University of Stavanger

**FACULTY OF SCIENCE AND TECHNOLOGY**

# MASTER'S THESIS

| | |
|---|---|
| Study program/specialization:<br>Computer Science | Spring semester, 2018<br><br>Open / ~~Confidential~~ |
| Author: Tomasz Gliniecki | ..............................................<br>(signature of author) |
| Programme coordinator:<br><br>Supervisor(s): Tomasz Wiktorski, Rui Esteves (external Bouvet) | |
| Title of Master's Thesis: **Categorising offshore wind turbine installations, and investigating the weather impact during installations.** | |
| Credits: 30 ECTS | |
| big data, data mining, drilling, windcarrier, windmill, machine learning, offshore | Number of pages: 36<br>+ supplemental material/other:<br>- attachmentTomaszGliniecki.zip<br><br>Stavanger, June 15, 2018 |

# Categorising offshore wind turbine installations, and investigating the weather impact during installations.

Tomasz Gliniecki

Universitetet i Stavanger

**Abstract.** This paper describes the process of analysis of the dataset provided by Fred. Olsen Windcarrier. Analysis of the dataset is supposed to help in to capturing the installations of windmills in the north sea, from the diverse dataset of different sources, and find a correlation between weather conditions such as wind, with installation times of the windmills, where self-elevating, self-propelled jack-up vessels are used to perform installations. Capturing installations means, finding the data points that correspond to a given installation, where data points are coming from different sensors on the ship, combining them to form a description of each operation with as much detail as possible. This paper is a continuation of the semester project, in which a simple way to capture the installations were found. The results of the work went from finding the installations based on some thresholds like the actual load on the main hook on the crane that is lifting the load, to later in this thesis fully analysing the patterns of the installations to be able to make predictions using machine learning techniques. The dataset for this thesis is bigger and comes from more sources compared to earlier work in the semester project, it allowed for more ideas for analysing the dataset and to finally incorporate machine learning into the whole analysis.

**Keywords:** big data, data mining, drilling, windcarrier, windmill, machinelearning

**Acknowledgements** I would first like to thank my supervisor Tomasz Wiktorski who made the work with the thesis possible, through his help I was able to get a hold of the datasets, and his advice put me in the right path for finishing the thesis. My external supervisor from Bouvet, Rui Esteves who was helping me with in collaboration with the dataset providers Fred. Olsen Windcarrier/Dolphin drilling. I would also thank my family and friends for understanding and support me through this journey. It was a hectic time in my life, and I had constant supporting words from all around me which motivated me to do as much as I could during the last five and a half months.
Tomasz Gliniecki
University of Stavanger.

# Table of Contents



## 1 Introduction and motivation

In recent years Big Data began to come forth like a hot topic for people across all different fields. It is being used in the academic environments as well as in the industry and different sectors. We finally start to have technology advanced enough to utilise the data that has been gathered from across multiple fields. It is the developing technology that allowed people to gather the data, which then created opportunities for people to investigate deeper into what has been gathered. Using people's behaviour pattern on the Internet, like tracking clicks and analysing cookies, to sell more ads, as well as logging of numerous other events stretching from operations on oil rigs to how to optimise sales in a corner store. In the case of this thesis, data collected by Fred. Olsen Windcarrier were the basis for analysis. Fred. Olsen Windcarrier operates a fleet of two ships that perform offshore wind turbine installations. One of the important aspects while offshore performing installations is the weather conditions like wind, during the installation process. The assumption was made that there would be a positive correlation between the wind conditions and the amount of time used for the installation of the different windmill component, especially the rotor and turn blades. If the weather conditions are not good enough, the installations are not being started, and the fleet always waits for the best conditions it can perform in given the deadlines. There is plenty of things one can investigate having access to the data from the fleet. The description of some of the problem statements comes in the next paragraph, followed by the focus of the thesis.

First of all, there is an obvious question if the weather window can be expanded in any way, that means given current wind constraints in which the fleet operates, can we based on the data say something about working in the worse conditions. Given the data and knowing that the fleet follows the weather window when installing, there is probably no information about the offshore installations under harsh weather conditions, that are outside of the constraints, so there is probably just a few data points or no data from these kinds installations. The other problem statement could be described as to what degree the weather affects each of the installations the fleet performs. It is also important to notice that the fleet, consists of two ships but they are working independently, so there are potentially two different sources of similar installations. The latter problem statement could be addressed by finding out the weather conditions in the data and put them together with the captured installations. To investigate the correlation between the wind and the duration of installations, wind data and operation data (like duration of installation) need to be put together. The thesis went from simple clustering based on simple constraints, and calculating correlation and regression to fully scaled clustering based on Pythons Scikit-learn libraries and HDBSCAN [11], and predicting installation times with machine learning classifiers.

## 1.1 Windcarrier background research

The self-elevating, self-propelled jack-up vessels operated by Fred. Olsen, work with the installation of offshore windmills. Here are the steps performed under each operation with slight variations, simplified to a certain degree for the sake of analysis.

Discussion with the engineers that work on board and analysis of the data give these general functions performed with each operation.

1. Jacking up in Harbor (Jacking legs extension <35 m)
2. Lifting multiple wind turbines onboard
   Turbines have different configurations for different projects. Typical configurations

   - one tower, one nacelle and one rotor-star
   - one tower, one nacelle and three blades
3. Transit to off-shore field
4. Jacking up in the off-shore field (Jacking legs extension >35m)
5. Installing one wind turbine by lifting tower, nacelle and rotor-star or three blades
   The process of installation is performed as the following:

   - starting with activation of the crane to lift, the tower is being first supported by the crane, and loosen from the ship. Here we see a small spike in the data, crane activated, and low weight is being recorded.
   - next the main weight is lifted, and the tower hangs on the crane, here we see a huge spike in the data, and continuous high readouts until next step.
   - the tower is being lifted to place, the main weight is now resting on the base, and crane acts as a support, here we can see a big dip on data when the main weight is released.
   - in general all of the three items (tower, nacelle, and rotor) have a similar pattern in the dataset, possible to distinguish by the size of spikes and readouts.
6. Jack down and continue to next foundation ...

Typical weights and durations of lifts:

- Tower: Weight 340-370 ton, including tool
- Nacelle: 250-270 ton

– Rotor: 150 ton

Installation time is anything from 30-120 minutes in normal conditions.

## 1.2 Captured installations and related work

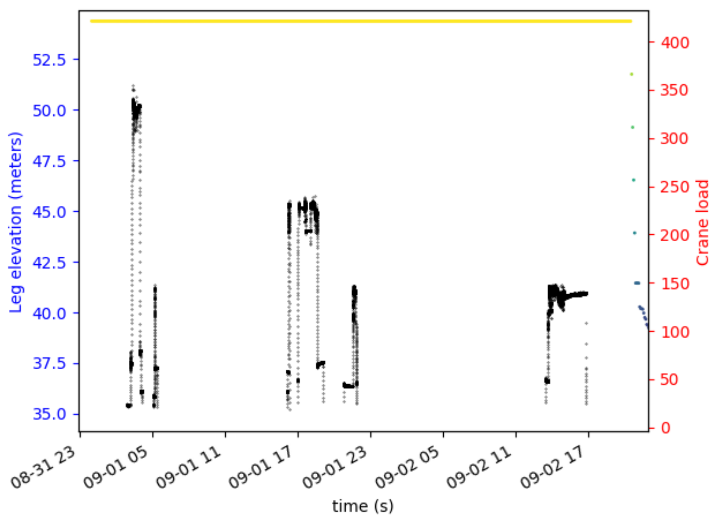
The offshore wind installations were to some degree captured and classified prior to this master thesis as mentioned earlier. Software was developed to take the available datasets and create an output that would consist of just the relevant data points that cover the installations, and categorise the different installation procedures during one operation based on mainly the weight on the main hoist. Throughout the dataset that is covering about three years of operations, very few operations were found and categorised, the main reason being holes in data, and not enough data to programmatically or in any other way find the correct timestamps of the operations. The categorisation of the installations is mainly finding beginning and end dates of the different components that the vessel installed, during one jack-up session. Jacking up is when the vessel is elevating itself above the sea level to stabilise itself. The beginning and end dates were based on the jack-up times Three main categories were settled upon first

1. installing the tower
2. installing the nacelle
3. installing the rotor (or individual turn blades)

This sums up the work before the thesis: Finding out when a vessel was jacked up, and based on the weight on the main hoist trying to pinpoint when and what kind of component was being installed.

The first thing on the list for this thesis was to try to get a hold of more data from the company Fred. Olsen Windcarrier as well as fine tune the categorisation. This is covered in the rest of the thesis. Fine tuning was time-consuming, with more data new ideas came and more installations have been captured, clustered, to the precise beginning and end dates of each component. This did add more complexity to how the installations are described, and helped to find more installations throughout the dataset. and made each installation more detailed. As more data was provided which included another vessel, it gave more events of installations. The clustering also made the events unique from each other, this led to difficulties considering regression and correlation analysis, as the events are less similar, for this case only the most similar events had to be compared to each other. The installations based on elevation times where now more dissimilar, this is where fine tuning of installations came useful. Now we have more unique events as a whole, but if we split them into the categories, and look at the installing of the towers by itself. For example, we can separate just the tower installations across the vessels, find the most similar ones based on the weights and hoist loads.

