

HafenCity Universität Hamburg
Department Geomatics

**Generation and Provision of Ship's Master Track
Data and Metadata for Standardized Access**

Master Thesis
by
Ammar Aljuhne

In partial fulfillment of the requirements
for the degree of
Master of Science in Geomatics

1st Examiner: Prof. Dr. Delf Egge
2nd Examiner: Dr. Ralf Krockner

November 2012

Declaration §23(4) PSO

I declare, that this Master Thesis – in the case of group work the respective marked parts of the work – has been completed independently without outside help and only the defined sources and study aids were used.

Literally or copied passages or passages analogous in sense from different text sources are marked by referencing the respective sources.

Hamburg, November 15, 2012

(Ammar Aljuhne)

Table of contents:

Declaration §23(4) PSO	ii
Table of contents:	iv
Abstract	vi
Acknowledgment	viii
Table of figures:	x
List of tables:	xi
Chapter1	1
Introduction	1
1.1 Positioning of georeferenced measurements at sea.....	2
1.2 Artifacts and errors observed in the data	4
1.3 Thesis objectives and methodology	8
1.4 Outline.....	9
Chapter2	10
Description of sensors and data (State of the art)	10
2.1 The description of the situation onboard R/V Polarstern.....	10
2.1.1 Positioning	11
2.1.2 Orientation	13
2.1.3 D-Ship and input data files	15
2.1.4 How the components work and interact.....	17
2.2 Outline.....	19
Chapter3	20
Methods for data analysis	20
3.1 Plausibility and domain tests	23
3.2 Outliers detection and identification	24
3.3 Outlier correction	32
3.4 Master track generation and PANGAEA standards.....	41
3.5 outline	44
Chapter4	45
Software description and implementation	45
4.1 Online mode	49
4.2 Offline mode	50
4.3 Implementation onboard R/V Polarstern	51
4.3 Outline.....	52
Chapter5	53
Results and discussions	53
5.1 Quantitative results of the final product.....	53

5.2 meeting PANGAEA specifications.....	61
5.3 assessment of the performance	63
5.3 Outline.....	64
Chapter6.....	65
Conclusions	65
Bibliography	67
Appendix A.....	69
Contents of the CD.....	69

Abstract

This thesis deals with the navigation tracks of the Research Vessel Polarstern and it provides a software application for statistical analysis of these tracks.

The scope of this program follows two requirements. Firstly the program provides an assessment functionality for post processing of old tracks of R/V Polarstern to include all these tracks in long term storage and archiving system of AWI, the PANGAEA network. Secondly, it provides an online assessment tool to analyze the navigation information onboard the vessel in real time.

The software application has been developed using Borland C++ Builder 6 environment under Windows 7 operating system. The numeric functions of this application are following ANSI standard of C++ programming language, and thus it could be implemented in any developing environments using this standard.

A full description of the navigation systems onboard the R/V Polarstern has been reviewed to investigate how the possible connections and relations between different sensors act, to understand how the cruise track is produced and to find appropriate analysis methods.

In particular, a statistical method is introduced for analyzing the navigational data of the vessel from different sources onboard. This method consists of several tests for detecting and identifying the outliers in the data. It follows a decision based filter that keeps the original data if it successfully passes the outlier tests. In addition, the filter replaces the outliers with appropriate solutions that are calculated using different routines such as transformation from different devices as well as interpolation and extrapolation procedures. This method provokes smoothing of the original data as well.

As results of the online and the offline mode, the final product is a file, called “Master track”, which consists of seven columns. These columns are: the date and time of the records with one second interval, the evaluated position of the vessel given in geographical coordinates, the heading of the ship, the roll and pitch and finally a quality number that indicates the precision of the position. In addition a generalized version of the Master track is provided applying Ramer–Douglas–Peucker algorithm.

At the end of this work some results are presented to show the improvements that have been achieved using the application. A sample Master track and the corresponding generalized track were finally published in PANGAEA.

