# Improving Decision Making in Ocean Race Sailing Using Sensor Data

*Full Paper*

**Jos van Hillegersberg**
University of Twente
j.vanhillegersberg@utwente.nl

**Mark Vroling**
University of Twente
markvroling@gmail.com

**Floris Smit**
University of Twente
florissmit10@gmail.com

## Abstract

The increasing availability of sensor data and analytics tools and techniques brings new opportunities to enhance decision making in professional sports. However, achieving higher sports performance using decision support systems is not straightforward. Sports professionals have often relied for decades on their experience and intuition only. While in some sports, experiences have been gained using traditional information and decision support systems, using large sensor datasets for sports analytics is a recent phenomenon. Using sensor data to arrive at effective decision support for sports encompasses various challenges: (1) Sensor data needs to be understood, processed, cleaned and efficiently stored and (2) appropriate data analytics and visualization techniques need to be selected and evaluated with the sports professionals. Few elaborate case studies are available that report on the development of decision support systems for professional sports teams. No comprehensive set of generically applicable design principles has been devised to develop analytics support for sports teams based on sensor data. We deploy a design science methodology to arrive at a sailing analytics architecture and race evaluation dashboard in close collaboration with a professional sailing team participating in an ocean race around the world. Using sensor data, we design a sensor data storage architecture and deploy a variety of analytics methods and visualizations to arrive at effective support for the team's navigator and captain. The architecture and analytics dashboard have been designed in several iterations using evaluation sessions with the team's navigator and captain. The study demonstrates how sensor based analytics could be used to achieve better performance in the hyper competitive environment of ocean race sailing. Based on the evaluations of the analytics dashboard we derive lessons learned. The paper concludes by outlining future directions in studying sensor based analytics for sports.

**Keywords**

Case study, sports analytics, sensor data, business analytics, decision support

## Introduction

### *The Volvo Ocean Race*

The Volvo Ocean Race is the world's pre-eminent round-the-world yacht race and one of the most coveted prizes in the sport. The race has been organized since 1963. The Race is owned by Volvo Cars and Volvo Group and managed by a dedicated team working from headquarters in Alicante, Spain. The 2014-15 race route tested the world's best professional sailors to the limit in a race around the world lasting almost nine months and 38,739 nautical miles. For the first time, the race adopted a "one-design" boat (all teams sailed with an identical boat) (Figure 1). The new one-design Volvo Ocean 65 that will contest the next two editions of the Volvo Ocean Race is delivered with everything included and "ready to sail", on a strict one-design rule with the latest technology on board (Volvo Ocean Race 2016). The sensor data generated by the boat give the sailing teams a range of opportunities to make better decisions based on detailed real-time sensor data and aim at improving performance and learn the decisive details of the boat in the variety conditions they encounter during the race. However, the race only provides the team with raw sensor data and no actual business analytics tools are offered to perform analysis on the complex data the Ocean 65 generates.


**Figure 1. The Ocean 65 Boat and Race Route**

The authors had the opportunity to team up with one of the seven participating teams and became scientific partner of the team with the task to design and apply data analytics on the sensor data the boat generates. The goal was to facilitate decisions by the team and accelerate the learning of the team to achieve optimal performance. The Volvo Ocean race allows 3rd party support during training and while teams halt in the various seaports at the end of each leg of the race.

The seven highly experienced and professional navigators that participate in the Volvo Ocean Race are all very knowledgeable in ocean navigation and in selecting optimal routes depending on weather conditions and forecasts. Navigation software and weather data have been used on board in sailboat races for several years. However, the new Ocean 65 is equipped with 160 sensors on board that offer new opportunities to understand the behavior of the boat in various conditions. More data is available than ever before and this data can be stored, aggregated and analyzed to facilitate understanding and learning and ultimately support both small operational decisions (such as trimming of sails or setting keel angle) and eventually more strategic decisions such as deciding to take an alternative route seeking wind and current conditions that fit the boat best (Figure 2).