**Fig. 1.** Example 1 of whole operation based on elevation depth:




On the figure 1 the yellow line represents the elevation, with the scale to the right. And the crane load with *red* in tonnes. We can see that the first group of data points (black) along the x-axis x=09-01 05, we read the highest point to be around 350 tonnes, this means its a tower installation event. The next group of points around x=09-01 23, crane load at the top point reads between 200 and 250 tonnes, based on the criteria we can conclude that this is a nacelle installation event. The last group of points is most likely rotor installation event. The order of these graphs going from high to medium to low represents a typical installation of the whole tower.

**Fig. 2.** Example 2 of whole operation based on elevation depth:

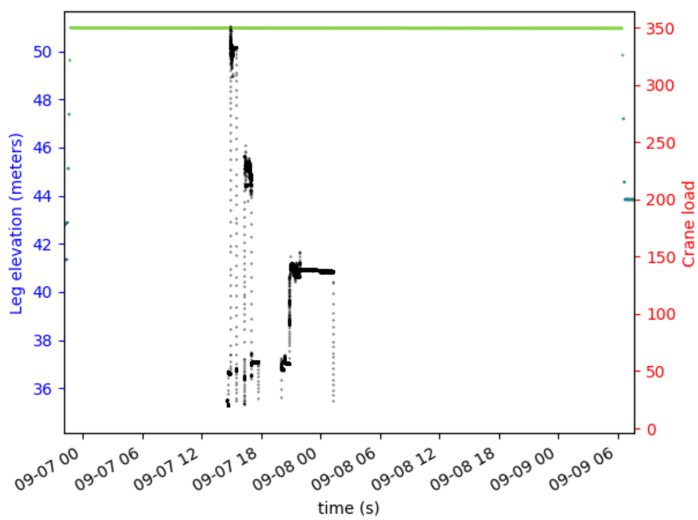
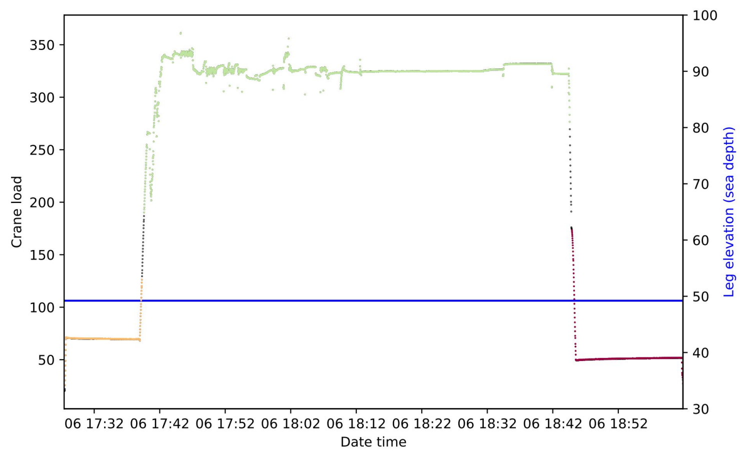



Figure 1 and 2 represents the whole installation process, the capturing of this is based on the leg elevation, sea depth, and crane load.

We can compare these two operations to some degree as they are describing similar events "windmill installation" but if we split it further, we can get for example:

**Fig. 3.** Example 1 of categorized tower installation:




In figure 3 we can see the separated tower installation and categorised events with the help of clustering. The green cluster represents the time where the tower was hanging with its whole weight on the crane, the orange to the left and red to the right are the points where the tower is partially supported by the crane, representing the events described in Section 1.1. The blue horizontal line represents the elevation at the time of this event.

**Fig. 4.** Example 2 of categorized tower installation:

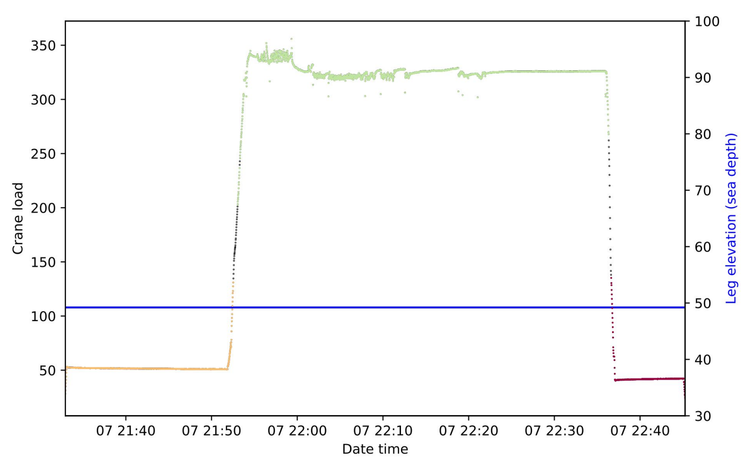


Figure 3 and 4 are just two examples of the same category of installation. The events are from two different locations, with different conditions.

**Fig. 5.** Example 1 of nacelle installation:

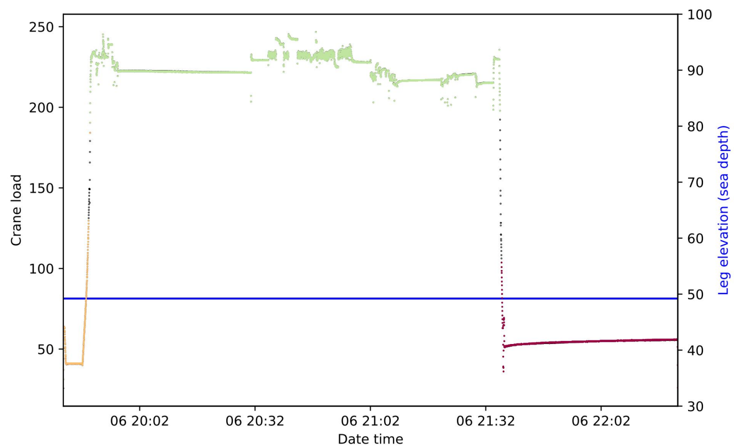



**Fig. 6.** Example 2 of nacelle installation:

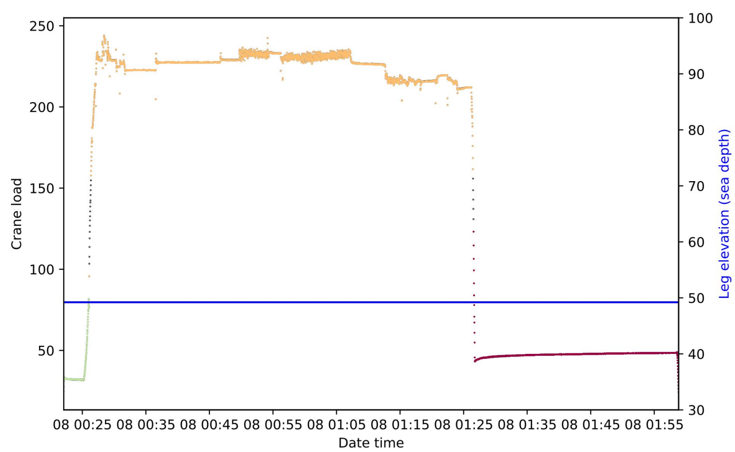


**Fig. 7.** Example 1 of rotor installation:

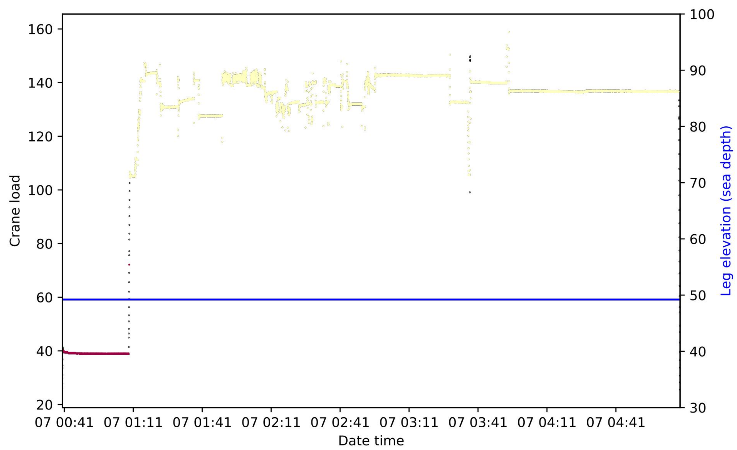



**Fig. 8.** Example 2 of rotor installation:

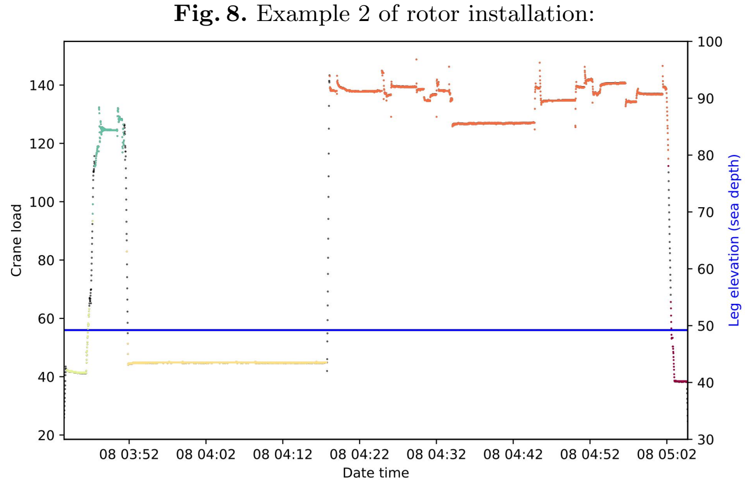



Figure 1 and 2 represented both rotor installation events and compared to the two towers and two nacelle installations, the graphs look more dissimilar then in the other examples. The categorisation of these data points was correct, but it is challenging to determine the source and reason behind the spikes in the data. Fortunately, the script is focusing on the most distinct characteristics.