Zusammenfassung

Die vorliegende Arbeit befasst sich mit den Fahrprofilen (cruise tracks) des Forschungsschiffes Polarstern und liefert ein Software-Programm zur statistischen Analyse dieser Profile.

Mit dieser Software werden zwei Anforderungen abgedeckt. Zum einen liefert die Software Funktionalitäten zur Nachbearbeitung der vorhandenen Fahrprofile von FS Polarstern, um diese im Langzeitarchiv des AWI, dem PANGAEA-Netzwerk, zu hinterlegen. Zum anderen können mit dem Programm die Navigationsdaten direkt an Bord online analysiert werden.

Die Software wurde mit dem Borland C++ Builder 6 auf einem Windows 7 Betriebssystem geschrieben. Die numerischen Funktionen des Programms entsprechen dem ANSI Standard für die Programmiersprache C++, so dass der Quelltext in jeder Entwicklungsumgebung implementiert werden kann, die diesem Standard folgt.

Das Navigationssystem an Bord FS Polarstern wird beschrieben, um zu untersuchen, wie die Sensoren miteinander agieren, um zu verstehen, wie das Fahrprofil generiert wird und um daraus die angemessenen Methoden zur Analyse abzuleiten.

Insbesondere wird eine statistische Methode vorgestellt, die die Navigationsdaten der verschiedenen Sensoren an Bord analysiert. Diese Methode besteht aus mehreren Tests zum Finden und Identifizieren von Ausreißern. Sie fußt auf einem fallbasierten Filter, der die originalen Daten unverändert übernimmt, wenn sie in den Ausreißertests keine Auffälligkeiten zeigen. Hingegen werden Ausreißer durch entsprechend verbesserte Werte ersetzt, die durch Transformation aus anderen Sensoren, sowie durch Interpolation oder Extrapolation berechnet werden. Mit diesen Funktionen geht jeweils eine Glättung einher.

Als Produkt wird im Online- und Offline-Modus eine Datei, nämlich der sogenannte „Master Track“ generiert, der aus folgenden sieben Spalten besteht: Datum/Uhrzeit des Rekords im Sekundenintervall, die evaluierte Position des Schiffes in geographischen Koordinaten, das Heading, der Roll- und der Stampfwinkel des Schiffes und schließlich ein Wert zur Angabe der Positionsgenauigkeit. Zusätzlich wird mit Hilfe des Ramer-Douglas-Peucker Algorithmus eine generalisierte Version des Master Tracks berechnet und zur Verfügung gestellt.

Zum Abschluss der Arbeit werden einige Beispiele präsentiert, die die Verbesserungen in den Positionen verdeutlichen, wenn das Programm zur Anwendung kommt. Beispielhaft wurde schließlich ein Master Track, sowie der zugehörige generalisierte Track in PANGAEA veröffentlicht.

Acknowledgment

Working for my master thesis at Alfred Wegener Institute was a great and wonderful experience for me, and I am indebted to many people for making the time working on my M.Sc. unforgettable.

First of all I would like to express my deep gratitude mixed with feelings of sadness to the person, without whom I would never have been able to complete this achievement. Thank you Professor Böder for all the contributions and supports you have given me during my study in Germany, and may your soul rest in peace.

Furthermore, I owe my deepest appreciation to Prof. Dr. Delf Egge for his support and valuable advice, and for being supportive all the way during the master study.

Moreover, I am very grateful to my supervisor Dr. Ralf Krockner, without whom it would have been impossible to complete this work. Dr. Krockner had supported me with all the time and experience needed in this work regarding the technical and the programming issues, and therefore many thanks to his great support.

I would like also to thank the staff of the Bathymetric group at AWI for the warm working environment during my work in this thesis: Dr.-Ing. H.W. Schenke (for his help and support), Dr. Hannes Grobe and Dr. Rainer Sieger (for their grateful contribution regarding PANGAEA implementation).

Moreover, special thanks to Dr. Saad El Naggar for his great support during my study.