**Figure 2. Example of Alternative Decisions Based on Estimates of Wind Conditions and Optimal Speed in the Route Leg to Auckland**

### Research Objective and Questions

The project was initiated by the captain of the team. Facing the challenge of getting the optimal performance out of the newly designed Ocean 65 boat, he contacted our research group for assistance in using the sensor data it generated. At the start of the research project, the ocean race team was using traditional navigation software for its route planning and pre-set factory tables to determine optimal route and sail configuration. Effective Storage and Business Analytics techniques are required to find relations in the sensor data and to turn the data in to useful information for the boat's navigator and captain. The model behind the sensors and the performance of the boat is currently unknown and no predictive models exist today.

Few elaborate case studies are available that report on development of decision support systems for professional sports teams. No comprehensive set of generically applicable design principles has been devised to develop analytics support for sports teams based on sensor data. We deploy a design science methodology to arrive at a sailing analytics architecture and race evaluation dashboard in close collaboration with a professional sailing team participating in an ocean race around the world. We design a sensor data storage architecture and deploy a variety of analytics methods and visualizations to arrive at effective support for the team's navigator and captain. The architecture and analytics dashboard are designed in several iterations using evaluation sessions with the team's navigator and captain. We aim to enhance our understanding of how sensor data and analytics can ultimately lead to better decisions in professional sports.

A related stream of research is focused on understanding sailing in order to build autonomous sailboats using artificial intelligence techniques (Adriaans 2003; Alves and Cruz 2008; Warden 1991). Our research is not striving for fully automating the work of the navigator in a race team. Given the large number of decision variables, unexpected events and unknown number of external factors that influence the ultimate performance, it is highly unlikely that a fully automated navigator will be able to outperform an effective interaction of a human navigator and a business analytics environment that supports him/her. We take a broader view than only focusing on the modelling aspects. We are foremost interested in supporting the human expert with proper analytics and visualizations, in this case the team navigator. We build on the large body of knowledge on business analytics and data mining (Duan and Xu 2012; Watson and Wixom 2007). By studying their application in the extremely competitive environment of race sailing where many decisions have to be made under high pressure, we aim to evaluate current techniques for tactical and strategic decision making. Lessons learned in extreme conditions like top-sports races can be transferred to and inspire organizational settings in business and non-profit sectors.

Two research questions guide our study: (1) How can an effective architecture be built to clean, store and analyze/visualize big sensor data generated by the Ocean 65 sailboat and (2) what analytic and visualization techniques are perceived most usable to enhance sailing performance. The remainder of the paper is structured as follows: The next section will briefly explain our research method and describe the architecture built for data storage and analytics and the application of various analytics and visualization techniques on the data. It also discusses the evaluation and perceived usefulness to enhance performance. Finally, we present conclusions and discussion.

## Design Science Research Process

### Design Science Methodology

To address the research questions and develop a Sailing Race Analytics Architecture (SRAA) we apply design science (Hevner et al. 2004; Peffers et al. 2007). In particular, we selected to use the Design Science Research Methodology (DSRM) of Peffers et al. (2007) (See Figure 3).

