Mining Sensor Data to Overcome Decision Making Challenges in Race Sailing

Floris Smit
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
f.smit-1@student.utwente.nl

ABSTRACT
In the 2014-2015 edition of the Volvo Ocean Race, the teams in the Volvo Ocean race all use the same boat, the Volvo Ocean 65. The Volvo Ocean 65 is equipped with around 160 sensors on board. The model behind the sensors and the performance of the boat is not clear. Data mining techniques are necessary to find relations and to turn the data into useful information for the boat’s navigator. This research aims to identify and compare data mining techniques that are useful to providing information to the navigator. And it is a case study on setting up a data mining project in a competitive environment. The data analyst has to work together with an expert to look for a few percent efficiency to gain an advantage over the competition. The data mining techniques which look most promising are regression and visualization. Other techniques could also be useful should certain specific questions about the data arise. Visualization with specialized software is promising tool in increasing the data understanding.

Keywords
Sailing, Data mining, business intelligence, sensor data, visualization, decision making

1. INTRODUCTION
In professional sailing a small advantage can mean the difference between victory and defeat. Knowing and understanding the environment is essential for sailing from one point to another in the shortest amount of time. The crew of the race sailing boats has a lot of experience in doing this as they are the among best in the world.

Knowing and understanding the environment and the boat is a type of problem that is well known in other areas of society. In businesses knowing and understanding the environment is important for being able to adapt to changes and maximizing profit. Businesses have been adopting IT and business intelligence to assist them with this. By collecting data generated by their internal processes and combining this data with other external sources they gain new insights.

When competing in a sailing race, a new insight could provide a significant advantage over your competitors. In the Volvo Ocean Race (VOR) all teams race with the same boat. This makes the race a competition on teamwork and tactics. The boat type Volvo Ocean 65 has around 160 on board sensors collecting data about the boat. The boat’s navigator uses some of this data to plot the optimal route. Not all the boat’s sensor data is used by the sailing team. Therefore, the Brunel team and the University of Twente started a collaboration. The goal of this collaboration is to find new uses for the sensor data that may assist the Brunel team.

The software company SAS has also been involved in this collaboration. They are eager to share their expertise on mining large datasets on the problems at hand. They have made software licenses and online training for their products available to help the project around the Volvo Ocean Race.

Another collaboration was started at the time of this research. Infotopics is a company which focusses on helping organizations leverage their data by using business intelligence. They use software called Tableau for visualizing data. Tableau is known for being user friendly and facilitating quick, explorative analyses. During this research Tableau was used to do an explorative analysis of the data generated by the Volvo Ocean 65.

Applying data mining techniques to provide benefit to the business has been researched thoroughly. This falls under the larger concept called business intelligence. According to Watson there are two main activities that fall under business intelligence: Getting the data in, and getting the data out. Getting the data in is the part where data is put through an ETL (Extract, Transform, Load) process. The result of that ETL process is data that is loaded in to a data warehouse. The data warehouse enables the further analysis: getting the data out in the organization to assist in decision making. This is the step that generates the valuable information[14].

According to [10] the field of data mining has been extensively studied. There are 6 types of task that fall under data mining. They all produce different results and can be used for different types of problems. Association is the process of finding relations between variables. Classification is a common task that classifies each record to a certain predefined class. Clustering is the process of separating a homogenous dataset in to 2 or more subsets which have something in common. Regression is mapping one or more variable on another variable, creating a model that can predict a variable based on the values of other variables. Sequence discovery is the discovery of patterns.
which change in time. Visualization is a method to make complex data understandable to users by presenting the data in an image or graph.

Applying data mining techniques to provide benefit to the business has been researched thoroughly. This falls under the larger concept called business intelligence. According to Watson there are two main activities that fall under business intelligence: Getting the data in, and getting the data out. Getting the data in is the part where data is put through an ETL (Extract, Transform, Load) process. The result of that ETL process is data that is loaded in to a data warehouse. The data warehouse enables the further analysis: getting the data out in the organization to assist in decision making. This is the step that generates the valuable information[14]

A more business focused definition of data mining is given by [5]. They state that data mining is the process of turning raw data into useful knowledge. The raw data is abundant in most organizations as the majority of the business processes generate data. The extracted knowledge can be used to improve the business processes.