**Fig. 9.** Example 1 of individual blade installation:

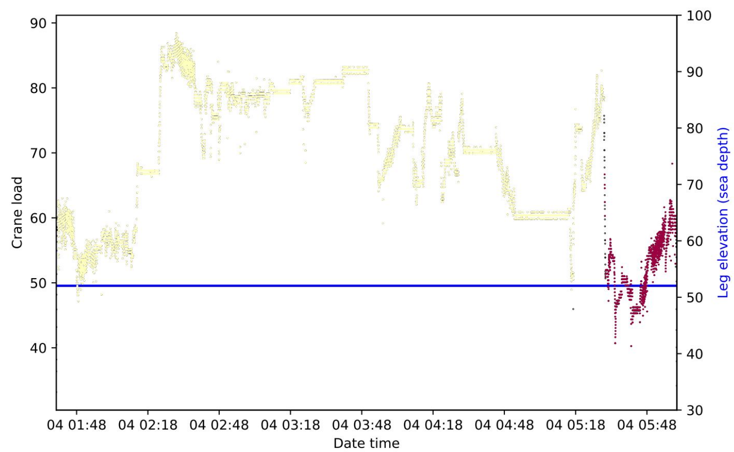
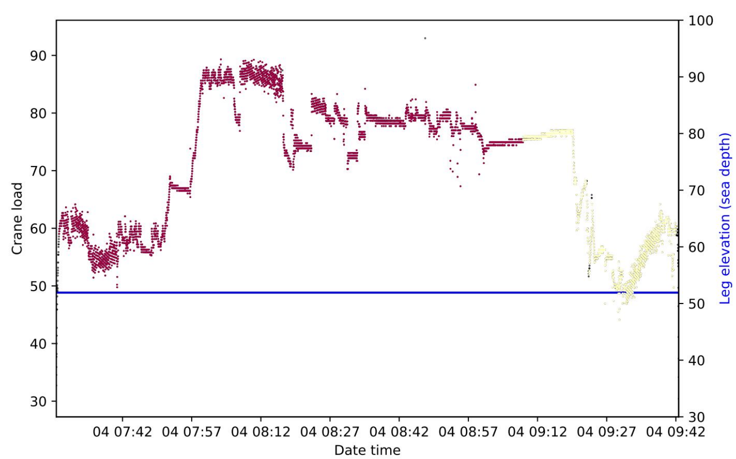



**Fig. 10.** Example 2 of individual blade installation:




By splitting the installation into three categories, we have made the basis for the analysis much richer, as now we can maybe differentiate which of the components is affected the most by the weather. Another differentiation is when the rotor is either installed as a whole, or as three separate turn blades. This comes from the fact that there are many different types of windmills. Summarizing the different components we have:

1. installing tower
2. installing nacelle
3. installing rotor
4. installing of 3 turn blades individually

### 1.3 Windcarrier

Currently, the decision whether or not to start installation is determined by the weather window, and this is based on the experience of the engineers. Knowing the weather window is irrelevant, as it was concluded that the installations are constrained within the weather window, so this can be determined by looking what were the worst weather conditions during any operation. The dataset provided is coming from two vessels, there has been no prior analysis done by the Fred. Olsen Windcarrier or it has not been a part of this thesis. Combining the different data sources has been time-consuming. What is expected from this thesis is a good visualisation of the dataset from different sources, and capturing the operations that are of interest in the dataset. Within the dataset some events do not correspond to the installation processes, there is loading the windmills on the ship from the harbour, and lifting other smaller things like equipment and personal items. This had to be filtered out, which was one of the first steps in beginning the analysis, except learning about where the data comes from, and how to understand the datasets in itself.

The data sources consist of, jackup elevation logs with different frequencies spanning from one second, to five-second intervals, the crane logs either with one second intervals or 5 minute intervals, and wind logs with an average of one second intervals, wind logs are coming from different sides of the vessels which are not time synchronised.

Preparation of the dataset includes combining it all to capture events step-by-step and adding more information with each step. This leads to having clearer and clearer picture of the history of the installations. As the investigation into the data progresses, it became possible to create the model that will try to predict the outcome of installation (mainly the duration), based on the current or forecasted wind, and the sea depth.

## 2 Datasets overview

This chapter gives an overview of the datasets used in this thesis, focusing on the most important parts of the datasets. It serves as a starting point for someone who is interesting in reproducing the results of this thesis.

### 2.1 Understanding the data

The dataset in this project consists of data from at least three different sources, for each of the self-elevating, self-propelled jack-up vessels. In each of the vessels, we have to deal with data from the jackup system, the crane lift, and of course, the weather conditions monitored on each ship individually. The files are organised with usually 1 hour of data recorded per file, and the files are separated into months and years, data spanning from 2012 from some sources, till the beginning of the year 2017. To be able to understand the log files, its necessary to learn about all the different attributes in each of the sources, and then figure out what is necessary and what attributes might be omitted. Having reduced the about of attributes, its possible to slowly figure out which of these describe the installation and what are the values that correspond to an event of interest. For example the elevation of the ship, can tell us whether the ship was in the harbor or jacked up in the sea, depending on the depth, with this constraint we can look closer and find what has been logged during this elevation, which in the end will lead us to figure out the correct places to look for.

### 2.2 File structure

The dataset comes from two vessels Brave Tern and Bold Tern, both have individual weather data associated with their operations, and sensor data from the ship equipment itself. Short description:

1. Crane load data
2. Jacking elevation data
3. Wind data

Each ship has three main sources of data, the jackup data, which is the length of the extended adjustable legs. The position of the extension of the legs indicates how deep water the ship is currently jacked-up. This information is used to determine whether the ship is in the harbour or in the deep sea where wind installations are performed. This is the starting point in categorising the operations, which is talked about later in more detail. The second source of data is the crane operation, it has much information about the positioning of the crane rotation angle, hoist torque, and hoist load, this information answers what was being lifted at any time during the wind installations. The third source of data is weather data which has the direction and speed of the wind, and sensors are mounted on three sides of the vessel. The wind dataset is used later in the project for correlation and regression analysis and classification. The first two are used for categorisation and clustering of the operations.

### 2.3 File structure

Files are located in multiple folders corresponding to the equipment that has created the files.

Parent directory of the files:

```
fowic-iot
 ├── bot_crane_load
 ├── bot_crane_on
 ├── bot_jack
 ├── bot_wind
 ├── brt_crane_load
 ├── brt_crane_on
 ├── brt_jack
 ├── brt_jack_on
 └── brt_wind
```

**Fig. 11.** Root of the dataset

The parent directory represents all the different data sources.

1. bot_crane_load
   - This source contains Crane load data from Bold Tern
2. bot_crane_on
   - This source contains the same type of data as above, except it, has shorter logging intervals. It is useful combined with the

other Crane data, to make the event description more detailed for machine learning algorithms.

3. bot_jack

   – This source contains the vessel elevation data that comes to form the leg elevation system, Load on jacking system for all four legs, whether Jacking system is switched on, vessel heel angle, vessel draft, water depth (not very reliable), Leg length below the hull.

4. bot_wind & brt_wind

   – This dataset source contains the wind data, which comes from 3 or 4 different sensors per the ship.

     - 1) BRT - Crane Boom
     - 2) BRT - Starboard bridge
     - 3) BRT - Port Bridge
     - 4) BOT - Crane Boom A
     - 5) BOT - Starboard bridge
     - 6) BOT - Port bridge
     - 7) BOT - Crane boom B (added later, only on BOT)

   The rest of the folders contains similar information, except form another ship "Brave Tern", the folders correspond to *bot* prefixed folders.

5. brt_crane_on
6. brt_jack
7. brt_jack_on
8. brt_wind

The folders are split into years and with subfolders for each month. Each month had filed with on average one hour of data each, going from the first day of the month till last, in some cases skipping days when no logs were saved.

```
fowic-iot
├── bot_crane_load
├── bot_crane_on
├── bot_jack
├── bot_wind
│   ├── 2015
│   ├── 2016
│   └── 2017
│       ├── 01
│       └── 02
│           ├── bot_wind_20160201_000000_000.csv
│           ├── bot_wind_20160201_001000_000.csv
│           ├── bot_wind_20160201_002000_000.csv
│           ├── bot_wind_20160201_003000_000.csv
│           └── [...]
├── brt_crane_load
│   ├── 2016
│   └── 2017
│       ├── 04
│       ├── 05
│       │   ├── [brt_crane_load_20170505_220000_000]
│       │   ├── [brt_crane_load_20170505_230000_000]
│       │   ├── [brt_crane_load_20170506_000000_000]
│       │   └── [...]
│       └── 06
├── brt_crane_on
├── brt_jack
├── brt_jack_on
└── brt_wind
```

**Fig. 12.** The structure of where data is stored and located

When looking at the structure, it possible to combine all the datasets based on the dates. Unfortunately, this would create a huge file, with many empty columns and half filled rows. Such colossal file would be complicated to load into memory and work with. The splitting of the logs, on the other hand, makes it difficult to have an overview of what is where and what is relevant.

