζ. However, as is evident from the mid and bottom figures, the fit is less convincing when only 100 samples are available. Based on the three bins in the figure, it seems that there is a positive correlation between σ_ζ and V_ζ. This tendency is confirmed by the scatter plot in Figure 3.7 on the next page of all the $(\hat{\sigma}_\zeta, \hat{V}_\zeta)$ pairs.

Figures 3.8 on page 34 to 3.11 on page 36 display the entire 2-dimensional discretisation, i.e. the all combinations of bins of $\hat{\sigma}_\zeta$ and $\hat{V}_\zeta$. Each »column« in Figure 3.8 on page 34 is a histogram like the ones shown in Figure 3.6 on the next page whereas each »column« of Figure 3.9 on page 35 shows the corresponding fitted distributions at Teufelsberge. The probability densities are coded according to the colour scale on the right. Figures 3.10 on page 35 and 3.11 on page 36 depicts the corresponding histograms and fitted distributions at Snaresmoor M1 60.

Several interesting points can be made from these figures. It seems that the log-normal distribution is a reasonable model for $\hat{\sigma}_\zeta$. The mean and 90 % percentiles look very similar between the histograms and the fitted distributions. Also the structure of the colours look alike, even though it is clear that the relative low number of samples in each bin makes the histograms appear less smooth. Especially the Snaresmoor M1 60 data is

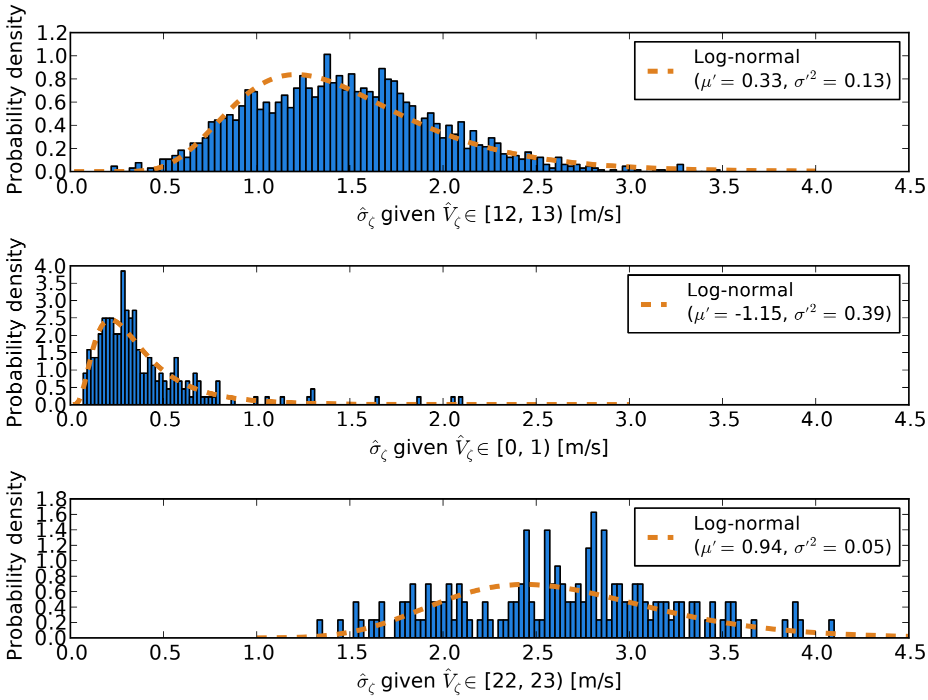

Figure 3.6. Probability density histograms of the wind speed standard deviation in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The mean wind speed is divided into bins. Top: The wind speed bin with the most data samples. Middle: The smallest wind speed with at least 100 samples in the bin. Bottom: The largest wind speed with at least 100 samples in the bin. The orange dashed curves show ML estimates of the assumed distributions **[Teufelsberge]**.

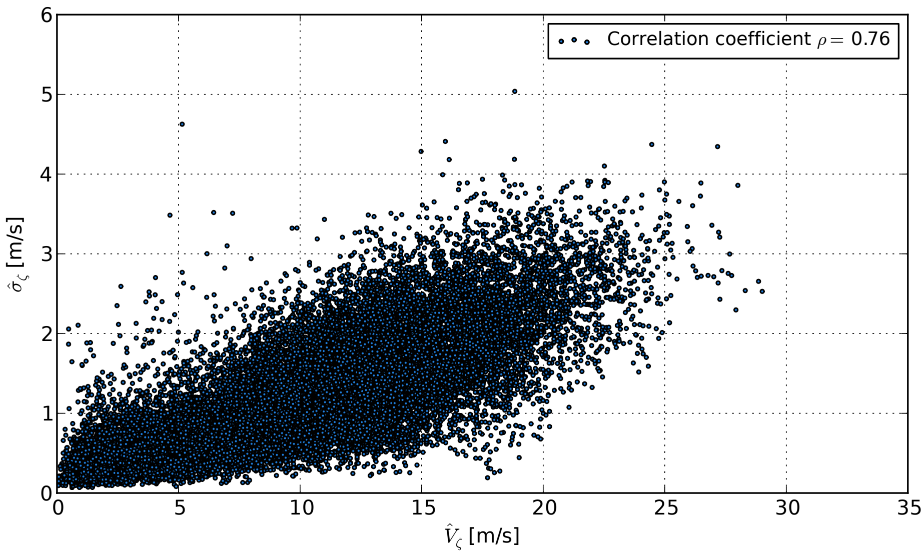

Figure 3.7. Scatter of wind speed standard deviation versus mean wind speed in 10-minute intervals **[Teufelsberge]**.

quite scatted across the entire set of bins. Also the 90 % percentile at Snaresmoor M1 60 shows differences between the histograms and the fitted distributions. Again, this is most likely a result of the rather low number of samples in each bin, especially at very low or very high values of $\hat{V}_\zeta$.

The strong positive correlation between $\hat{\sigma}_\zeta$ and $\hat{V}_\zeta$ is also evident from these figures. It seems reasonable to approximate this relation between $\hat{\sigma}_\zeta$ and $\hat{V}_\zeta$ with an affine function at Teufelsberge. At Snaresmoor M1 60 an affine relation is less convincing, though.

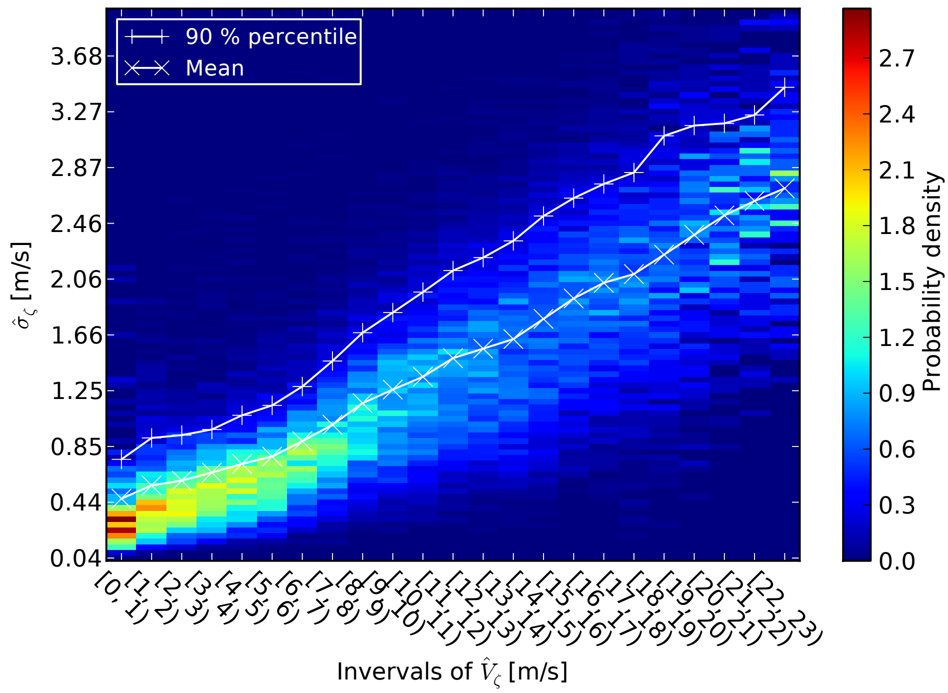

Figure 3.8. Probability density histograms of the wind speed standard deviation in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The mean wind speed is divided into bins. Only bins that contain at least 100 samples are shown. The density is colour coded according to the scale on the right. Superimposed on the density is the calculated mean and 90 % percentile of the samples. **[Teufelsberge]**.

In the IEC 61400 standard, the representative value of σ_ζ is chosen as the 90 % percentile at the given hub height wind speed. For the wind turbine standard classes this value is assumed to be described by the affine function [4]:

$$\sigma_\zeta = I_{\text{ref}}(0.75V_\zeta + 5.6) \tag{3.31}$$

Figures 3.12 on page 37 and 3.13 on page 37 show the 90 % percentiles of both $\hat{\sigma}_\zeta$ and $\hat{I}$ versus $\hat{V}_\zeta$. The solid blue lines of $\hat{\sigma}_\zeta$ are the same 90 % percentile lines as shown in Figures 3.8 and 3.10 on the facing page (the 90 % percentiles estimated directly from data) whereas the dashed orange lines are the same as shown in Figures 3.9 on the next page and 3.11 on page 36 (the 90 % percentiles calculated from the fitted log-normal distribution). Also shown are the IEC subclass lines based on (3.31). Note that the

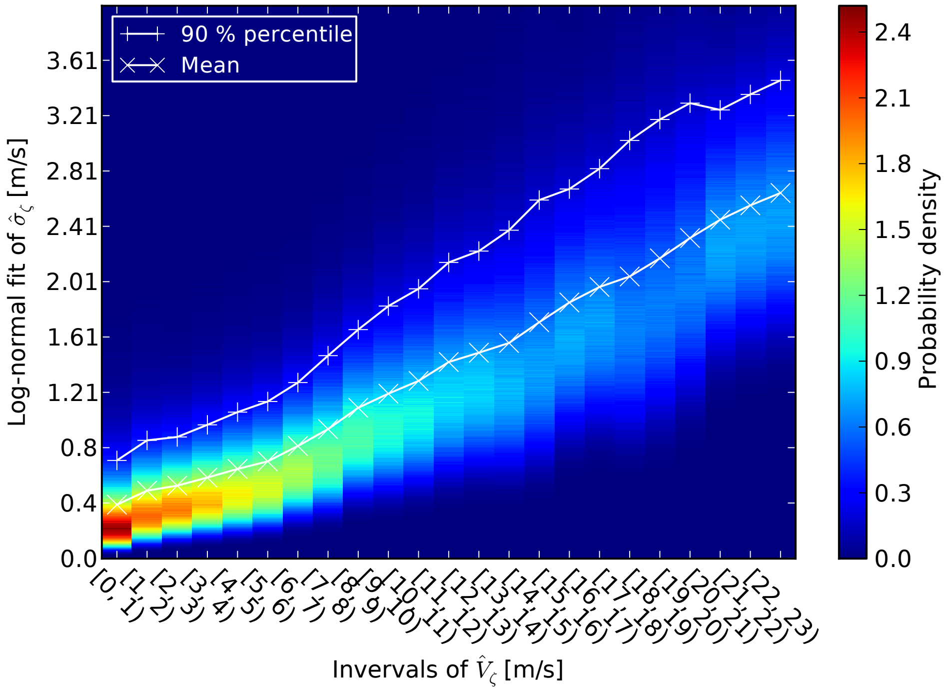

Figure 3.9. Log-normal ML fits calculated from the wind speed standard deviation in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The mean wind speed is divided into bins. Only bins that contain at least 100 samples are shown. The density is colour coded according to the scale on the right. Superimposed on the density is the mean and 90 % percentile resulting from the fit **[Teufelsberge]**.

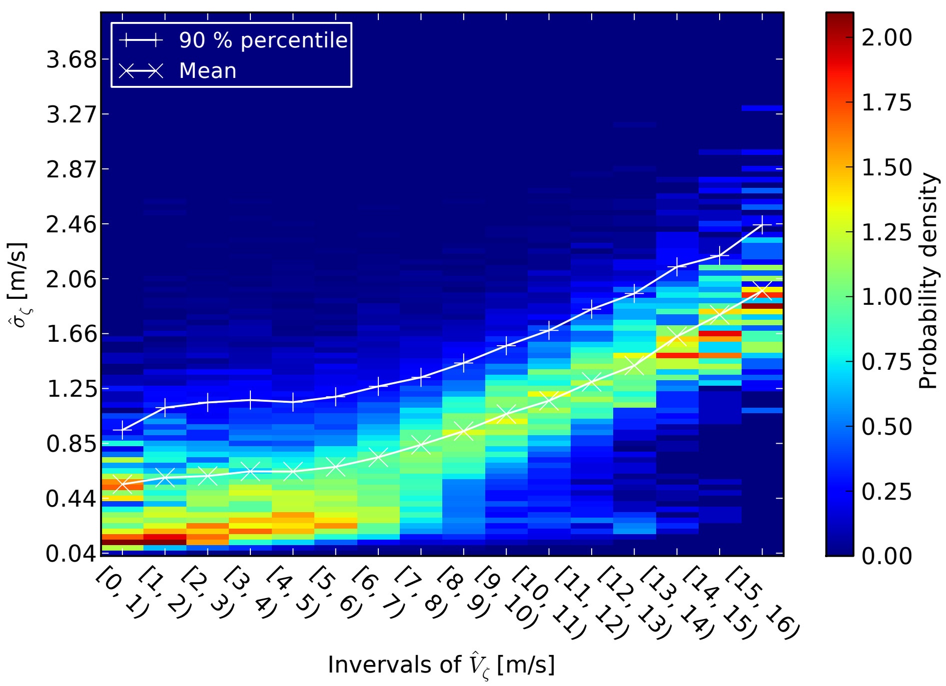

Figure 3.10. Probability density histograms of the wind speed standard deviation in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The mean wind speed is divided into bins. Only bins that contain at least 100 samples are shown. The density is colour coded according to the scale on the right. Superimposed on the density is the calculated mean and 90 % percentile of the samples. **[Snaresmoor M1 60]**.

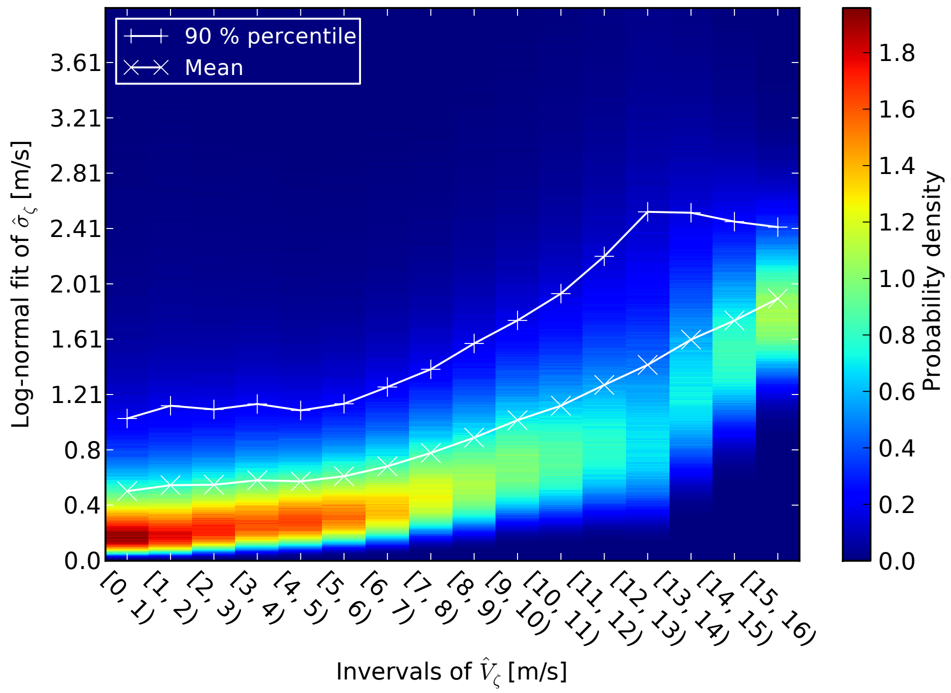

Figure 3.11. Log-normal fits calculated from the wind speed standard deviation in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The mean wind speed is divided into bins. Only bins that contain at least 100 samples are shown. The density is colour coded according to the scale on the right. Superimposed on the density is the mean and 90 % percentile resulting from the fit **[Snaresmoor M1 60]**.

corresponding figures showing the subclass lines in the IEC 61400 standard have been updated in the amendment A1 to IEC 61400 [25].

Again, it seems that the affine relation between σ_ζ and V_ζ is a better approximation at Teufelsberge than at Snaresmoor. The 90 % percentiles of $\hat{\sigma}_\zeta$ estimated directly from the data are generally lower than the corresponding 90 % percentiles of $\hat{\sigma}_\zeta$ calculated from the fitted log-normal distribution. The corresponding 90 % percentiles of $\hat{I}$ do, however, seem to coincide. The choice of either IEC subclass (A, B, or C) is not obvious from the figures due to the dependence on the mean wind speed. This is properly best illustrated by the 90 % percentile of $\hat{\sigma}_\zeta$ calculated from ML fit to the Snaresmoor M1 60 data and shown in Figure 3.13 on the next page. This 90 % percentile crosses all three IEC class lines throughout the sequence of mean wind speed bins. Also, for some mean wind speeds, the two ways of estimating the 90 % percentiles yields lines that cross different IEC class lines.

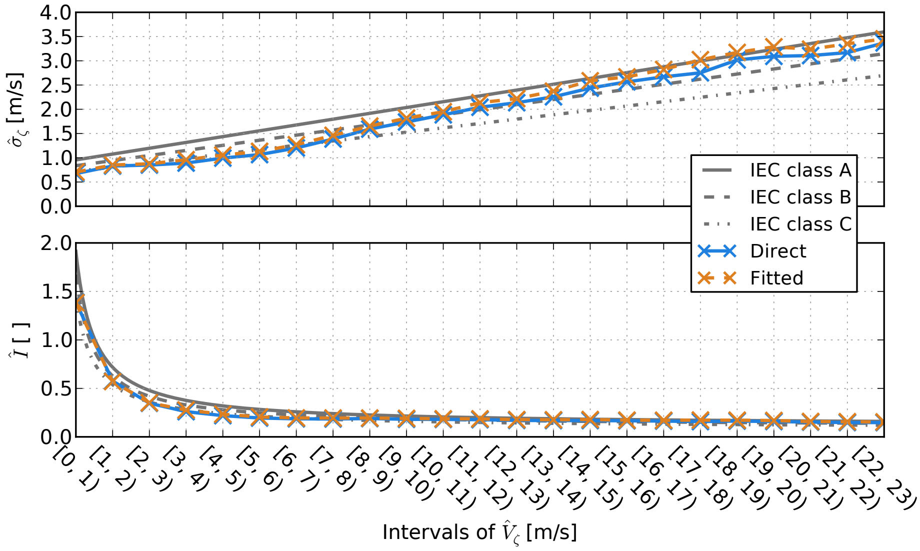

Figure 3.12. 90 % percentiles (blue and orange curve) of the wind speed standard deviation and turbulence intensity in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The 90 % percentiles marked *fitted* are calculated from a ML fit of a log-normal distribution to the samples whereas the 90 % percentiles marked *direct* are estimated directly from the samples. The mean wind speed is divided into bins. Only bins that contain at least 100 samples are shown. Also shown are the corresponding curves of the IEC 61400 standard classes based on the midpoint of the mean wind speed bins **[Teufelsberge]**.

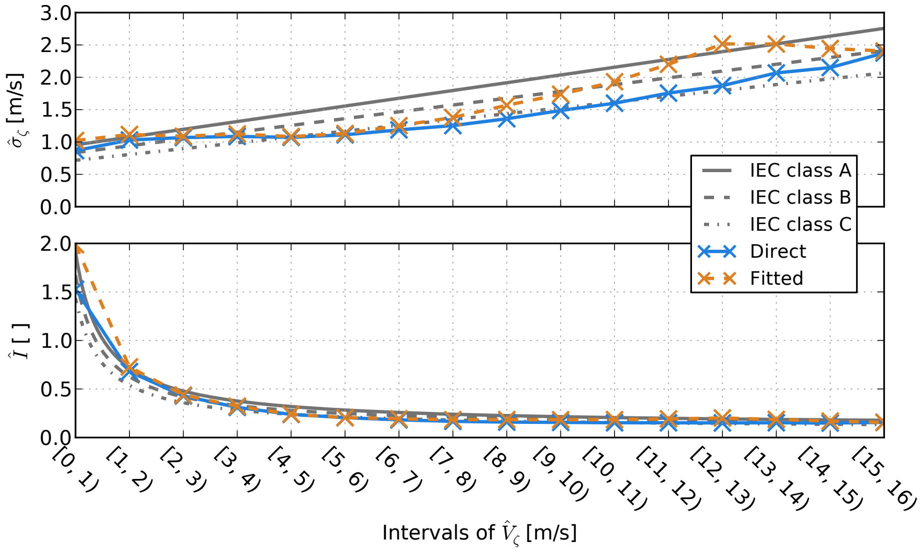

Figure 3.13. 90 % percentiles (blue and orange curve) of the wind speed standard deviation and turbulence intensity in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The 90 % percentiles marked *fitted* are calculated from a ML fit of a log-normal distribution to the samples whereas the 90 % percentiles marked *direct* are estimated directly from the samples. The mean wind speed is divided into bins. Only bins that contain at least 100 samples are shown. Also shown are the corresponding curves of the IEC 61400 standard classes based on the midpoint of the mean wind speed bins **[Snaresmoor M1 60]**.

3.2.5 Convergence Analysis

When choosing representative values of V_ζ and σ_ζ based on aggregated statistics, it is important to have »enough« data to ensure that the distributions of the statistics have converged to a stable state, i.e. the distributions only exhibit negligible change if more data is taken into account. This section provides an assessment of the number of 10-minute intervals that must be used in order to obtain stable estimates of V_ζ and σ_ζ. The »stability« of the estimated distributions of V_ζ and σ_ζ is judged by the stability of their first and second moment, i.e. mean and variance. Other (more complex) measures of stability could have been used. However, since the use of measures as simple as the mean and variance results in the need for examining statistics of statistics of statistics, it seems (at least to the authors) nearly impossible to cope with the interpretation of more complex measures.

When looking across different samples, i.e. different sets of 10-minute statistics, a stable state is consequently reflected in a small variance of the mean of the 10-minute statistic and small variance of the variance of the 10-minute statistic. These measures are plotted versus samples size, i.e. number of 10-minute intervals used in their evaluation, in Figures 3.14 on the facing page to 3.17 on the next page. The decay in the stability measures closely follows a straight line in the log-log plots. Hence, the decay seems to be described by a power law. The fluctuating behaviour seen in the figures for larger sample sizes is probably a consequence of the limited number of samples available for evaluation of such large sample sizes. It generally, seems that around 3000 – 4000 samples are needed in order have the stability measures drop below 10^{-1}. It can always be argued when this measure is small »enough«. However, 10^{-1} seems to be a reasonable threshold given that $\hat{V}_\zeta$ is typically in the range of 10 to 15 and $\hat{\sigma}_\zeta$ is typically in the range of 0.5 to 2. Nevertheless, 3000 – 4000 samples is only a small percentage of the nearly 25000 available samples a Teufelsberge and the nearly 34000 available samples at Snaresmoor M1 60.

The figures have been created as follows: In order to obtain as uncorrelated samples as possible, 20000 ten minute intervals have been chosen at random from the given location. The mean and variance of successively larger portions (up to one fourth) of the samples have been calculated. All possible portions have been taken into account, e.g. a sample size of 20 results in 1000 portions whereas a sample size of 50 results in 400 portions. The variance of the mean and variance across the portions have then been plotted versus sample size. The crosses in the figures mark the sample sizes that have been evaluated. The lines are a linear interpolations between the crosses.