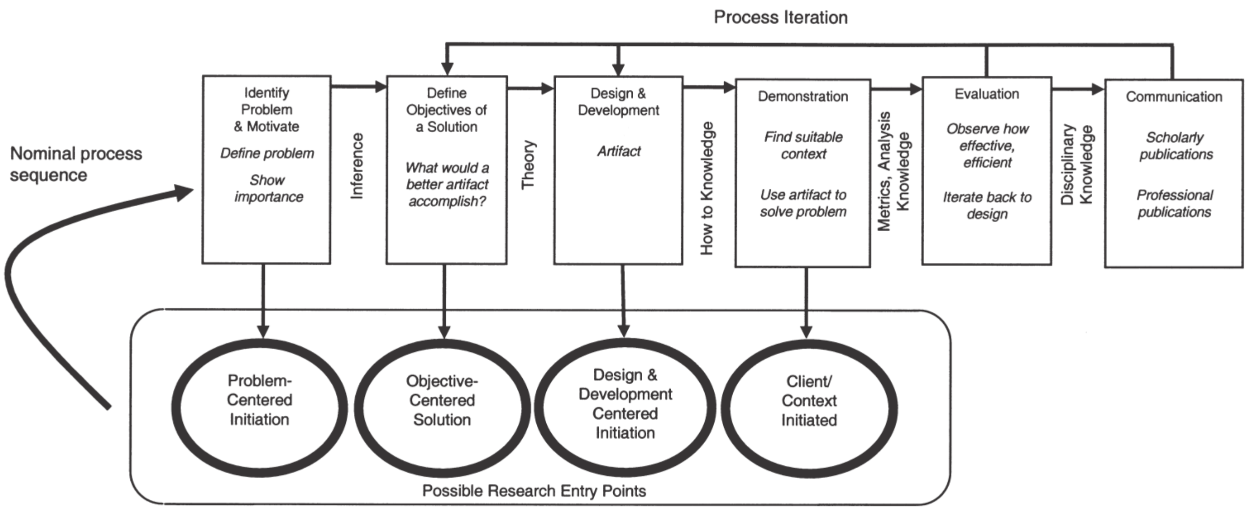


**Figure 3. Design Science Process (Peffers et al. 2007)**

DSRM fits the problem centered approach required to iteratively design a sailing analytics and support environment and learn lessons and derive design principles through artefact demonstration and evaluation. DSRM consists of 6 stages that we have applied to iteratively design the SRAA. We describe the 6 stages below:

## *Motivation and Objectives for the Solution (Stage 1&2)*

The motivation for this project is both theoretical and practical. As analyzed in (Chen et al. 2012; Phillips-Wren et al. 2015), new large and complex data sets resulting from sensor data pose new challenges to designing effective analytics and decision support. The authors outline a range of challenging research questions including: "How can cause and effect relationships between various big data measures (i.e., analytical pathways) be discovered/validated? What visualization methods are most useful for decision support with big data?". Through this design science research using complex real world data from a sailing range and evaluating an analytics architecture and various analytic methods and visualizations we aim to address these timely questions. From a practical perspective, we address the need of the ocean race team to better understand the opportunities that business analytics of sensor data brings to their performance improvement ambitions. More generically, we aim to develop an inspiring and informative sample study to any team that aims to effectively use sensor data to enhance its capabilities through evaluation and learning from analytics.

Based on an interview with the Sailing Team navigator several insights on the decision-making challenges of navigating a sailboat were identified. During the race, there are few limitations on where the boats can go as long as it stays within 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. During training sessions, 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 trying to get the highest speed possible for that wind angle and speed. If the boat is not reaching the theoretical maximum speed the crew will take action and try a new strategy. The navigator monitors the instruments and checks if the speed improves. Wind is obviously the most important factor to achieve the maximum speed possible. Weather data is used to analyze different wind speeds and directions. The navigator uses the theoretical maximum boat speeds in various circumstances to find 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. He/she 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. Predicting the velocity of the boat at various conditions is one of the major challenges. This is something where analyzing and mining the sensor data collected can help. In between the race legs the team evaluates the results of the most recent route leg on shore. They summarize race data and decisions made and use these for their debriefings. Using analytics can increase the effectiveness of these debriefings and allow the team to learn more from looking back at the race leg.

## *Design and Development of the Sailing Race Analytics Architecture (Stage 3)*

This stage comprises the development of the data and analytics architecture and sailing race evaluation dashboard. The SRAA was developed in several iterations. For the data storage platform, the university's MySQL server was used. For creating the visualizations and evaluation dashboard the data analytics tool Tableau was selected based on availability, skills and resources and requirements to quickly build various visualizations on in-memory datasets. Tableau is a software package that lets users create a wealth of visualizations. It is well known to be user friendly and relatively easy to use. This makes Tableau suitable for doing exploratory data visualizations in this project. The data is transmitted from the on-board sensor data collection software to the SRAA in csv format. Using Python, the csv files are converted to several SQL files. The SQL files are imported using MySQL Workbench. Then Tableau is configured to extract the data to store it in a more efficient in-memory format before being ready for visualizing (Figure 4).