According to [10] the field of data mining has been extensively studied. There are 6 types of task that fall under data mining. They all produce different results and can be used for different types of problems. Association is the process of finding relations between variables. Classification is a common task that classifies each record to a certain predefined class. Clustering is the process of separating a homogenous dataset in to 2 or more subsets which have something in common. Regression is mapping one or more variable on another variable, creating a model that can predict a variable based on the values of other variables. Sequence discovery is the discovery of patterns which change in time. Visualization is a method to make complex data understandable to users by presenting the data in an image or graph.

A lot of companies say they use business intelligence because they realize that they have to use it to stay competitive. Truly getting benefit from business intelligence and analytics has proved to be challenging. Chen et. al. speak of three levels of evolution in business intelligence: BI&A 1.0, BI&A 2.0 and BI&A 3.0. In these levels of evolution the semantics of Web 1.0, 2.0 and 3.0 is used. BI&A 1.0 is the traditional business intelligence: structured data in a DMBS with various reporting tools. Although this has proven useful, the majority of the data that exists is not structured and often it is difficult to store everything. The next step BI&A 2.0 encompasses the analysis of web based, unstructured content such as social networks. This source of information is much richer but the unstructured nature makes the analysis more complex. The last level BI&A 3.0 is paired with the coming of the internet of things and the semantic web. It encompasses the vast amounts of data generated by mobile devices and sensor systems linked to the internet. Using these devices would allow for a location or person aware analysis. The challenge in this is the complexity and the size of the data. [4]

The size of data keeps increasing and there is plenty of research on big data. Big data means big in volume[large amounts], variety [lots of different variables] and velocity[high turnover][9]. Machines have physical limits in terms of memory, cpu and storage. The solution to these limits is letting multiple machines work together in tackling a problem. There are multiple tiers in an application stack where distributed systems are used. For storage a distributed database management system is required. This lets users query the system as though the data is present on one system. The database management system provides an abstraction layer for the distributed nature of the database. There are several technologies available that provide distributed storage systems. There are two major classes in these systems: systems that are optimized for update heavy workloads and systems that are optimized for quick read access. The latter are systems that are most suitable for data analysis. [1]

According to Turban, "Decision making is a process of choosing among alternative courses of action for the purpose of attaining a goal or goals." [13] Often there is an uncertainty in the different alternatives: the boundary between different goals and/or constraints are sometimes not clearly defined. This uncertainty in decision making is called fuzzy nature of the environment. Another form of uncertainty is incompleteness: some important data or alternatives for decision making may be missing.[11] The sailing domain is certainly fuzzy. The relations between the adjustable parameters, the environmental conditions and the time it takes to get from one point to another are not mathematically defined.

1.1 Problem statement
The Volvo Ocean 65 generates a lot of data. It is well known that large quantities of relevant data can be mined to create valuable information. This information is considered even more valuable if it leads to improved decision making. How information can be used to improve decision making is still unclear. This research aims to study how sensor data can be mined to improve decision making in professional sailing.

This problem applies to a lot of other fields and domains. The amount of data that is available increases, and often companies just collect data hoping it will add value in the future. Mining this data for impact on decision making is often perceived as difficult. The outcome of this research could provide handles to other domains on how to tackle this problem.

2. RELATED WORK
There is previous work done on this project: An analysis of data storage solutions and data mining software. Results of that analysis show that the data storage solution that seems most suitable is the PostgreSQL database. The dataset is small enough to be performing well on a traditional database. If the dataset would grow larger a distributed data storage solution like Hadoop would be better. For the analytics part SAS software seems most suitable. This previous work covered a lot data mining software, but tableau was not included. In this research tableau will be included to find out how useful it is.[8]

2.1 Research questions
Q1) What are key decision making challenges in navigating a professional sailing boat?
Q2) What data is generated by the Volvo Ocean 65?
Q3) What are the most promising data mining techniques to address Q1 and Q2?
Q4) Does application of the most promising mining techniques lead to better (perceived) results?
3. RESEARCH METHODS
This research aims to find how mining sensor data can lead to better (perceived) results in professional sailing. The research method will be design science: An architecture for data mining will be created. This architecture will include a script for importing data, data storage platform and a visualization platform.

3.1 Design science
Design science is a research method focused on gaining knowledge to solve field problems. This is achieved by creating tools and artifacts. In this research we will be using the three cycle variant of design science of Hevner et al (see figure 1) [6]. This design science variant consists of 3 cycles:

Relevance cycle Represents the interaction with the domain. It contains the mapping of domain needs and making requirements, as well as testing the created artifact and getting user feedback.
Design cycle The core of the design science method. It consists of creating or expanding artifacts and evaluating them.
Rigor cycle Represents the interaction with the scientific community: attaining knowledge to help in creating new artifacts and expanding the scientific knowledgebase.

