

UNIVERSITY OLDENBURG

WIND PHYSICS MEASUREMENT PROJECT

Exercise 1 - Handling and preprocessing of measurement data

Author:

Jan KÄMPER

Florian BÖRGEL

Supervisor:

Mathias WÄCHTER

April 29, 2016

Contents

1	Importing Data into Matlab	2
2	Marking invalid data	2
3	Generating continuous time axis	2
4	Computing 10min means and stddev	3
5	Extreme values	4
6	Increment PDF	5

Introduction

The goal of this exercise was to perform some basic processing steps of raw measurement data from a met mast. The data used here originates from the FINO 1 platform in the German Northsea which includes two wind vanes at heights of 33m and 90m as well as eight anemometers at heights 33m, 40m, 50m, 60m, 70m, 80m, 90m and 100m. The given time period of 1-Hertz data is of one month (January 2013). The six tasks described in the following reach from standard data treatment to a first simple analysis by looking at the increment probability density function in terms of wind speed fluctuations.

1 Importing Data into Matlab

The first step was loading the data given as ASCII file into Matlab-readable data structures. For the first task we used the function `readtable()` to import the data with the corresponding delimiter as parameter. The advantage of using a table structure instead of a matrix is that the column headers like 'u100' or 'Time' are stored within the data structure. In this way we can easily separate the actual measurement data from the time stamp by splitting the table up in two matrices.

```
time_stamp = raw_data{:, {'Time'}};  
raw_data = raw_data{:, {'d90', 'd33', 'u100', 'u90', 'u80', ...  
    'u70', 'u60', 'u50', 'u40', 'u33'}};
```

2 Marking invalid data

In order to mark invalid data which is provided with a value of -999 by the measurement system the next section of our Matlab script converts all values -999 to *NaN*. Matlab checks if there is any invalid Data and replaces it with *NaN*. This is necessary for some remaining tasks when means and standard deviations will be computed which must not consider invalid values.

```
raw_data(raw_data==-999) = NaN;
```

3 Generating continuous time axis

To avoid gaps in the time axis we first converted our time t with `datenum()` to a numeric value. The numeric values represent elapsed time in units of days. Hence, 1 second corresponds to the fraction of $\frac{1}{24 \cdot 60 \cdot 60}$ of these numeric values. So after multiplication with the inverse value of that

and rounding we obtain unique integer IDs for every occuring timestamp. Afterwards, we created the continuous time axis, by initializing a vector with length equaling exactly the number of seconds in January. Then, we filled the corrected data matrix *data_pp* with *NaN* values and overwrote the file with our existing data at all indices where values are given.

```

disp( 'Creating_continuous_time_axis' )
tnew=[t(1):1:t(end)]';
data_pp = NaN(length(tnew),10);
disp( 'Writing_preprocessed_Data...' )
for i = 1:length(raw_data(:,1))
    data_pp(t(i)-t(1)+1, :) = raw_data(i, :);
end
time = (1:length(data_pp))';
data_pp = [time, data_pp];
save( 'data_pp.mat', 'data_pp', 'raw_data' );
clear;

```

4 Computing 10min means and stddev

This task is a first step of statistical analysis of the given data. We split our time axis into intervals of 600 seconds and for each interval we compute the mean and standard deviation for all ten variables for the . Invalid data can be ignored by using the commands *nanmean()* and *nanstd()*. Considering the wind directions a special treatment of the angles is required in order to handle the circular data, e.g. the mean value of 350° and 10° is not 180° but 0° . Some trigonometric functions can be employed in order to cope with this.

We plotted the ten minute means of the *u100* anemometer for one specific day and added the standard deviation to it. The outcome is depicted in figure 1.

For this specific day we observe a very strong fluctuation of the wind over the day. During the early morning the wind is steadily strong with windspeeds greater then $20m/s$ then drops by almost 60% untill 8am, then picks up again to reach a maximum of around $24m/s$ in the evening. Indeed a very stormy day. The mean value and standard deviation are obviously not stationary because then mean is high in the morning and evening and low inbetween. The fluctuations and thus the standard deviation picks up in the evening and therefore is also not stationary.