I also appreciate the support of all the staff of HafenCity University for their great support, and special thanks to the coordinator of Geomatics department frau Rosi Garcia for her great help.

Many thanks with love to my family in Syria; parents, sisters and brother for being beside me all the way even during their tough situation. I do not have words to express my gratefulness, and I hope you and all Syrians live in peace and happiness again.

Also big thanks to all friends who supported me during my study. Special thanks to the best friend Alaa Memari for his time and discussions. Also thank you Ammar Naggar for your help.

Last, but not least, I would like to thank my love Layal for her understanding and love during this work. Her support and encouragement were in the end what made this thesis possible.

Table of figures:

Figure 1: The intensity of marine data at AWI which emphasizes the need of an appropriate storage and archiving strategies (Schäfer, 2011).	2
Figure 2 Error in the plotted tracks from different sensors onboard R/V Polarstern	4
Figure 3 Displacement of the distance between the GPS Trimble-1 and the MINS.	5
Figure 4: the correlation between speed and distance displacement.	6
Figure 5 the difference from the built in position of the MINS and the recorded position	6
Figure 6 Jump in the longitude position of the MINS	7
Figure 7: observed error in the positioning of Trimble-1 receiver onboard R/V Polarstern	12
Figure 8 MINS components	13
Figure 9 Ring Laser Gyroscope diagram	14
Figure 10: The configuration of the downloaded input files in D-Ship.	15
Figure 11: Flow chart of the interaction of all systems.	17
Figure 12: Installation of the sensors onboard R/V Polarstern	18
Figure 13 Single and sequence sliding windows used in the filtering algorithm	22
Figure 14: The speed and acceleration tests for the MINS position	25
Figure 15: The shift error type occurs in MINS position.	26
Figure 16: The parameters used in the geodetic calculations (Ghilani, 2005).	28
Figure 17: Angle computations and the thresholds consideration.	28
Figure 18: The computation of the angles B1 and B2 between the heading of the ship and the azimuth from the MINS to the GPS location.	29
Figure 19: The threshold calculation for the angle test of Trimble 2 antenna.	31
Figure 20: single epoch window test for the angle test.	32
Figure 21: Bad Position produced by Trimble-(1-2) GPS antennas.	36
Figure 22: Flow chart shows all the functions used in the algorithms.	38
Figure 23: the simulator of D-Ship server.	40
Figure 24: header information for the Master track file with metadata description.	42
Figure 25: The principle of Ramer–Douglas–Peucker algorithm. (Peucker, 2012).	43
Figure 26: the header file of the Analysis functions class used in the application	46
Figure 27: The header file of the Analysis functions class that shows the different variables used in the software.	48
Figure 28: The online processing window while processing a dataset produced as stream using the D-ship simulator.	50
Figure 29: the offline window while processing an input file.	51
Figure 30: Part of the expedition (ARK-XXV/1) used for results visualizations.	53
Figure 31: The effect of the geometric distance on visualizing the track of the devices.	54
Figure 32: Error in the track produced by MINS.	55
Figure 33: the smoothing routines of small deviations	55
Figure 34: error produced by Trimble antenna and a similar behavior followed by MINS.	56
Figure 35: the correction applied to the MINS by the transformation.	56
Figure 36: Shifted positions in the MINS track.	57
Figure 37: the correction of the shifting positions produced by the algorithms.	58
Figure 38: Missing data generated in the MINS positions.	59
Figure 39: the replacement of the missing data produced by the application.	60
Figure 40: information that is presented in the log file at the end of the process,	61
Figure 41: The implementation of the Master track of the cruise (ANT-XXVI/4) into	

PANGAEA.	62
Figure 42: The generalized track of the cruise (ANT-XXVI/4) and its implementation in PANGAEA.	63

List of tables:

Table 1: The navigation systems and GPS receivers onboard R/V Polarstern	11
Table 2: The performance of the MINS as defined by the manufacturer.	14
Table 3: The parameters in PODAS file that have been used in the software	16
Table 4: The average positions that had been used in the computation of the angles for the angle test.	29
Table 5: The lever arms of different devices in the ship's reference system	30
Table 6: the header of the columns in the Master track file.....	41
Table 7: the classes that have been used in the software.	45

Chapter1

Introduction

Geographical Information Systems are widely used worldwide to visualize various kinds of spatial related data. This concept has a variety of applications, and day after day the techniques used in such field are becoming more advanced and sophisticated. Moreover, the scientific research at sea has always high interests towards including the scientific measurements in GIS environments. In addition, the information about the navigation tracks of scientific expeditions is drawing more attentions in the GIS area for its importance.

This information takes part in most of the scientific measurements at sea where nearly every research needs to allocates the place and time, where and when the measurement had been taken, and the navigation information that we are describing is simply not more than this information about the position and the time of the ship during the research cruise. For the fact that this information is shared by most researches at sea, including this information in GIS environments is a high recommended goal that will give any user the possibility of having the cruise track information at any time he needs it.

Eventually this objective comes with different tasks that need to be achieved in order to reach this goal. The evaluation of this information in term of errors is necessary to have good elements to share, and in order to give different users with different interest of the accuracy of the ship the necessary information regarding this manner.

We should also emphasize that many standards have been established to provide standardized metadata elements that help in documentation, contribution to catalogues, finding existed datasets and understanding the contents of datasets within an organization. In the US for example, they follow the FGDC standard “Federal Geographic Data Committee” to develop procedures that assist in the implementation and distribution of national digital geospatial data. Also, the ISO 19115 is another base standard for the description of geographical information systems that are used in Europe (FGDC, 2011).

1.1 Positioning of georeferenced measurements at sea

Worldwide there are extremely tremendous amounts of marine environmental data from different sources that are represented or could be represented in GIS environments. The sources of the data vary from ship based instruments, sensor networks, water column and seafloor or air based instruments. Figure 1 shows the intensity of the marine data gathered by different research vessels controlled by Alfred Wegener Institute for Polar and Marine research (AWI), and this amount of data needs an appropriate storage strategy with standardized access to fully capture the basic characteristics of data and its resources.