Figure 1. The 3 cycle view of design science [6]

3.2 Data mining methods
The classical software engineering approach is not suitable for data mining and business intelligence projects. [3] Therefore a more suitable methodology is needed for this research project. There are multiple methodologies that are suited for data mining: KDD(Knowledge Data Discovery), SEMMA(Sample, Modify, Model, Explore) and CRISP-DM (Cross Industry Standard Process for Data Mining). The CRISP-DM methodology seems the most complete methodology so in this research CRISP-DM will be used. [2] CRISP-DM is an iterative data mining method that is based on input from over 200 data mining users. It divides the research process into six phases: [12]

Business understanding Mapping business needs and converting it to a data mining problem
Data understanding Collecting data to assess data quality and increase familiarity with the data
Data preparation Transform the data to get the final dataset which will be used in modelling
Modelling Selecting different modelling techniques, optimize these different techniques
Evaluation Evaluate whether the data mining model achieves the business goals
Deployment Present the gained knowledge so that the customer can use it

The CRISP-DM method does not impose strict order of phases. It is encouraged to go to other phases of the cycle if necessary. The arrows in figure 2 indicate common dependencies rather than strict sequence rules. [12]

Figure 2. The CRISP-DM research method cycle [12]

The different research questions will have a place in the CRISP-DM cycle.
- Q1: Business understanding Understanding the challenges in the domain.
- Q2: Data understanding Understanding the data generated in the domain.
- Q3: Modeling Finding the right modeling techniques.
- Q4: Evaluation Evaluating and validating the results.

3.3 Interview
In both design science and CRIPS-DM some form of interaction with the domain is required. An interview with Team Brunel’s navigator was held using Skype to provide this interaction. The architecture created for this research was evaluated using Teamviewer. The interview was also used to increase data and business understanding.

4. DEVELOPING AN ARCHITECTURE
For this research a data and analytics architecture was required. The choice of software was mainly based on the availability of existing architecture and licenses. For the data storage platform the university MySQL server was used. This way it saved time by not having to set up a new database server. For creating visualizations Tableau was used.

The data is supplied to us in csv format. Using python the csv files are converted to several sql files. The sql files are manually imported using MySQL Workbench. Then Tableau extracts the data to store it in a more efficient format before being ready for visualizing. See figure 3. In the following paragraphs follows a report of creating the infrastructure for analysis.

Figure 3. Architecture

4.1 Data import
The data is collected and preprocessed by the B&G WTP3 Wave Technology Processor. The WTP3 has several options for saving the data. The Brunel team uses csv files
which are stored in a different directory every day. As a first step each file type was loaded in its own table. Using a python script the files are parsed and two sql files are generated: one for creating tables and one for inserting the data. The data turned out to have some inconsistencies: one type of file destined for one table turned out to have different columns, and some of the files lacked headers. To compensate the multiple columns were duplicated to prevent loss of data, and the import script was enhanced with predefined header files. The result of this enhanced script was a database with several tables containing the data.

In the first exploratory analysis it turned out that the relations between the tables where hard to edit later. In the second iteration the script was edited to aggregate all of the different tables in one table containing all columns. This aggregated table was used to perform an exploratory analysis using tableau. When trying to plot the data on a map, it turned out that the coordinates where in a different format than tableau expected. It is possible to transform the data within tableau and SAS Enterprise, but it seemed cleaner to transform the coordinates during the import. This way the data is stored in the correct format in the database. During the third aggregation the correction of gps coordinates was added to the import script. The analytical software used to visualize the data cannot handle time stamps in an efficient manner. It is optimized to handle database dumps which have their relations expressed using ID’s and foreign keys. In the data generated there are two tables which represent intervals rather than a single measurement in time. The analytical software doesn’t have functions to match these records on date. These tables contain vital information: for example which mainsail is used. Filtering data based on such parameters will lead to graphs and other visualizations which give a good view of the situation at one glance. So the last step was to add relations between the tables containing single measurements and the ones representing an interval. To do this two extra columns were added in each of the single measurements tables referencing the interval tables. This way the analytical software doesn’t have to repeat the calculations.