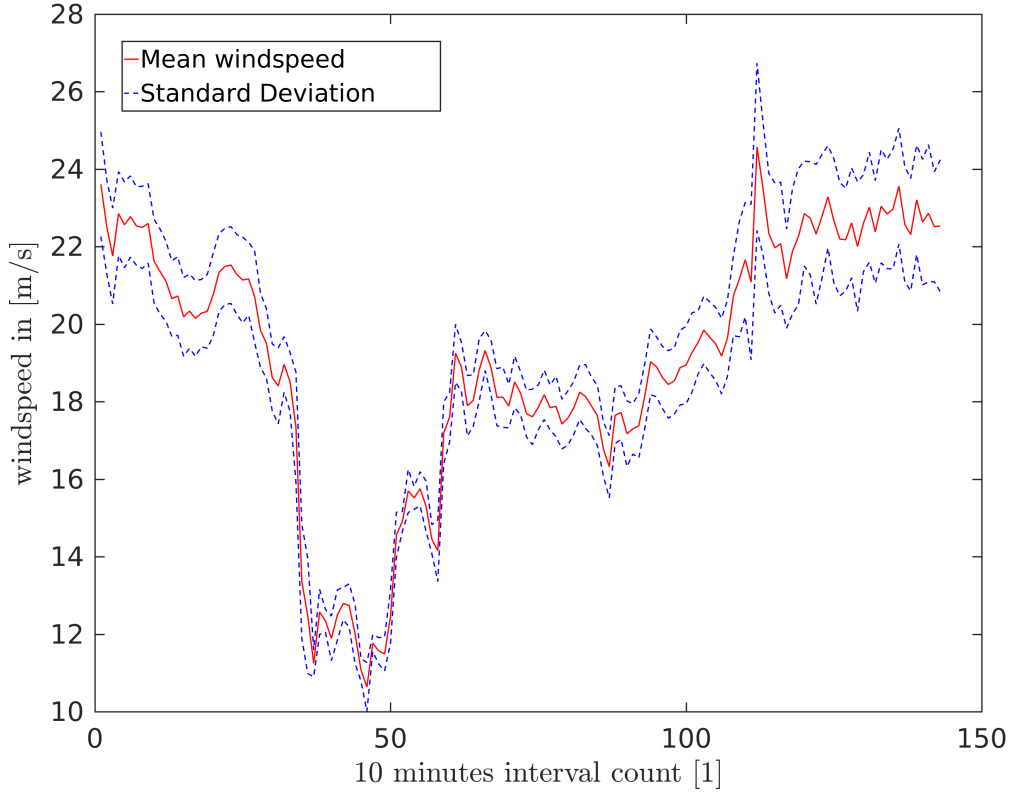


Figure 1: 10 minute intervals for Jan 30th 2013

5 Extreme values

In this task we looked for so called spikes. These are extremely high or low values within their immediate neighborhood, here 10 minutes intervals. A standard procedure to define spikes is to find values which exceed $[\mu - 5\sigma, \mu + 5\sigma]$ in these intervals where μ is the mean and σ is the standard deviation.

Two such intervals and contained spikes where found by our algorithm and are plotted in figures 2 and 3. The first shows a spike which appears to be of unphysical nature due to a relatively monotonous behaviour with only small fluctuations except one single value. This single value is below the $\mu - 5\sigma$ line. So here we could talk of a measurement error. The second plot however exhibits a more physical behaviour. Here we have stronger fluctuation overall and the spike which is below $\mu - 5\sigma$ matches the fluctuational pattern because the values before and after the spike also deviate a lot from the mean over a considerable period of time.