### 2.4 Combining datasets

From all the attributes present in the Jacking data, it is concluded to stick with the following attributes, and keep then flowing throughout the other output files:

1. Jacking data

- Time
- S7 Leg SBFore Data Data Leg Load
- S7 Leg Sb Aft Data Data Leg Load
- S7 Leg PSAft Data Data Leg Load
- S7 Leg PSFore Data Data Leg Load
- Mem Main Cycle Bridge Control On
- ICSSRecieve Waterdepth
- S7 Leg SBFore Data Data Legtipbelowship
- S7 Leg Sb Aft Data Data Legtipbelowship
- S7 Leg PSAft Data Data Legtipbelowship
- S7 Leg PSFore Data Data Legtipbelowship

Notice in Jacking data, there is no date just time in hours, the full date has to be derived based on the placement of the file in the folder, and the filename.

This is the summary of attributes kept in the Crane load data:

1. Crane load data

- Date
- Time
- Actual_load_main_hoist
- Actual_load_aux_hoist
- Actual_load_whip_hoist
- Actual_load_radius_main_hoist

The Wind data has much fewer attributes, and all of them are kept and listed below:

1. Wind data

- refid:
- time (Unix epoch time (int))
- current:

- average2:
- average10:
- direction:

Whats interesting about these attributes especially *refid* and *time* is that *refid* which is an integer from 1 to 7 represents the unique id of the sensor, that means all the wind data is combined into one place and needs to be separated. The time is given in Unix epoch format, which is the number of seconds passed since Thursday 1. January 1970 00:00:00. Combining the dataset works as following, firstly a python script computes the elevation dates and duration of an elevation, that correspond to the correct depth which is over 35 meters. With this information that is based on the Jacking data, another script uses the found dates to find the information from Crane load and puts them together. When the script is done producing the data, it outputs it to a file for later use. This process is creating outputs for all the data available, or a user can add a constraint on the time he is interested in. The output that the script produces at this point looks like in the example below:

- Legtipbelowship.csv: Is a collection of data points derived from the Jacking data, consisting only of the points that cover the criteria of depth over 35 meters.

```
Time; Leg_load_1; Leg_load_2; Leg_load_3; Leg_load_4; Water depth;
    Leg_tip_below_ship; [...]
2017-01-08 16:35:05  -610   -672   -729   -660   3479
    35.4166717529297 [...]
2017-01-08 17:20:05  -462   -524   -385   -189   3520
    36.6154899597168 [...]
2017-01-08 17:25:05  654   1165   1062   840   3629   38.2485771179199
    [...]
2017-01-08 17:30:05  818   1208   1126   1160   3640   38.4234886169434
    [...]
2017-01-08 17:35:05  2016   2625   2304   2023   3729   39.34716796875
      [...]
.
.
.
```

- Elevation_dates.csv: This output has the start dates and end dates of the elevations with criteria over 35 meters below the sea level, and estimated duration of each elevation.

```
Starting_time; Finish_time;Duration;Elevation; [...]
2017-01-08 17:20:05  2017-01-18 15:10:04;9 days, 21:49:59; 53.62842[...]
2017-01-18 20:30:04  2017-01-21 06:15:05;2 days, 9:45:01; 52.45221[...]
2017-01-21 09:40:05  2017-01-22 23:45:05;1 day, 14:05:00; 51.81839[...]
2017-01-23 04:55:05  2017-01-24 13:35:05;1 day, 8:40:00; 51.37732[...]
2017-01-30 13:20:04  2017-01-31 10:20:04;21:00:00; 51.12780[...]
.
.
.
```

– Actual_load_main_hoist.csv: This output has only filtered data from *Crane load* where only data points representing over 30 tonnes match.

```
Time; Actual_load_main_hoist; [...]
2017-01-01 23:16:45  30.579 [...]
2017-01-01 23:16:46  31.046 [...]
2017-01-01 23:16:47  31.951 [...]
2017-01-01 23:16:48  32.397 [...]
2017-01-01 23:16:49  32.398 [...]
2017-01-01 23:16:50  33.303 [...]
2017-01-01 23:16:51  35.109 [...]
.
.
.
```

– Filtered_load_main_hoist.csv: This output combines the *Jacking data* with *Crane load* where only loads that are within the Elevation_dates.csv file.

```
Time; Actual Load; average depth; Duration; [...]
2016-05-02 10:47:03  30.269   47.93196928346312  5 days, 8:10:00 [...]
2016-05-02 10:47:04  30.743   47.93196928346312  5 days, 8:10:00 [...]
2016-05-02 10:52:16  31.648   47.93196928346312  5 days, 8:10:00 [...]
2016-05-02 10:52:17  35.297   47.93196928346312  5 days, 8:10:00 [...]
2016-05-02 10:52:18  37.575   47.93196928346312  5 days, 8:10:00 [...]
2016-05-02 10:52:19  39.852   47.93196928346312  5 days, 8:10:00 [...]
2016-05-02 10:52:20  43.897   47.93196928346312  5 days, 8:10:00 [...]
.
```

For these output files, there is a simple way to add more Attributes. Each row in this output file represents an event from the *Crane load* dataset that happened during actual elevation over 35 meters, in this case, 47.93 meters sea depth. *Crane load* and Jacking data attributes that the script put together, Time and Actual Load come from Crane load while average depth and Duration come from the computed duration of elevation and elevation depth. In addition to *Actual load*, other attributes are kept as described at the beginning of this Subsection 2.4, for checking the full extended outputs it is necessarily to look into attachments of this thesis, which has its own instructions.

## 3 Working environment

This chapter is about presenting the working environment for the project, what tools have been used and technical data to reproduce the results from this master thesis.

### 3.1 Python setup and libraries

The most simple and easy to use language for quick data processing is Python, in this thesis, there was almost no question about what kind of environment to use. Python provides built-in methods for both simple and complicated data processing. For this thesis, Python provided all the necessary tools for manipulating and processing the dataset. The dataset was local and small enough to be able to process it in a reasonable time. When starting to deal with scientific tasks in Python, one comes for help to Python's SciPy stack, which is a collection of libraries designed for scientific computing in Python.

The majority of work in this thesis is manipulating arrays and matrices of data, combining them, and preparing them for passing on to other libraries for further analysis. That is where NumPy (a library in the SciPy stack) is most useful as it provides features for operations on n-arrays and matrices in Python. This library also implements vectorisation of mathematical operations on the NumPy array type, which betters the performance and speeds up the execution.

The SciPy stack includes the following

- NumPy
- SciPy library
- Matplotlib
- IPython
- pandas
- Sympy

The reference to these libraries can be found in The SciPy Stack specification [10]

### 3.2 Machine learning libraries

For the machine learning part of the thesis, which was the last part that was developed, Scikit-learn provided libraries for predicting the time duration of operation based on the wind. The model for predicting the installation times is based on the final output from the python script, which combines the datasets and creates a clear description of all individual installations together with the wind data.

## 4 Windcarrier analysis results and approach

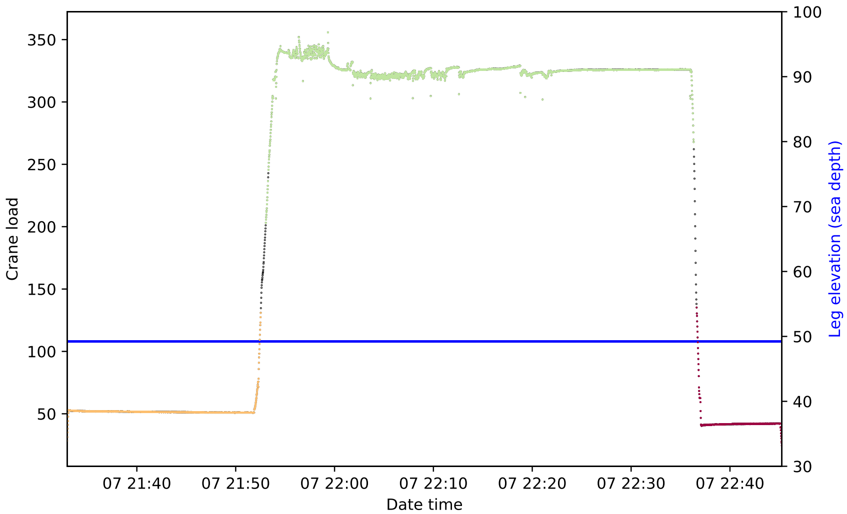
This chapter will cover the results of analysing the datasets, starting with visualisation of the datasets, and explaining different approaches that were used to understand the datasets better. After the visual analysis is done, it continues with programmatically combining and extracting useful information from the datasets and reviewing the quality of the datasets.

### 4.1 Visualizing data

The visualisation of the data properly is a significant step to the understanding of the whole dataset, and it also requires some prior knowledge about the source of the dataset to be able to interpret what we see on the graphs. After studying the attributes in the datasets, and picking up important ones many iterations of plots were made to make as much sense and start recognising the graphs, following the guidelines of how the operations look like described in Section 1.1. After getting to know the datasets better it was possible to create set of Python scripts that would find the operations automatically, and with time it was possible to fine-tune the capturing of the installations.