Table 1. Data platform

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Multiple tables</td>
</tr>
<tr>
<td>2</td>
<td>Aggregated table</td>
</tr>
<tr>
<td>3</td>
<td>Aggregated Table, corrected gps coordinates</td>
</tr>
<tr>
<td>4</td>
<td>Multiple tables, corrected gps coordinates</td>
</tr>
<tr>
<td>5</td>
<td>Multiple tables, corrected gps coordinates, added relations</td>
</tr>
</tbody>
</table>

4.2 Tableau

Tableau is a software package which lets users create visualizations. It is well known to be really user friendly and easy to use. This makes Tableau suitable for doing exploratory data visualizations. Tableau can connect with numerous data sources including MySQL. But using tableau in this manner puts a lot of pressure on the resources of the database server. This results in reduced performance and long load times when making visualizations. Therefore Tableau data extracts where made which store an entire tables in a single file. By making several optimizations to the way that data is stored Tableau is able to respond much faster when applying filters.

5. RESULTS

5.1 Challenges

Based on the interview done with the Brunel navigator some insights on the decision making challenges of navigating a sailboat where found. During the race there are not many limitations on where the boats can go as long as they reach certain waypoints. This leaves a wide range of routes that can be sailed. Choosing which route to go is the navigator’s job. The teams make decisions they think will get them an advantage over their competitors.

The crew has collected a vast amount of information on the boat’s behavior in different conditions. For every wind angle and wind speed they know what the theoretical maximum speed is. This is dependent on physical properties of the boat and on the different sails they use. The crew is constantly looking to get the highest speed possible for that wind angle and speed. If the boat is not going the theoretical maximum speed the crew will try something new. The navigator monitors the instruments and checks if the speed improves. Wind is obviously the highest influence in the maximum speed possible. They use weather equipment to look at the different wind speeds and directions. Since they know how fast the boat will go in all circumstances the navigator is capable of finding the fastest route. The navigator relies on a wide array of information sources as well as years of experience. Changing course to try for more favorable conditions is a tactical aspect of the race. When a boat is behind it will stay behind if the same course and route is followed. The navigator has to find a route which might turn out to be faster. They can make an educated guess to predict how the weather will change and what effect that will have on the different routes. Making a good guess lets you win the race. Some of it is luck but a great deal of it is based on having correct information. Predicting the velocity of the boat at certain conditions is one of the major challenges. This is something where mining the generated data can help.

In between the races the team evaluates the results on shore. They analyze the data further and use it for their debriefings. Increasing the effectiveness of these debriefings and allowing the team to learn more from looking back at the race.

5.2 Data generated

The Volvo Ocean 65 has over 160 sensors on board which generate data for around 40 variables. The navigator performs analysis on a daily basis using visual basic macros and excel. In the current workflow the files are shared using dropbox each time the team enters a harbor where they have a stable internet connection. They are not allowed to use internet during the race. The data that is generated by the Volvo Ocean 65 fall one of 3 categories: wind data, navigation data and boat data. See appendix A for a list of the most important data that is generated.

Data about wind is very important for sailing. In the sailing domain there are two wind types: true wind and apparent wind. The true wind is the wind that would be measured when not moving. The apparent wind is the result of the true wind combined with the wind as shown in Figure 4. The boat has a certain speed (V). As a result of this speed there is a headwind (HW), this head wind would be present even if there is no true wind. The apparent wind (AW) is the wind that is experienced and measured on the moving boat. It is the result of the two
vectors head wind and true wind. The B&G WTP3 processes the wind data and calculates true wind speed and angle using the following formula:

\[ TWS = \sqrt{AWS^2 + V^2 - 2 \times AWS \times V \cos (AWA)} \]

Where TWS is the True Wind Speed, AWS is the Apparent Wind Speed, AWA is the Apparent Wind Angle and V is the boat speed. The AWS and the AWA are measured with sensors on the boat. The boat speed is calculated using GPS.

![Figure 4. Wind types][7]

### 5.3 Promising Techniques

There have not been any data mining techniques applied to the data yet. Therefore, the main criterion for evaluating the data mining techniques was the complexity of applying them to the sailing data. If they are too complex there is a risk that the navigator of the boat doesn’t understand what has been done. The results and the techniques’ effectiveness can’t be validated when the end users don’t understand what has been done. Applying a few relatively simple data mining techniques and evaluating them can increase the data and business understanding. This will increase the likelihood of more complex data mining techniques proving useful in the future.

The different techniques require different data types as input. It is possible to transform the data to the correct format, but there is a risk of losing or distorting information in doing this. The team working on this project needs to enlarge the data understanding before attempting these transformations.