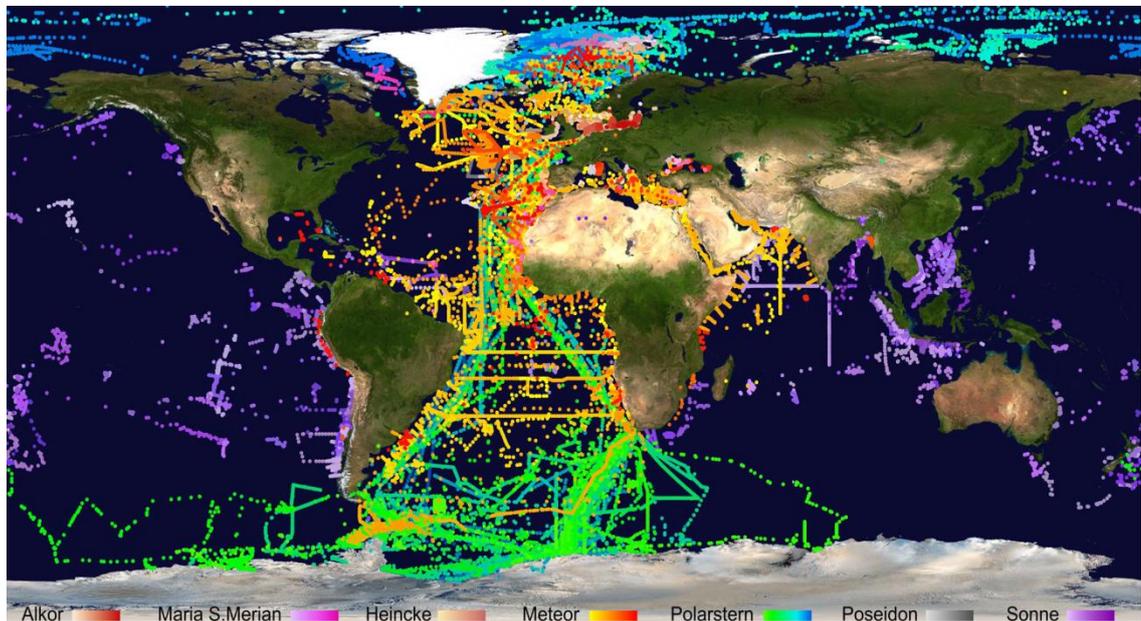


Figure 1: The intensity of marine data at AWI which emphasizes the need of an appropriate storage and archiving strategies (Schäfer, 2011).

The AWI is providing the data collections of their expeditions through GIS environments. One of these environments used is an open access library for archiving, publishing and distributing georeferenced data collections called PANAGEA which stands for Publishing Network for Geoscientific and Environmental Data. The main purpose of this archiving system is to guarantee long term storage availability (www.pangaea.de, 1998). In addition to PANGAEA, another system is the project MaNIDA “Marine Network for Integrated Data Access” which has the aim to sufficiently provide quality-controlled marine data at national and international levels and education. This system is used as a parallel portal for providing better investigation services for the marine data, and better handling services like the downloading features provided.

As shown in the previous figure, varieties of scientific measurements are gathered during the expeditions of marine research vessels. Some of them are gathered in stationary work which needs the ship to be maneuvering in one position, while other measurements are collected when the ship is sailing or underway. In PANGAEA, for example, a variety of datasets are included in the system from the expeditions of the research vessels controlled by AWI. These datasets include for example the sediment cores, cruise reports and data, Oceanographic observations, metrological measurements, etc.

R/V Polarstern, a research vessel operated by AWI, is equipped for biological, geological, geophysical, glaciological, chemical, oceanographic and meteorological research, and contains nine research laboratories (AWI, 1980). Some of stationary measurements onboard R/V Polarstern are the CTD (stands for conductivity, temperature and depth) and the geological sampling. CTD measurement measures the characteristics of the water columns like temperature, salinity and pressure. This is done by a vertical profiler that could be deployed until 6000 m, while Geological sampling provides sediment sampling from seabed up to 10000 m depth.

In other hand there are underway measurements when the ship is navigating through its navigation track. VaMoS II radar, for example, is a device for wave and current monitoring system that measures and displays all the essential wave field parameters such as significant wave height, periods and directions as well as surface current speed and direction in real time. The Acoustic Doppler Current Profiler (ADCP) also provides three-dimensional current vectors while the ship is moving and to a depth of 150 m (El Nagggar, 2007).

Each measurement could be provided in a GIS environment to share its characteristics and values, and this sharing process is mandatory in the scientific work especially if the sharing procedures are controlled and structured in a specific standard as we figured out earlier. All examples mentioned above about the stationary measurements and the underway measurements onboard R/V Polarstern, show us different characteristics of different measurements. Nevertheless a common parameter that usually takes part as an attribution in most scientific measurements at sea is the information about when and where the measurements have been recorded. In other words, the time and the position of the measurement are common features for the vast majority of the scientific records at sea.

Of course when we are dealing with scientific research the more accurate our data is, the better results we achieve, but this wish will acquire more efforts, technologies and costs from the producers of the desired results. Additionally, let us take the navigation information of the sea measurements (time and position): The accuracy of the position plays an important role for the underway measurements and the more accuracy we can achieve in the absolute position the more valuable our results would be. But it is not that important to achieve a high accuracy when it is related to stationary measurements like the geological sampling and the CTD measurements, and the uncertainty of the position would not affect the measurements whether it is in centimeters or 10 meters scale. This means that the accuracy demands of the position vary in scientific measurements at sea, and they depend on the characteristic of the measurement itself. Nevertheless providing a quality indication of the position would serve different scientists in different fields.