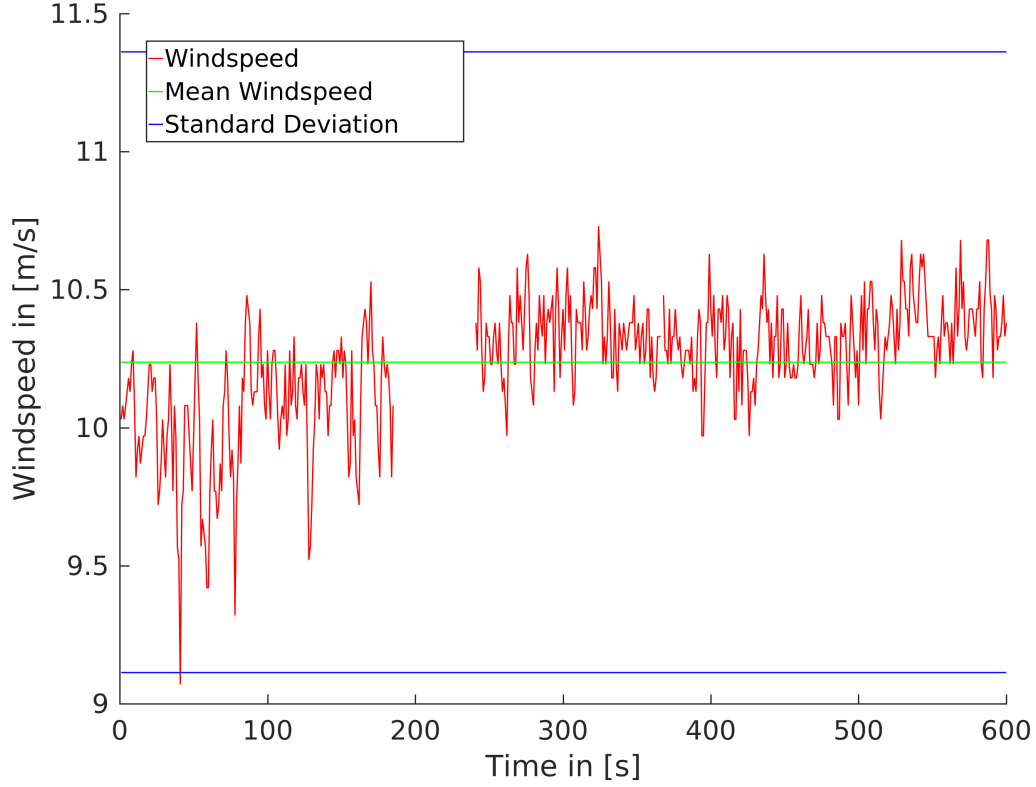


Figure 2: Physical cause

All in all we can say that the 5σ criterion is only an estimation of unusable values and relies on experience. By fixing the threshold to 5σ one states that all fluctuation within this interval are statistically significant. It implies reliability of the measurement system and is vulnerable enviromental influences, e.g. weather conditions.

6 Increment PDF

In this last task the increment PDF was to be computed. This probability density defines the distribution of short term changes on the time axis in the data. The formalism is $\delta u_\tau = u(t+\tau) - u(t)$ for a time lag of $\tau = 1s$. We collected the data in a histogram with 100 bins and normalized the values to unit standard deviation.