The data is collected and preprocessed by the B&G WTP3 Wave Technology Processor (part of the Ocean 65 infrastructure). The WTP3 has several options for saving the data. The Sailing Team uses csv files which are stored in a new 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.

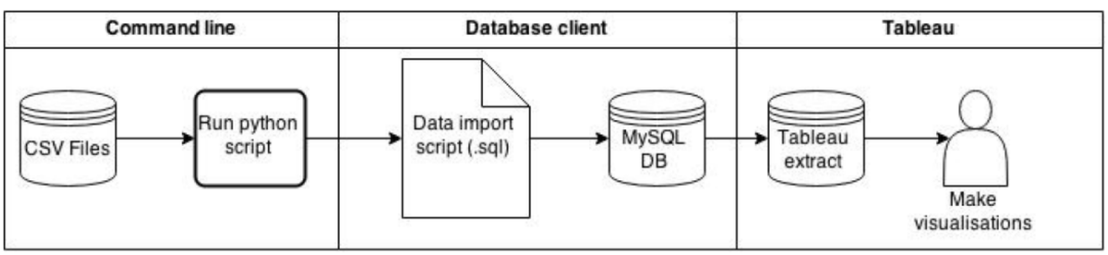


**Figure 4. Sailing Race Analytics Architecture (SRAA)**

In the first exploratory analysis it turned out that the relations between the tables w

ere 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 were in a different format than Tableau expected. Thus, the coordinates are transformed during the import. 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 does not have to repeat the calculations.

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 were made which store an entire table 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.

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 synchronized shared directories (using Dropbox) each time the team enters a harbor where they have a stable internet connection. Teams 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. Clearly, wind data 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. 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:

$$TWS = \sqrt{AWS^2 + V^2 - 2 * AWS * \cos(AWA)}$$

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.

The main criteria for evaluating the data mining techniques were their effectiveness and complexity in applying them to the sailing data. If they are too complex, there is a risk that the navigator of the boat does not understand the steps taken in the data analytics. 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 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? We review the potential of various data analytics techniques to the research objectives below:

*Association* is the process of finding relations between discrete variables. 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 in data mining is highly dependent on this transformation.

*Classification* is a common task that classifies each record to a certain predefined class. Classification brings along a similar problem as association. One could divide the data into segments which have similar conditions or split them based on points in time where decisions are made. This could enable new insights that 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.

*Regression* is mapping one or more variables 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. Only focusing on the speed of the boat is not viable. Making a detour where the wind is more favorable can result in higher boat speed, but doesn't necessarily let the team arrive at a good result.

*Sequence discovery* is the discovery of patterns which change over time. Sequence discovery is similar to association, but it also takes the order of the records in to account. Because of the similarity to association the same issues are encountered. All 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 at making associations based on visual presentations of data. 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, it is possible to use different amounts of data transformation. This makes visualization well suitable as a data mining technique. 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 high effectiveness and limited 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 listed could also prove useful, but as complex transformations of the data would be required it was decided to first focus on visualization of current data and performing regression analysis.