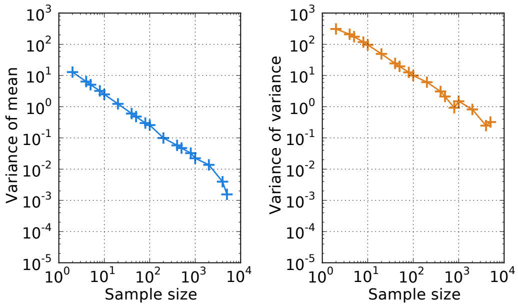

Figure 3.14. Statistics of statistics of the 10-minute mean statistic $\hat{V}_\zeta$ versus sample size, i.e. the number of 10-minute intervals. **[Teufelsberge]**

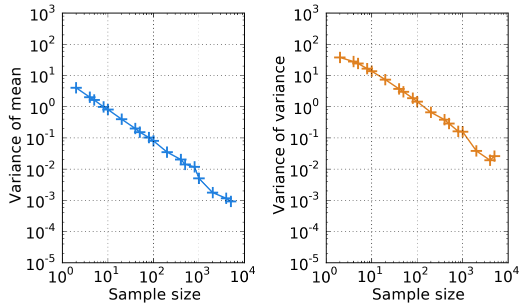

Figure 3.15. Statistics of statistics of the 10-minute mean statistic $\hat{V}_\zeta$ versus sample size, i.e. the number of 10-minute intervals. **[Snaresmoor M1 60]**

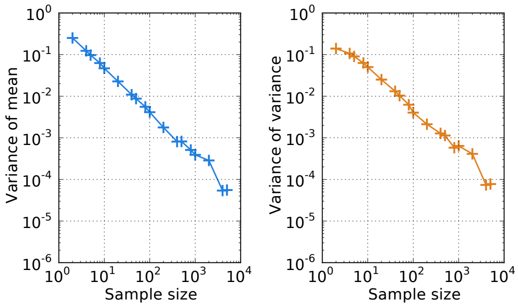

Figure 3.16. Statistics of statistics of the 10-minute mean statistic $\hat{\sigma}_\zeta$ versus sample size, i.e. the number of 10-minute intervals. **[Teufelsberge]**

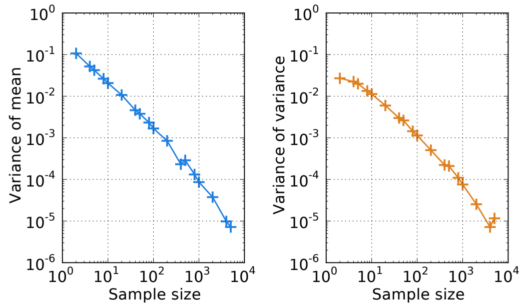

Figure 3.17. Statistics of statistics of the 10-minute mean statistic $\hat{\sigma}_\zeta$ versus sample size, i.e. the number of 10-minute intervals. **[Snaresmoor M1 60]**

3.3 Observations Within 10-Minute Intervals

The wind model (3.1) presented in Section 3.1.1 on page 21 is based on observing the mean wind speed in a time span of 10 minutes. However, one may ask whether the estimates of V_ζ and σ_ζ change if the aggregation period is chosen differently?

If changes in the weather in a time span of 10 minutes are so minimal that they do not have an influence on the statistics $\hat{V}_\zeta$ and $\hat{\sigma}_\zeta$, then the distribution of $\hat{V}_\zeta$ and $\hat{\sigma}_\zeta$ should stay the same if aggregated to, say, 5 minute intervals instead of 10-minute intervals. That is, if all the high-rate measurements were aggregated to 5 minute intervals coinciding with the 10-minute intervals used up until now, then the resulting distributions of $\hat{V}_\zeta$ and $\hat{\sigma}_\zeta$ should be invariant to this change of aggregation period. To evaluate if this is actually true, statistics resulting from the non-uniformly spaced aggregation periods $t_{\text{agg}} \in [30, 40, 50, 60, 75, 100, 120, 150, 200, 300, 600]$ seconds are compared. Note that all of the aggregation periods are integral multiples of 600 seconds. The comparison is done using 2-dimensional probability density plots like the ones from Section 3.2.4 on page 32 - the only difference is that the abscissa now displays t_{agg}.

In analysis of the impact of the choice of aggregation period on the mean wind speed and turbulence standard deviation, the following notation (which is illustrated in Figure 3.18 on the following page) is used: N is used to denote the number of samples of a given quantity in a 10-minute interval. $x[k]$ denotes the k'th sample of the sequence x for $k = 0, \ldots, M$ of a given quantity. This sequence x is a concatenation of samples in 10-minute intervals. Thus, M is an integer times N, more specifically M is N times the number of available 10-minute intervals. The number of samples in a sub-interval of a 10-minute interval is denoted by L, e.g. if a 5-minute sub-interval is considered, then $L = N/2$. Only sub-intervals that are part of a 10-minute interval are used in the analysis.

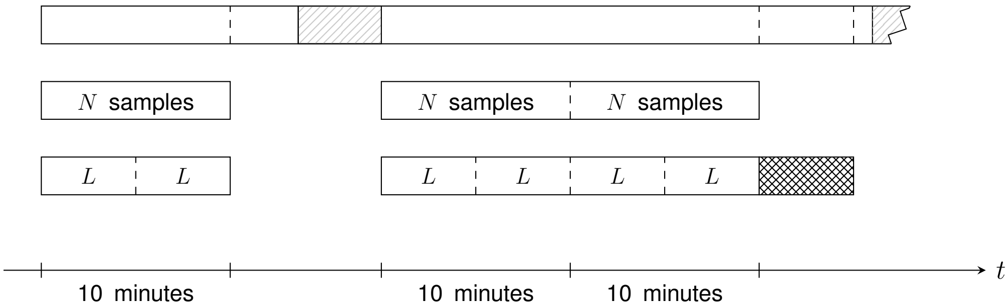

Figure 3.18. Illustration of intervals and sub-intervals used in the discussion of the impact of the choice of aggregation period. The first row illustrates all available samples with the gray hatched parts being missing samples. All the samples are divided into 10-minute intervals containing N samples each. These are then divided into 5-minute intervals containing L samples each. The black crosshatched 5-minute interval is not used since it is not part of a 10-minute interval.

3.3.1 Mean Wind Speed

Figure 3.19 on the next page indicates that the distribution of V_ζ does indeed seem invariant to change in the aggregation period at Teufelsberge. Especially, the mean value of $\hat{V}_\zeta$ across all considered intervals of a given aggregation period does not change with aggregation period. This comes as no surprise since:

$$\bar{x} = \frac{1}{M}\sum_{k=0}^{M-1} x[k] = \frac{N}{M}\sum_{m=0}^{\frac{M}{N}-1} \frac{L}{N}\sum_{n=0}^{\frac{N}{L}-1} \frac{1}{L}\sum_{k=0}^{L-1} x[k + nL + mN] \tag{3.32}$$

That is, calculating the mean value is a linear operation. Thus, theoretically, taking the average of a sequence of averages of averaged sub-intervals yields the same result as taking the average of all the samples in all the sub-intervals. Looking closely at the *mean*-markers in Figure 3.19 on the facing page, it is seen that their values decrease for increasing aggregation period. The specific change with aggregation period is displayed in Table 3.1 on the next page. Due to the theoretical result in (3.32), the difference in mean value must be due to numerical issues in computing the mean values.

The less smooth distribution displayed in Figure 3.19 on the facing page towards the higher values of t_{agg} is likely explained by the fewer available intervals for such high values compared to the lower values of t_{agg}, e.g. twice as many 5-minute intervals as 10-minute intervals are available from a fixed number of samples.

The same conclusions as stated above for the Teufelsberge data can be drawn from the Snaresmoor M1 60 data which is why only the result from Teufelsberge is depicted.

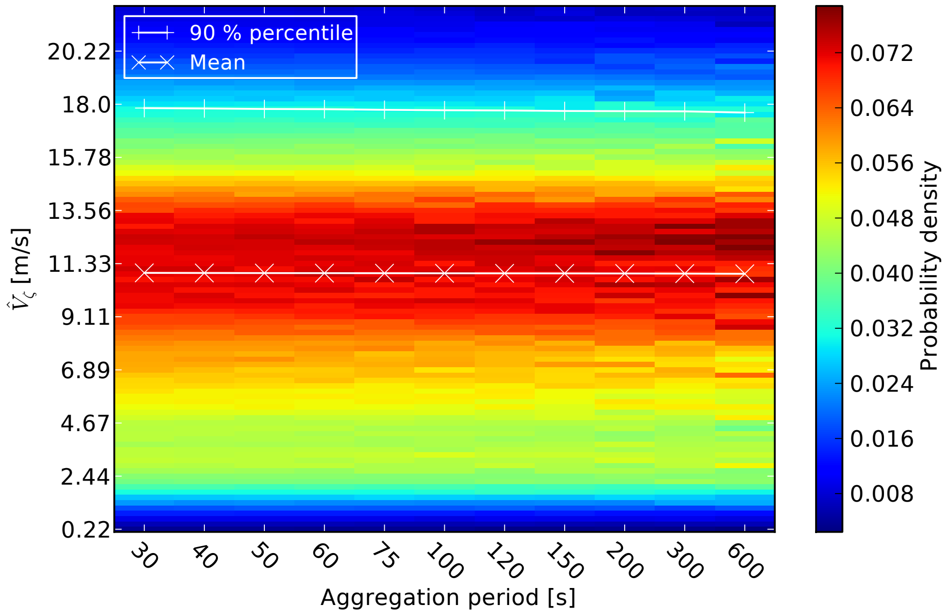

Figure 3.19. Probability density histograms of the mean wind speed versus aggregation period. The density is colour coded according to the scale on the right. Superimposed on the density is the calculated mean and 90 % percentile of the samples. Note that the aggregation period scale is not uniform **[Teufelsberge]**.

Aggregation period [s]	Mean Value $[\frac{\mathrm{m}}{\mathrm{s}}]$
»Raw«	10.467
30	10.502
40	10.498
50	10.494
60	10.492
75	10.488
100	10.485
120	10.482
150	10.480
200	10.477
300	10.472
600	10.464

Table 3.1. The marker values of the *mean* line in Figure 3.19. That is, mean value of the estimated mean wind speed versus aggregation period. »Raw« denotes the mean calculated directly from all M samples **[Teufelsberge]**.

3.3.2 Turbulence Standard Deviation

Figures 3.20 to 3.22 on the facing page show the change in the distribution of the turbulence standard deviation statistics $\hat{\sigma}_\zeta$, $\hat{\sigma}_\xi$, and $\hat{\sigma}_\psi$ across the different aggregation periods. Only the results from Teufelsberge are shown since the results are essentially the same for the data from Snaresmoor M1 60. A clear tendency in the data for all three components σ_ζ, σ_ξ, and σ_ψ is that the mean values of their distributions increase as the aggregation period increase. Also, the density seems to be less clustered when the aggregation period increase.

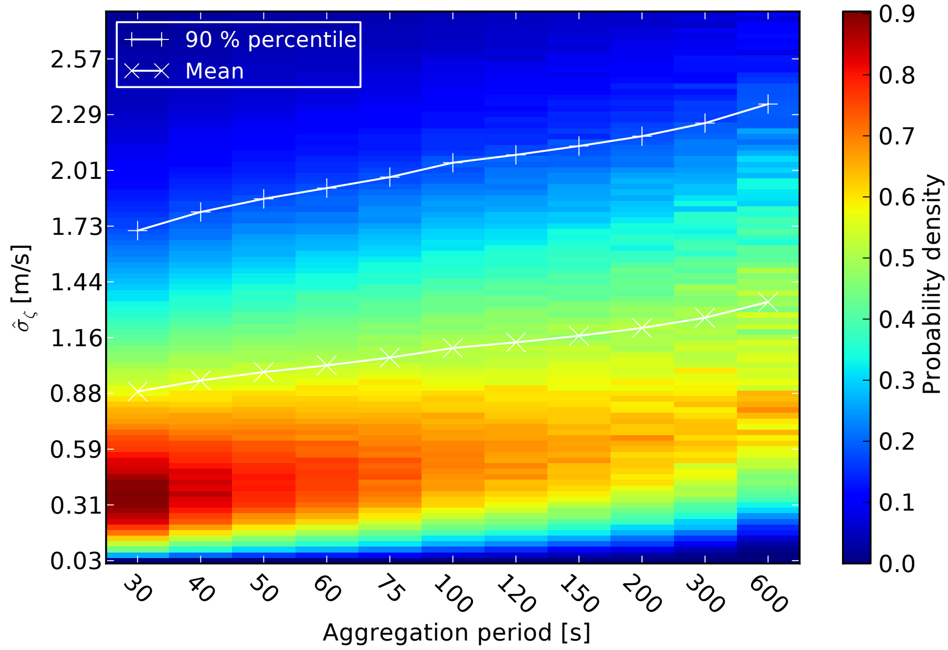

Figure 3.20. Probability density histograms of the wind speed standard deviation (of $\hat{\sigma}_\zeta$) versus aggregation period. The density is colour coded according to the scale on the right. Superimposed on the density is the calculated mean and 90 % percentile of the samples. Note that the aggregation period scale is not uniform **[Teufelsberge]**.

In the following, bounds on the mean values of the statistics $\hat{\sigma}_\zeta$, $\hat{\sigma}_\xi$, and $\hat{\sigma}_\psi$ are derived. These bounds are then used to give an interpretation of the increase in the means of the statistics $\hat{\sigma}_\zeta$, $\hat{\sigma}_\xi$, and $\hat{\sigma}_\psi$ with increasing aggregation period.
Let $\bar{x}_{k_0,k_1}$ denote the mean of $x[k]$ for $k = k_0, \ldots, k_1 - 1$:

$$\bar{x}_{k_0,k_1} = \frac{1}{k_1 - k_0} \sum_{k=k_0}^{k_1-1} x[k] \tag{3.33}$$

Define the function $d(q)$ as the sum of squared errors between the samples x and the value q in the interval L :

$$d(q) = \frac{1}{L} \sum_{k=0}^{L-1} (x[k] - q)^2 = q^2 - 2q \cdot \bar{x}_{0,L} + \frac{1}{L} \sum_{k=0}^{L-1} x[k]^2, \qquad q \in \mathbb{R} \tag{3.34}$$

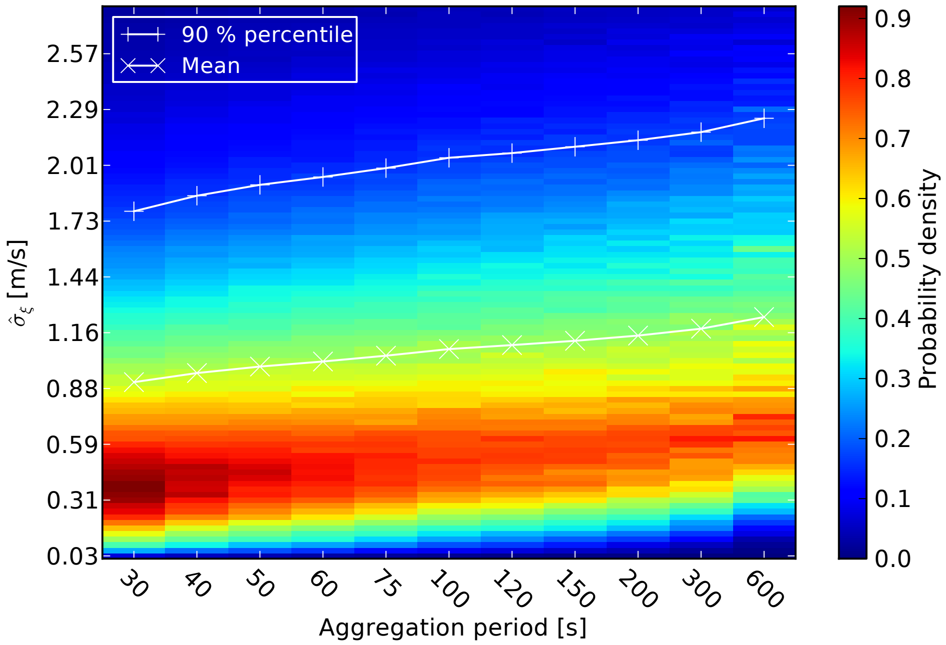

Figure 3.21. Probability density histograms of the wind speed standard deviation (of $\hat{\sigma}_\xi$) versus aggregation period. The density is colour coded according to the scale on the right. Superimposed on the density is the calculated mean and 90 % percentile of the samples. Note that the aggregation period scale is not uniform **[Teufelsberge]**.

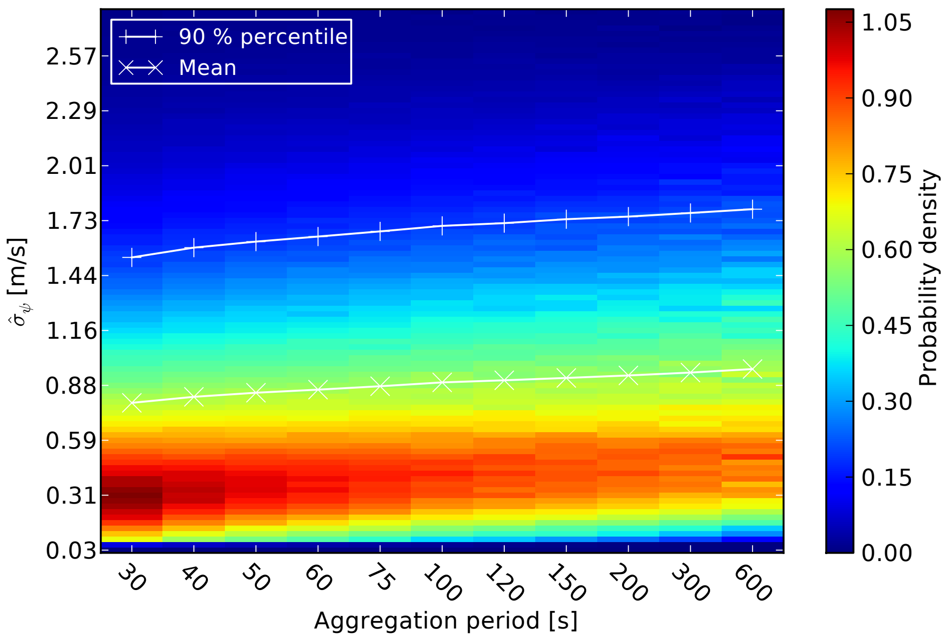

Figure 3.22. Probability density histograms of the wind speed standard deviation (of $\hat{\sigma}_\psi$) versus aggregation period. The density is colour coded according to the scale on the right. Superimposed on the density is the calculated mean and 90 % percentile of the samples. Note that the aggregation period scale is not uniform **[Teufelsberge]**.

Observe that $d(\bar{x}_{0,L})$ is the sample variance of $x[k]$ for $k = 0, \dots, L-1$:

$$d(\bar{x}_{0,L}) = -\bar{x}_{0,L}^2 + \frac{1}{L}\sum_{k=0}^{L-1} x[k]^2 \tag{3.35}$$

Now for any offset $r \in \mathbb{R}$ from $\bar{x}_{0,L}$:

$$\begin{aligned} d(\bar{x}_{0,L} + r) &= \bar{x}_{0,L}^2 + r^2 + 2r\bar{x}_{0,L} - 2\bar{x}_{0,L}^2 - 2r\bar{x}_{0,L} + \frac{1}{L}\sum_{k=0}^{L-1} x[k]^2 && (3.36)\\ &= -\bar{x}_{0,L}^2 + r^2 + \frac{1}{L}\sum_{k=0}^{L-1} x[k]^2 && (3.37)\end{aligned}$$

Which leads to the following result:

$$d(\bar{x}_{0,L} + r) - d(\bar{x}_{0,L}) = r^2 \tag{3.38}$$

That is, the difference between the »variance« $d(\bar{x}_{0,L} + r)$ calculated using a mean value offset by r and the sample variance $d(\bar{x}_{0,L})$ is r^2. The unique minimiser of this difference is $r = 0$. Consequently, the following bound on the estimated variance in a sub-interval of L samples emerges:

$$\frac{1}{L}\sum_{k=0}^{L-1}(x[k] - \bar{x}_{0,N})^2 \geq \frac{1}{L}\sum_{k=0}^{L-1}(x[k] - \bar{x}_{0,L})^2 \tag{3.39}$$

with equality if and only if $\bar{x}_{0,L} = \bar{x}_{0,N}$, i.e. if and only if the mean value in a sub-interval of the 10-minute interval is equal to the mean value in the 10-minute interval. This leads to the following bounds on the sample standard deviation of x in a 10-minute interval:

$$\sqrt{\frac{1}{N}\sum_{k=1}^{N}(x[k] - \bar{x}_{0,N})^2} = A \overset{①}{\geq} B \overset{②}{\geq} C \tag{3.40}$$

where:

$$A = \sqrt{\frac{L}{N}\sum_{n=0}^{\frac{N}{L}-1}\frac{1}{L}\sum_{k=0}^{L-1}(x[k+nL] - \bar{x}_{0,N})^2} \tag{3.41}$$

$$B = \sqrt{\frac{L}{N}\sum_{n=0}^{\frac{N}{L}-1}\frac{1}{L}\sum_{k=0}^{L-1}(x[k+nL] - \bar{x}_{nL,(n+1)L})^2} \tag{3.42}$$

$$C = \frac{L}{N}\sum_{n=0}^{\frac{N}{L}-1}\sqrt{\frac{1}{L}\sum_{k=0}^{L-1}(x[k+nL] - \bar{x}_{nL,(n+1)L})^2} \tag{3.43}$$

Here, A is the sample standard deviation of x in a 10-minute interval, B is the square root of the average sample variance of x in each of the N/L sub-intervals of the 10-minute interval, and C is the average sample standard deviation of x in each of the N/L sub-intervals. The first bound ① follows from (3.39) and the fact that the square root is a monotonically increasing function. The second bound ② follows from the use of the Cauchy-Schwarz Inequality (see the proof of remark 8.7 ii in [26] for the details).

With respect to the mean values of the estimated wind speed turbulence standard deviation $\hat{\sigma}_\zeta$ shown in Figure 3.20 on page 42, the following can now be concluded:

1. Due to the inequality ①, the mean of the estimated turbulence standard deviation $\hat{\sigma}_\zeta$ increases for longer aggregation periods if the mean value of v_ζ is not the same in all sub-intervals.
2. Due to the inequality ②, the mean of the estimated turbulence standard deviation $\hat{\sigma}_\zeta$ increases for longer aggregation periods if the standard deviation σ_ζ is not the same in all sub-intervals.
3. The mean value of $\hat{\sigma}_\zeta$ is constant across different aggregation periods if and only if both the mean of v_ζ and the standard deviation σ_ζ is the same in all sub-intervals.

The same conclusions hold for σ_ξ and σ_ψ in Figures 3.21 on page 43 and 3.22 on page 43. Thus, the increasing mean of the estimated turbulence standard deviation for longer aggregation periods is due to either the mean V_ζ, the standard deviation σ_ζ, or both not being the same in all sub-intervals. The mean values of B and C for the components v_ζ, v_ξ, v_ψ for the Teufelsberge data are shown in Table 3.2. Comparing the values in the table, it is seen that none of the inequalities ① or ② are satisfied with equality. Thus, for all three components v_ζ, v_ξ, and v_ψ, both the mean and the standard deviation are different across sub-intervals.

Measure		30 s	40 s	50 s	60 s	75 s	100 s	120 s	150 s	200 s	300 s	600 s
v_ζ	B	0.87	0.93	0.97	1.00	1.03	1.08	1.10	1.13	1.17	1.21	1.28
	C	0.83	0.89	0.93	0.96	1.00	1.05	1.08	1.11	1.15	1.20	1.28
v_ξ	B	0.90	0.94	0.97	0.99	1.02	1.04	1.06	1.08	1.10	1.13	1.18
	C	0.85	0.90	0.93	0.96	0.99	1.02	1.04	1.06	1.09	1.12	1.18
v_ψ	B	0.77	0.79	0.81	0.82	0.84	0.85	0.86	0.87	0.88	0.89	0.90
	C	0.73	0.76	0.78	0.80	0.82	0.84	0.85	0.86	0.87	0.89	0.90

Table 3.2. Mean values (across all intervals of the respective period) of the measures B and C in (3.42) and (3.43), respectively, calculated from the data illustrated in Figures 3.20 on page 42 (v_ζ), 3.21 on page 43 (v_ξ), and 3.22 on page 43 (v_ψ) **[Teufelsberge]**.

3.4 Assumption Summary

Having read Chapters 1 to 3 of the present report, the reader will likely agree that load estimation for wind turbines is a very involved task in which attention must be payed to the details in order to fully comprehend the subject. In this section an overview of most critical parts of the analysis that has been presented up until now is provided.

Section 3.1 on page 21 has introduced the setup used in estimation of loads on wind turbines. This setup is based on the model of the wind regime stated in 3.1 on page 22, which is reproduced below:

$$\mathbf{v}(t) = \mathbf{V}(t) + \tilde{\mathbf{v}}(t) \tag{3.1}$$

Here, the mean wind velocity $\mathbf{V}(t)$ is modelled by a constant in short time spans of 10 minutes whereas $\tilde{\mathbf{v}}(t)$ is a wide sense stationary Gaussian process described by Mann's spectral tensor. The wind velocities are described in a relative rectangular coordinate system, (v_ζ, v_ξ, v_ψ), aligned with the mean wind velocity such that $\mathbf{V}(t) = [V_\zeta, 0, 0]$ in 10-minute intervals.

Based on the available conditioned wind velocity measurements described in Chapter 2 on page 7, Sections 3.1 on page 21 to 3.3 on page 39 provides an analysis of underlying »mechanics« of the model in (3.1). These are:

1. The assumption of constant mean wind speed V_ζ and constant turbulence standard deviation components σ_ζ, σ_ξ, and σ_ψ in 10-minute intervals (Assessed in Sections 3.3.1 on page 40 and 3.3.2 on page 42).
2. The assumption of $\tilde{v}_\zeta$, $\tilde{v}_\xi$, and $\tilde{v}_\psi$ being marginally zero mean Gaussian distributed (Not assessed in this report).
3. The assumption of the 10-minute mean wind speed V_ζ being Rayleigh distributed (Assessed in Section 3.2.3 on page 31).
4. The assumption of the 10-minute turbulence standard deviation conditioned on the mean wind speed $\sigma_\zeta | V_\zeta$ being log-normal distributed (Assessed in Section 3.2.4 on page 32).
5. The positive correlation between 10-minute turbulence standard deviation σ_ζ and 10-minute mean wind speed V_ζ (Assessed in Section 3.2.4 on page 32).