1.2 Artifacts and errors observed in the data

The AWI has the plan to localize all scientific measurements in the information system PANGAEA. This system will insure long term storage of these data as we said earlier. Thus it follows a fully standardized data model in its storage procedures of archiving datasets. These procedures contain different levels, tables, references, items, parameters and methods for archiving a dataset. When archiving a dataset in this system, the combination of the data and parameter tables is an essential part (Grobe, 2006).

The navigation information from all previous cruises of the research vessel R/V Polarstern will be implemented in PANGAEA. In pre-investigations of the tracks of old expeditions some outliers have been recognized with the vessel track during its cruise. Therefor this will need an evaluation of this information, and needs to analyze it to detect and correct any possible outliers associated with it and to provide it to be compatible with standards of metadata environments. Figure 2 shows a big distance between the plotted tracks of the ship from different sources onboard.

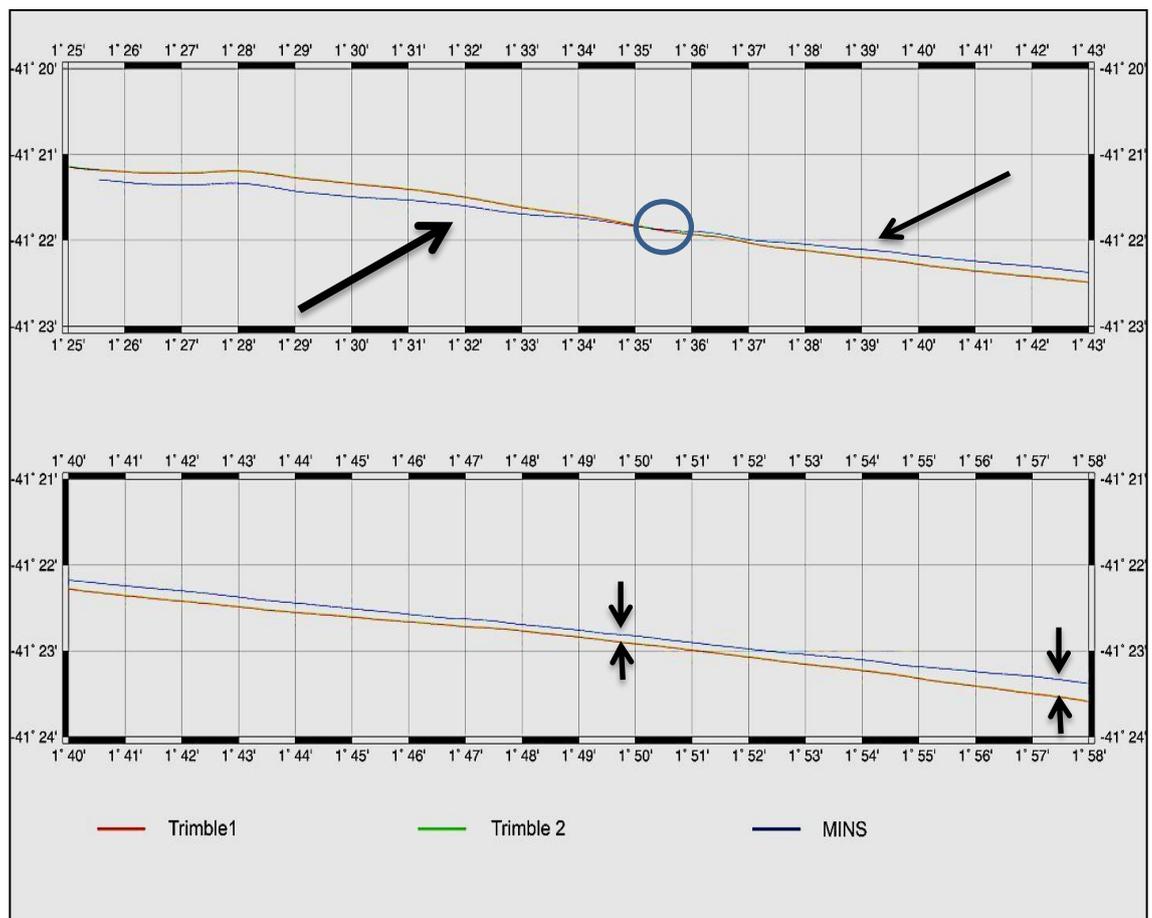


Figure 2 Error in the plotted tracks from different sensors onboard R/V Polarstern