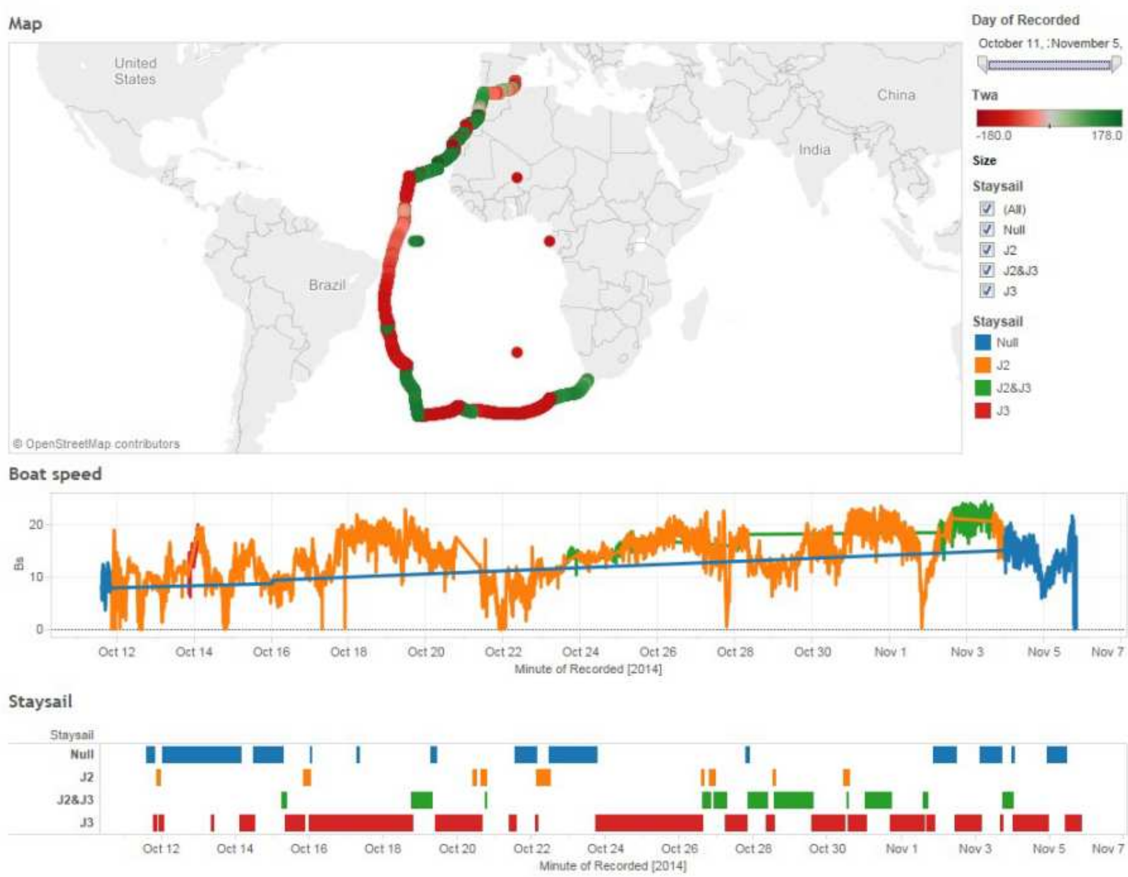


**Figure 5. Map Speed Dashboard Showing Speed and Sails Used**

The business analytics tool Tableau was used to create visualizations. The GPS coordinates of the measurements were used to generate 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 particular segment of the route. Figure 5 shows the dashboard containing 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.

In the other graph the boat's properties concerning wind were visualized. Unlike the last dashboard time was not included in the graph. Normally Tableau aggregates and summarizes the data to a level which is suitable for the analysis, but for this graph no aggregation was applied. In these two graphs the true wind speed is shown on the x axis and the boat speed is shown on the y axis. The data is separated based on the angle of the wind and shown in different colors. In the bottom graph a selection of three wind directions is made to show the difference. The top red line shows a favorable wind angle: The line is quite steep so a lot of speed is gained with little wind. The bottom two lines show much less favorable conditions. In some cases, the wind speed is actually much higher although the boat speed stays the same.