Here follows 4 pages of raw output from the script, that categorizes and clusters then plots the output file, giving a short summary of the event that has occured, and how it categorized it as:

**Fig. 13.** Direct output from the categorization python script: this shows the TOWER installation

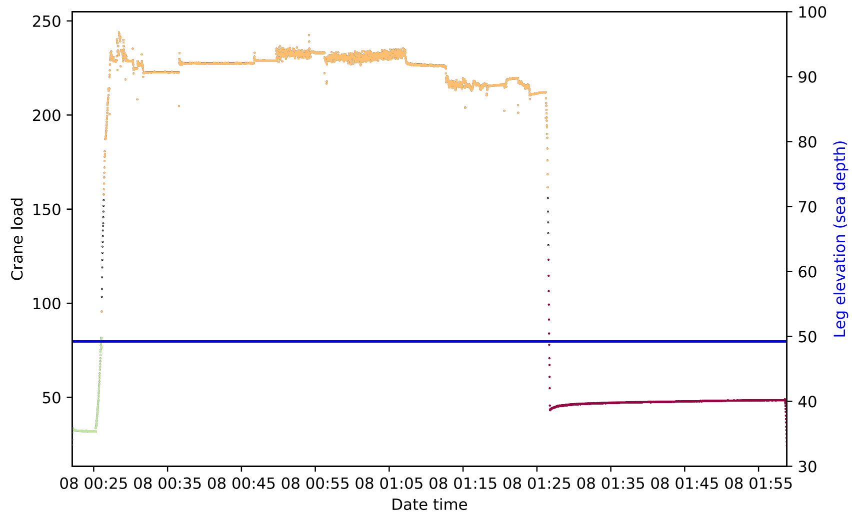



BLUE: Elevation depth
Colored dots: Different colors are separating clusters
Black dots are not clustered noise
Start time: 2017-05-07 21:32:56
End time: 2017-05-07 22:45:15
Duration: 1:12:19
Max weight: 355.78
Average weight: 215.94160682185068
Clusters: 3
Category:
Average wind: 4.268451419149688

cluster means:
43.832835907335905+
52.32442431972789+
325.02908320493066+
182.77959183673468+

max cluster : 325.02908320493066
TOWER

**Fig. 14.** Direct output from the categorization python script: this shows the NACELLE installation

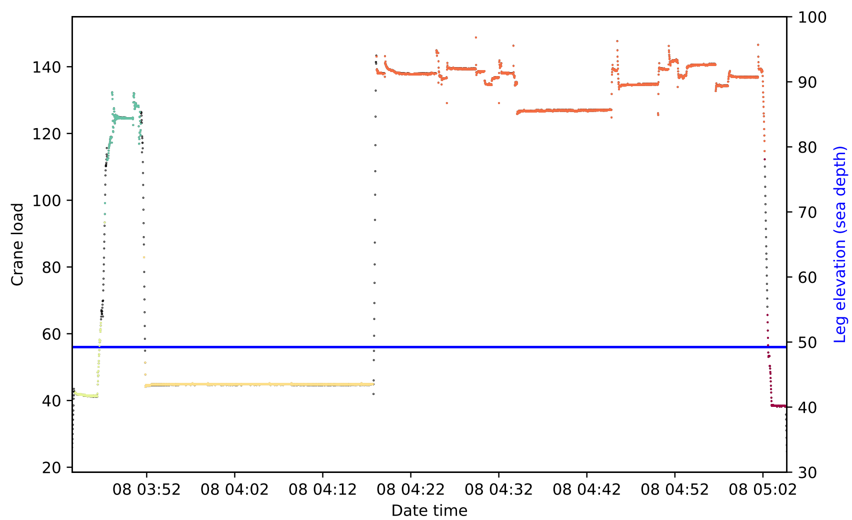



BLUE: Elevation depth
Colored dots: Different colors are separating clusters
Black dots are not clustered noise
Start time: 2017-05-08 00:22:04
End time: 2017-05-08 01:58:50
Duration: 1:36:46
Max weight: 243.92
Average weight: 157.8141744746786
Clusters: 3
Category:
Average wind: 2.5297074505649717

cluster means:
47.61655240061951+
225.1717987804878+
36.604479166666664+
134.80523809523808+

max cluster : 225.1717987804878
NACELLE

**Fig. 15.** Direct output from the categorization python script: this shows the ROTOR installation

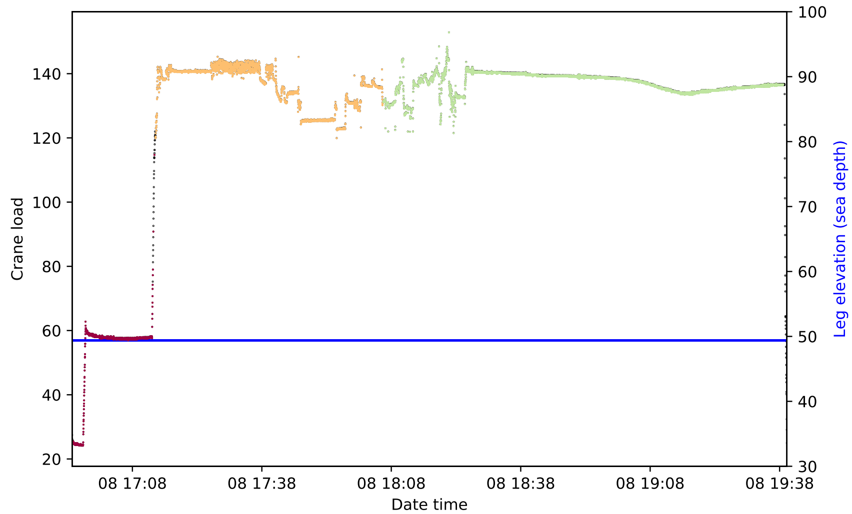



BLUE: Elevation depth
Colored dots: Different colors are separating clusters
Black dots are not clustered noise
Start time: 2017-05-08 03:43:31
End time: 2017-05-08 05:04:43
Duration: 1:21:12
Max weight: 148.77
Average weight: 98.45610324302353
Clusters: 5
Category:
Average wind: 4.092951368795905

cluster means:
41.92550806451613+
135.08063114134544+
44.86808810289389+
43.52392737430168+
123.68593991416309+
79.08546666666666+

max cluster : 135.08063114134544
ROTOR

**Fig. 16.** Direct output from the categorization python script: this shows ROTOR installation



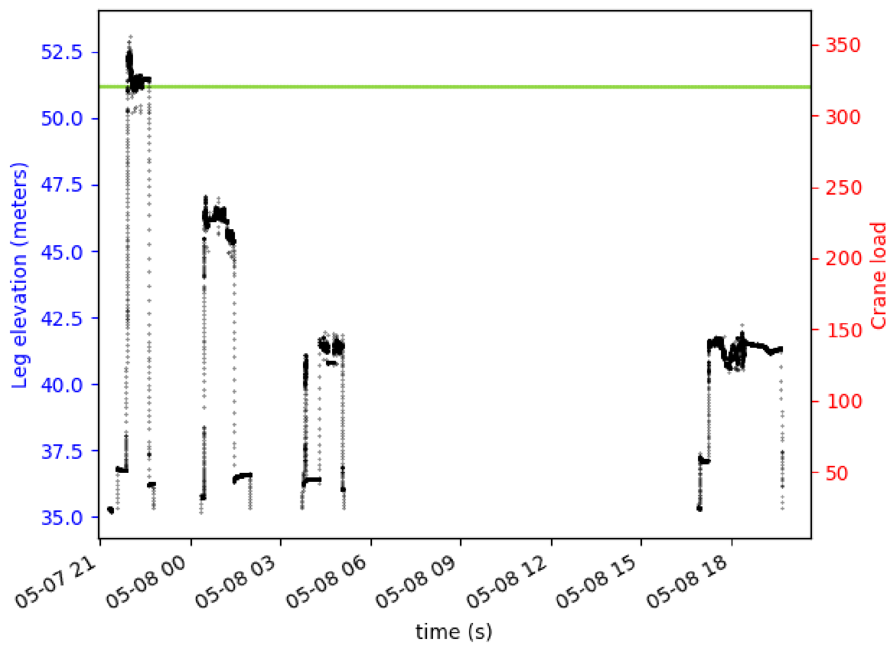

BLUE: Elevation depth
Colored dots: Different colors are separating clusters
Black dots are not clustered noise
Start time: 2017-05-08 16:54:01
End time: 2017-05-08 19:39:41
Duration: 2:45:40
Max weight: 152.86
Average weight: 126.81834872722534
Clusters: 3
Category:
Average wind: 3.473918864826029

cluster means:
53.51810298102981+
135.7729469361971+
137.05520488856936+
85.98743137254903+

max cluster : 137.05520488856936
ROTOR

All these four events come from under the same jacking moment, that is this are all the events that are of interest while the vessel was jacked, represented by the blue line. It is strongly recommended to look into the attached files and look through the output files from the scripts where the rest of the 210 operations are presented as well as the scripts themselves that create them. It is infeasible to attach the source dataset so the scripts can only be run locally with the dataset present locally.

**Fig. 17.** Direct output from the categorization python script:




In the Figure 17 we can see the output from the categorisation based on the elevation, it gives a zoomed out view of the four previous categorised events. These kinds of scripts can be produced for any range of dates, and as long as the scripts find actual installations in the date range in the data set, it will produce an output like the one shown above. The description under the graphs is also auto-generated.