Items 1. and 2. are related to an analysis *within* 10-minute intervals whereas items 3. to 5. examine the behaviour *across* different 10-minute intervals. Looking across different 10-minute intervals provides for a single *representative value* of a given quantity for the entire dataset. However, the justification of such a representative value is based on the validity of the model in (3.1) and its underlying assumptions. Thus, it is interesting to note that analysis revealed that the assumption of constant mean and turbulence standard deviation in 10-minute intervals is not fulfilled.

4 Trend Classification

The problem thesis stated in Section 1.3 on page 5 raises the question: "what information about »trends« in the mean wind speed V and turbulence standard deviation σ_{v} can be extracted from high-rate wind measurements?" However, if there are no trends in the high-rate sampled wind velocities, obviously, no information about trends can be extracted from these measurements. By visually inspecting parts of the measurements, examples of both time intervals with and without visible trend are observed. Figure 4.1 shows an example of a time interval without any apparent trend whereas Figure 4.2 on the next page shows an example with apparent trend.

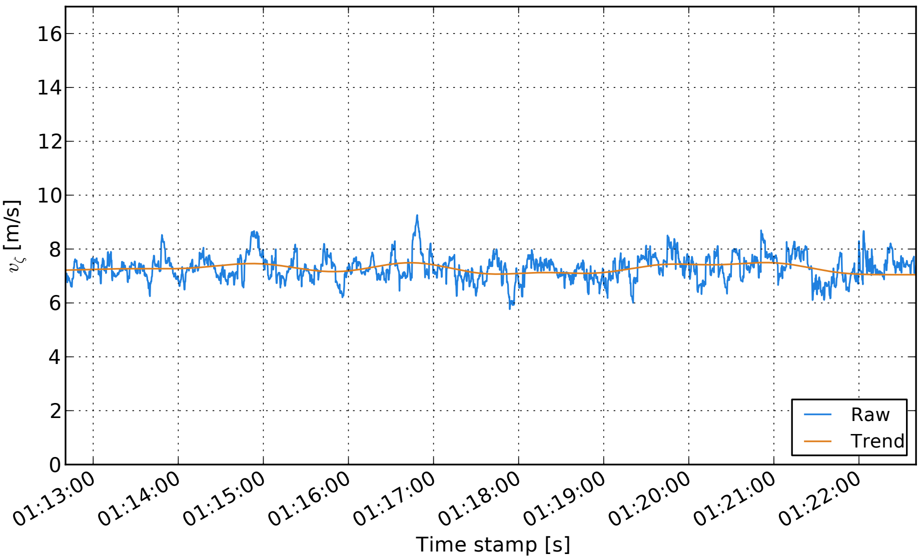

Figure 4.1. Example of measured horizontal wind speed in a 10-minute interval. This example contains no trend. The data is from Teufelsberge (07-11-2006).

However, visual inspections of parts of the measurements cannot reveal the full picture about the presence of trend. Section 3.4 on the facing page concludes that, in general, the mean wind speed and turbulence standard deviation are not constant in 10-minute intervals, which further indicates that the sampled wind velocities may contain trends. This chapter is concerned with determining if this is indeed the case. Thus the non-overlapping 10-minute intervals investigated in the previous chapter are classified as containing either a significant trend or an insignificant trend.

Section 4.1 on the next page first introduces formal notions of the types of trends considered and the significance of these trends. These notions are used in designing a classification system in Section 4.2 on page 52 which can automatically perform the required classification. Finally, the results of the classification are presented in Section 4.3 on page 57 to conclude if there are indeed trends in the high-rate sampled wind velocities.

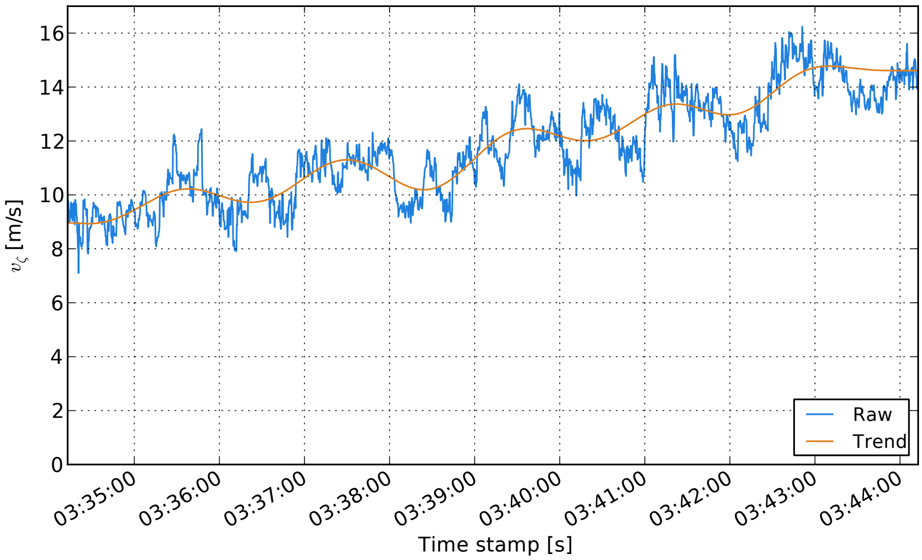

Figure 4.2. Example of measured horizontal wind speed in a 10-minute interval. This example contains a linear trend with a periodic trend superimposed. The data is from Teufelsberge (29-11-2006).

4.1 Classes of Trends

As stated, the trend classification covered in this chapter must classify 10-minute intervals of measurements as containing either a significant trend or an insignificant trend. However, before this can be done, it must be established what significant trends and insignificant trends are. Furthermore, a distinction between linear trends and periodic trends is made. This section is thus concerned with introducing the notions of significant trend, insignificant trend, linear trend, and periodic trend.

4.1.1 Significant and Insignificant Trends

Obviously, the distinction between significant trends and insignificant trends depends entirely on the application in mind. In the case of this report, the application is the estimation of loads on wind turbine components and this must somehow be reflected in the notions of significant trend and insignificant trend. One way to define the significance of trend would be by the significance of its impact. As an outset, assume that the wind speed v_ζ consists of the two components in (3.1) and an additional trend component $T_\zeta \in \mathbb{R}$:

$$v_\zeta = V_\zeta + \tilde{v}_\zeta + T_\zeta \tag{4.1}$$

where V_ζ is the constant mean wind speed component, $\tilde{v}_\zeta$ is the zero-mean turbulence, and T_ζ is a zero-mean trend. Furthermore, to simplify the following expressions, assume that $\tilde{v}_\zeta$ and T_ζ are uncorrelated. While the latter assumption may seem unjustified, this is addressed in Section 4.2.2 on page 54. If the trend is neglected and the turbulence standard deviation is calculated then the following is obtained:

$$\sigma_\zeta = \sqrt{E[(v_\zeta - E[v_\zeta])^2]}$$
$$= \sqrt{E[(V_\zeta + \tilde{v}_\zeta + T_\zeta - V_\zeta)^2]}$$
$$= \sqrt{E[\tilde{v}_\zeta^2] + E[T_\zeta^2]} \tag{4.2}$$

where $E[\ ]$ denotes the expectation operator. That is, the turbulence standard deviation is overestimated if there is indeed a trend and this is neglected. This overestimation is an impact of trend, and thus, if the turbulence standard deviation is significantly overestimated, the trend must be significant. To determine when this is the case, consider Figure 3.12 on page 37: the 90 % percentiles of the wind speed standard deviation lie below the IEC class B curve for the lower mean wind speeds and roughly on the same curve for the upper mean wind speeds. If the actual turbulence standard deviation followed the IEC class C curve but a trend component caused the wind speed standard deviation to follow the IEC class B curve then the turbulence standard deviation is considered significantly overestimated in the context of this report. This choice is merely a proposal that is related to the application in wind turbine load estimation - other choices may be more descriptive for other applications. For overestimation to occur, the variance of the trend would have to be at least:

$$\sigma_{\zeta,\mathrm{B}} = \sqrt{\sigma_{\zeta,\mathrm{C}}^2 + E[T_\zeta^2]} \tag{4.3}$$

$$E[T_\zeta^2] = \sigma_{\zeta,\mathrm{B}}^2 - \sigma_{\zeta,\mathrm{C}}^2$$
$$= (0.14 \cdot (0.75 \cdot V_\zeta + 5.6))^2 - (0.12 \cdot (0.75 \cdot V_\zeta + 5.6))^2$$
$$= 0.0052 \cdot (0.75 \cdot V_\zeta + 5.6)^2 \tag{4.4}$$

Thus if the variance of the trend is greater than or equal to the derived expression in (4.4) then the trend is considered significant. Otherwise the trend is considered insignificant.

4.1.2 Linear and Periodic Trends

The above distinction between significant trend and insignificant trend is independent of the definition of trend; the distinction is merely based on a decomposition of the wind speed into the three components in (4.1). In practice, this decomposition requires the trend to be defined: Chapter 1 on page 1 loosely introduces trend as wind velocity components with periods larger than around 1 minutes to 2 minutes. Furthermore, the methods for removing trend that are presented in Section 5.2 on page 64 suggest more concrete definitions of trend in the form of trend models. Specifically, the modelling of trend by a linear function and the modelling of trend by low-frequent components. In this report, trends that can be modelled by a linear function are denoted linear trends; whereas trends that can be modelled by low-frequent components are denoted periodic trends. These two classes of trends are discussed in the following sections.

Linear Trends

A linear trend $T_\zeta(t) \in \mathbb{R}$ in a 10-minute interval can be described as [27]:

$$T_\zeta(t) = \frac{\Delta v_\zeta}{600\,\mathrm{s}} \cdot t + V_\zeta - \frac{\Delta v_\zeta}{2} \tag{4.5}$$

where $\Delta v_\zeta \in \mathbb{R}$ is the linear increment or decrement in the mean wind speed v_ζ which occurs over the course of 10 minutes. Here, as well as in the following sections, an increment denotes a monotonic, positive change whereas a decrement denotes a monotonic, negative change. Only the trend in v_ζ is shown in (4.5). However, similar expression are valid for v_ξ and v_ψ. Figure 4.3 shows an example which contains a near-linear decrement in mean wind speed.

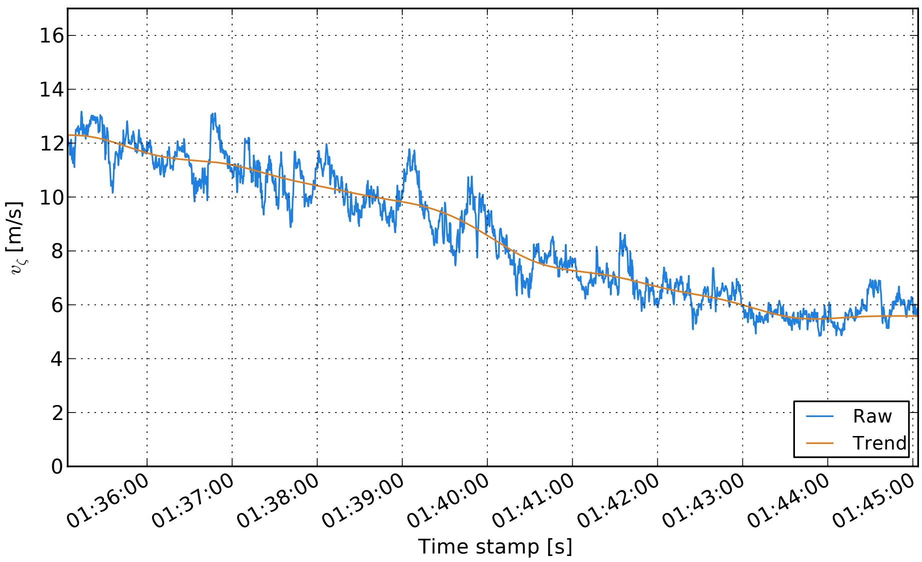

Figure 4.3. Example of measured horizontal wind speed in a 10-minute interval. This example contains a linear trend with no periodic trend superimposed. The data is from Teufelsberge (06-08-2006).

Basically, the above description may be unable to describe some trend for two reasons: 1) the change in mean wind speed may not be linear or 2) the change may occur over the course of a period different from 10 minutes. In the first case, intuitively, the trend cannot be a linear trend. However, in the second case, intuitively, the trend can be a linear trend. Figure 4.4 on the next page shows an example which contains a somewhat linear increment in the mean wind speed but over the course of less than 5 minutes. Thus, in this report, a linear trend is generalised to encompass linear increments and decrements in the mean wind speed which occur over the course of arbitrary periods. That is, a linear trend can be described as

$$T_\zeta(t) = \begin{cases} V_\zeta - \frac{\Delta v_\zeta}{2} & t < t_0 \\ V_\zeta - \frac{\Delta v_\zeta}{2} + \frac{\Delta v_\zeta}{\Delta t} \cdot t & t < t_0 + \Delta t \\ V_\zeta + \frac{\Delta v_\zeta}{2} & \textit{otherwise} \end{cases} \tag{4.6}$$

where t_0 is the time of the beginning of the trend and Δt is the duration of the trend.

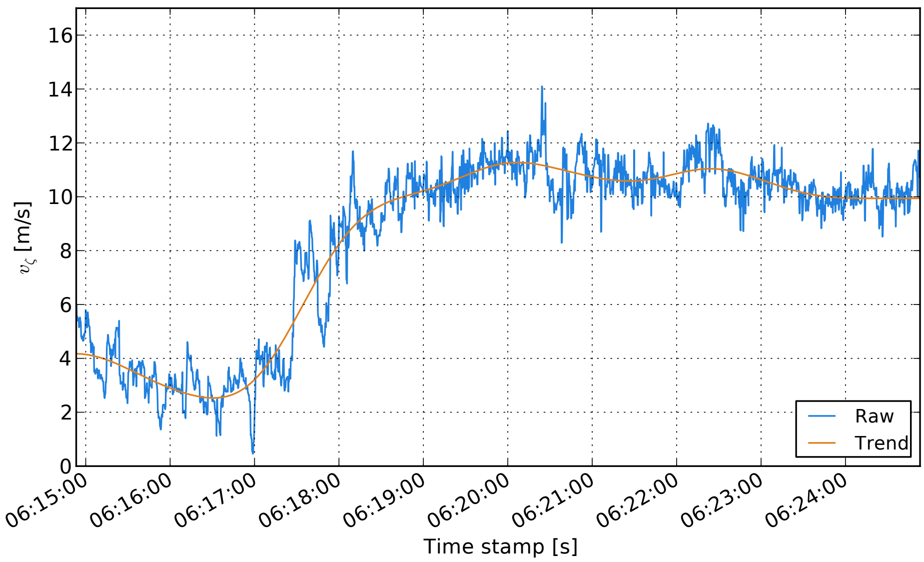

Figure 4.4. Example of measured horizontal wind speed in a 10-minute interval. This example contains a linear trend with the trend having an unknown duration. The data is from Teufelsberge (22-10-2006).

Periodic Trends

Figure 4.5 on the following page shows an example which contains low-frequent oscillations in the wind speed. However, to denote these oscillations as low-frequent, the term »low-frequent« must be defined. According to Chapter 1 on page 1, components with periods larger than around 1 minutes to 2 minutes can be considered low-frequent. This chapter is concerned with establishing the presence of significant trend in the sampled wind velocities. If components with period larger than 2 minutes are considered trend and the measurements contain significant trend then this will also be the case if, additionally, smaller periods are considered trend. Thus, throughout this chapter, a periodic trend refers to components in the signal with frequencies between 0 Hz and $\frac{1}{120}$ Hz.

The Distinction

From the considerations presented in Chapter 1 on page 1, the definition of periodic trends must necessarily be the definition of interest to this report. Thus in the classification system, the wind speed will be decomposed into mean wind speed, turbulence, and trend based on this definition. Thus, obviously, the trend component of the decomposition can be modelled as a periodic trend.

However, instead of abandoning the definition of linear trends and focussing only on periodic trends, a distinction between the two types of trend is made. That is, the 10-minute intervals classified as containing a significant trend are, furthermore, classified as containing either a linear trend or a periodic trend. The purpose of this distinction is *not* to determine which model most accurately models the trend. Rather, the purpose is

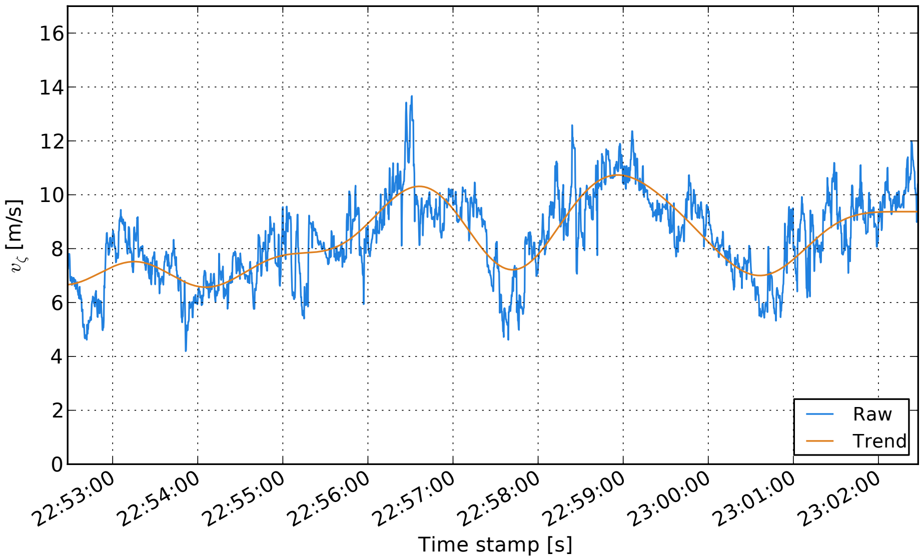

Figure 4.5. Example of measured horizontal wind speed in a 10-minute interval. This example contains a periodic trend. The data is from Teufelsberge (28-09-2006).

to determine if the significant trend can be modelled as a linear trend or not. This is of interest becuase of the mathematical simplicity and the few parameters of the linear trend model. Consequently, the specific distinction must be established.

As a starting point, the trend is considered linear if the trend consists of one dominating increment or decrement. However, a periodic trend may be superimposed on a linear trend as shown in Figure 4.2 on page 48. In this case, the problem is to decide which of a linear trend and a periodic trend is more significant than the other. To retain simplicity and visual intuition in the distinction, the trend is considered linear if the increment in the linear trend is larger than the largest increment or decrement in the periodic trend. Otherwise the trend is considered periodic.

4.2 Classification System

The trend classification task is approached as a pattern recognition problem leading to the use of a pattern recognition system as described in [28]. Typically, the pattern recognition system contains the blocks of sensing, pre-processing, feature extraction, classification, and post-processing. In the context of the present task, these blocks cover the following:

- The sensing block covers the process of acquiring measurements of the wind regime. Since the measurements have already been acquired, the sensing block is not included in the pattern recognition system.
- The pre-processing block covers the process of conditioning the acquired measurements and segmenting these into 10-minute intervals. Since the acquired measurements have already been conditioned as described in Chapter 2 on page 7 and can trivially be segmented into 10-minute intervals, the pre-processing block is not included in the pattern recognition system.

- The feature extraction block covers the process of characterising the 10-minute intervals in terms of a number of features.
- The classification block covers the process of classifying 10-minute intervals as containing either a significant or an insignificant trend.
- The post-processing block usually *"uses the output of the classifier to decide on the recommended action"* [28]. However, the purpose of the current task is to establish the ratio between the number of occurrences of the two classes. Since there is no recommended action, the post-processing block is not included in the pattern recognition system.

Consequently, the pattern recognition system comprises the feature extraction and classification blocks, takes the conditioned, acquired measurements as input, and outputs the classes of the inherent 10-minute intervals. The two blocks are discussed in the following sections.

4.2.1 Feature Extraction

With the notions of significant trends, insignificant trends, linear trends, and periodic trends established, features that characterise the trend contained in a 10-minute interval must be chosen. These features must allow the classification block to distinguish between significant trends and insignificant trends and, furthermore, between linear trends and periodic trends.

However, to be able to extract features that characterise the trend contained in 10-minute intervals, the trend must first be separated from the turbulence. Per definition, any periodic trend may be separated from turbulence using a low-pass filter. Consequently, the trend is estimated from 10-minute intervals using a low-pass filter with a 1 dB cut-off frequency of $\frac{1}{120}$ Hz. For specific information about the filter used, see Section 5.3 on page 64.

With the trend separated from the turbulence, the features must be chosen. Regardless of the type of the trend, the first two moments of the trend are obvious suggestions for features due to the emphasis on mean and standard deviation throughout the analysis of the assumptions in Chapter 3 on page 21. Thus the first two features are:

1. The mean value of the trend denoted s_{mean}
2. The variance of the trend denoted s_{var}

The remaining features arise as a consequence of the two types of trend: linear and periodic. With the linear trend, the mean wind speed changes from one »level« to another. Thus, the difference between these two »levels«, i.e. the increment or decrement, characterises a linear trend. This difference gives rise to another feature:

3. The difference between the minimum and maximum value of the trend denoted s_{diff}

With the periodic trend, the wind speed fluctuates slowly around the mean wind speed. If these fluctuations had constant amplitude, the amplitude of the trend would have characterised a periodic trend. However, the fluctuations do not have constant amplitude as seen from Figure 4.5 on page 52. Instead, the amplitudes of the individual fluctuations characterise a periodic trend. These amplitudes give rise to a list of features:

4. The amplitude of the n^{th} largest monotonic increment or decrement in the trend denoted $s_{\mathrm{mon,1}}$, $s_{\mathrm{mon,2}}$, $\ldots$, $s_{\mathrm{mon,n}}$

Thus, four features that characterise the types of trend have been determined.

4.2.2 Classification

To be able to implement the classification block, a classification scheme based on the chosen features must be decided on. This classification scheme must be able to distinguish between significant and insignificant trends and, furthermore, between linear and periodic trends. Since the classification system is merely intended to provide an overall idea about the presence of significant trend and the type of such significant trend, the type and design of the classifier are chosen to retain simplicity.

An extensive number of classification techniques exist such as those of Bayesian decision theory, maximum-likelihood estimation, k_n-nearest-neighbour estimation, discriminant functions and neural networks [29]. However, no knowledge of the underlying probability distributions of the classes is immediately available and establishing such distributions is beyond the scope of this project. Thus, the method of directly establishing a discriminant function is chosen. Furthermore, the criterion in (4.4) suggests an outset to establishing such a discriminant function.

Significant versus insignificant trend

In the case of significant trend versus insignificant trend, the following discriminant function can be derived from (4.4):

$$g_1(s_{\mathrm{mean}}, s_{\mathrm{var}}) = \frac{(0.75 \cdot s_{\mathrm{mean}} + 5.6)^2}{s_{\mathrm{var}}} + w_1 \tag{4.7}$$

where $w_1 \in \mathbb{R}$ is the bias or threshold weight. The trend in the sample is considered significant if $g_1(f_{\mathrm{mean}}, f_{\mathrm{var}}) > 0$ and insignificant otherwise. The derivation leading up to (4.4) suggests that $w_1 = -0.0052$. However, this derivation is based on the assumption that $\tilde{v}_\zeta$ and T_ζ are uncorrelated which may not be the case. Thus, the classifier is first trained and then tested.

The training is done based on the Teufelsberge dataset. According to the discussion in the previous section, a 10-minute interval (hereafter denoted as a sample) is considered to contain significant trend if the removal of the trend reduces the standard deviation such that the sample moves from the IEC class B curve to the IEC class C curve. However, no or very few samples will lie exactly on either of these curves. Thus, the samples which are closer to the IEC class B than to any of the two other classes are found instead. That is, the samples for which

$$0.13 < \frac{\hat{\sigma}_\zeta}{0.75 \cdot \hat{V}_\zeta + 5.6} < 0.15 \tag{4.8}$$

where $\hat{V}_\zeta$ is the estimated mean value and $\hat{\sigma}_\zeta$ is the estimated standard deviation. These samples must, next, be »manually« classified to support supervised learning and testing. This is done by estimating the standard deviation of the raw wind speed, $\hat{\sigma}_{\zeta,\mathrm{raw}}$, and of the raw wind speed with the estimated trend subtracted, $\hat{\sigma}_{\zeta,\mathrm{detrended}}$. Each sample is, then, »manually« classified as containing a significant trend if

$$\frac{\hat{\sigma}_{\zeta,\mathrm{raw}} - \hat{\sigma}_{\zeta,\mathrm{detrended}}}{0.75 \cdot \hat{V}_\zeta + 5.6} \geq 0.02 \tag{4.9}$$

Every other sample is chosen for training, and the remaining samples are chosen for testing. Furthermore, the same procedure is repeated for the Snaresmoor M1 60 dataset except all the selected samples are used for testing. The training chooses the threshold weight which minimises the sum of percentage of incorrectly classified training samples with significant trend and the percentage of incorrectly classified training samples with insignificant trend. The threshold weight was trained to $w_1 = -0.0055$ which is quite close to the theoretical value of $w_1 = -0.0052$ which indicates that the assumption that $\tilde{v}_\zeta$ and T_ζ are uncorrelated is almost complied with.