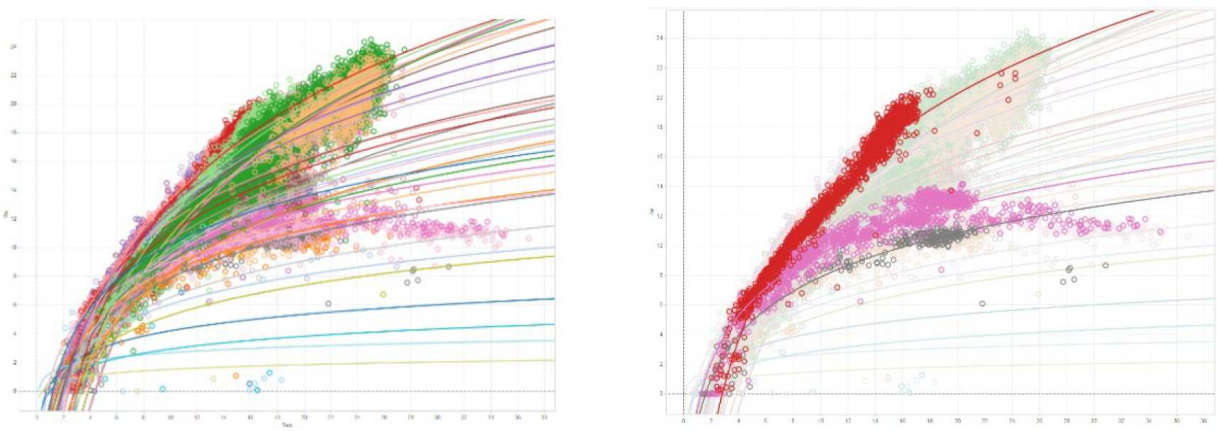

**Figure 6. Wind Graphs Regression Boat Speed versus True Wind Speed**

Regression has shown to be a promising technique. 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. Using these visualizations, the team started using historical data and some simple plotting to determine maximum speed at different conditions.

## *Demonstration, Evaluation and Communication (Stage 4, 5&6)*

During interviews with the navigator these visualizations were shown. The goal of these visualizations was to show the possibilities to the navigator and to find out whether 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 upfront. After the first iterations and processing his feedback, he liked the way the visualizations looked and also how easy it was to make them and change them. Next, the SRAA and dashboards were presented in a workshop to a group of sailing experts, trainers and researchers including sailing teams preparing for the Rio 2016 Olympics. Feedback on the visualization of the data was collected. Moreover, alternative analyses opportunities were suggested such as comparison of route legs and the use of qualitative data to annotate the current data e.g. explaining why certain decisions were made.

Finally, an interview with the team manager and team captain were held. They suggest more enhancements that would make the SRAA visuals suitable to be a central tool in the race evaluations after each leg. One important element mentioned was including the possibility to create a matrix that would show which sail would lead to best performance under a certain wind angle and wind speed. These enhancements will be included in future versions of the SRAA.

## Conclusion and Discussion

We aimed at studying how sensor data including geo-and performance data can enhance decision making in the highly competitive environment of professional sailboat racing. Two research questions guided our studies: (1) How can an effective architecture be built to clean, store and analyze/visualize big sensor data generated by the Ocean 65 sailboat and (2) what analytic and visualization techniques are perceived most usable to enhance sailing performance.

A combination of custom made scripts, modern database and analytics tools was used for setting up an effective architecture for analyzing big sensor data. However, there is no single tool that can do the job and the wealth of software solutions available in cleaning, storing and analyzing/visualizing makes selecting and integrating software tools and building a robust and user friendly architecture non-trivial. A design approach applying multiple cycles and incorporating the latest knowledge on methods and tools and evaluating frequently is recommended here as theory and technologies are evolving and each field context may require unique architecture configurations. Regarding the effectiveness of various analytics and visualization techniques, we also conclude that an iterative design approach in close collaboration with the expert user is effective and prevents overly complex or unneeded analyses or visualizations.

Like in other complex expert domains, the expert user conducts a difficult task (navigating a sailboat). The navigator relies on different information sources for making decisions. All the information combined with years of experience allows him to make good decisions. The most promising data mining techniques identified to support the navigator are regression and visualization. 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. 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 in the domain for decades. They might see the potential added value in mining data and they might cooperate because they like it or 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.