In some cases when there is much noise in the data, and therefore it is still important to carefully study the output and see if the automatically generated description is describing the output graph as correctly in some cases.

In some cases, we can see that the categorisation script fails to find some parts of the tower installation. This is also one of the reasons that pushed towards developing scripts that would split the categorisation to distinguish between installing different parts. The outputs given are either lacking turn blades installations or other major parts of the whole windmill. This makes the installation time of the whole windmill unreliable, and therefore after splitting all events into four categories, tower, nacelle, rotor, and blades, another script can combine all the tower components that come from the same elevation operation and use this information for further analysis.

## 4.2 Correlations

When looking at the operations as a whole unit of from starting of installing the tower, to finishing the whole windmill, instead of the individual operations, we see no correlation with the weather data provided.

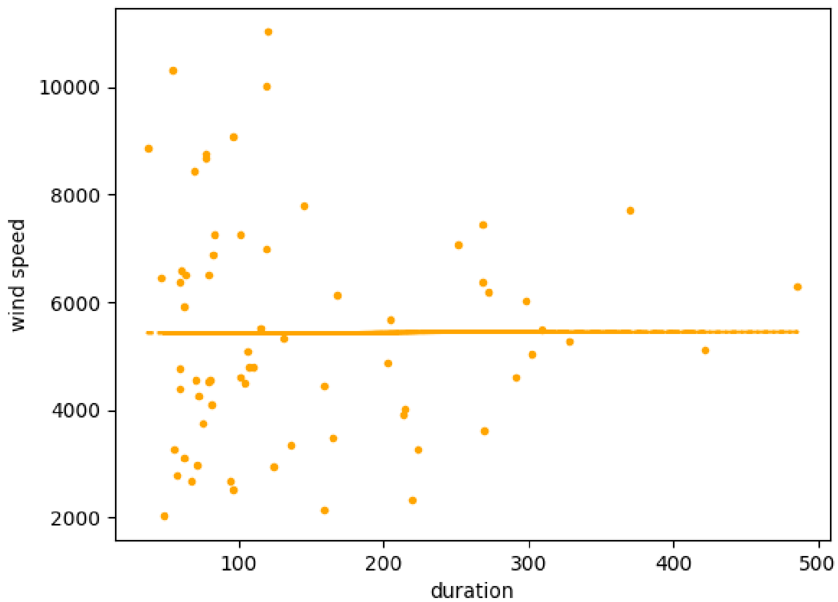



**Fig. 18.** Mounting the windmill (whole) 3 parts, the line in the middle is Least squares polynomial fit

```
slope of the regression line: 0.03790295917989172
intercept of the regression line: 5429.425917566802
correlation coefficient: 0.0018231321043794767
Standard error of the estimate : 2.578680753652249
```

The objective was to put the installations of the windmill and put them together with the wind data, to find whether the wind is affecting the installations and to what degree. We know that the weather has a major impact on the installations, but as mentioned previously, it is unsure if this is visible in the data, as the installations are not performed in the ”extreme” conditions. Se we are only bound to the weather that is good enough to work in, which is usually pretty calm when it comes to wind. With so much randomness that is visible on the graph, we can either conclude with that there is no correlation or that the installations differ to a greater degree than previously anticipated. This leads us to the next point, where we try to split the installations into parts, and look for correlations with collecting as much similar installations as possible.

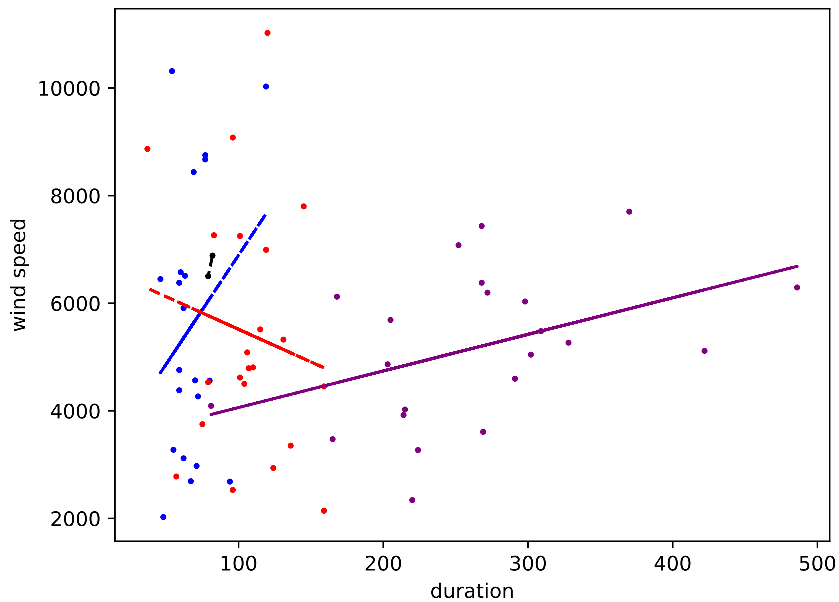



**Fig. 19.** Time vs wind correlation. blue: tower, red: nacelle, purple: rotor, black: blades

On this figure, we can see that there is some correlations between the length of any type of operation vs the wind. This graph combines and shows correlation of tower, nacelle, rotor and individual blades. This is all operations combined on one graph to wisually represent which part has higher correlation with wind. As it turns out the tower has the highest of correlation coefficient.

Here next, separate all tower installations across general installations of whole windmills, and do the same with all other parts.

- tower
- nacelle
- rotor (all rotors)

Working on the correlation of the individual parts times, vs the weather might have given us a little different picture, but unfortunately as shown on the graphs, no correlations are visible in the graph.

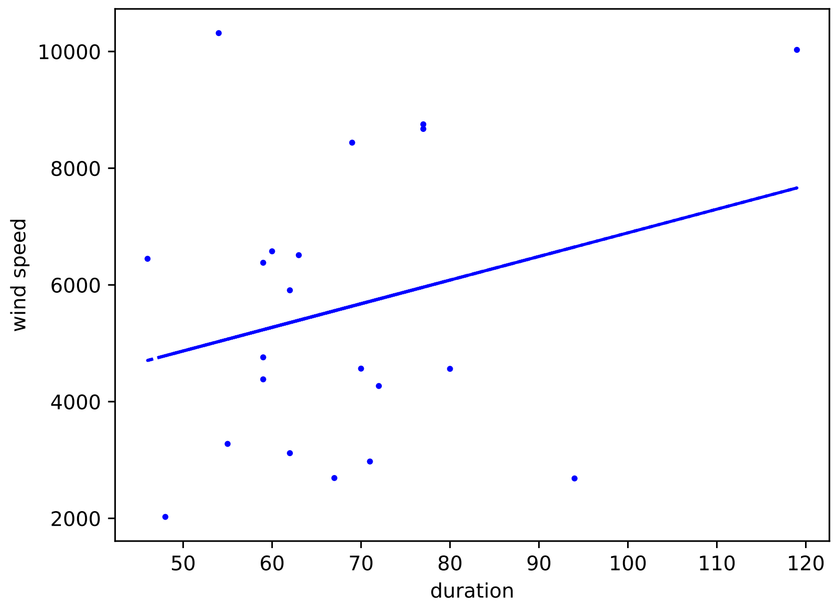



**Fig. 20.** tower time vs wind correlation, where the time needed for tower installation is along the X-axis

```
slope of the regression line: 40.49504440143965
intercept of the regression line: 2843.216753178637
correlation coefficient: 0.2609703660507878
Standard error of the estimate: 34.36507144288896
```

Tower time installation is usually very consistent, compared to for example rotor blades installation, as described in next Figure 4.2 4.2. Although the times are consistent, we cannot give any reasonable explanation to the different wind measures. We would not see much influence of the wind on the tower, as it is most resistant to the wind effects of all the parts, and is also the heaviest part, on top of it there are installations along the tower to minimise the wind effects. This makes the task of looking for wind effects even harder, and without enough examples and data points it impossible to give a reliable answer to that question.

To begin with, we are dealing here with a minimal amount of data. These data points come from the categorisation part, where installations of individual parts were categorised and times of each installation was estimated. Going from gigabytes of data, left us with only a few installations that were combined into just a few data points. As the correlation here is non-existent, we have to conclude that with that amount of data we cannot surely confirm weather correlation with the time installations.