To evaluate how well the trained classifier performs, the samples that were selected for testing are classified using the trained classifier. The results are shown in Table 4.1 on the following page and Table 4.2 on the next page. In the case of Teufelsberge, it can be seen that 90.5 % of the samples with significant trend are correctly classified; likewise, 96.9 % of the samples with insignificant trend are correctly classified. In the case of Snaresmoor, it can be seen that 97.9 % of the samples with significant trend are correctly classified; likewise, 81.2 % of the samples with insignificant trend are correctly classified. The reader should note the following:

- The number of samples that contain insignificant trend is less than one tenth of the number of samples that contain significant trend. Thus the percentage of incorrectly classified samples that contain insignificant trend may be inaccurate.
- For Snaresmoor, the larger difference between the percentage of incorrectly classified samples that contain significant and insignificant trend, respectively, suggest that another threshold weight might provide better results. This is not unlikely, since the threshold weight was trained on the Teufelsberge dataset. However, to retain generality, the same threshold weight is used for all datasets.

	Insignificant		Significant	
	Occurrences	%	Occurrences	%
Insignificant	63	96.9	2	3.1
Significant	91	9.5	869	90.5

Table 4.1. Results from the evaluation of the trained trend significance classifier. The rows denote the classes of the samples input to the classifier whereas the columns denote the output of the classifier **[Teufelsberge]**.

	Insignificant		Significant	
	Occurrences	%	Occurrences	%
Insignificant	13	81.2	3	18.8
Significant	17	2.1	804	97.9

Table 4.2. Results from the evaluation of the trained trend significance classifier. The rows denote the classes of the samples input to the classifier whereas the columns denote the output of the classifier **[Snaresmoor M1 60]**.

Linear versus periodic trend

In the case of linear trend versus periodic trend, no discriminant function is analytically derived. Instead, a combination of the following three discriminant functions were experimentally found to yield the best results:

$$g_2(s_{\text{mon},1}, s_{\text{diff}}) = \frac{s_{\text{mon},1}}{s_{\text{diff}}} + w_2 \tag{4.10}$$

$$g_3(s_{\text{mon},2}, s_{\text{diff}}) = \frac{s_{\text{mon},2}}{s_{\text{diff}}} + w_3 \tag{4.11}$$

$$g_4(s_{\text{mon},3}, s_{\text{diff}}) = \frac{s_{\text{mon},3}}{s_{\text{diff}}} + w_4 \tag{4.12}$$

where $w_2 \in \mathbb{R}$, $w_3 \in \mathbb{R}$, and $w_4 \in \mathbb{R}$ are the threshold weights. If fewer than three increments and decrements are present, the absolute value of the remaining increments or decrements are set to zero. The first term of each of the three discriminant functions describes the absolute value of the largest, second largest, and third largest increment or decrement, respectively, relative to the difference between the maximum and minimum value. The trend in a sample is considered linear if *either* $g_2(s_{\text{mon},1}, s_{\text{diff}}) < 0$, $g_3(s_{\text{mon},2}, s_{\text{diff}}) < 0$, or $g_4(s_{\text{mon},3}, s_{\text{diff}}) < 0$ and periodic otherwise.

As with the first classification, the training is done based on the Teufelsberge dataset. Thus, 200 samples are randomly selected from this dataset. These samples must be manually classified to support supervised learning and testing. Section 4.1.2 on page 49 discusses when trend should be considered linear and when it should be considered periodic. Each selected sample is manually classified as containing linear trend or periodic trend based on this discussion. Again, every other sample is chosen for training and the remaining samples are chosen for testing. Furthermore, the same procedure is repeated with 100 randomly selected samples from the Snaresmoor M1 60 dataset except all the selected samples are used for testing. The training chooses the three threshold weights,

w_2, w_3, and w_4, which minimise the sum of percentage of incorrectly classified training samples with linear trend and the percentage of incorrectly classified training samples with periodic trend. The threshold weights were trained to $w_2 = 0.65$, $w_3 = 0.60$, $w_4 = 0.40$.

To evaluate how well the trained classifier performs, the remaining manually classified samples are classified using the trained classifier. The results are shown in Table 4.3 and Table 4.4. In the case of Teufelsberge, it can be seen that 84.4 % of the samples with linear trend are correctly classified; likewise, 82.4 % of the samples with periodic trend are correctly classified. In the case of Snaresmoor, it can be seen that 76.3 % of the samples with significant trend are correctly classified; likewise, 71.0 % of the samples with insignificant trend are correctly classified. Clearly, a substantial percentage of the samples are incorrectly classified. While this indicates that additional features, another classification scheme, or both could be considered, only the approximate ratio between samples with linear trend and periodic trend is of interest. Thus the trained classifier is maintained although the reader should keep in mind that the ratio between samples with linear trend and periodic trend presented in the following section might be somewhat inaccurate.

	Linear		Periodic	
	Occurrences	%	Occurrences	%
Linear	27	84.4	5	15.6
Periodic	12	17.6	56	82.4

Table 4.3. Results from the evaluation of the trained trend type classifier. The rows denote the classes of the samples input to the classifier whereas the columns denote the output of the classifier **[Teufelsberge]**.

	Linear		Periodic	
	Occurrences	%	Occurrences	%
Linear	29	76.3	9	23.7
Periodic	18	29.0	44	71.0

Table 4.4. Results from the evaluation of the trained trend type classifier. The rows denote the classes of the samples input to the classifier whereas the columns denote the output of the classifier **[Snaresmoor M1 60]**.

4.3 Results

A classification of 10-minute intervals of the wind velocity measurements from Teufelsberge and Snaresmoor have been carried out. Furthermore, the robustness of the classification to time offset of the 10-minute intervals have been evaluated. The results are given in the following.

4.3.1 Classification Results

The available wind velocity measurements with consecutive time spans of at least 10 minutes from Snaresmoor and Teufelsberge (see Table 2.5 on page 19 for the specific size of the datasets) have been classified according to the discussion in Section 4.2.2 on page 54. The number of occurrences of 10-minute intervals of the wind velocity measurements containing insignificant (I), linear (L), and periodic (P) trends is presented in Table 4.5. Generally, around 10 % to 12 % of the 10-minute intervals contain a linear trend. Thus, the percentage of 10-minute intervals that contains a linear trend is very stable across all the datasets. A greater variation in the percentage of 10-minute intervals containing a periodic trend is observed. Specifically, between 6 % and 16 % of the 10-minute intervals contain a periodic trend. Comparing the different observation heights at Snaresmoor, a larger percentage of the 10-minute intervals contain a periodic trend for lower heights compared to higher heights. Generally, 70 % to 80 % of the 10-minute intervals contain no significant trend.

	Insignificant (I)		Linear (L)		Periodic (P)	
Dataset	Occurrences	%	Occurrences	%	Occurrences	%
/Snaresmoor/Raw_M1_20	18573	72.0	3072	11.9	4138	16.1
/Snaresmoor/Raw_M1_40	19398	77.8	2852	11.4	2680	10.8
/Snaresmoor/Raw_M1_60	25193	80.8	3353	10.8	2629	8.4
/Snaresmoor/Raw_M1_80	24801	82.3	3167	10.5	2160	7.2
/Snaresmoor/Raw_M2_20	18603	73.4	3175	12.5	3586	14.1
/Snaresmoor/Raw_M2_40	21649	77.6	3319	11.9	2931	10.5
/Snaresmoor/Raw_M2_60	5034	83.5	616	10.2	380	6.3
/Snaresmoor/Raw_M2_80	23313	80.9	3226	11.2	2261	7.9
/Snaresmoor/Raw_M4_20	19166	73.9	3266	12.6	3504	13.5
/Snaresmoor/Raw_M4_40	25099	78.5	3652	11.4	3227	10.1
/Snaresmoor/Raw_M4_60	24501	79.8	3485	11.4	2696	8.8
/Snaresmoor/Raw_M4_80	25753	80.5	3615	11.3	2623	8.2
/Snaresmoor/Raw_M5_20	19685	75.9	3182	12.3	3074	11.8
/Snaresmoor/Raw_M5_40	22339	78.6	3386	11.9	2707	9.5
/Snaresmoor/Raw_M5_60	21782	81.4	2869	10.7	2096	7.9
/Snaresmoor/Raw_M5_80	18390	80.7	2564	11.2	1838	8.1
/Teufelsberge/T07_HF	17106	71.2	3012	12.5	3915	16.3

Table 4.5. The results of classification of 10-minute intervals of data into the classes of trend: insignificant (I), linear (L), and periodic (L). The results for all available dataset from Teufelsberge and Snaresmoor are shown.

The results in Table 4.5 show the overall distribution of 10-minute intervals containing insignificant, linear, and periodic trends independent of the mean wind speed V_ζ and turbulence standard deviation σ_ζ. To gain insight into the distribution between insignificant, linear, and periodic trends for different levels of V_ζ and σ_ζ, the classification results have been conditioned on intervals of $\hat{V}_\zeta$ and $\hat{\sigma}_\zeta$. The results for Teufelsberge are shown in Table 4.6 whereas the results for Snaresmoor M1 60 are shown in Table 4.7. These results show that an increasingly larger percentage of the 10-minute intervals contain linear and periodic trends as $\hat{\sigma}_\zeta$ increases. This is especially true if at the same time $\hat{V}_\zeta$ is small. Thus, popularly, the close one comes to the upper right corner of the tables the larger percentage of the 10-minute intervals contain a linear or periodic trend.

$\hat{\sigma}_\zeta$	[0, 1)			[1, 2)			[2, 3)			[3, ∞)			Sum
$\hat{V}_\zeta$	I	L	P	I	L	P	I	L	P	I	L	P	
[5, 10)	4427	100	35	1365	786	835	1	82	69	0	5	0	7705
[10, 15)	1773	0	0	4174	625	939	128	463	775	0	28	20	8925
[15, 20)	203	0	0	1813	40	46	742	393	720	2	86	159	4204
[20, 25)	3	0	0	140	0	0	373	48	84	30	44	92	814
[25, ∞)	0	0	0	0	0	0	16	2	2	10	4	8	42
Sum	6406	100	35	7492	1451	1820	1260	988	1650	42	167	279	21690

Table 4.6. The results of classification of 10-minute intervals of data conditioned on mean wind speed and turbulence standard deviation. The columns indicate the turbulence standard deviation intervals whereas the rows indicate the mean wind speed intervals. I, L, and P denotes the trend classes: insignificant, linear, and periodic, respectively **[Teufelsberge]**.

$\hat{\sigma}_\zeta$	[0, 1)			[1, 2)			[2, 3)			[3, ∞)			Sum
$\hat{V}_\zeta$	I	L	P	I	L	P	I	L	P	I	L	P	
[5, 10)	14910	218	93	1894	1698	1396	0	78	32	0	3	1	20323
[10, 15)	1086	0	0	1827	418	410	12	105	84	0	3	0	3945
[15, 20)	0	0	0	100	14	8	10	26	39	0	0	0	197
[20, 25)	0	0	0	0	0	0	1	3	3	0	0	0	7
[25, ∞)	0	0	0	0	0	0	0	0	0	0	0	0	0
Sum	15996	218	93	3821	2130	1814	23	212	158	0	6	1	24472

Table 4.7. The results of classification of 10-minute intervals of data conditioned on mean wind speed and turbulence standard deviation. The columns indicate the turbulence standard deviation intervals whereas the rows indicate the mean wind speed intervals. I, L, and P denotes the trend classes: insignificant, linear, and periodic, respectively **[Snaresmoor M1 60]**.

4.3.2 Robustness Evaluation Results

The classification system must be robust to time offsets of the 10-minute intervals that are classified. If this is not the case, the classification results describe the chosen partitioning of the measurements rather than the sampled wind velocities in general.

To evaluate the robustness to time offsets, chunks of consecutive valid measurements are extracted from a dataset. Each of these chunks is then truncated such that the time span of the chunk is an integral multiple of 10 minutes, since the chunk is later partitioned into 10-minute intervals for classification. Denote the start time of the chunk t_0 and the end time of the chunk t_1; both given in seconds. For every offset $\Delta t \in \{0, 120, 240, 360, 480\}$, a sub-chunk from $t_0 + \Delta t$ to $t_1 - 600 + \Delta t$ is extracted and partitioned into 10-minute intervals. For every offset, the 10-minute intervals partitioned from all of the selected chunks are then classified.

Due to the offset, the extracted sub-chunks do not include the exact same measurements. Thus only chunks of at least 101 consecutive valid 10-minute intervals are used. With a chunk consisting of 101 10-minute intervals, the sub-chunks will consist of 100 10-minute intervals. Thus more than 99 % of the measurements in the sub-chunks of two different offsets will be the same. The only available dataset that contains large enough chunks for this evaluation is the Teufelsberge dataset as can be seen from Table 2.5 on page 19.

The results of the robustness evaluation are shown in Table 4.8. Based on the small variation of the percentages of 10-minute intervals that fall into the different classes across the time offsets, the classification system is concluded to be sufficiently robust to offsets. Thus the results presented so far do describe the sampled wind velocities and not only the specific choices of partitioning of the measurements.

	Insignificant (I)		Linear (L)		Periodic (P)	
Offset	Occurrences	%	Occurrences	%	Occurrences	%
0 minutes	16326	71.9	2758	12.2	3608	15.9
2 minutes	16383	72.3	2703	11.9	3584	15.8
4 minutes	16268	71.7	2822	12.4	3591	15.8
6 minutes	16211	71.5	2885	12.7	3587	15.8
8 minutes	16247	71.6	2901	12.8	3542	15.6

Table 4.8. The results of classification of 10-minute intervals of the data into the classes of trend: insignificant (I), linear (L), and period (P) for different offsets of the 10-minute intervals **[Teufelsberge]**.

5 Trend Modelling

In Section 4.3 on page 57, it is concluded that, in general, 20 to 30 % of the 10-minute intervals of each dataset contain significant trend. This violates the assumption of constant mean wind speed, V_ζ, in 10-minute intervals, presented in Chapter 3 on page 21. Thus a new model, with fewer assumptions, is sought after.

Section 5.1 is concerned with establishing an extended wind model which is capable of modelling the trend. As the name suggest, there is a clear link between the basic model and the extended model: if the trend modelled by the extended model is removed then the resulting signal can be modelled by the basic model. The removal of trend is called detrending, and Section 5.2 on page 64 briefly presents the state-of-the-art detrending methods. Following this, a new detrending method based on the extended wind model is introduced by Section 5.3 on page 64. Finally, Section 5.5 on page 77 presents the results of the detrending.

5.1 An Extended Wind Model

One way to characterise a model is by its *purpose*, its *description*, and its *limitations*. Thus, these three are considered consecutively throughout the following sections.

5.1.1 Model Purpose

First, the purpose of the extended model must be decided on. The purpose of the model is to model the three instantaneous wind speed components v_x, v_y, and v_z of the wind regime observed in a single point in space. Note that this first part of the purpose states that: 1) The model is only required to model the wind regime in a single point in space; that is, it should not concern itself with correlations with other points in space. 2) The model is required to model the 3-dimensional instantaneous wind velocity, in the given point in space. 3) The choice of coordinate system is that of an absolute, rectangular coordinate system; that is, a rectangular coordinate system with a static or constant orientation throughout the modelling period.

However, a description such as $\mathbf{v}(t) = [v_\mathrm{x}(t), v_\mathrm{y}(t), v_\mathrm{z}(t)]^T$ adheres to the purpose, although it does not reveal any information about the wind regime beyond what was stated in the purpose. Likewise, the basic model described by Equation 3.1 on page 22 adheres to the purpose, and so would any number of additional models. Thus the purpose has to be made more specific. Given that the model should be an extension of the basic model, the purpose of the model must, necessarily, be to decompose the instantaneous wind velocity into a number of meaningful components. The basic model decomposes the instantaneous wind speed into mean wind speed and turbulence. To ensure a link between the extended wind model and the basic model, these components are chosen for the extended model. Furthermore, the extended model is motivated by the presence of trend. To ensure that this is captured by the extended wind model, a trend component is chosen in addition to the first two components.

To summarise: *The purpose of the extended model is to model the three instantaneous wind speed components v_{x}, v_{y}, and v_{z} of the wind regime observed in a single point in space. This must be done by decomposition into a mean wind speed component, a trend component, and a turbulence component.*

5.1.2 Model Description

With the purpose settled, a suitable description for the fulfilment of the purpose must be decided on. The purpose requires a decomposition into three components: a mean wind speed component, a trend component, and a turbulence component. This decomposition is described as

$$\mathbf{v}(t) = \mathbf{V}(t) + \mathbf{T}(t) + \tilde{\mathbf{v}}(t) \tag{5.1}$$

where $\mathbf{V}(t) \in \mathrm{R}^{3\times1}$ is the mean wind component, $\mathbf{T}(t) \in \mathbb{R}^{3\times1}$ is the trend component, and $\tilde{\mathbf{v}}(t) \in \mathrm{R}^{3\times1}$ is the turbulence component. In accordance with the purpose, this description models the three instantaneous wind speed components in a single point in space. Furthermore, the chosen coordinate system is an absolute, rectangular coordinate system with components v_{x}, v_{y}, and v_{z}. That is,

$$\mathbf{V}(t) = \begin{bmatrix} V_{\mathrm{x}}(t) \\ V_{\mathrm{y}}(t) \\ V_{\mathrm{z}}(t) \end{bmatrix} \quad , \quad \mathbf{T}(t) = \begin{bmatrix} T_{\mathrm{x}}(t) \\ T_{\mathrm{y}}(t) \\ T_{\mathrm{z}}(t) \end{bmatrix} \quad , \quad \tilde{\mathbf{v}}(t) = \begin{bmatrix} \tilde{v}_{\mathrm{x}}(t) \\ \tilde{v}_{\mathrm{y}}(t) \\ \tilde{v}_{\mathrm{z}}(t) \end{bmatrix} \tag{5.2}$$

The choice of coordinate system does, however, pose one critical problem: in the IEC61400 standard [4], the turbulence standard deviation is defined as the "standard deviation of the longitudinal component of the turbulent wind velocity at hub height." Furthermore, the longitudinal component is defined as the component "along the direction of the mean wind velocity." With the extended model, it is chosen to define the longitudinal component as the component along the direction of the horizontal part of $\mathbf{V}(t) + \mathbf{T}(t)$. This is chosen to ensure that the direction of the longitudinal component is continuous between successive time intervals that have been modelled individually. The problem arises in the modelling of the turbulence: the Mann uniform shear turbulence model assumes that the turbulence is WSS throughout the modelling period, and the turbulence is allowed to be anisotropic. That is, in general,

$$\sigma_\zeta(t) = \sigma_\zeta \quad , \quad \sigma_\xi(t) = \sigma_\xi \quad , \quad \sigma_\psi(t) = \sigma_\psi \tag{5.3}$$

$$\sigma_\zeta \neq \sigma_\xi \neq \sigma_\psi \tag{5.4}$$

However, with the choice of definition of the longitudinal component, the assumption that $\tilde{v}_\zeta(t)$, $\tilde{v}_\xi(t)$, and $\tilde{v}_\psi(t)$ are WSS does not imply that $\tilde{v}_\mathrm{x}(t)$, $\tilde{v}_\mathrm{y}(t)$, and $\tilde{v}_\mathrm{z}(t)$ are WSS. To see this, consider a hypothetical example where

$$\sigma_\zeta = 1 \quad , \quad \sigma_\xi = 0 \quad , \quad \sigma_\psi = 0 \tag{5.5}$$

$$\mathbf{V}(t) + \mathbf{T}(t) = \begin{bmatrix} \cos\left(\frac{\pi}{1200} \cdot t\right) \\ \sin\left(\frac{\pi}{1200} \cdot t\right) \\ 0 \end{bmatrix} \tag{5.6}$$

At the beginning of the first 10-minute interval, $\mathbf{V}(0) + \mathbf{T}(0) = [1, 0, 0]^T$ and thus $\sigma_\mathrm{x} = 1$ and $\sigma_\mathrm{y} = 0$. At the end of the first 10-minute interval, $\mathbf{V}(600) + \mathbf{T}(600) = [0, 1, 0]^T$ and thus $\sigma_\mathrm{x} = 0$ and $\sigma_\mathrm{y} = 1$. Clearly, the standard deviation of the components in the absolute, rectangular coordinate system are not constant, and the turbulence components are thus not WSS. To allow for a description of the model where the components of the turbulence follow the same assumptions as in the basic model, the description of the model is updated to

$$\mathbf{v}(t) = \mathbf{V}(t) + \mathbf{T}(t) + \mathbf{R}_{\mathbf{V}+\mathbf{T}}(t)\tilde{\mathbf{v}}(t) \tag{5.7}$$

where $\mathbf{R}_{\mathbf{V}+\mathbf{T}}(t) \in \mathbb{R}^{3x3}$ is a rotation matrix that transforms $\tilde{\mathbf{v}}(t)$ from a relative, rectangular coordinate system oriented in the direction of the horizontal part of $\mathbf{V}(t)+\mathbf{T}(t)$ to the absolute, rectangular coordinate system. This matrix $\mathbf{R}_{\mathbf{V}+\mathbf{T}}(t)$ is given by

$$\mathbf{R}_{\mathbf{V}+\mathbf{T}}(t) = \begin{bmatrix} \frac{V_\mathrm{x}+T_\mathrm{x}}{\sqrt{(V_\mathrm{x}+T_\mathrm{x})^2+(V_\mathrm{y}+T_\mathrm{y})^2}} & -\frac{V_\mathrm{y}+T_\mathrm{y}}{\sqrt{(V_\mathrm{x}+T_\mathrm{x})^2+(V_\mathrm{y}+T_\mathrm{y})^2}} & 0 \\ \frac{V_\mathrm{y}+T_\mathrm{y}}{\sqrt{(V_\mathrm{x}+T_\mathrm{x})^2+(V_\mathrm{y}+T_\mathrm{y})^2}} & \frac{V_\mathrm{x}+T_\mathrm{x}}{\sqrt{(V_\mathrm{x}+T_\mathrm{x})^2+(V_\mathrm{y}+T_\mathrm{y})^2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}(t) \tag{5.8}$$

To explicitly indicate the coordinate system,

$$\mathbf{V}(t) = \begin{bmatrix} V_\mathrm{x}(t) \\ V_\mathrm{y}(t) \\ V_\mathrm{z}(t) \end{bmatrix} \quad , \quad \mathbf{T}(t) = \begin{bmatrix} T_\mathrm{x}(t) \\ T_\mathrm{y}(t) \\ T_\mathrm{z}(t) \end{bmatrix} \quad , \quad \tilde{\mathbf{v}}(t) = \begin{bmatrix} \tilde{v}_\zeta(t) \\ \tilde{v}_\xi(t) \\ \tilde{v}_\psi(t) \end{bmatrix} \tag{5.9}$$

5.1.3 Model Limitations

Finally, with the purpose and description settled, the limitations of the model must be established. In the beginning of this chapter, it was chosen to abandon the assumption of a constant mean wind speed component. However, with the introduction of the additional component, $\mathbf{T}(t)$, the mean wind speed component, $\mathbf{V}(t)$, is still required and thus assumed to be constant in the relative, rectangular coordinate system over 10-minute intervals. Consequently, the limitations of the extended model are the assumptions of the basic model. There are no additional limitations of the extended model.

5.2 State-of-the-art Detrending

The process of removing the (undesired) trend from wind velocity measurements is generally termed "detrending". Two distinct approaches to detrending are well established as the current state-of-the-art methods. These are subtraction of a linear trend and high-pass filtering [30].

Linear Trend Subtraction

The premise behind the linear trend subtraction method is that a trend is well described by a straight line. An example of such as trend is given in Figure 4.3 on page 50. The idea is then to use a least squares fit to determine a straight line that represents the trend in the signal [27]. The fitted straight line can then be subtracted from the signal in order to yield the detrended signal. A large study of the effect of linear trend subtraction has shown that the mean turbulence intensity is reduced with 4 % to 11 % depending on the specific site [27]. The largest reduction is, not surprisingly, obtained at sites that are subject to frequent front passages. Another study of linear trend subtraction has indicated that it may be possible to do detrending of 10-minute statistics using the linear trend subtraction approach by exploiting statistical information in the 10-minute statistics [31].

High-Pass Filtering

The idea of the separating the low frequency trend from the high frequency turbulence in the frequency domain is the reasoning behind the use of a high-pass filter on the signal. High-pass filtering is the method recommended in the IEC 61400 Part 13 standard. However, the IEC 61400 does not provide any guidance towards type and design of the filter. It only states that it *"shall be chosen so that similar filtering applied to the wind model used in the design does not significantly affect the resulting design loads"* [32]. As an alternative, the IEC 61400 standard suggests the pragmatic approach of simply discarding 10-minute intervals that contain a trend. However, when used, the most common filter type seems to be a finite impulse response (FIR) moving average filter due to its simplicity and possibility of linear phase [30].