There are several future research directions. Professional sailing teams will increasingly start using data visualizations to evaluate race leg results. Observing these evaluations and mapping the information needs could result in eliciting new requirements. Expanding the infrastructure to reduce the amount of time data management takes would be useful. More complex data mining techniques can also be studied. Experimenting with machine learning and intelligence amplification/cognitive computing offer new opportunities.

## Acknowledgements

The authors wish to acknowledge the team captain, navigator and crew members for the inspiring collaboration and insights they have offered us in the world of race sailing.

## REFERENCES

Adriaans, P. 2003. "From Knowledge-Based to Skill-Based Systems: Sailing as a Machine Learning Challenge," in *Knowledge Discovery in Databases: PKDD 2003*, Lecture Notes in Computer Science, N. Lavrač, D. Gamberger, L. Todorovski, and H. Blockeel (eds.), Springer Berlin Heidelberg, pp. 1–8 (available at http://link.springer.com/chapter/10.1007/978-3-540-39804-2_1).

Alves, J. C., and Cruz, N. A. 2008. "FASt - an autonomous sailing platform for oceanographic missions," in *OCEANS 2008*, Presented at the OCEANS 2008, , September, pp. 1–7 (doi: 10.1109/OCEANS.2008.5152114).

Chen, H., Chiang, R. H., and Storey, V. C. 2012. "Business Intelligence and Analytics: From Big Data to Big Impact.," *MIS quarterly*, (36:4), pp. 1165–1188.

Duan, L., and Xu, L. D. 2012. "Business Intelligence for Enterprise Systems: A Survey," *IEEE Transactions on Industrial Informatics*, (8:3), pp. 679–687 (doi: 10.1109/TII.2012.2188804).

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design science in information systems research," *MIS quarterly*, (28:1), pp. 75–105.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems*, (24:3), pp. 45–77 (doi: 10.2753/MIS0742-1222240302).

Phillips-Wren, G., Iyer, L. S., Kulkarni, U., and Ariyachandra, T. 2015. "Business analytics in the context of big data," *Communications of the Association for Information Systems*, (37) (available at https://asu.pure.elsevier.com/en/publications/business-analytics-in-the-context-of-big-data-a-roadmap-for-resea).

Volvo Ocean Race. 2016. "Volvo Ocean 65," (available at http://www.volvooceanrace.com/en/boat.html; retrieved June 15, 2016).

Warden, W. H. 1991. "A control system model for autonomous sailboat navigation," in *, IEEE Proceedings of Southeastcon '91*, Presented at the , IEEE Proceedings of Southeastcon '91, , April, pp. 944–947 vol.2 (doi: 10.1109/SECON.1991.147900).

Watson, H. J., and Wixom, B. H. 2007. "The Current State of Business Intelligence," *Computer*, (40:9), pp. 96–99 (doi: 10.1109/MC.2007.331).

# Appendix A. Generated Sensor Data Categories

| **Wind Sensor Data** | | **Navigation Sensor Data** | | **Boat Sensor Data** | |
|---|---|---|---|---|---|
| True wind speed | Speed of the wind, without the motion of the boat. | Boat speed | Speed of the boat including currents. | Heel | The angle that the boat heals to one side. |
| True wind direction | Direction of the wind. | Compass heading | Course indicated by the compass. | Trim | The angle that the boat heals forward/backward. |
| True wind angle | Direction of the wind relative to the ship. | Speed over ground | Speed of the boat excluding current. | Rudder | The angle of the rudder. |
| Apparent wind speed | Speed of the wind as experienced on the moving boat. | Course over ground | Course relative to the land. | Keel | The angle of the adjustable keel. |
| Apparent wind angle | Angle of the wind as experienced on the moving boat | Position GPS | Coordinates of the boat's position. | | |

**Table 1. Sensor Data Categories**