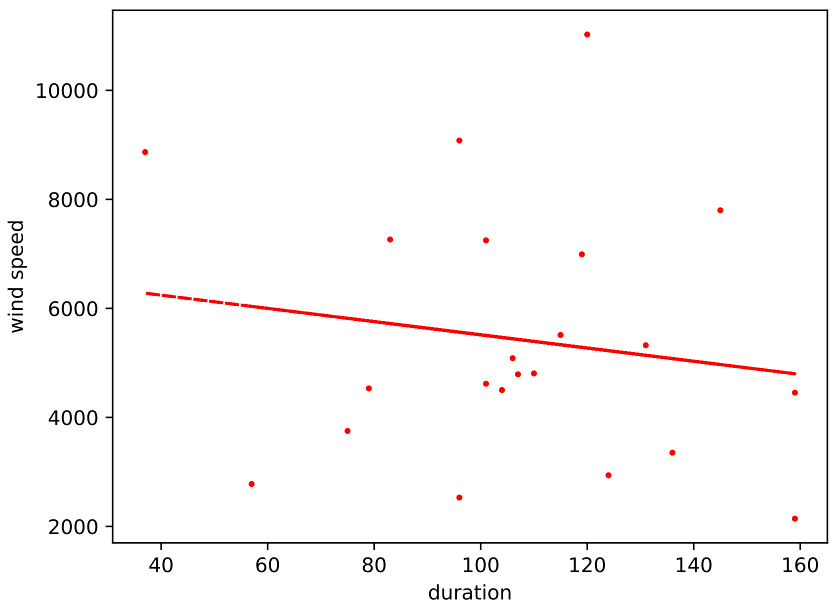



**Fig. 21.** nacelle time vs wind correlation, where the time needed for nacelle installation is along the X-axis

```
slope of the regression line: -12.121214579535582
intercept of the regression line: 6727.366654895635
-correlation coefficient: 0.15602367590558322
Standard error of the estimate: 17.158887712884233
```

Nacelle time Figure 21 correlates a little bit with the wind strength, but in a way that one would not expect as the slope is reversed indicating higher times with calmer wind. Nacelle installation should be one of the most similar right after tower installations. So we expect similar installation duration and wind effect on it. The data points show us otherwise, and we see no positive wind correlations, there have to be other things that influence the time of installation. This has to be investigated further with the help of more data.

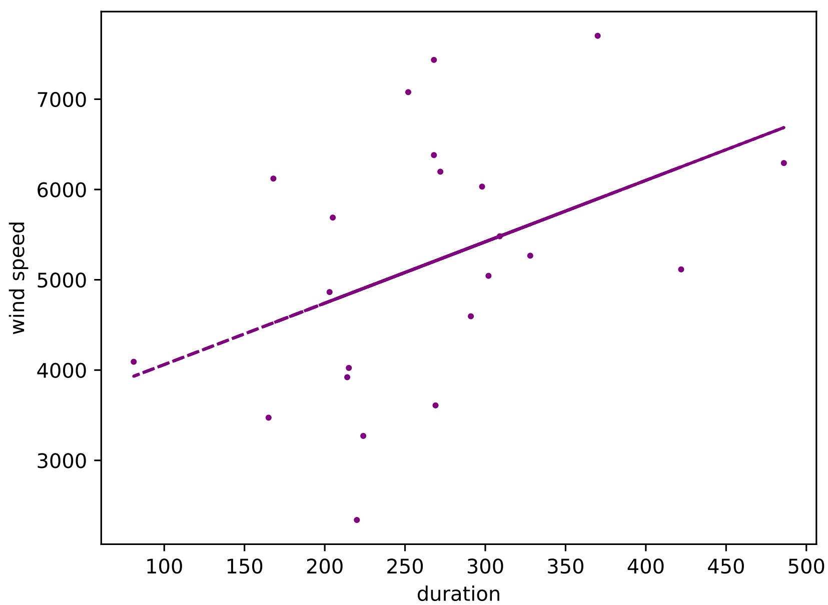



**Fig. 22.** rotor 1 vs wind correlation, where the time needed for rotor no.1 installation is along the X-axis

```
slope of the regression line: 6.79957997085512
intercept of the regression line: 3381.111307723393
correlation coefficient: 0.42204374476870415
Standard error of the estimate: 3.2659804644189023
```

Looking at the individual rotor installation we see that the correlation is positive, in the expected direction. In figure 4.2, we can also see correlation just by visual inspection. In all these examples we are dealing with a small amount of data, and what would be very interesting to see is more of this and to confirm the results with more data.

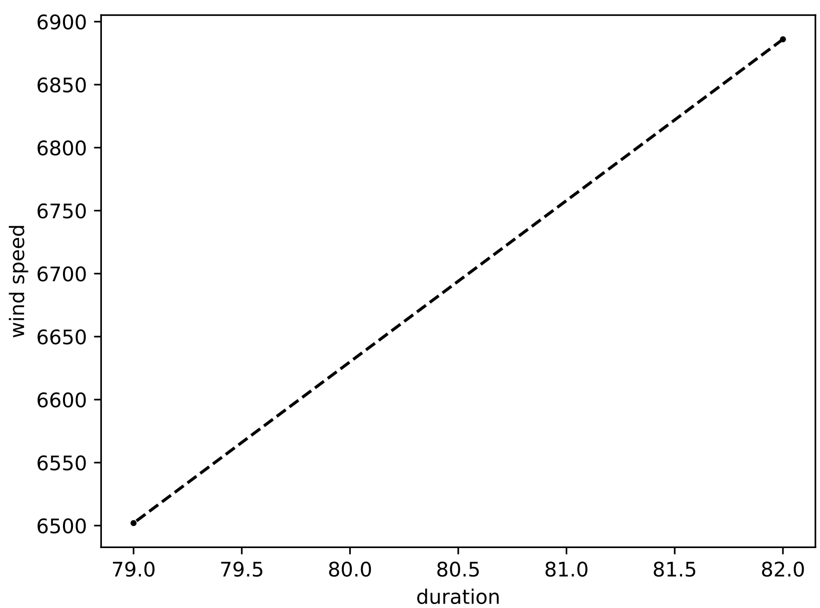



**Fig. 23.** Individual blade installations (only two that had wind data associated with them) vs wind correlation, where time needed for rotor no.2 installation is along the X-axis

```
slope of the regression line: 128.0
intercept of the regression line: -3610.0
correlation coefficient: 1.0
Standard error of the estimate : 0.0
```

We hypothesized that we would see a positive correlation with the duration of installations, and to some degree this has been confirmed. The data doesnt give us enough information to say anything more than confirming the correlation, and hoping for more data as the days progress and the program developed for this purpose can be utilized further.

## 5 Modeling and predicting

In order to run meaningful machine learning algorithms we have to be in possession of lots of data and descriptions of individual events. In this case what we are concerned with is the installations. We have four different categories of installations. Across the dataset, there has been categorize and captured 210 different installations. Each of those keeps track of all the events that happened from the start time to the end time of each operation. That means, the vectors that describe crane load, the pressure on the legs, the wind data, averaged weights and clustered from Section 1.2.

## 6 Conclusion

The main problem while working with a new dataset, is that much time is needed to learn about it, especially without any prior knowledge about the topic that the data concerns. Also trying to figure out what value the data set has, to begin with. Logging everything is relatively easy, and very valuable in theory, but the task of analysing it is not of the simple ones. The dataset in itself is of relatively good quality, as it is structured, on the other hand, it is not well organised. It is possible to develop further on the scripts and tools developed during the thesis.