5.3 Detrending

This section presents the design of the detrending method used in this report. Throughout the design phase a number of parameter choices are made. In some of these cases, choices can be made which are based on considerations that apply to detrending in general. In the remaining cases, the choices are somewhat arbitrary in the sense that they must be made on a per applicattion basis. In such cases, choices have been made that seem reasonable to the authors for the purpose of investigating the influence of trend. Note, however, that the focus is on the methodology and not on the specific parameter values. Thus, the authors have pursued an approach of developing procedures, methods, and tools which can automatically adapt to changes in the assumptions. This makes it easy to further refine and update the assumptions while easily getting the relevant results.

First of all, the overall detrending approach must be chosen. In the previous section, the approaches of linear trend subtraction and high pass filtering were introduced as the state-of-the-art methods. As mentioned in Chapter 1 on page 1, wind speed components with periods larger than a threshold somewhere in between 1 and 2 minutes contribute insignificantly to wind turbine loads. Such wind speed components can thus be considered as trend which suggests detrending by filtering. Furthermore, Section 4.3 on page 57 established that 1) a significant trend is present in the sampled wind velocities and that 2) generally, the identified trend can be modelled as periodic trend but not as linear trend. This puts detrending by filtering over subtraction of a linear trend. Consequently a filter solution is chosen.

5.3.1 The Filter Expression

In Chapter 4 on page 47, the trend in the wind velocity component v_ζ was investigated. However, if only one of the three wind velocity components is detrended, any trend in the two other wind velocity components is inevitably left unchanged. Consequently all three components of the wind velocity must be detrended to ensure consistency in the concept of detrending. Such 3-dimensional detrending can be done using 1-dimensional detrending on the individual wind velocity components, since nothing about the relation between the three components is known in general. This solution does not reduce the quality of the detrended signal and retains all relevant information. Thus this solution is chosen for simplicity. The 1-dimensional digital filtering can be expressed as the difference equation [5]:

$$a_0 \cdot q[i] = \sum_{k=0}^{N} b_k \cdot p[i-k] - \sum_{k=1}^{M} a_k \cdot q[i-k] \tag{5.10}$$

where $p[i] \in \mathbb{R}$ is the input, $q[i] \in \mathbb{R}$ is the output, and $\mathbf{a} \in \mathbb{R}^{M+1\times 1} = [a_0, \ldots, a_M]^T$ and $\mathbf{b} \in \mathbb{R}^{N+1\times 1} = [b_0, \ldots, b_N]^T$ are the filter coefficient vectors. Thus, $\max\{M, N\}$ is the filter order. For infinite length input and output sequences, $i \in \mathbb{Z}$. In practice, the input and output sequences have finite length which means that the initial conditions of the filter have to be considered. However, the initial conditions are considered in Section 5.3.3 on page 71 so, for now, ignore the necessity of these considerations.

The chosen 3-dimensional digital filtering can be expressed as:

$$\begin{bmatrix} a_{1,0} \cdot q_1[i] \\ a_{2,0} \cdot q_2[i] \\ a_{3,0} \cdot q_3[i] \end{bmatrix} = \sum_{k=0}^{N} \begin{bmatrix} b_{1,k} \cdot p_1[i-k] \\ b_{2,k} \cdot p_2[i-k] \\ b_{3,k} \cdot p_3[i-k] \end{bmatrix} - \sum_{k=1}^{M} \begin{bmatrix} a_{1,k} \cdot q_1[i-k] \\ a_{2,k} \cdot q_2[i-k] \\ a_{3,k} \cdot q_3[i-k] \end{bmatrix} \tag{5.11}$$

where $\mathbf{p}[i] \in \mathbb{R}^3$ is the input, $\mathbf{q}[i] \in \mathbb{R}^3$ is the output, $\mathbf{a}_1 \in \mathbb{R}^{N+1\times 1}$ and $\mathbf{b}_1 \in \mathbb{R}^{M+1\times 1}$ are the filter coefficient vectors for the first component, $\mathbf{a}_2 \in \mathbb{R}^{N+1\times 1}$ and $\mathbf{b}_2 \in \mathbb{R}^{M+1\times 1}$ are the filter coefficient vectors for the second component, and $\mathbf{a}_3 \in \mathbb{R}^{N+1\times 1}$ and $\mathbf{b}_3 \in \mathbb{R}^{M+1\times 1}$ are the filter coefficients for the third component. Note the use of $[q_1, q_2, q_3]^T$ rather than $[q_\zeta, q_\xi, q_\psi]^T$ or $[q_\mathrm{x}, q_\mathrm{y}, q_\mathrm{z}]^T$: the above can be used to describe 3-dimensional digital filtering in either of the two coordinate systems.

Since, in general, nothing is known about any statistical differences between the three components, the definition of trend is identical for the three components. Intuitively, these should thus be removed in the same way and so the three 1-dimensional filters should be identical. Note that this simplifies the design task, as only one filter must be designed. Also, with $\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{a}_3 = \mathbf{a}$ and $\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{b}_3 = \mathbf{b}$, (5.11) simplifies to:

$$a_0 \cdot \mathbf{q}[i] = \sum_{k=0}^{N} b_k \cdot \mathbf{p}[i-k] - \sum_{k=1}^{M} a_k \cdot \mathbf{q}[i-k] \tag{5.12}$$

Thus, if some rotation matrix, $\mathbf{R} \in \mathbb{R}^{3\times 3}$, is applied the expression can be rearranged as:

$$\mathbf{R} \cdot a_0 \cdot \mathbf{q}[i] = \mathbf{R} \cdot \left(\sum_{k=0}^{N} b_k \cdot \mathbf{p}[i-k] - \sum_{k=1}^{M} a_k \cdot \mathbf{q}[i-k] \right) \tag{5.13}$$

$$= \sum_{k=0}^{N} b_k \cdot \mathbf{R} \cdot \mathbf{p}[i-k] - \sum_{k=1}^{M} a_k \cdot \mathbf{R} \cdot \mathbf{q}[i-k] \tag{5.14}$$

That is, rotating the coordinate system and applying the filter yields the same result as applying the filter and rotating the coordinate system. Or simply put, the filtering is invariant to static rotations of the coordinate system. However, the two coordinate systems that have been defined are not static rotations of one another. Thus, filtering in the relative, rectangular coordinate system does not necessarily yield the same result as filtering in the absolute, rectangular coordinate system. The difference is apparent if the expression for high-pass filtering the wind velocity to obtain the turbulence is stated for both of the coordinate systems. To simplify the notation, the filter coefficients are normalised such that $a_0 = 1$ and the notation of the rotation matrix is abbreviated such that $\mathbf{R}[i] = \mathbf{R}_{\mathbf{T}[i]+\mathbf{V}[i]}$. In the absolute, rectangular coordinate system, the decomposition of the wind velocity is given by (5.7) as

$$\mathbf{v}(t) = \mathbf{V}(t) + \mathbf{T}(t) + \mathbf{R}_{\mathbf{V}+\mathbf{T}}(t) \cdot \tilde{\mathbf{v}}(t)$$

That is, in the absolute, rectangular coordinate system, the wind velocity is given by $\mathbf{v}[i]$, and the turbulence is given by $\mathbf{R}[i] \cdot \tilde{\mathbf{v}}[i]$. Thus by filtering $\mathbf{v}[i]$, ideally $\mathbf{R}[i] \cdot \tilde{\mathbf{v}}[i]$ is recovered as

$$\mathbf{R}[i] \cdot \tilde{\mathbf{v}}[i] = \sum_{k=0}^{N} b_k \cdot \mathbf{v}[i-k] - \sum_{k=1}^{M} a_k \cdot \mathbf{R}[i-k] \cdot \tilde{\mathbf{v}}[i-k] \tag{5.15}$$

In (5.15), every term on the right hand side is known and the computation of $\mathbf{R}[i] \cdot \tilde{\mathbf{v}}[i]$ is straight forward. In the relative, rectangular coordinate system, on the other hand, the decomposition of the wind velocity is given by (5.7) rotated by $\mathbf{R}_{\mathbf{V}+\mathbf{T}}(t)^{-1}$ as

$$\begin{aligned} \mathbf{R}_{\mathbf{V}+\mathbf{T}}(t)^{-1} \cdot \mathbf{v}(t) &= \mathbf{R}_{\mathbf{V}+\mathbf{T}}(t)^{-1} \cdot (\mathbf{V}(t) + \mathbf{T}(t) + \mathbf{R}_{\mathbf{V}+\mathbf{T}}(t) \cdot \tilde{\mathbf{v}}(t)) && (5.16) \\ &= \mathbf{R}_{\mathbf{V}+\mathbf{T}}(t)^{-1} \cdot (\mathbf{V}(t) + \mathbf{T}(t)) + \tilde{\mathbf{v}}(t) && (5.17) \end{aligned}$$

That is, in the relative, rectangular coordinate system, the wind velocity is given by $\mathbf{R}[i]^{-1} \cdot \mathbf{v}[i]$, and the turbulence is given by $\tilde{\mathbf{v}}[i]$. Thus by filtering $\mathbf{R}[i]^{-1} \cdot \mathbf{v}[i]$, ideally $\tilde{\mathbf{v}}[i]$ is recovered as

$$\tilde{\mathbf{v}}[i] = \sum_{k=0}^{N} b_k \cdot \mathbf{R}[i-k]^{-1} \cdot \mathbf{v}[i-k] - \sum_{k=1}^{M} a_k \cdot \tilde{\mathbf{v}}[i-k] \tag{5.18}$$

Here, the right hand side includes the term $\mathbf{R}[i]$ which depends on $\mathbf{T}[i] + \mathbf{V}[i]$ which is unknown. Moving this term to the left hand side of the expression yields

$$\tilde{\mathbf{v}}[i] - b_0 \cdot \mathbf{R}[i]^{-1} \cdot \mathbf{v}[i] = \sum_{k=1}^{N} b_k \cdot \mathbf{R}[i-k]^{-1} \cdot \mathbf{v}[i-k] - \sum_{k=1}^{M} a_k \cdot \tilde{\mathbf{v}}[i-k] \tag{5.19}$$

Now, every term on the right hand side is known and the right hand side will be denoted as $\mathbf{rhs}[i]$ to simplify the following expressions. That is,

$$\mathbf{rhs}[i] = \sum_{k=1}^{N} b_k \cdot \mathbf{R}[i-k]^{-1} \cdot \mathbf{v}[i-k] - \sum_{k=1}^{M} a_k \cdot \tilde{\mathbf{v}}[i-k] \tag{5.20}$$

However, to compute $\tilde{\mathbf{v}}[i]$, $\mathbf{R}[i]$ still needs to be computed. As can be seen from (5.8), $\mathbf{R}[i]$ depends on $\mathbf{T}[i] + \mathbf{V}[i]$ which can be expressed as

$$\mathbf{T}[i] + \mathbf{V}[i] = \mathbf{v}[i] - \mathbf{R}[i] \cdot \tilde{\mathbf{v}}[i] \tag{5.21}$$

Isolating $\tilde{\mathbf{v}}[i]$ in (5.19) and substituting this into (5.21) yields

$$\tilde{\mathbf{v}}[i] = b_0 \cdot \mathbf{R}[i]^{-1} \cdot \mathbf{v}[i] + \mathbf{rhs}[i] \tag{5.22}$$

$$\mathbf{T}[i] + \mathbf{V}[i] = (1 - b_0) \cdot \mathbf{v}[i] - \mathbf{R}[i] \cdot \mathbf{rhs}[i] \tag{5.23}$$

That is, $\mathbf{T}[i] + \mathbf{V}[i]$ depends on $\mathbf{R}[i]$ which depends non-linearly on $\mathbf{T}[i] + \mathbf{V}[i]$. It seems unlikely that an analytic expression for $\mathbf{T}[i]+\mathbf{V}[i]$ can be found. Consequently, computing the filtered sequence of $\tilde{\mathbf{v}}[i]$ in the relative, rectangular coordinate system entails solving a system of non-linear equations numerically for every index i. For comparison, computing the filtered sequence of $\mathbf{R}[i] \cdot \tilde{\mathbf{v}}[i]$ in the absolute, rectangular coordinate system merely entails computing a sum of products. Thus the absolute, rectangular coordinate system is chosen to avoid a greatly increased computational complexity. This choice is only for convenience and does not reduce the generality of the methods or results to be presented.

Looking at the expression in (5.15), the $\mathbf{R}[i] \cdot \tilde{\mathbf{v}}[i]$ and $a_k \cdot \mathbf{R}[i-k] \cdot \tilde{\mathbf{v}}[i-k]$ terms indicate that the filter should be a high-pass filter since the output of the filter is the turbulence which is the high-frequent component. However, another approach is to low-pass filter the wind velocity to obtain the trend and mean wind speed. These components are then subtracted from the wind velocity to obtain the turbulence. In Section 5.3.3 on page 71 it is concluded that the use of a high-pass filter is not an option and so the low-pass filtering approach is chosen which yields:

$$\mathbf{T}[i] + \mathbf{V}[i] = \sum_{k=0}^{N} b_k \cdot \mathbf{v}[i-k] - \sum_{k=1}^{M} a_k \cdot (\mathbf{T}[i-k] + \mathbf{V}[i-k]) \tag{5.24}$$

$$\tilde{\mathbf{v}}[i] = \mathbf{v}[i] - \mathbf{T}[i] - \mathbf{V}[i] \tag{5.25}$$

This approach does, however, require the filter to be a zero-phase filter as $\tilde{\mathbf{v}}[i]$ in (5.25) contains some of the trend and mean wind speed if these components are phase offset before they are subtracted. This can be avoided by applying the filter as described in (5.24) and then applying the same filter backwards. The actual filtering is performed by the `filtfilt` function in the python package `scipy.signal`[1] which, as the name suggests, applies the filter forward and then backwards. A side-effect of this approach is that the attenuation in decibel is twice that of the designed filter, since the filter is applied twice.

5.3.2 The Filter Coefficients

With the expression in (5.24) in place, the filter coefficients must be determined. Generally, digital filters can be divided into two types: 1) Finite Impulse Response (FIR) filters which have $M = 0$ and 2) Infinite Impulse Response (IIR) filters which have $M > 0$ [5]. To see the difference between the two filter types, consider the following. A digital filter must be designed for a signal with a sampling frequency of 10 Hz. The passband edge should be at

[1] Scipy is a state-of-the-art library for scientific computing with Python.

$\frac{1}{120}$ Hz, the stopband edge should be at $\frac{1}{12}$ Hz, the maximum passband attenuation must be 3 dB, and the minimum stopband attenuation must be 20 dB. Figure 5.1 shows the filter characteristics of three digital filters that adhere to these requirements: 1) an FIR filter designed by the window method (see e.g. [5]) which has $N = 531$ based on a rectangular window, 2) an FIR filter designed by the window method which has $N = 370$ based on a sinc function, and 3) a Butterworth (see e.g. [33]) IIR filter which has $N = M = 1$ derived by use of the bilinear transformation (see e.g. [5]). Because of the significantly fewer coefficients, an IIR filter is chosen.

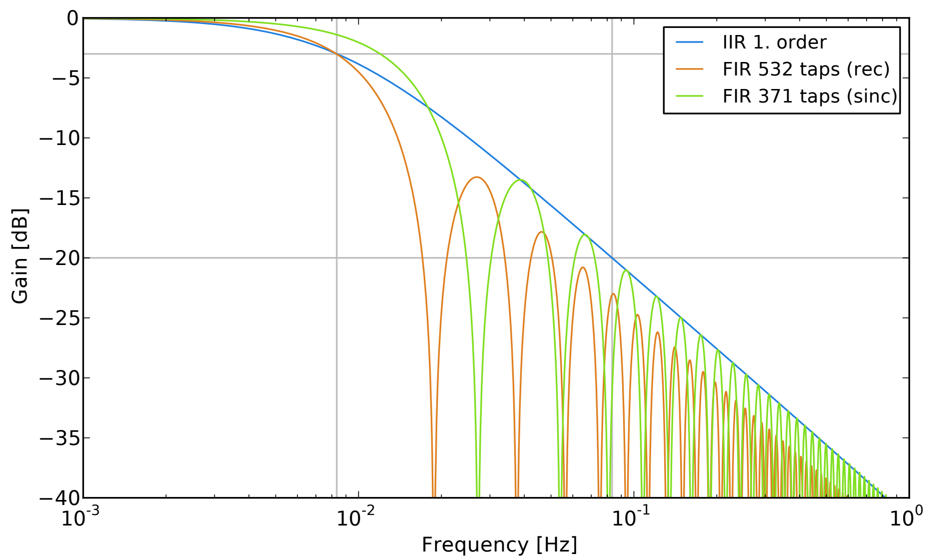

Figure 5.1. Illustration of the amplitude frequency response for different FIR and IIR filters. The FIR filters have been designed using the window method. FIR (rec) is obtained using a rectangular window whereas FIR (sinc) is obtained using a sinc window. The IIR filter is a Butterworth filter obtained by bi-linear transformation.

There exist a number of IIR filter types of which the following four types are considered (see e.g. [33] for detail about the different types): 1) Butterworth which has neither passband ripple nor stopband ripple, 2) Chebyshev type 1 which has passband ripple, 3) Chebyshev type 2 which has stopband ripple, and 4) elliptic which has both passband and stopband ripple. Figure 5.2 on the next page shows examples of the four types. For the purpose of separating the trend and turbulence, ripples in the passband and stopband are undesired and Butterworth is thus chosen due to its property of being maximally flat [33].

Before the filter coefficients can be calculated for the Butterworth filter, the requirements for this must be decided on. From a wind turbine load estimation perspective, the transition from trend to turbulence happens somewhere between 1-minute periods and 2-minute periods[2]. Thus everything with a frequency below $\frac{1}{120}$ Hz must be attributed to trend and everything with a frequency over $\frac{1}{60}$ Hz must be attributed to turbulence. If, for some application, the transition from trend to turbulence happens in another frequency interval, the frequencies mentioned can be changed accordingly. In

[2]Based on personal communicaiton with specialist Frede Aakmann Tøgersen, Vestas, and specialist Jesper Graugaard, Vestas, December 2012.

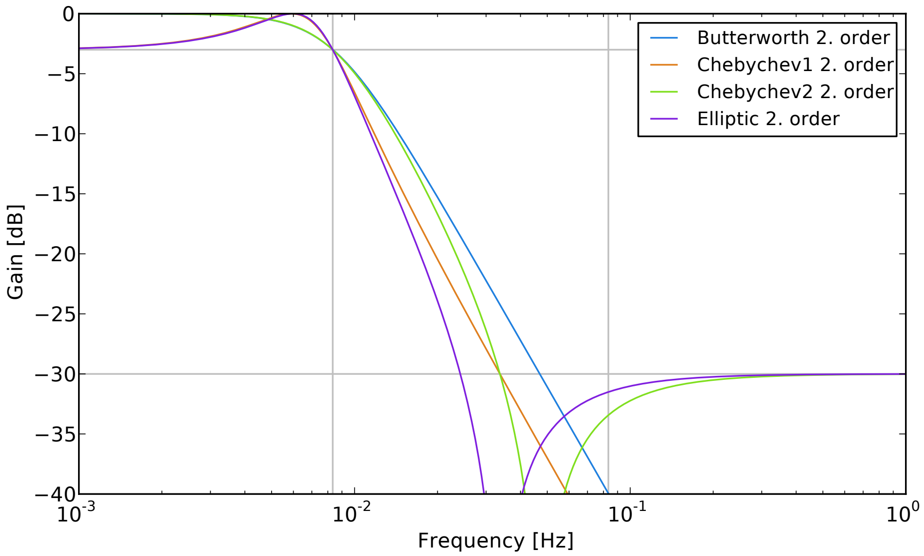

Figure 5.2. Illustration of the amplitude frequency response for different IIR filter types.

practice, the filter cannot have a gain of exactly 1 in some frequency band and a gain of exactly 0 in some other frequency band. Thus, a minimum gain for frequencies less than $\frac{1}{120}$ Hz of 0.9 and a maximum gain for frequencies greater than $\frac{1}{60}$ Hz of 0.1 are chosen. If, for some application, more or less restrictive gains are desired, the gains chosen can be changed accordingly. The current minimum requirements simply ensure that at most 10 % of a frequency component within the »trend frequency band« is included in the turbulence component. And, likewise, that at most 10 % of a frequency component within the »turbulence frequency band« is included in the trend component. To summarise, the requirements are the following when taking into account that the filter is applied twice:

- The passband edge must be at $\frac{1}{120}$ Hz
- The stopband edge must be at $\frac{1}{60}$ Hz
- The minimum gain in the passband must be at least $\sqrt{0.9} \approx 0.949$
- The maximum gain in the stopband must be at most $\sqrt{0.1} \approx 0.316$

With the requirements in place, the filter order can be chosen. To make the developed software applicable to new datasets with different sampling frequencies, the task is carried out dynamically in the developed software. Specifically, the choice of filter order and natural frequency, or 3 dB cut-off frequency, are performed by the `buttord` function in the python package `scipy.signal`. Basically, this function does the following:

1. The band edge frequencies are pre-warped as the bilinear transformation is later used to convert the analogue filter to a digital filter
2. The required order of the analogue filter is determined
3. The natural frequency of the analogue filter is determined (chosen such that the gain at the stopband edge is exactly the maximum gain allowed)

When the filter order and natural frequency have been chosen, the value of the filter coefficients can be computed. As with the choice of filer order and natural frequency, the task is carried out dynamically in the developed software. Specifically, the computation of the filter coefficients is performed by the `butter` function in the python package `scipy.signal`. Basically, this function does the following:

1. The natural frequency is pre-warped
2. The zeroes of the normalised, analogue filter coefficients are computed
3. The scaled, analogue filter coefficients are determined
4. The scaled, digital filter coefficients are determined using the bilinear transformation

5.3.3 Remaining Issues

There are two remaining issues. These are numerical issues with the filter coefficients and the considerations regarding the initial conditions of the filter. These two issues are discussed in the following.

Numerical Issues

With a sampling frequency of 10 Hz and the above requirements and design choices, `scipy` raises a warning that there are badly conditioned filter coefficients in the numerator. This warning is raised when one or more of the coefficients have an absolute value smaller than 10^{-14}. The following informal discussion explains the issue with small filter coefficients. Consider (5.10) which is restated here:

$$a_0 \cdot q[i] = \sum_{k=0}^{N} b_k \cdot p[i-k] - \sum_{k=1}^{M} a_k \cdot q[i-k] \tag{5.10}$$

Assume that a constant value is input to and output by the filter, i.e. $q[i] = q = p[i] = p$. While this is, obviously, not the case in practice, the input and output values easily have the same order of magnitude for wind speeds above $5\,\frac{\text{m}}{\text{s}}$. And especially so for a few consecutive measurements equal to the filter order plus one. Furthermore, with the given sampling frequency, requirements, and design choices, $a_0 = 1$ and $a_1, \ldots, a_M$ sums to -1 when rounding to six decimals. If the two sums are calculated before being subtracted, the right hand side of (5.10) can, thus, be approximated by

$$\sum_{k=0}^{N} b \cdot p + q \tag{5.26}$$

Next, ignore all but one b coefficient (this simplifies the discussion but leaves the conclusion unchanged) such that the expression is reduced to $b \cdot p + q$. With standard 64-bit floating-point values, the machine epsilon is $2.22 \cdot 10^{-16}$. This means that any number in the interval $[q - 1.11 \cdot 10^{-16} \cdot q, q + 1.11 \cdot 10^{-16} \cdot q)$ is rounded to q for the usual round-to-nearest kind of rounding. If $b = 10^{-14}$ then the expression becomes $10^{-14} \cdot p + q$ in which case the result will remain unchanged for all $p \in [q - 0.01 \cdot q, q + 0.01 \cdot q]$. That is, the input value can change by approximately 1 % without the output value changing. If b is made even smaller this percentage increases.

Since the problem with the small coefficients is the machine epsilon of 64-bit floating-point values, one possible solution might be to switch to 80-bit floating-point values which have a machine epsilon of $1.08 \cdot 10^{-19}$. However, the threshold of 10^{-14} is hard-coded in the source code of `scipy.signal` and filter coefficients below this threshold are discarded by `butter`. Thus, this solution would either require modifying `scipy.signal` or finding a suitable replacement which is undesirable. Another possible solution is to ease the requirements for the filter. However, the 0.9 and 0.1 gain requirements are not unreasonable requirements if any conclusions about the consequences of trend are to have any merit. Also, for the software to retain general applicability, the addition of new datasets with sampling frequencies greater than 10 Hz would necessitate the requirements to be eased even further.