The type of result a certain technique generates is also important to take in consideration. Are they suitable for the type of problems that are encountered in sailing? Transforming the results or input data could change the suitability of a certain technique, but the complexity increases.

Association is the process of finding relations between discrete variables. For example mining association rules of a student database we could discover that 80% of all students following a data mining course has also followed a course about databases. In a supermarket transaction database we could discover that 75% of customers buying flour and apples also buy sugar. This technique is very suitable to study and predict the occurrence of certain events. The data that is generated by the Volvo Ocean 65 is continuous data. Before association can be used the data needs to be transformed in a meaningful way. The usefulness of association as a data mining is highly dependent on this transformation.

Classification is a common task that classifies each record to a certain predefined class. For example a database of insurance claims could be classified in fraudulent and non-fraudulent claims or one could classify emails as spam or not spam. Classification brings along a similar problem as association. The technique itself could be useful if the data is transformed in a meaningful way. You could divide the data in to segments which have similar conditions or split them based on points in time where decisions are made. Then classification can be used to validate the decisions.

Clustering is the process of separating a homogenous dataset in to 2 or more subsets which have something in common. Clustering is similar to classification. The difference between the two techniques is that classification requires predefined classes. The clustering technique determines the different classes or clusters by itself. Clustering also requires that the data generated by the Volvo Ocean 65 be transformed before something useful can be done.

Regression is mapping one or more variable on another variable, creating a model that can predict a variable based on the values of other variables. The result of a regression analysis is the weight each variable has. The combined result of the variables multiplied by the weight is the predicted value. Most of the sensor data generated by the Volvo Ocean 65 is continuous. This kind of data is quite suitable for regression analysis. The only thing necessary which might pose problems is the predicted value. The speed of the boat is not suitable for this. Making a detour where the wind is more favorable can result in higher boat speed, but doesn’t necessarily let the team make a good time.

Sequence discovery is the discovery of patterns which change in time. Sequence discovery is similar to association, but it also takes the order of the records in to account. For example, sequence discovery could be applied to the database transactions of a webshop. For each customer there is a sequence of purchases in time. Because of the similarity to association the same issues are encountered. A data transformation has to be done for the technique to have a meaningful result. All the data generated by the Volvo Ocean 65 has a timestamp. After transforming the data in a meaningful way sequence discovery could prove to be useful.

Visualization allows presentation of complex data using an image or graph. The human brain is very good in making associations. When color and shape are used correctly the human brain can see patterns and anomalies in the data at a glance. Most data mining techniques result in a series of numbers. Visualization is often used as a final step to show the results of a data mining analysis. When visualizing data its possible to use different amounts of data transformation. This makes visualization well suitable as a data mining technique. The proper visualizing tools allow exploring data in a quick manner. Visualization could be used to display the exact route that was taken during a race. Together with graphs about variables like wind and boat speed, a quick evaluation of the race would be possible.

With the criteria of complexity in mind visualization and regression look the most promising. Visualization is a great way of exploring data and gaining more insight in
the data. Regression could be used to predict the velocity of the boat. The current models used for predicting the boat speed use much less variables than a regression would allow.

The other techniques could also prove useful. But at this point applying them requires major transformations of the data and sometimes the result. Given the limited time it would seem wise to begin exploring visualization and regression first.

5.4 Improved results
The approach to address the last research question was explorative. The data and business understanding needed to get higher to be able to do more complex mining techniques. Setting up the environment for data mining and parsing the data in the correct way has taken up more time than expected. The most promising techniques could still be applied: visualizations were made and one regression analysis was done. The effect of applying these techniques on the decision making challenges has not been measured. Tableau was used to make some example visualizations. The gps coordinates of the measurements where used to make a map view. This map view could be used to make selections in the data on an interactive dashboard. This allows the navigator to select data based on a portion of the route. In figure 5 you can see the example dashboard with two graphs: one has the boat speed and the other has which staysail was used. The route in the map is colored based on the true wind angle.

During the interview with the navigator these visualizations where shown. The goal of these visualizations was to show the possibilities to the navigator and to find out of these kinds of visualizations could be of use. The reason for this is that the navigator didn’t have any specific questions or information requests. He liked the way the visualizations looked and also how easy it was to make them and change them.

Regression has shown to be a promising technique but due to time constraints only a simple analysis has been done. In figure 6 there are lines drawn through the scatterplot. These lines are regression models to determine boat speed at different wind speeds and wind angles.