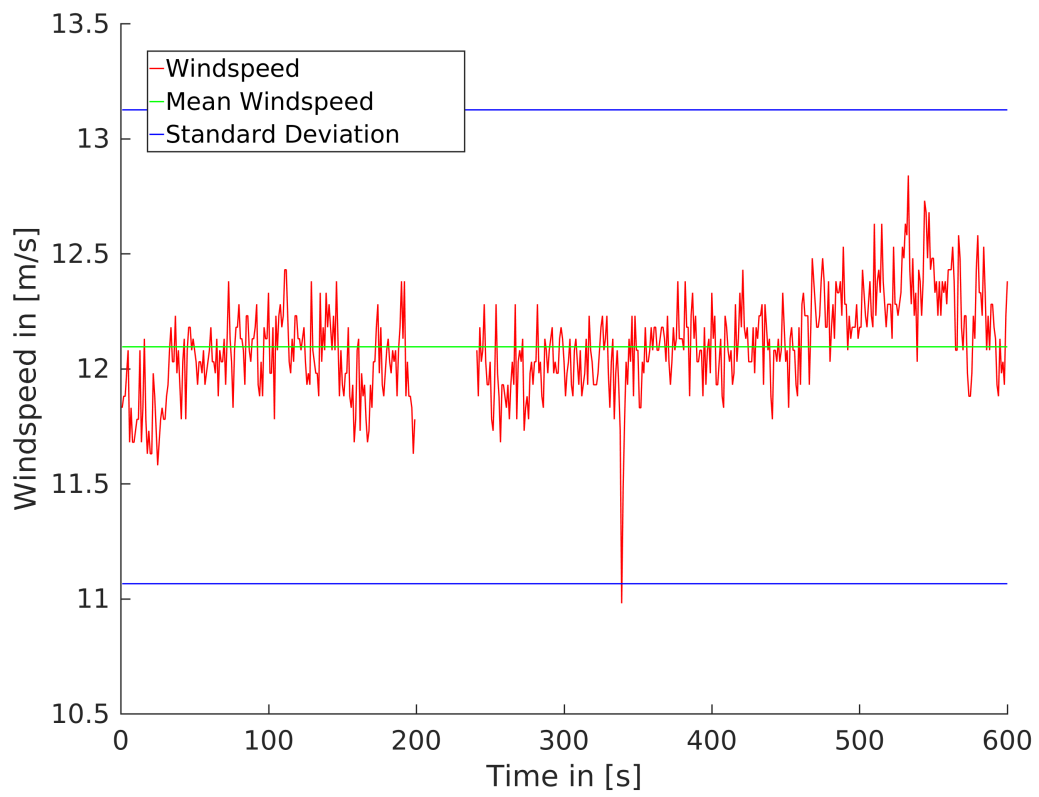


Figure 3: Unphysical cause

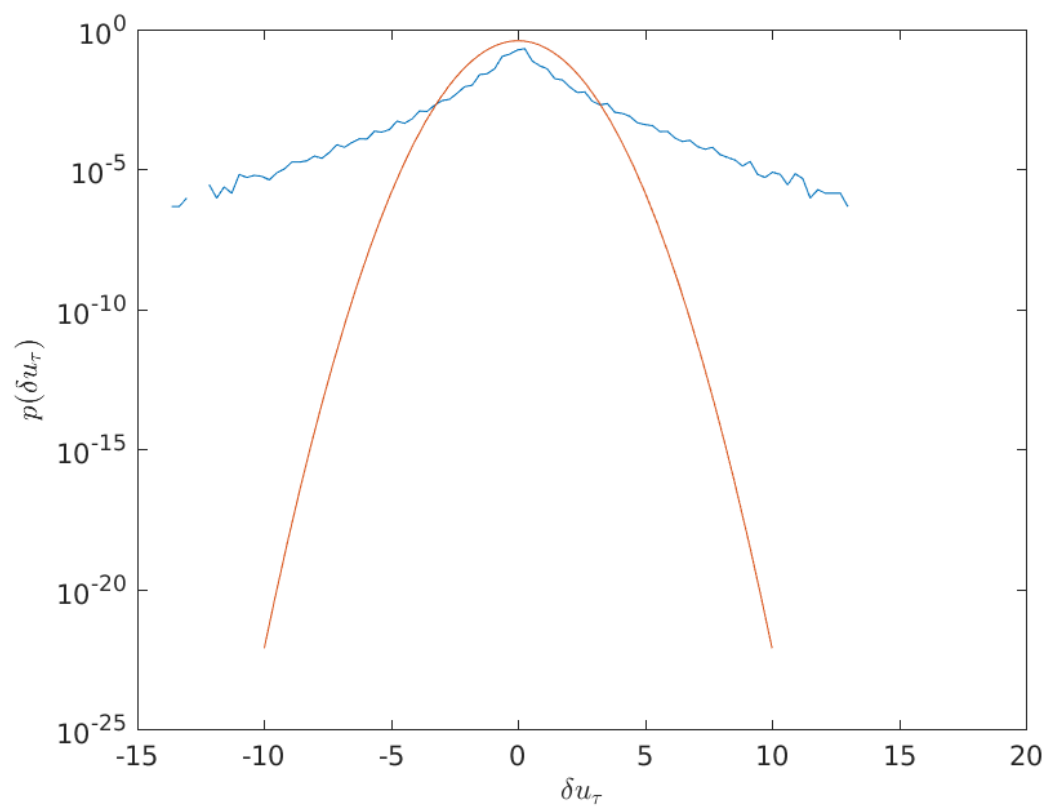


Figure 4: Unphysical cause