Instead, it is noted that the smallest coefficient is in the order of 10^{-7} if a sampling frequency of 1 Hz is used instead. Thus, the chosen solution is to downsample the signal, apply the designed filter, and upsample the signal to the original sampling frequency. Furthermore, the signal must be low-pass filtered before being downsampled and after being upsampled to avoid aliasing. Specifically for a signal with sampling frequency f_s,

- The signal is low-pass filtered. The filter used is designed with passband edge at $\frac{f_\mathrm{s}}{200}$ Hz, stopband edge at $\frac{f_\mathrm{s}}{20}$ Hz, maximum passband attenuation of 1 dB, and minimum stopband attenuation of 60 dB. Furthermore, the python function `filtfilt` is used and, thus, the filter is zero-phase and has twice the stated attenuation.
- The signal is downsampled by a factor of 10. This is done by picking out every 10^{th} sample.
- The signal is low-pass filtered. The filter used is the one designed in the preceding part of this section.
- The signal is upsampled by a factor of 10. This is done by multiplying the signal by 10 and adding 9 zeros after each sample.
- The signal is low-pass filtered. The filter used is identical to the first filter used.

The reasoning behind the choices above are as follows. The first filter should not attenuate frequencies in the passband of the second filter. Since the second filter has its passband edge at $\frac{1}{120}$ Hz, it is chosen that the first filter must have its passband edge at, at least twice that frequency, i.e. at $\frac{1}{60}$ Hz. Next, it is chosen that the stopband edge should lie one decade higher than the passband edge. Furthermore, the stopband edge must lie no higher than at half the sampling frequency if aliasing is to be avoided. Thus, with the above choices, the minimum allowed sampling frequency after the downsampling is

$\frac{1}{3}$ Hz. The available high-rate dataset with the lowest sampling frequency has a sampling frequency of 4 Hz. Thus, to keep the method general, the downsampling factor must be less than or equal to 12. In practice, a downsampling factor of 10 was chosen. The maximum passband attenuation and minimum stopband attenuation were chosen as a compromise to avoid attenuating any trend and to avoid any aliasing.

Assuming that the signal was not oversampled in the first place, information is lost when the signal is low-pass filtered and downsampled. However, note that the frequency of the stopband edge of the second low-pass filter is lower than that of the first low-pass filter. Thus, the information which is lost when low-pass filtering and downsampling the signal is undesired and should be removed by the second filter otherwise. Note that the approach of downsampling and upsampling cannot be used if a high-pass filtering approach is taken to detrending, as suggested in Section 5.3.1 on page 65.

Initial Conditions

As stated in the beginning of this section, the initial conditions of the filter have to be considered. If each of the datasets had consisted of one uninterrupted time series, the initial conditions would have been less important as the beginning and end of the series could have simply been ignored. This is, however, not the case as can be seen from Table 2.5 on page 19: the Snaresmoor datasets consist of data chunks which consist of only one or two consecutive 10-minute intervals for the most part. Teufelsberge, on the other hand, contains longer chunks. To enable a »fair« comparison of the available datasets, the 10-minute intervals are detrended individually.

The task of determining the initial conditions of the filters is performed by the `lfilter_zi` called by `filtfilt` in the python package `scipy.signal`. Specifically, the function "finds the initial condition for which the response to an input of all ones is a constant"[3]. The initial conditions are then modified such that the response to a constant padding value is the constant padding value. The padding value preceding the chunk of samples is chosen as the mean value of the first $0.443 \cdot \frac{f_s}{f_p}$ (rounded to the nearest integer) measurements where f_p is the passband edge frequency of the filter applied afterwards. For an elaboration of the choice of number of samples averaged, note that the 3 dB cut-off frequency of a rectangular window with l taps is $0.443 \cdot \frac{f_s}{l}$ Hz.

The effect of having to filter individual 10-minute intervals rather than being able to filter a large number of consecutive 10-minute intervals can be seen from the Teufelsberge dataset. The largest number of consecutive 10-minute intervals in this dataset is 3391. First, the sum of the trend and mean wind speed is estimated on the entire chunk of measurements and the result is denoted $T_{\text{full}}[i] + V_{\text{full}}[i]$. Next, the sum of the trend and mean wind speed is estimated on the individual chunks of measurements before being pieced together and the result is denoted $T_{\text{parts}}[i] + V_{\text{parts}}[i]$.

[3] `http://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.lfilter_zi.html#scipy.signal.lfilter_zi`

The following statistics describe the difference between the two:

$$\frac{1}{M}\sum_{i=0}^{M-1}|(T_{\text{parts}}[i]+V_{\text{parts}}[i])-(T_{\text{full}}[i]+V_{\text{full}}[i])|=0.07 \tag{5.27}$$

$$\frac{1}{M}\sum_{i=0}^{M-1}\frac{|(T_{\text{parts}}[i]+V_{\text{parts}}[i])-(T_{\text{full}}[i]+V_{\text{full}}[i])|}{|T_{\text{full}}[i]+V_{\text{full}}[i]|}=0.01 \tag{5.28}$$

where M is the number of measurements in the entire chunk. Thus, the average absolute difference between the two estimates is 0.07 $\frac{\text{m}}{\text{s}}$ while the average relative difference between them is 1 %. With these small differences, the effects of using detrending on individual 10-minute intervals instead of on consecutive 10-minute intervals are deemed negligible.

5.4 Detrending Examples

Figures 5.3 to 5.7 on page 76 show examples of measured wind speeds in 10-minute intervals before and after detrending. The examples are based on the same measurements as the examples in Figures 4.1 on page 47 to 4.5 on page 52.

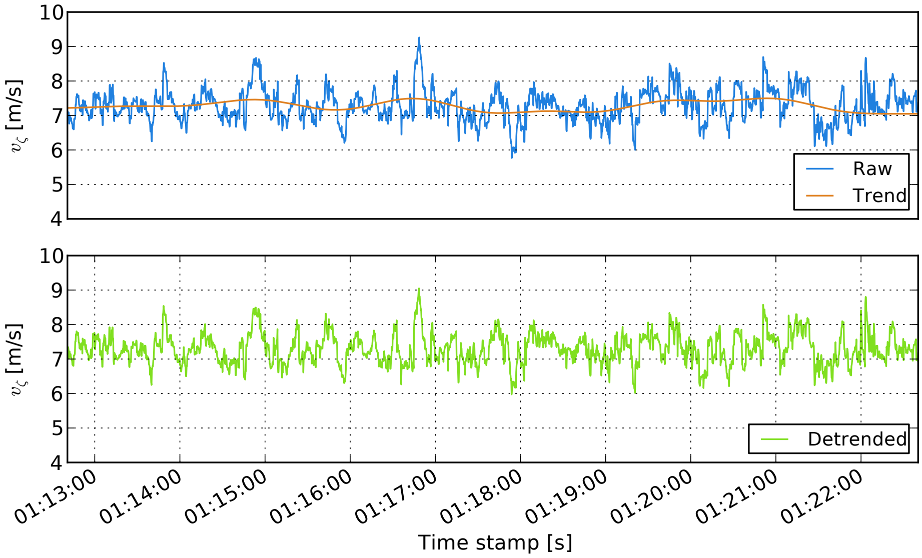

Figure 5.3. Example 1 of measured horizontal wind speed in a 10-minute interval before and after detrending. Upper figure: The raw signal and the trend signal from the example in Figure 4.1 on page 47. Lower figure: The detrended signal. The data is from Teufelsberge (07-11-2006).

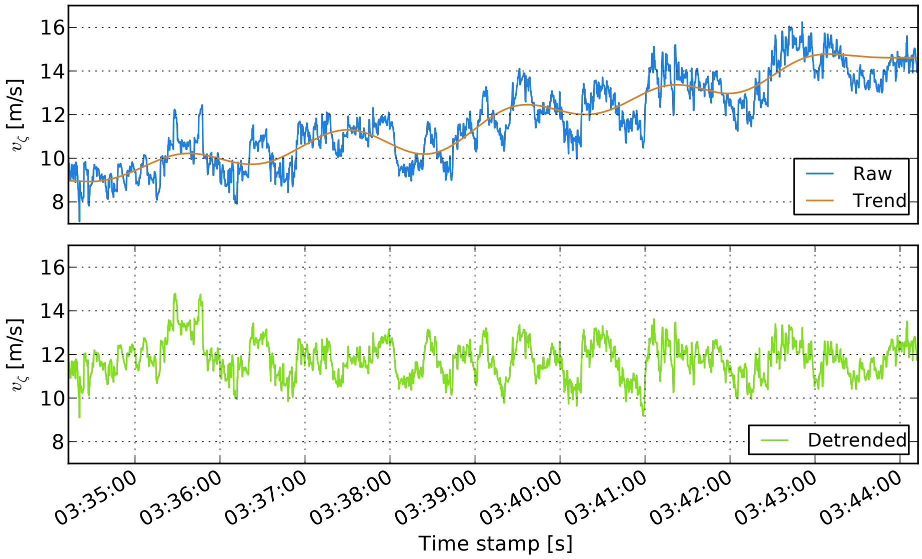

Figure 5.4. Example 2 of measured horizontal wind speed in a 10-minute interval before and after detrending. Upper figure: The raw signal and the trend signal from the example in Figure 4.2 on page 48. Lower figure: The detrended signal. The data is from Teufelsberge (28-11-2006).

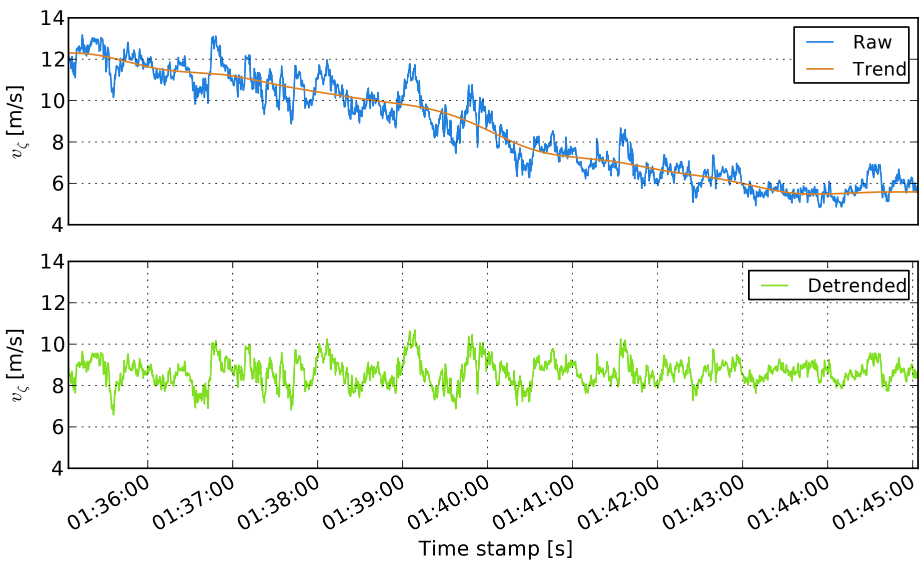

Figure 5.5. Example 3 of measured horizontal wind speed in a 10-minute interval before and after detrending. Upper figure: The raw signal and the trend signal from the example in Figure 4.3 on page 50. Lower figure: The detrended signal. The data is from Teufelsberge (06-08-2006).

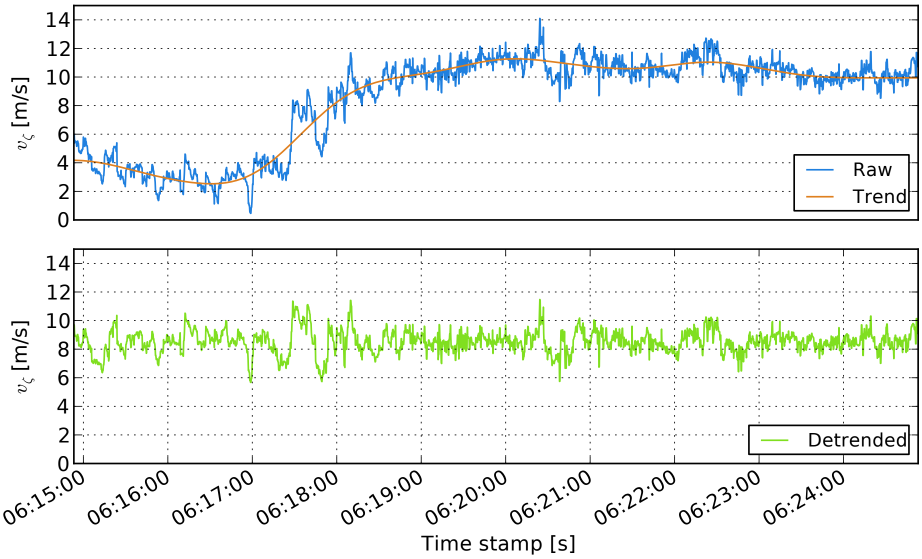

Figure 5.6. Example 4 of measured horizontal wind speed in a 10-minute interval before and after detrending. Upper figure: The raw signal and the trend signal from the example in Figure 4.4 on page 51. Lower figure: The detrended signal. The data is from Teufelsberge (22-10-2006).

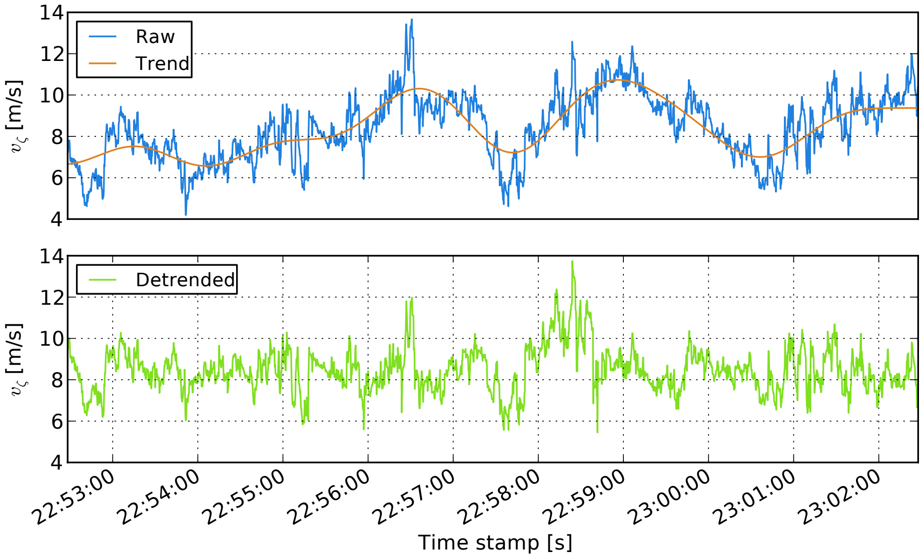

Figure 5.7. Example 5 of measured horizontal wind speed in a 10-minute interval before and after detrending. Upper figure: The raw signal and the trend signal from the example in Figure 4.5 on page 52. Lower figure: The detrended signal. The data is from Teufelsberge (29-09-2006).

5.5 Results

To assess the results of using detrending several of the measures and statistics discussed in the Chapters 3 on page 21 and 4 on page 47 are reevaluated after the high-rate wind measurements have been detrended according to the approach outlined in Section 5.3 on page 64. A comparison of the results before and after detrending is presented in this section.

5.5.1 Classification Results

The classification results before detrending is presented in Tables 4.5 on page 58 to 4.7 on page 59. The corresponding results after detrending is presented in Tables 5.1 on the next page to 4.7 on page 59.

Table 5.1 on the next page clearly shows that there are no significant trends left after the detrending has been performed. This is not really a surprise. However, note that the detrending has been performed in the *absolute rectangular coordinates* as described in Section 5.3.1 on page 65, whereas the classification result states that no trend is left when evaluated in the *relative rectangular coordinates*. Hence, since all trends are classified as *insignificant* the detrending method used ensures that there are no trends left that can influence the estimate of the turbulence standard deviation in such a way that an IEC subclass B wind turbine is required instead of an IEC subclass C when evaluated in 10-minute intervals (see Section 4.1.1 on page 48 for the definition of an insignificant trend).

The fact that no significant trends are left after detrending is also reflected in the (conditional) classification results in Tables 5.2 on the following page and 5.3 on page 79. However, a few other results are also worth noting from these tables. Comparing with the results before detrending (Tables 4.6 on page 59 and 4.7 on page 59), it can be seen that the estimated turbulence standard deviation $\hat{\sigma}_\zeta$ in 10-minute intervals is generally lower. That is, the number of occurrences of turbulence standard deviation in the intervals $[0, 1)$ and $[1, 2)$ has increased. Consequently, the number of occurrences in the intervals $[2, 3)$ and $[3, \infty)$ has decreased. In fact nearly no occurrences of »extreme« turbulence standard deviation in the interval $[3, \infty)$ are present after detrending. The fact that only low values of the turbulence standard deviation are left after detrending is particularly distinct for the Snaresmoor data as can be seen from Table 5.3 on page 79. Note that the slightly lower total number of 10-minute intervals (the sum-sum entry) in the tables with results after detrending compared to the tables with results before detrending is due to a few of the of the 10-minute intervals having a mean wind speed below 5 $\frac{\text{m}}{\text{s}}$ after detrending.

Dataset	Insignificant (I) Occurrences	%	Linear (L) Occurrences	%	Periodic (P) Occurrences	%
/Snaresmoor/Raw_M1_20	25933	100	0	0	0	0
/Snaresmoor/Raw_M1_40	25022	100	0	0	0	0
/Snaresmoor/Raw_M1_60	31285	100	0	0	0	0
/Snaresmoor/Raw_M1_80	30226	100	0	0	0	0
/Snaresmoor/Raw_M2_20	25512	100	0	0	0	0
/Snaresmoor/Raw_M2_40	28025	100	0	0	0	0
/Snaresmoor/Raw_M2_60	6040	100	0	0	0	0
/Snaresmoor/Raw_M2_80	28888	100	0	0	0	0
/Snaresmoor/Raw_M4_20	26064	100	0	0	0	0
/Snaresmoor/Raw_M4_40	32108	100	0	0	0	0
/Snaresmoor/Raw_M4_60	30774	100	0	0	0	0
/Snaresmoor/Raw_M4_80	32070	100	0	0	0	0
/Snaresmoor/Raw_M5_20	26067	100	0	0	0	0
/Snaresmoor/Raw_M5_40	28536	100	0	0	0	0
/Snaresmoor/Raw_M5_60	26819	100	0	0	0	0
/Snaresmoor/Raw_M5_80	22862	100	0	0	0	0
/Teufelsberge/T07_HF	24093	100	0	0	0	0

Table 5.1. The results of classification of 10-minute intervals of detrended data into the classes of trend: insignificant (I), linear (L), and periodic (L). The results for all available dataset from Teufelsberge and Snaresmoor are shown.

$\hat{\sigma}_\zeta$ / $\hat{V}_\zeta$	$[0,1)$ I	L	P	$[1,2)$ I	L	P	$[2,3)$ I	L	P	$[3,\infty)$ I	L	P	Sum
$[5,10)$	6923	0	0	768	0	0	1	0	0	0	0	0	7692
$[10,15)$	4086	0	0	4755	0	0	105	0	0	0	0	0	8946
$[15,20)$	462	0	0	3131	0	0	635	0	0	2	0	0	4230
$[20,25)$	7	0	0	385	0	0	422	0	0	10	0	0	824
$[25,\infty)$	0	0	0	6	0	0	34	0	0	3	0	0	43
Sum	11478	0	0	9045	0	0	1197	0	0	15	0	0	21735

Table 5.2. The results of classification of 10-minute intervals of detrended data conditioned on mean wind speed and turbulence standard deviation. The columns indicate the turbulence standard deviation intervals whereas the rows indicate the mean wind speed intervals. I, L, and P denotes the trend classes: insignificant, linear, and periodic, respectively **[Teufelsberge]**.

$\hat{\sigma}_\zeta$	[0, 1)			[1, 2)			[2, 3)			[3, ∞)			Sum
$\hat{V}_\zeta$	I	L	P	I	L	P	I	L	P	I	L	P	
[5, 10)	19814	0	0	601	0	0	0	0	0	0	0	0	20415
[10, 15)	2484	0	0	1495	0	0	0	0	0	0	0	0	3979
[15, 20)	6	0	0	194	0	0	2	0	0	0	0	0	202
[20, 25)	0	0	0	7	0	0	0	0	0	0	0	0	7
[25, ∞)	0	0	0	0	0	0	0	0	0	0	0	0	0
Sum	22304	0	0	2297	0	0	2	0	0	0	0	0	24603

Table 5.3. The results of classification of 10-minute intervals of detrended data conditioned on mean wind speed and turbulence standard deviation. The columns indicate the turbulence standard deviation intervals whereas the rows indicate the mean wind speed intervals. I, L, and P denotes the trend classes: insignificant, linear, and periodic, respectively **[Snaresmoor M1 60]**.

5.5.2 Assumption Related Results

Chapter 3 on page 21 presents an evaluation of the Teufelsberge data and Snaresmoor data in the light of the assumptions behind the current state-of-the-art model of turbulent wind fields. This section presents a equivalent evaluation of the assumptions based on the detrended data.

The horizontal mean wind speed V_ζ in 10-minute intervals is assumed to be Rayleigh distributed, whereas the vertical mean wind speed V_ψ in 10-minute intervals is assumed to be zero as stated in Chapter 3 on page 21. From Figures 5.8 on the next page and 5.9 on the following page it can be seen that the influence of the detrending on the estimate $\hat{V}_\zeta$ is limited. The histograms before and after detrending are almost equal and the ML fitted Rayleigh parameter σ is 8.23 before detrending versus 8.25 after detrending for the Teufelsberge data, whereas it is the same before and after detrending for the Snaresmoor data. This is to be expected since the removed trend is defined as being zero-mean (according to the definition in Section 5.1 on page 61).

Due to the trend in the vertical plane being non-zero by definition (see Section 5.1 on page 61), the detrending should force $\hat{V}_\psi$ to zero. However, due to the non-ideal filtering, it is to be expected $\hat{V}_\psi$ is not exactly zero. This is indeed is also the case as can be seen from figures 5.8 on the next page and 5.9 on the following page. It should be noted though, that the assumption of V_ψ being zero is very close to being fulfilled after the data has been detrended.

The probability density histograms of the detrended turbulence standard deviation condition on the mean wind speed in Figures 5.10 on page 81 to 5.13 on page 82 reveal that both the mean value and 90 % percentile of $\hat{\sigma}_\zeta$ decrease substantially compared to before detrending. In fact the 90 % percentile after detrending is close to coinciding with the mean value before detrending. This result is evident from both the directly estimated mean values and 90 % percentiles and the values estimated from the ML fitted log-normal distributions. The »strange« behaviour of the estimated mean value and 90 % percentiles in Figure 5.13 on page 82 for larger values of $\hat{V}_\zeta$ is likely due to a limited number of 10-minute intervals available for calculating the ML fit of a log-normal distribution.

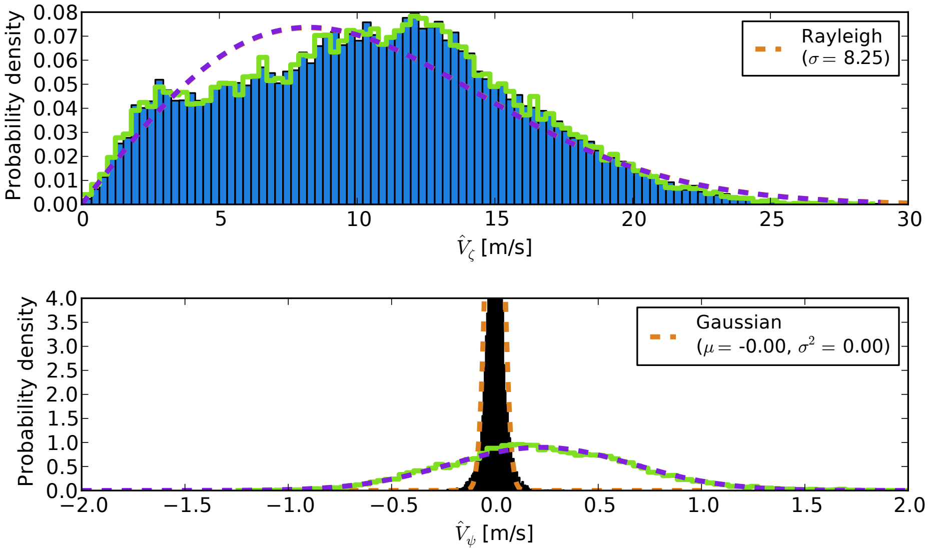

Figure 5.8. Probability density histograms of the detrended mean wind speed in 10-minute intervals. The orange dashed curves show ML estimates of the assumed distributions. For comparison, the histogram envelopes and ML estimates before detrending (from Figure 3.4 on page 31) are depicted as the green and purple curves, respectively **[Teufelsberge]**.