**Fig. 24.** The basis for the machine learning algorithm, captured installations

| Start_time | End_time | Durat... | Leg_ele... | Max_... | Avg_... | Timestamps | Weight_array | weight_clust... | Average_wi... | Wind_data | Wind_time... | Category | ID_operati... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2017-05-05 ... | 2017-05-06 ... | 4:29:50 | [49.205... | 171.51 | 134.55... | ['2017-05-05... | [29.529, 35.0... | [[163.55, 126... | 3609.04121... | [6943, 40... | ['2017-05-... | ROTOR | 22:30:00 |
| 2017-05-06 ... | 2017-05-06 ... | 1:34:40 | [49.205... | 361.23 | 242.62... | ['2017-05-06... | [20.539, 20.5... | [[145.83, 144... | 2684.40909... | [2367, 23... | ['2017-05-... | TOWER | 22:30:00 |
| 2017-05-06 ... | 2017-05-06 ... | 2:39:48 | [49.205... | 246.71 | 168.25... | ['2017-05-06... | [25.65, 31.18... | [[25.65, 31.1... | 4454.50138... | [2638, 32... | ['2017-05-... | NACELLE | 22:30:00 |
| 2017-05-07 ... | 2017-05-07 ... | 4:28:47 | [49.225... | 158.91 | 125.96... | ['2017-05-07... | [26.094, 27.8... | [[87.561, 82.... | 6381.33931... | [6138, 35... | ['2017-05-... | ROTOR | 1 day, 12:3... |
| 2017-05-07 ... | 2017-05-07 ... | 1:12:19 | [49.225... | 355.78 | 215.94... | ['2017-05-07... | [28.287, 31.3... | [[60.95, 62.2... | 4268.45141... | [4772, 15... | ['2017-05-... | TOWER | 1 day, 12:3... |
| 2017-05-08 ... | 2017-05-08 ... | 1:36:46 | [49.225... | 243.92 | 157.81... | ['2017-05-08... | [24.703, 27.4... | [[24.703, 27.... | 2529.70745... | [4538, 11... | ['2017-05-... | NACELLE | 1 day, 12:3... |
| 2017-05-08 ... | 2017-05-08 ... | 1:21:12 | [49.225... | 148.77 | 98.456... | ['2017-05-08... | [25.934, 27.1... | [[25.934, 27.... | 4092.95136... | [4940, 49... | ['2017-05-... | ROTOR | 1 day, 12:3... |
| 2017-05-08 ... | 2017-05-08 ... | 2:45:40 | [49.395... | 152.86 | 126.81... | ['2017-05-08... | [27.466, 26.1... | [[92.496, 89.... | 3473.91886... | [6625, 65... | ['2017-05-... | ROTOR | 3 days, 8:4... |
| 2017-05-11 ... | 2017-05-11 ... | 1:10:23 | [49.395... | 353.58 | 269.03... | ['2017-05-11... | [27.734, 30.5... | [[146.62, 135... | 4566.41335... | [5400, 58... | ['2017-05-... | TOWER | 3 days, 8:4... |
| 2017-05-11 ... | 2017-05-11 ... | 1:47:58 | [49.395... | 237.03 | 146.61... | ['2017-05-11... | [28.776, 33.5... | [[31.292], [31... | 4791.05361... | [4140, 54... | ['2017-05-... | NACELLE | 3 days, 8:4... |
| 2017-05-11 ... | 2017-05-11 ... | 1:22:01 | [49.395... | 55.045 | 53.595... | ['2017-05-11... | [31.537, 36.9... | [[36.05, 30.0... | 6886.22211... | [5904, 59... | ['2017-05-... | INDIVIDU... | 3 days, 8:4... |
| 2017-05-14 ... | 2017-05-14 ... | 3:34:12 | [49.400... | 148.85 | 111.81... | ['2017-05-14... | [25.717, 25.7... | [[61.504, 49.... | 3921.00826... | [2666, 32... | ['2017-05-... | ROTOR | 17:40:00 |
| 2017-05-14 ... | 2017-05-14 ... | 0:55:58 | [49.400... | 360.49 | 246.91... | ['2017-05-14... | [20.035, 20.0... | [[138.66, 135... | 3276.06177... | [4903, 18... | ['2017-05-... | TOWER | 17:40:00 |
| 2017-05-14 ... | 2017-05-15 ... | 1:46:15 | [49.400... | 246.25 | 154.16... | ['2017-05-14... | [25.142, 31.6... | [[34.702, 36.... | 5086.51936... | [5904, 57... | ['2017-05-... | NACELLE | 17:40:00 |
| 2017-05-15 ... | 2017-05-15 ... | 4:58:36 | [49.410... | 151.03 | 117.18... | ['2017-05-15... | [27.855, 30.3... | [[27.855, 30.... | 6032.58226... | [6699, 76... | ['2017-05-... | ROTOR | 15:40:00 |
| 2017-05-15 ... | 2017-05-15 ... | 1:09:14 | [49.410... | 352.33 | 216.10... | ['2017-05-15... | [22.729, 22.9... | [[109.84, 106... | 8438.70373... | [8000, 78... | ['2017-05-... | TOWER | 15:40:00 |
| 2017-05-15 ... | 2017-05-16 ... | 1:23:31 | [49.410... | 247.87 | 151.16... | ['2017-05-15... | [24.954, 28.4... | [[39.246, 36.... | 7264.04648... | [7890, 82... | ['2017-05-... | NACELLE | 15:40:00 |
| 2017-05-16 ... | 2017-05-16 ... | 2:48:34 | [49.521... | 148.77 | 132.13... | ['2017-05-16... | [25.934, 27.6... | [[90.866, 87.... | 6121.27910... | [7050, 87... | ['2017-05-... | ROTOR | 3 days, 7:5... |
| 2017-05-19 ... | 2017-05-19 ... | 1:59:32 | [49.521... | 357.01 | 270.75... | ['2017-05-19... | [29.494, 32.2... | [[200.08, 197... | 10028.4948... | [10220, 1... | ['2017-05-... | TOWER | 3 days, 7:5... |
| 2017-05-19 ... | 2017-05-19 ... | 1:41:26 | [49.521... | 246.16 | 159.83... | ['2017-05-19... | [34.064, 46.1... | [[34.064, 46.... | 7249.11448... | [4837, 54... | ['2017-05-... | NACELLE | 3 days, 7:5... |
| 2017-05-19 ... | 2017-05-19 ... | 8:06:52 | [49.495... | 163.89 | 122.02... | ['2017-05-19... | [28.773, 32.5... | [[75.743, 73.... | 6293.02258... | [1760, 71... | ['2017-05-... | ROTOR | 23:25:00 |
| 2017-05-22 ... | 2017-05-22 ... | 1:02:08 | [49.495... | 361.78 | 222.30... | ['2017-05-22... | [20.538, 21.7... | [[67.58, 73.6... | 5908.97827... | [3817, 72... | ['2017-05-... | TOWER | 23:25:00 |
| 2017-05-22 ... | 2017-05-22 ... | 1:50:58 | [49.495... | 248.86 | 167.86... | ['2017-05-22... | [22.631, 24.1... | [[121.48, 120... | 4807.03682... | [6961, 73... | ['2017-05-... | NACELLE | 23:25:00 |
| 2017-05-22 ... | 2017-05-22 ... | 4:51:29 | [49.544... | 151.74 | 115.45... | ['2017-05-22... | [28.865, 32.1... | [[55.736, 50.... | 4596.26785... | [6381, 63... | ['2017-05-... | ROTOR | 1 day, 0:10... |
| 2017-05-23 ... | 2017-05-23 ... | 1:11:36 | [49.544... | 353.93 | 200.02... | ['2017-05-23... | [25.269, 27.5... | [[61.492, 64.... | 2974.50375... | [5371, 20... | ['2017-05-... | TOWER | 1 day, 0:10... |
| 2017-05-23 ... | 2017-05-23 ... | 0:57:07 | [49.544... | 245.68 | 170.99... | ['2017-05-23... | [25.211, 29.7... | [[71.296, 72.... | 2779.70931... | [2423, 22... | ['2017-05-... | NACELLE | 1 day, 0:10... |
| 2017-05-23 ... | 2017-05-23 ... | 1:19:39 | [49.544... | 242.67 | 145.00... | ['2017-05-23... | [226.4, 226.4... | [[176.73, 172... | 4532.83889... | [3620, 32... | ['2017-05-... | NACELLE | 1 day, 0:10... |
| 2017-05-23 ... | 2017-05-23 ... | 5:02:21 | [49.602... | 152.02 | 108.57... | ['2017-05-23... | [26.188, 27.6... | [[78.791, 68.... | 5044.56228... | [5492, 60... | ['2017-05-... | ROTOR | 1 day, 4:55... |

This work aimed to be able to tell if the wind data has any impact on the installations durations, the weather correlation with the historical data gave minimal results, but it does not mean that if we can get a hold of more data points, we can draw more insightful results.

Each row in this file represent a captured categorized installation. Data from all data sources has been appended to each installation according to previous steps and constrains described in section 1.2

## 7 Future work

This chapter is about the ideas that came up during the process of writing this thesis, but the scope of the work had to be limited because of time constraints as well as to keep the focus on specific tasks from start to finish.

### 7.1 Adding plugins and expanding the data for machine learning predictions

With the set of scripts developed during this thesis, it is possible to add new scripts that will append new exciting data to the already gathered and categorised datapoints. Based on the dates of the events of the installations, we can add more data if available, like wave hight on the sea, how many workers were on duty, the temperature, anything that can have a significant impact on the duration of the installation can be very easily added. Another source of weather could also be used, historical weather data from the area could potentially significantly improve the results. Those additional attributes are placed there for the case of needing it in predicting the amount of time needed for the installation and training the model. This would also be useful information to have on an installation report as well.

### 7.2 Waiting for more data

Much work has been done to create an environment that processes the dataset, and more data will give a better grounds for making the correlation with wind and installation time more precise. As the environment is set up, we only need more data and keep computing until we end up with meaningful results and insights. There is much potential for the use of the work for the Fred. Olsen Windcarrier. So far the visualisation itself gives much insight with little work to produce it, as all of the data cleanings is automated, and categorisation outputs are easily readable.

### 7.3 Creating automatic categorisation of datasets

When it comes to the value that this master thesis gives to anyone, the provider of the dataset will have the most use of it. As the result of this thesis is a kind of automatic categorisation and finding exciting points in the data (visualising the installations based on data). It is possible to automate reporting of each installation/operations that the vessels perform. If the uploading of the data is automated to cloud storage, we can automate downloading the data to the script and produce an output with logs and descriptive information about each installation. This functionality will give much information to the engineers working on the vessels, and maybe provide even more insight into the daily routine.

*Notes and Comments.*

## References


1. Songcheng Li: Tension Loss of Hydro Pneumatic 3 Proceedings of the ASME 2013 32nd International Conference on Ocean, Offshore and Arctic Engineering https://www.scribd.com/document/345916037/Tension-Loss-of-Hydro-Pneumatic-Riser-Tensioners
2. Fundamentals of Marine Riser Mechanics: Basic Principles and Simplified Analyses Section Principle Reasons for confusion about Effective Tention, page 58
3. http://windcarrier.com/fleet/#jack-ups
   Fred. Olsen Windcarrier vessels
4. http://www.encyclopedia.com/science-and-technology/technology/technology-terms-and-concepts/windmill
   Windmills
5. https://docs.python.org/3/library/statistics.html
   Python libraries
6. https://jaxenter.com/top-5-machine-learning-libraries-java-132091.html
   Python libraries
7. https://www.surveysystem.com/correlation.htm
   Correlation and statistics
8. https://sciencing.com/windmills-used-today-7220114.html
   Windmills and offshore technology
9. https://docs.scipy.org/doc/numpy-1.14.0/reference/ NymPy ref.
10. https://www.scipy.org/stackspec.html NymPy ref.
11. http://hdbscan.readthedocs.io/en/latest/how_hdbscan_works.html NymPy ref.