The team already uses historical data and some simple plotting to determine maximum speed at different conditions. Although this could be sufficient for their purpose it is still worth trying out regression to learn if better velocity prediction can be achieved.

6. DISCUSSION AND CONCLUSION
Modern technology and trends stimulate use of data mining techniques in a lot of areas of society. In all these area’s there are experts that have been working there for decades. They might see the potential added value in mining data and they might cooperate because they have to. But the key aspect in working with these experts is the fact that they are an expert in their area, and the data analyst is an expert with data. The challenge for data analysts is to walk the thin line between showing things that are trivial to the expert and making the analysis too complex for the expert to understand. In both extremes the expert will dismiss the result and continue to rely on his own expertise instead of combining their expertise with statistical evidence.
The first research question aimed to find the decision making challenges of navigating a race sailboat. Based on the interview with the navigator there are conclusions that can be drawn. Navigating the boat is one of the most important tasks. Finding the route that lets them reach their destination in the shortest amount of time is a difficult task. The navigator relies on different information sources for making decisions. All the information combined with years of experience allows him to make good decisions. If an competitor is ahead trying a different route is the only way to get ahead of them.

The goal of the second research question was to increase the data understanding. In the beginning of the research some variables were unknown. During the interview most of the variables and their units where identified. Roughly the data can be divided in three groups: Wind data, navigation data and boat data.

The result of the third research question are the most promising data mining techniques. The most promising data mining techniques are regression and visualization. They are most compatible with the data type of the generated data. Applying these techniques require little data transformation than the other techniques. Visualization is different compared to other data mining techniques since it has much more applications and can be seen as a multi tool. Visualizing data using Tableau turned out to allow for a quick and explorative analysis. Although regression is a promising technique caution is advised in applying this technique. It is very important that the results are properly validated in a realistic situation or with test data. Good communication with the sailing team is vital and developing a validation technique before making the model could be beneficial. Another reason it is important to validate the result with the sailing team is that correlation doesn’t have to mean causation. When a correlation is found or when testing the model it is important to try it out on in the real world.

Every data mining technique can be seen as a tool, which is good at a specific job. The other techniques are valuable assets to the toolbox of a data analyst. The conclusions in this research regarding the different data mining techniques apply to other projects that are just starting and/or have time constraints. In general it’s wise to pick the right tool for the job by looking at the nature of the problem at hand.

The last research question was to find out how applying the most promising data mining techniques would lead to better (perceived) results. The two most promising data mining techniques have been tried out and results were shown to Brunel’s navigator. The navigator was positive about the results, but due to time constraints it was not possible to study if applying them would lead to better (perceived) results. This leaves the last research question only partially answered.

7. FUTURE WORK
There are several things that could prove useful when doing future work on this topic. The team now uses data to evaluate leg results. Making a dashboard and setting up an environment will allow them to have more effective evaluations. Mapping the information needs could be done by doing interviews and forming requirements. Expanding the infrastructure to reduce the amount of time data management takes would be useful if more complex data mining techniques are to be studied. When doing more complex analysis it is important to have an overview of the data. During this research it appeared that the analysis done to calculate the theoretical maximum speed is pretty basic. More research could be done on how to improve this model. A regression model could be a solution to this.

8. ACKNOWLEDGEMENTS
I would like to thank Prof. Dr. J. Hillegersberg for his guidance and support, without him this research would not have been possible. Thanks to team Brunel for their time and sharing the data with us. I hope the cooperation will bear its fruits for them in the future. Thanks to all the people who have helped me set up the infrastructure.

9. REFERENCES
APPENDIX

A. GENERATED DATA

A.1 Wind data

**True wind speed** Speed of the wind, without the motion of the boat.

**True wind direction** Direction of the wind.

**True wind angle** Direction of the wind relative to the ship.

**Apparent wind speed** Speed of the wind as experienced on the moving boat.

**Apparent wind angle** Angle of the wind as experienced on the moving boat.

A.2 Navigation data

**Boat speed** Speed of the boat including currents.

**Compass heading** Course indicated by the compass.

**Speed over ground** Speed of the boat excluding current.

**Course over ground** Course relative to the land.

**Position** GPS coordinates of the boat’s position.

A.3 Boat data

**Heel** The angle that the boat heels to one side.

**Trim** The angle that the boat heels forward/backward.

**Rudder** The angle of the rudder.

**Keel** The angle of the adjustable keel.