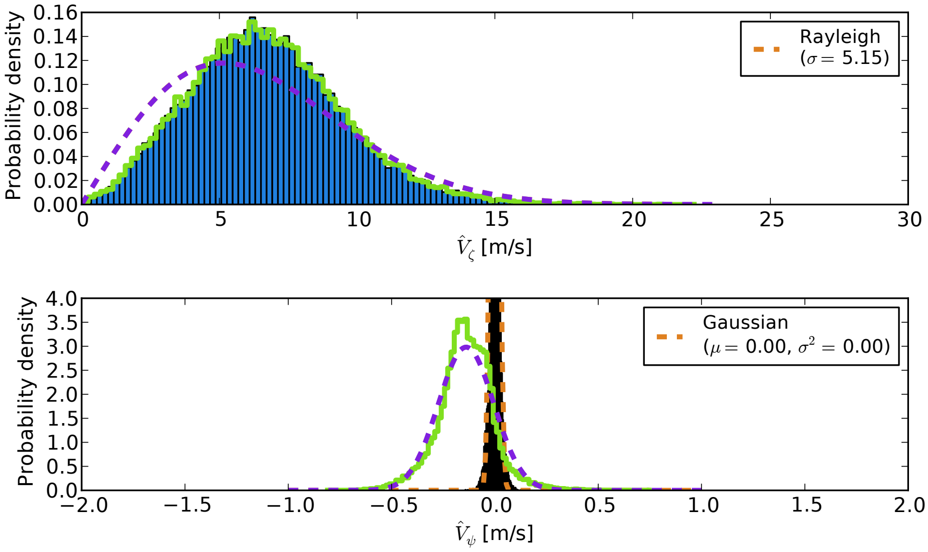

Figure 5.9. Probability density histograms of detrended the mean wind speed in 10-minute intervals. The orange dashed curves show ML estimates of the assumed distributions. For comparison, the histogram envelopes and ML estimates before detrending (from Figure 3.5 on page 32) are depicted as the green and purple curves, respectively **[Snaresmoor M1 60]**.

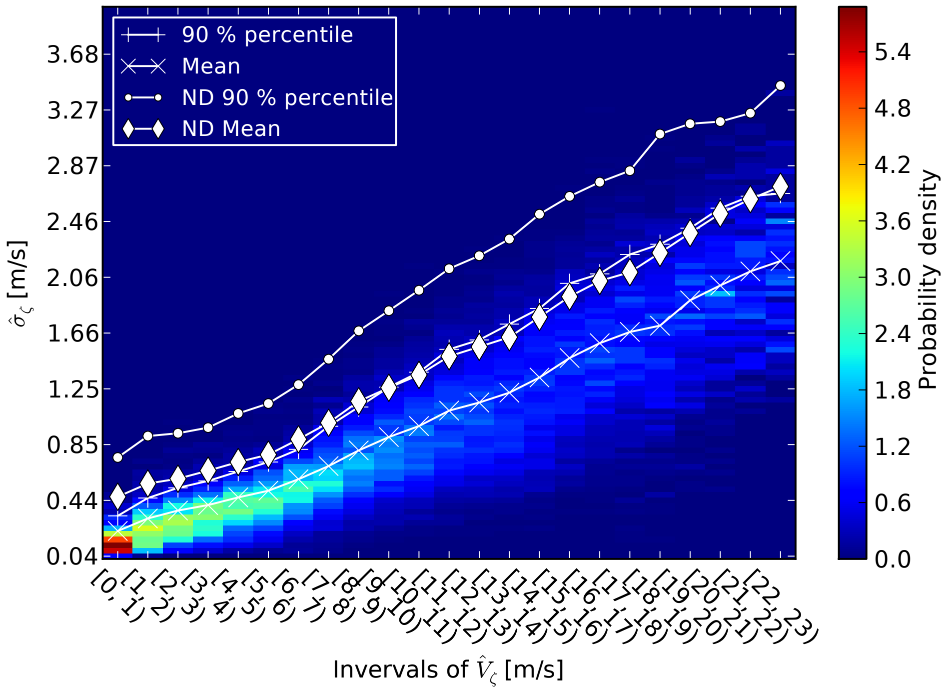

Figure 5.10. Probability density histograms of the detrended wind speed standard deviation in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The mean wind speed is divided into bins. Only bins that contain at least 100 samples are shown. The density is colour coded according to the scale on the right. Superimposed on the density is the calculated mean and 90 % percentile of the samples. For comparison, the curves marked *ND* depict the mean and 90 % percentile before detrending (from Figure 3.8 on page 34) **[Teufelsberge]**.

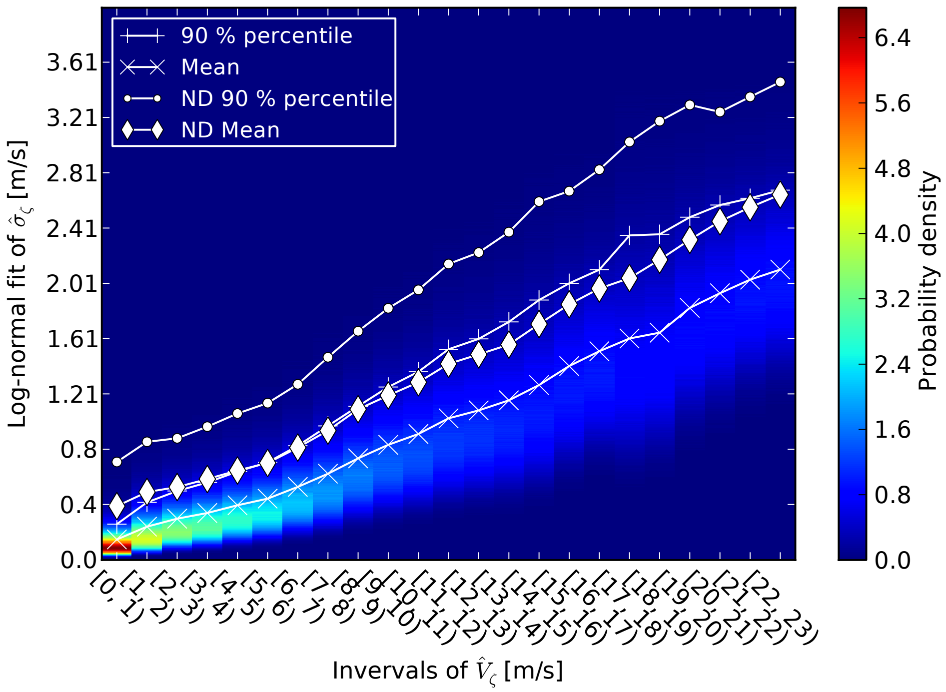

Figure 5.11. Log-normal ML fits calculated from the detrended wind speed standard deviation in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The mean wind speed is divided into bins. Only bins that contain at least 100 samples are shown. The density is colour coded according to the scale on the right. Superimposed on the density is the mean and 90 % percentile resulting from the fit. For comparison, the curves marked *ND* depict the mean and 90 % percentile before detrending (from Figure 3.9 on page 35) **[Teufelsberge]**.

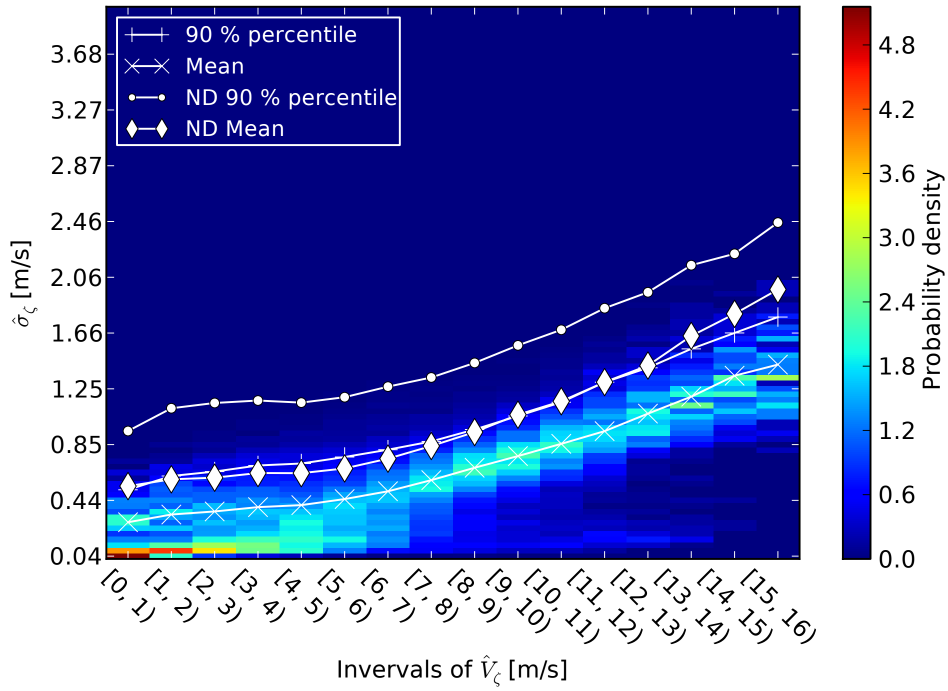

Figure 5.12. Probability density histograms of the detrended wind speed standard deviation in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The mean wind speed is divided into bins. Only bins that contain at least 100 samples are shown. The density is colour coded according to the scale on the right. Superimposed on the density is the calculated mean and 90 % percentile of the samples. For comparison, the curves marked *ND* depict the mean and 90 % percentile before detrending (from Figure 3.10 on page 35) **[Snaresmoor M1 60]**.

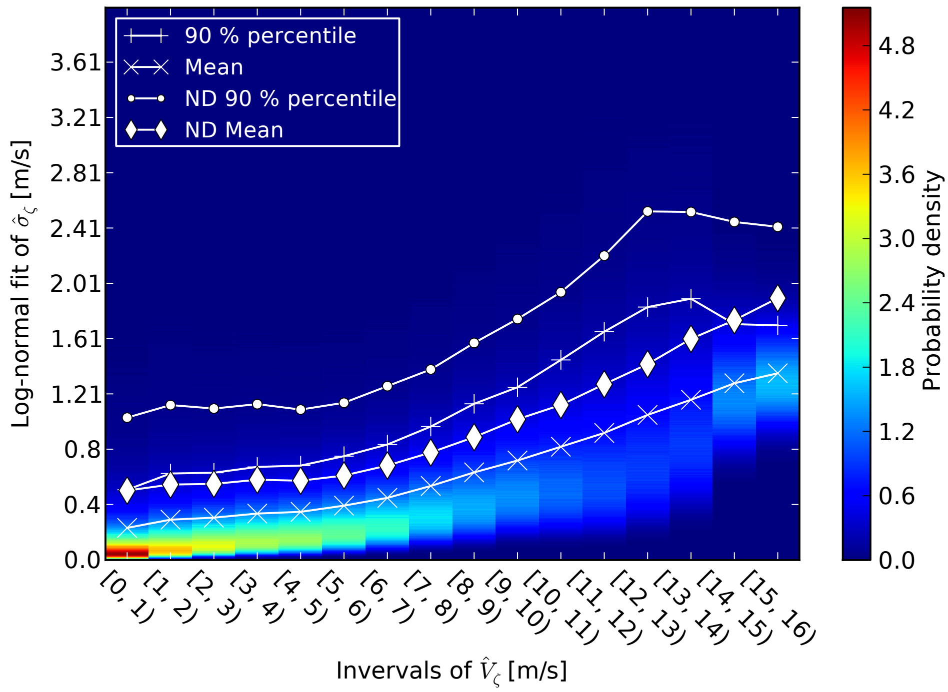

Figure 5.13. Log-normal ML fits calculated from the detrended wind speed standard deviation in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The mean wind speed is divided into bins. Only bins that contain at least 100 samples are shown. The density is colour coded according to the scale on the right. Superimposed on the density is the mean and 90 % percentile resulting from the fit. For comparison, the curves marked *ND* depict the mean and 90 % percentile before detrending (from Figure 3.11 on page 36) **[Snaresmoor M1 60]**.

A further comparison of the 90 % percentiles from Figures 5.10 on page 81 to 5.13 on the preceding page with the IEC standard subclasses A, B, and C is shown in Figures 5.14 and 5.15 on the next page. From the figures it can be concluded that the use of detrending makes the estimated turbulence standard deviation $\hat{\sigma}_\zeta$ decrease to a level that corresponds to at least one step down the IEC standard subclass hierarchy, e.g. comparing the blue and green lines in Figure 5.14 shows that $\hat{\sigma}_\zeta$ decreases from a level around the IEC subclass B to a level around the IEC subclass C. This result for the 90 % percentile across all 10-minute intervals is, thus, in line with the classification results of the individual 10-minute intervals presented in Tables 5.1 on page 78 (the result that the trends after detrending are classified as insignificant).

Another interesting aspect to note from Figures 5.14 and 5.15 on the next page is that the estimated turbulence intensity $\hat{I}$ conditioned on the mean wind speed versus the mean wind speed is closer to being a horizontal line after detrending, i.e. that estimated turbulence intensities for smaller mean wind speeds after detrending is substantially lower than the corresponding values before detrending.

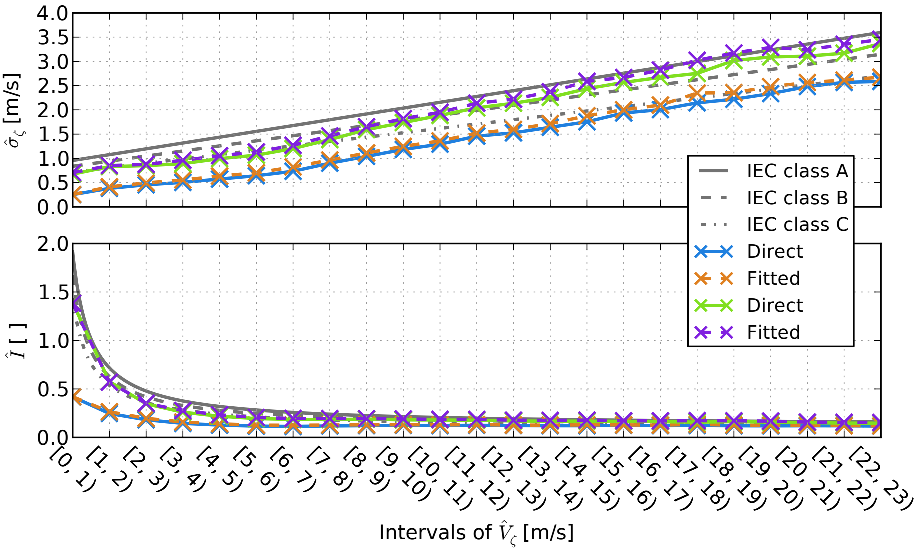

Figure 5.14. 90 % percentiles (blue and orange curve) of the detrended wind speed standard deviation and detrended turbulence intensity in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The 90 % percentiles marked *fitted* are calculated from a ML fit of a log-normal distribution to the samples whereas the 90 % percentiles marked *direct* are estimated directly from the samples. For comparison, the corresponding curves before detrending (from Figure 3.12 on page 37) are depicted as the green and purple lines. The mean wind speed is divided into bins. Only bins that contain at least 100 samples are shown. Also shown are the corresponding curves of the IEC 61400 standard classes based on the midpoint of the mean wind speed bins **[Teufelsberge]**.

The use of detrending has a substantial influence on the connection between estimated turbulence standard deviation ($\hat{\sigma}_\zeta$, $\hat{\sigma}_\xi$, or $\hat{\sigma}_\psi$) and the choice of aggregation period. The specific results are depicted in Figures 5.16 on page 85 to 5.18 on page 86. Both the mean value and the 90 % percentile in all figures change from being clearly increasing throughout the range of aggregation periods before detrending to being increasing for

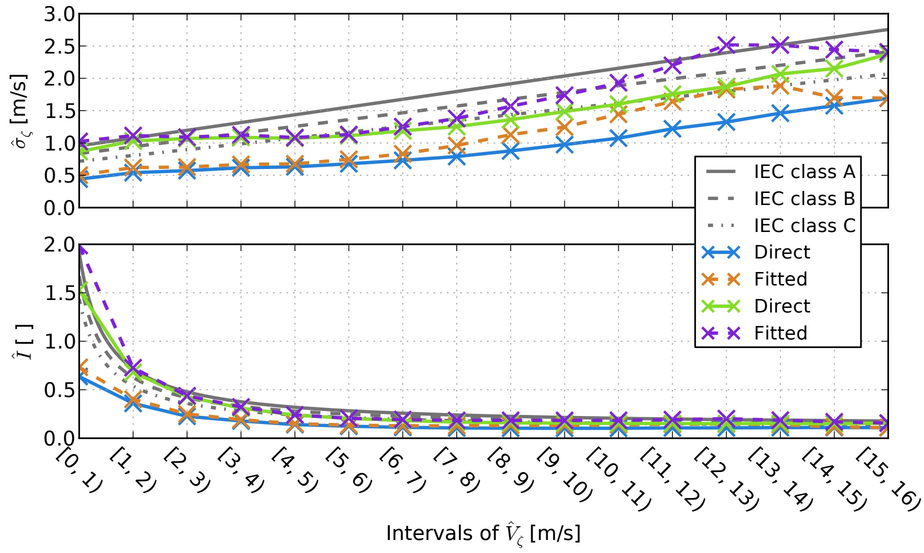

Figure 5.15. 90 % percentiles (blue and orange curve) of the detrended wind speed standard deviation and detrended turbulence intensity in 10-minute intervals conditioned on the mean wind speed in 10-minute intervals. The 90 % percentiles marked *fitted* are calculated from a ML fit of a log-normal distribution to the samples whereas the 90 % percentiles marked *direct* are estimated directly from the samples. For comparison, the corresponding curves before detrending (from Figure 3.13 on page 37) are depicted as the green and purple lines. The mean wind speed is divided into bins. Only bins that contain at least 100 samples are shown. Also shown are the corresponding curves of the IEC 61400 standard classes based on the midpoint of the mean wind speed bins **[Snaresmoor M1 60]**.

shorter aggregation periods and close to constant for long aggretation periods after detrending. The mean values before and after detrending are nearly the same for an aggregation period of 30 s. The tendency is the same for the 90 % percentile. It is nearly constant between longer aggregation periods. However, due to the »flat« lines toward longer aggregation periods after detrending, the difference between the estimated turbulence standard deviation before and after detrending is larger the longer an aggregation period, that is used.

As described in Section 5.3.2 on page 68, the low-pass filter used for detrending of the signals is designed such that the pass-band ends at a frequency of $\frac{1}{120}$ Hz whereas the stop-band starts at a frequency of $\frac{1}{60}$ Hz. It is interesting to note that the »bend«, that is the curve's deviation from being nearly constant, of the mean and 90 % percentile curves in Figures 5.16 on the next page to 5.18 on page 86 happens somewhere between 60 s and 120 s indicating a correlation between this »bend« and the choice of filter pass-band.

Comparing with the results before detrending (Figures 3.15 on page 39 to 3.22 on page 43), the probability density after detrending is more constant between the different aggregation periods longer than 120 s which is in line with the mean value and 90 % percentile also being more constant between these aggregation periods.

The measures in Table 5.4 on the next page show that neither ① nor ② in (3.40) are satisfied with equality. That is, both the mean wind speed and turbulence standard deviation are, despite the detrending, still different in subintervals of the 10-minute intervals. However, the smaller differences between the measures in Table 5.4 for longer aggregation periods compared to result before detrending (Table 3.2 on page 45) shows that ① and ② are closer to being satisfied with equality after detrending than before. Especially, the measure B is substantially closer to being equal between aggregation periods, i.e. ① is closer to being satisfied with equality. This seems reasonable, since ① relates to the mean value between sub-intervals which ought to become more constant when the trend is removed.

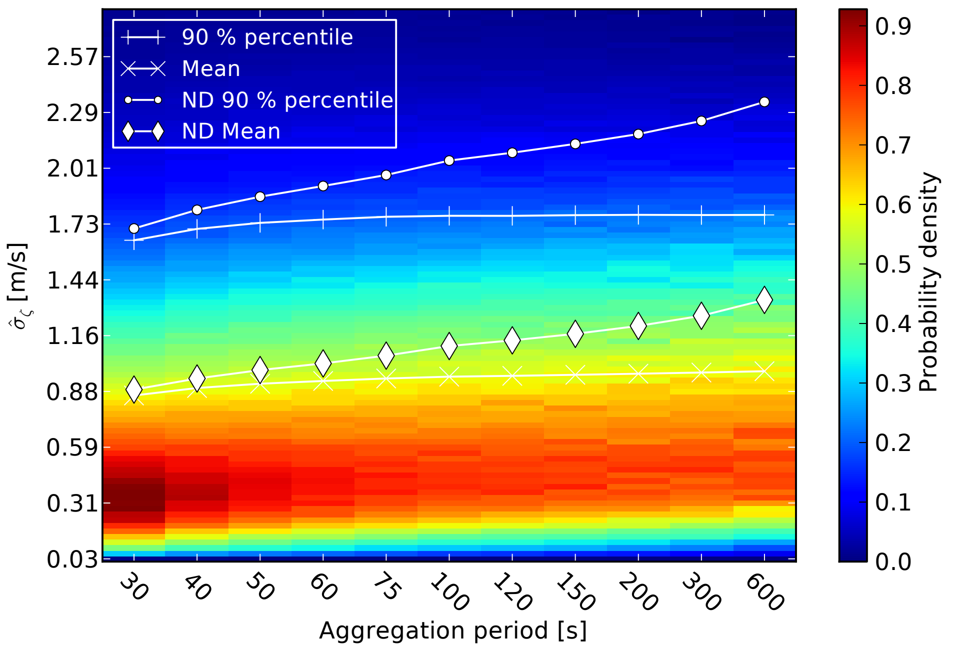

Figure 5.16. Probability density histograms of the detrended wind speed standard deviation (of $\hat{\sigma}_\zeta$) versus aggregation period. The density is colour coded according to the scale on the right. Superimposed on the density is the calculated mean and 90 % percentile of the samples. For comparison, the curves marked *ND* depict the mean and 90 % percentile before detrending (from Figure 3.20 on page 42). Note that the aggregation period scale is not uniform **[Teufelsberge]**.

Measure		30 s	40 s	50 s	60 s	75 s	100 s	120 s	150 s	200 s	300 s	600 s
v_ζ	B	0.76	0.77	0.78	0.79	0.79	0.79	0.79	0.79	0.80	0.80	0.80
	C	0.72	0.74	0.76	0.76	0.77	0.78	0.78	0.78	0.79	0.79	0.80
v_ξ	B	0.87	0.90	0.91	0.92	0.93	0.93	0.93	0.94	0.94	0.94	0.94
	C	0.83	0.86	0.88	0.89	0.91	0.91	0.92	0.92	0.93	0.93	0.94
v_ψ	B	0.84	0.87	0.89	0.90	0.91	0.91	0.91	0.92	0.92	0.92	0.92
	C	0.80	0.84	0.86	0.87	0.89	0.90	0.90	0.90	0.91	0.92	0.92

Table 5.4. Mean values (across all intervals of the respective period) of the measures B and C in (3.42) and (3.43), respectively, calculated from the data illustrated in Figures 5.16 (v_ζ), 5.17 on the next page (v_ξ), and 5.18 on the following page (v_ψ) **[Teufelsberge]**.

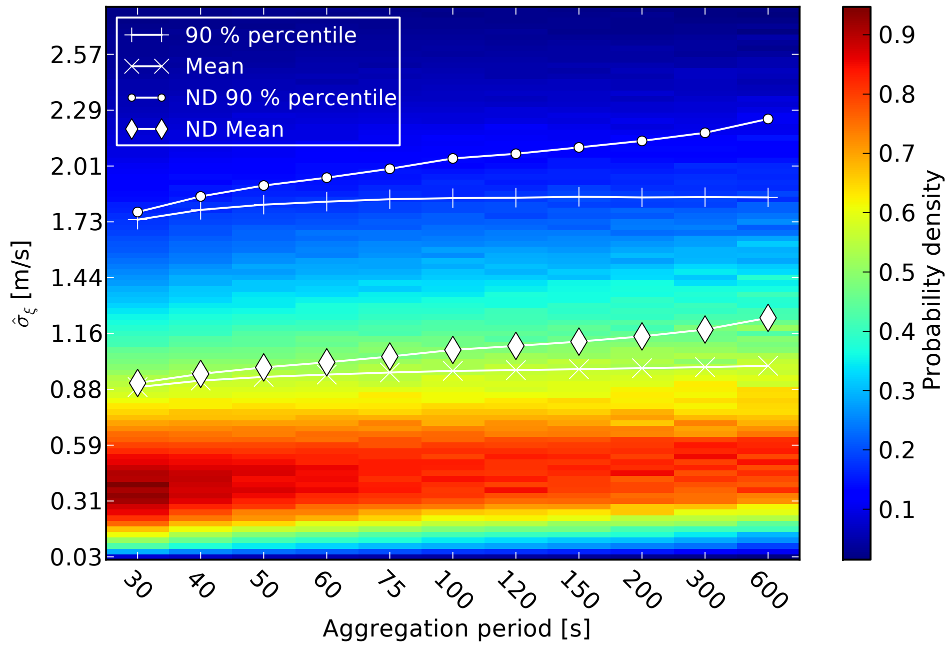

Figure 5.17. Probability density histograms of the detrended wind speed standard deviation (of $\hat{\sigma}_\xi$) versus aggregation period. The density is colour coded according to the scale on the right. Superimposed on the density is the calculated mean and 90 % percentile of the samples. For comparison, the curves marked *ND* depict the mean and 90 % percentile before detrending (from Figure 3.21 on page 43). Note that the aggregation period scale is not uniform **[Teufelsberge]**.

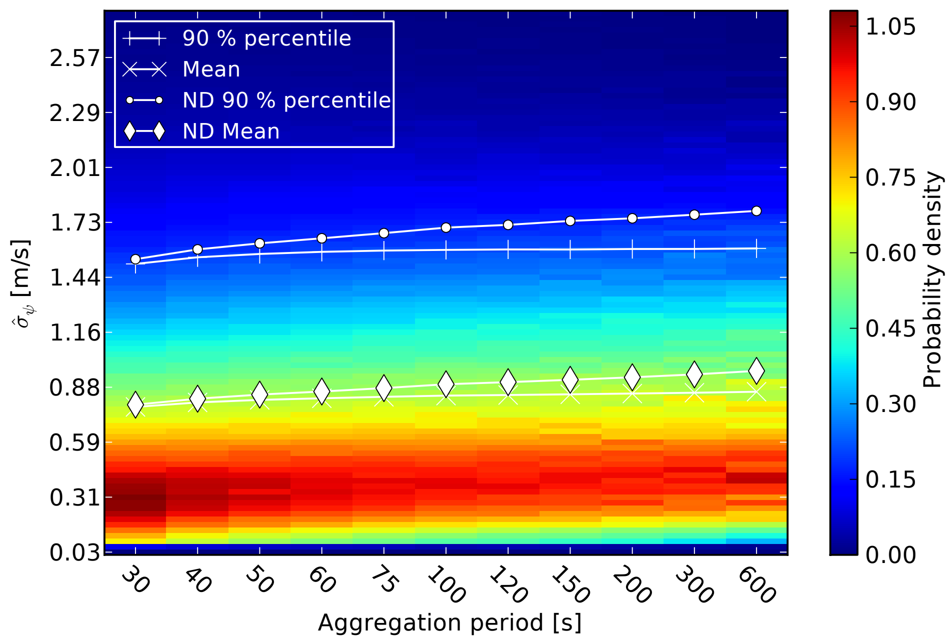

Figure 5.18. Probability density histograms of the detrended wind speed standard deviation (of $\hat{\sigma}_\psi$) versus aggregation period. The density is colour coded according to the scale on the right. Superimposed on the density is the calculated mean and 90 % percentile of the samples. For comparison, the curves marked *ND* depict the mean and 90 % percentile before detrending (from Figure 3.22 on page 43). Note that the aggregation period scale is not uniform **[Teufelsberge]**.

6 Developed Software

A Python package named *Wind Analysis Framework (WAF)* has been developed in connection with this report. This chapter provides an overview of structure of WAF in addition to an introduction to its basic usage.

6.1 An Overview of WAF

WAF consists of 10 subpackages. An illustration of this subpackage structure is given in Figure 6.1.

`waf`
- `aggregation` .... Aggregation functionality
- `analysis` .... Plotting and description functionality
- `classification` .... Trend classification functionality
- `detrending` .... Detrending functionality
- `downsampling` .... Downsampling functionality
- `io` .... Utilities for building an HDF database
- `multiprocessing` .... Utilities for exploiting parallel computations
- `postprocessing` .... Postprocessing functions
- `unittesting` .... Unit tests for WAF subpackages
- `utils` .... Functionality for coordinate change, input validation, and misc.

Figure 6.1. Illustration of the subpackage structure of the developed Python package Wind Analysis Framework (WAF).

The WAF Python package has been designed to be used with the Enthought Python Distribution (EPD) version 7.3-2[1] with the following upgraded »eggs«:

- pandas 0.10.1-1
- matplotlib 1.1.0-2

All data processing presented in this report have been done using on a computer featuring a 4-way Intel Xeon E7-4850 2.00 GHz based 64-bit system (a total of 40 CPU cores) with 512 GB memory running 64-bit Ubuntu 12.04.1 LTS Linux and a 64-bit version of the EPD setup outlined above. See the README file found in the WAF package for the specific details about requirements for WAF.

[1]For more details about the Enthought Python Distribution see `https://www.enthought.com/products/epd/`.

6.2 Basic Usage

The software directory on the enclosed CD contains a script named »build.py«. It may be used to perform the basic data processing described in this report. Specifically, the following is done by running the script:

1. An HDF database containing the wind velocity measurements is built according to the specification given in Chapter 2 on page 7.
2. Trend classfication features of the wind velocity measurements are calculated according to the specification in Chapter 4 on page 47.
3. The measurements are aggregated according to the specification in Chapter 3 on page 21.
4. The wind velocity measurements are detrended according to the specification in Chapter 5 on page 61 and stored in another HDF database.
5. Trend classfication features of the detrended wind velocity measurements are calculated according to the specification in Chapter 4 on page 47.
6. The detrended measurements are aggregated according to the specification in Chapter 3 on page 21.

See the »build.py« script for further details.

6.2.1 Specifying the Content of the HDF database

A prerequisite to building the HDF database of wind measurements is a specification of its outline. The structure of the HDF database must be specified, i.e. Groups and Tables must be declared. Additionally, the datasets (CSV-files) and meta-data information that is stored in each Table must be specified. An XML markup, which has been developed for the specific purpose, is used to describe this database outline. An example of such an outline is given in Listing 6.1 on the facing page.

6.2.2 Further Processing of the Measurements

Once the »build.py« script has been executed further processing of the wind velocity measurements may be done using the functionality in the *waf.analysis* subpackage or any other Python function the user finds appropriate for the task. For inspiration see the scripts used to produce the figures and tables in this report that contain data processing results. They are collected in the *report* Python package which is found on the enclosed CD.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="db.xsl"?>

<root title="Wind database version 1">
  <group name="Location1">
    <table name="UltraSonic" description="d3_wind_measurement"
        title="Ultra sonic wind data from Location1"
        rows_expected="100000">
      <info name="gath_period">2006-2008</info>
      <data dialect="meas1" relative_height="0" path="./">
        <file head="true">data_file_1.csv</file>
        <file head="false">data_file2.csv</file>
      </data>
    </table>
  </group>
  <group name="Location2">
    <table name="Lidar" description="d3_wind_measurement"
        title="Some lidar wind data from Location2"
        rows_expected="20000">
      <data dialect="meas2" relative_height="0" path="./">
        <file head="false">data_file_3.csv</file>
      </data>
    </table>
  </group>
</root>
```

Listing 6.1. Example of XML outline specification of database.

7 Conclusions

The topic of trends in the mean wind speed V and turbulence standard deviation σ_{v} is the main topic addressed in this report. As stated in the Problem Thesis in Section 1.3 on page 5 the objective is to gain knowledge about such trends from high-rate wind measurements. Furthermore, a connection between trends and the 10-minute statistics of wind measurements used in wind turbine load estimation is sought after. The main conclusions of the study of these topics may be stated from the following four *aspects of statistical trends in high-rate wind measurements*:

1. Establishing the context of trends
2. Assessing the measurements before detrending
3. Modelling and removing trends
4. Assessing the measurements after detrending

Before establishing the context of trends, it is necessary to first define the notion of it. The wind turbine load estimation setup, being the context, must then be represented in a way that allows for an analysis of the influence of trends. Thus, the load estimation setup has been discussed in Section 3.1 on page 21 whereas classes of trends have been discussed in Section 4.1 on page 48. The main conclusions are:

- The setup used in load estimation on wind turbines has been identified and discussed in detail including a presentation of the state-of-the-art model of wind velocities.
- The assumptions behind the state-of-the-art wind velocity model used in the load estimation setup have been outlined and their connection to the mean wind speed, V, and turbulence standard deviation, σ_{v}, has been established.
- Existing definitions of a trend have been reviewed. Based on a frequency threshold definition of a trend, three classes of trends have been introduced: insignificant trends, linear trends, and periodic trends.

The load estimation setup is outlined based on a discussion with specialists from Vestas. It provides the context for the determination of a clear connection between the assumptions behind the state-of-the-art wind velocity model, the mean wind speed V, the turbulence standard deviation σ_{v}, and the classes of trends introduced such that the influence of the trend may be evaluated.

An assessment of the conformity of the available wind measurements to the load estimation setup is done prior to any detrending, to form a baseline for a later comparison with results after detrending. The assessment is based on 10-minute statistics described in the relative, rectangular coordinate system, (v_ζ, v_ξ, v_ψ), to provide a clear link to the current state-of-the-art wind velocity model. Thus, the distributions of V_ζ and σ_ζ have been assessed in

Section 3.2 on page 27. An assessment of the impact of the choice of aggregation period has been given in Section 3.3 on page 39 whereas the classes of trends contained in the measurements have been assessed in Section 4.3 on page 57. The main conclusions are:

- The log-normal distribution of σ_ζ conditioned on V_ζ has been found to be well reflected in the wind measurements whereas the Rayleigh distribution has been found to be a less convincing model for V_ζ for the available measurements. No other clear model of V_ζ has been established.
- The choice of aggregation period has been found to have no impact on the estimate of V_ζ. However, it has been established that the mean value of σ_ζ across intervals increases as the aggregation period increases due to low-frequent components in σ_ζ.
- Results from a classification of trends contained in the available measurements in 10-minute intervals have shown that around 10 % of the intervals contain a linear trend, another 10 % contain a periodic trend whereas the remaining 80 % of the 10-minute intervals contain no significant trend.

To ensure the integrity of the assessment, the measurements have been conditioned prior to the assessment. The Rayleigh distribution and log-normal distribution assessed in the first main conclusion are established and standardised models of V_ζ and σ_ζ conditioned on V_ζ, respectively.

The state-of-the-art wind velocity model used in the load estimation setup is based on the assumption that there is no trend contained in wind measurements. Since 20 % of the measurements contain a significant trend, a new model that distinguishes between mean wind speed, trend, and turbulence and describes the interconnection between these must be established. Next, a method for removing the trend must be established in order to assess the consequences of the trend. Thus, the modelling of a trend has been discussed in Section 5.1 on page 61 whereas the detrending, i.e. the removal of the trend, has been discussed in Section 5.3 on page 64. The main conclusions are:

- A new model of wind velocities has been proposed. It has been derived as an extension to the state-of-the-art wind velocity model by introducing a zero-mean trend component.
- The definition of the trend component of the proposed wind velocity model allows for the isolation of the trend by applying an ideal 3-dimensional low-pass filter to the wind velocity signal.
- It has been shown that a 1-dimensional Infinite Impulse Response (IIR) filter may be applied to each of the wind velocity components as a practical filter solution that successfully removes the trend.

With a practical way to remove the trend established, it is now possible to extract information about the difference in V_ζ and σ_ζ introduced by detrending. This information may be directly related to the assumption of no trends in the state-of-the-art wind velocity model due to the new model being an extension of the state-of-the-art model.

The assessment of the measurements after detrending is the final part of the study described in this report. An overview of the impact of a trend on V_ζ and σ_ζ by directly comparing these with the baseline must be given. Thus, the results of the detrending have been presented in Section 5.5 on page 77. The main conclusions are:

- Comparing the results before and after detrending, the distribution of V_ζ in 10-minute intervals has been shown to be unchanged, whereas the distribution of σ_ζ in 10-minute intervals has been shown to be moved towards zero. Specifically, the 90 % percentile of σ_ζ conditioned on V_ζ after detrending has been shown to correspond to the mean value of σ_ζ conditioned on V_ζ before detrending.
- The choice of aggregation period has been shown to have a substantially smaller impact on the estimate of σ_ζ after detrending compared to be before detrending if the aggregation period is longer than the period used as the threshold in the definition of a trend.
- For the available wind measurements, the use of detrending has been shown to entail a decrease in σ_ζ corresponding to a step down the IEC 61400 wind turbine subclass hierarchy, e.g. if the site required a subclass B wind turbine before detrending, only a subclass C turbine is required after detrending.

The above stated conclusions are related to *some aspects* of statistical trends in high-rate wind measurements. In no way do they cover everything there is to say about the extensive topic »trends«. Due to the focus on 10-minute intervals they do, however, shed light on the relation between trends and current 10-minute statistics used in wind turbine load estimation. Thus, it is now up to the wind turbine designers to decide for a way to incorporate the findings presented about this relation in the load estimation setup.

References

[1] E. Hau, *Wind Turbines: Fundamentals, Technologies, Application, Economics.* Springer Berlin Heidelberg, 2013.

[2] J. F. Manwell, J. G. McGowan, and A. L. Rogers, *Wind Energy Explained*, 2nd ed. John Wiley & Sons ltd., 2009.

[3] T. Burton, D. Sharpe, N. Jenkins, and E. Bossanyi, *Wind Energy Handbook.* John Wiley & Sons, 2001.

[4] *Wind turbines Part 1: Design requirements*, IEC Std. IEC 61 400-1, Rev. 3, 2005.

[5] A. V. Oppenheim and R. W. Shafer, *Discrete-Time Signal Processing*, 3rd ed. Pearson, 2010.

[6] T. Laubrich and H. Kantz, "Statistical analysis and stochastic modelling of boundary layer wind speed," *The European Physical Journal Special Topics*, vol. 174, pp. 197–206, 2009.

[7] A. Morales, M. Wächter, and J. Peinke, "Characterization of wind turbulence by higher-order statistics," *Wind Energy*, vol. 15, pp. 391–406, April 2012.

[8] M. Nielsen, K. S. Hansen, and B. J. Pedersen, "Gyldighed af Antagelsen om Gaussisk Turbulens," Risø National Laboratory, Roskilde, Denmark, Tech. Rep. Risø-R-1195(DA), 2000, in Danish. [Online]. Available: http://www.risoe.dk/rispubl/vea/veapdf/ris-r-1195.pdf

[9] W. McKinney, *Python for Data Analysis*, J. Steele and M. Blanchette, Eds. O'Reilly, 2013.

[10] F. Alted, "Why Modern CPUs Are Starving and What Can Be Done about It," *Computing in Science & Engineering*, vol. 12, pp. 68–71, 2010.

[11] A. Pfeiffer, I. Bausch-Gall, and M. Otter, "Proposal for a Standard Time Series File Format in HDF5," in *9th International MODELICA Conference*, Munich, Germany, September 3-5 2012.

[12] E. M. Sibarani, P. Wagenaars, and G. Bakema, "A TYPICAL CASE OF RECOMMENDING THE USE OF A HIERARCHICAL DATA FORMAT," *International Journal of Database Management Systems*, vol. 4, no. 6, pp. 27–48, December 2012.

[13] H. A. Madsen and K. Thomsen, "Analysis of Wind Turbine Loads," in *Advances in Wind Energy Conversion Technology*, M. Sathyajith and G. S. Philip, Eds. Springer Berlin Heidelberg, 2011, pp. 133–157.

[14] M. Nielsen, G. C. Larsen, and K. S. Hansen, "Simulation of inhomogeneous, non-stationary and non-Gaussian turbulent winds," *Journal of Physics: Conference Series*, vol. 75, pp. 1–9, 2007.

[15] M. Shinozuka and C.-M. Jan, "Digital Simulation of Random Processes and Its Applications," *Journal of Sound and Vibration*, vol. 25 issue 1, pp. 111–128, 1972.

[16] G. E. Johnson, "Constructions of Particular Random Processes," *Proceedings of the IEEE*, vol. 82, no. 2, pp. 270–285, February 1994.

[17] J. Mann, "The spatial structure of neutral atmospheric surface-layer turbulence," *Journal of Fluid Mechanics*, vol. 273, pp. 141–168, 1994.

[18] ——, "Wind field simulation," *Probabilistic Engineering Mechanics*, vol. 13, no. 4, pp. 269–282, 1998.

[19] G. I. Taylor, "The Spectrum of Turbulence," *Proc. R. Soc. Lond. A*, vol. 164, no. 919, pp. 476–490, 1938.

[20] J. C. Kaimal, J. C. Wyngaard, Y. Izumi, and O. R. Coté, "Spectral characteristics of surface-layer turbulence," *Quart. J. R. Met. Soc.*, vol. 98, pp. 563–589, 1972.

[21] S. M. Kay, *Intuitive Probability and Random Processes using MATLAB*. Springer, 2006.

[22] K. S. Shanmugan and A. M. Breipohl, *Random Signals: Detection, Estimation and Data Analysis*. John Wiley & Sons, 1988.

[23] M. P. Wand, "Data-Based Choice of Histogram Bin Width," *The American Statistician*, vol. 51, no. 1, pp. 59–64, 1997.

[24] G. C. Larsen, "Offshore Fatigue Design Turbulence," *Wind Energy*, vol. 4, pp. 107–120, 2001.

[25] *Wind turbines - Part 1: Design requirements*, IEC Std. IEC61 400-1/A1, 2010.

[26] W. R. Wade, *An Introduction to Analysis*, 4th ed. Pearson, 2010.

[27] K. S. Hansen and G. C. Larsen, "Characterising Turbulence Intensity for Fatigue Load Analysis of Wind Turbines," *Wind Engineering*, vol. 29, no. 4, pp. 319–329, 2005.

[28] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, 2001.

[29] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[30] J. Moncrieff, R. Clement, J. Finnigan, and T. Meyers, "Averaging, Detrending, and Filtering of Eddy Covariance Time Series," in *Handbook of Micrometeorology: A Guide for Surface Flux Measurement and Analysis*, ser. Atmospheric and Oceanographic Sciences Library, X. Lee, W. Massman, and B. Law, Eds. Kluwer Academic Publishers, 2004, vol. 29, ch. 2, pp. 7–31.

[31] K. S. Hansen and G. C. Larsen, "De-trending of turbulence measurements," in *European Wind Energy Conference & Exhibition*. Milan, Italy: EWEA, 7-10 May 2007.

[32] *Wind turbine generator systems - Part 13: Measurement of mechanical loads*, IEC Std. IEC61 400-13, Rev. 1, 2002.

[33] K. Su, *Analog Filters*, 2nd ed. Kluwer Academic Publishers, 2002.

List of Symbols

$(v_\mathrm{x}, v_\mathrm{y}, v_\mathrm{z})$	Absolute rectangular wind velocity coordinate system.	25
$(v_\mathrm{h}, \theta, v_\mathrm{v})$	Horizontal plane polar wind velocity coordinate system. The vertical component is perpendicular to the horizontal plane.	10, 25
(v_ζ, v_ξ, v_ψ)	Relative rectangular wind velocity coordinate system.	24, 25, 46, 91
$(\lvert\mathbf{v}\rvert, \theta, \phi)$	Spherical wind velocity coordinate system.	14
α	Angle of rotation between absolute and relative coordinates [°]	25
ℓ	Scale parameter used in Mann's spectral tensor [m]	23
γ	Shear parameter used in Mann's spectral tensor [ ]	23
I	Turbulence intensity [ ]	3, 5
$\hat{I}$	Sample turbulence intensity of the wind speed component v_ζ [ ]	25, 34
I_ref	Expected turbulence intensity [ ]	2, 3, 27
ϕ	Total vertical wind direction [°]	14
$\mathbf{R}_{\mathbf{V}+\mathbf{T}}(t)$	Rotation matrix that transforms $\tilde{\mathbf{v}}(t)$ from a relative, rectangular coordinate system oriented in the direction of the horizontal part of $\mathbf{V}(t) + \mathbf{T}(t)$ to the absolute, rectangular coordinate system.	63
$\hat{\sigma}_a$	Sample standard deviation of the wind speed component (substitute a for any wind speed subscript) $[\frac{\mathrm{m}}{\mathrm{s}}]$	24
σ_ψ	Turbulence standard deviation perpendicular to the mean wind direction and vertical $[\frac{\mathrm{m}}{\mathrm{s}}]$	24, 26, 42, 46
σ_v	Turbulence standard deviation $[\frac{\mathrm{m}}{\mathrm{s}}]$	4, 5, 21, 24, 47, 91
σ_ξ	Turbulence standard deviation perpendicular to the mean wind direction and horizontal $[\frac{\mathrm{m}}{\mathrm{s}}]$	24, 26, 42, 46
σ_ζ	Turbulence standard deviation along the mean wind direction $[\frac{\mathrm{m}}{\mathrm{s}}]$	23, 26, 27, 32, 39, 42, 46, 58
t	Time [s]	10, 22, 23
t_agg	Aggregation period [s]	39, 40
T_ζ	Trend along the mean wind direction $[\frac{\mathrm{m}}{\mathrm{s}}]$	48, 54
θ	Total horizontal wind direction $[\frac{\mathrm{m}}{\mathrm{s}}]$	10, 14
$\mathbf{T}(t)$	Total trend in a point in space as a function of time $[\frac{\mathrm{m}}{\mathrm{s}}]$	62

$T_\zeta(t)$	Trend along the mean wind direction as a function of time $[\frac{\mathrm{m}}{\mathrm{s}}]$	50
V	Mean wind speed $[\frac{\mathrm{m}}{\mathrm{s}}]$	3, 5, 21, 24, 47, 91
$\lvert\mathbf{v}\rvert$	Total wind speed $[\frac{\mathrm{m}}{\mathrm{s}}]$	14
v_h	Total horizontal wind speed $[\frac{\mathrm{m}}{\mathrm{s}}]$	10, 13
$\hat{V}_a$	Sample mean of the wind speed component (substitute a for any wind speed subscript) $[\frac{\mathrm{m}}{\mathrm{s}}]$	24
$v_{a,\mathrm{max}}$	Sample maximum of the wind speed component (substitute a for any wind speed subscript) $[\frac{\mathrm{m}}{\mathrm{s}}]$	24
$v_{a,\mathrm{min}}$	Sample minimum of the wind speed component (substitute a for any wind speed subscript) $[\frac{\mathrm{m}}{\mathrm{s}}]$	24
V_ψ	Mean wind velocity perpendicular to the mean wind direction and vertical $[\frac{\mathrm{m}}{\mathrm{s}}]$	24, 31, 79
v_ψ	Total wind velocity perpendicular to the mean wind direction and vertical $[\frac{\mathrm{m}}{\mathrm{s}}]$	23, 45, 50
$\tilde{v}_\psi$	Turbulence wind velocity perpendicular to the mean wind direction and vertical $[\frac{\mathrm{m}}{\mathrm{s}}]$	46, 63
V_ref	Reference wind speed $[\frac{\mathrm{m}}{\mathrm{s}}]$	2, 3, 27
v_v	Total vertical wind velocity towards zenith $[\frac{\mathrm{m}}{\mathrm{s}}]$	10
V_x	Mean horizontal wind velocity towards North $[\frac{\mathrm{m}}{\mathrm{s}}]$	26
v_x	Total horizontal wind velocity towards North $[\frac{\mathrm{m}}{\mathrm{s}}]$	26, 61
$\tilde{v}_\mathrm{x}$	Horizontal turbulence wind velocity towards North $[\frac{\mathrm{m}}{\mathrm{s}}]$	63
V_ξ	Mean wind velocity perpendicular to the mean wind direction and horizontal $[\frac{\mathrm{m}}{\mathrm{s}}]$	24, 31
v_ξ	Total wind velocity perpendicular to the mean wind direction and horizontal $[\frac{\mathrm{m}}{\mathrm{s}}]$	23, 45, 50
$\tilde{v}_\xi$	Turbulence wind velocity perpendicular to the mean wind direction and horizontal $[\frac{\mathrm{m}}{\mathrm{s}}]$	46, 63
V_y	Mean horizontal wind velocity towards West$[\frac{\mathrm{m}}{\mathrm{s}}]$	26
v_y	Total horizontal wind velocity towards West $[\frac{\mathrm{m}}{\mathrm{s}}]$	26, 61
$\tilde{v}_\mathrm{y}$	Horizontal turbulence wind velocity towards West $[\frac{\mathrm{m}}{\mathrm{s}}]$	63
V_z	Mean vertical wind velocity towards zenith $[\frac{\mathrm{m}}{\mathrm{s}}]$	26
v_z	Total vertical wind velocity towards zenith $[\frac{\mathrm{m}}{\mathrm{s}}]$	26, 61
$\tilde{v}_\mathrm{z}$	Vertical turbulence wind velocity towards zenith $[\frac{\mathrm{m}}{\mathrm{s}}]$	63
V_ζ	Mean wind speed along the mean wind direction $[\frac{\mathrm{m}}{\mathrm{s}}]$	24, 27, 32, 39, 40, 46, 48, 58, 79

v_ζ	Total wind speed along the mean wind direction $[\frac{\mathrm{m}}{\mathrm{s}}]$	23, 45, 48, 50, 65
$\tilde{v}_\zeta$	Turbulence wind velocity along the mean wind direction $[\frac{\mathrm{m}}{\mathrm{s}}]$	46, 48, 54, 63
Δv_ζ	Linear increment or decrement in the mean wind speed along the mean wind direction over the course of 10-minutes $[\frac{\mathrm{m}}{\mathrm{s}}]$	50
$\mathbf{V}(t)$	Mean wind velocity component in a point in space as a function of time $[\frac{\mathrm{m}}{\mathrm{s}}]$	22, 46, 62
$\mathbf{v}(t)$	Total wind velocity in a point in space as a function of time $[\frac{\mathrm{m}}{\mathrm{s}}]$	10, 22, 23
$\tilde{\mathbf{v}}(t)$	Fluctuating wind velocity component in a point in space as a function of time $[\frac{\mathrm{m}}{\mathrm{s}}]$	22, 23, 46, 62

Lists of Figures and Tables

List of Figures

List of Tables