The sources of the plotted tracks shown in the figure are the positions taken from two GPS receivers, (Trimble-1) in red and (Trimble-2) in green, and the third one in blue color is taken from the marine inertial navigation system (MINS). In fact, the ideal results would be to have a zero difference in the distance between the sensors after subtracting the geometric distances, but clearly it is not the case here. The maximum geometric distance between the sensors is around 32 meters between the MINS and the first GPS antenna, but as shown in the previous figure where the small black arrows are,

an offset up to 150 meters exists. In other hand, the plotted track of the MINS is behaving irregularly, and it crosses from side to another as shown in the upper part of the figure, and it is clear that an error is associated with one of the devices for such behavior to appear.

Moreover, Figure 3 proves what is shown in the previous plot, and it shows the variance in the distances between the MINS and the Trimble-1 GPS receiver after subtracting the geometric distance. This variation is a clear indication that there is an error associated with the sensors and should be identified and corrected.

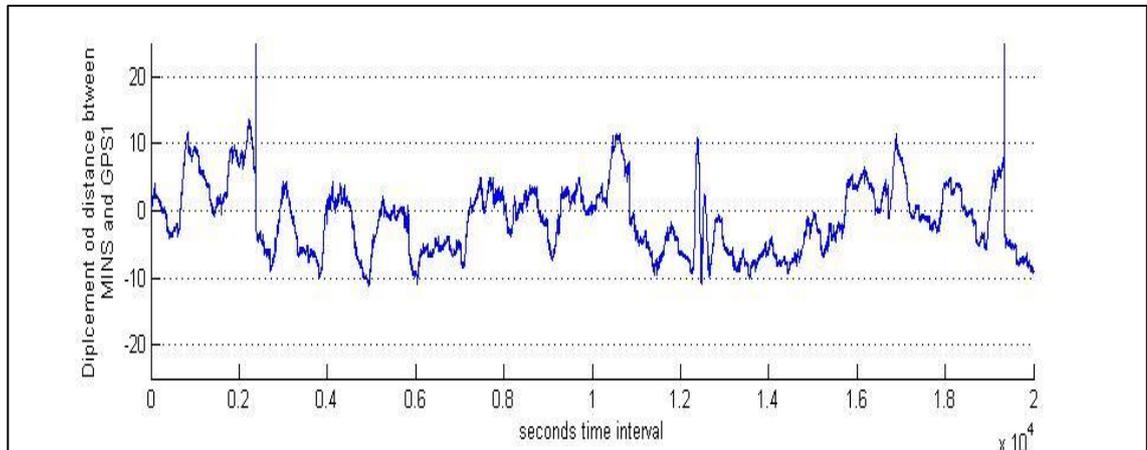


Figure 3 Displacement of the distance between the GPS Trimpl-1 and the MINS.

However, finding any kind of correlation between different parameters and the erroneous behavior show previously would give hints of the possible causes of such errors to exist. Another investigation has shown that there is no correlation between the displacements of the distance between the MINS and the GPS antennas and other parameters like the roll, pitch, heading or the number of satellites in view. In other hand an interesting correlation has been observed between the displacements and the speed of the vessel. The displacements between the MINS and the GPS receiver Trimble1 as well as between the MINS and Trimble2 have shown big values (about 2.5 meters) in the range between 4.3 to 11 knots, while the displacements in the distance between the both GPS receivers were much smaller, and this indicates that the MINS is producing as error (Bumke, 2011). Figure 4 clarifies this situation.

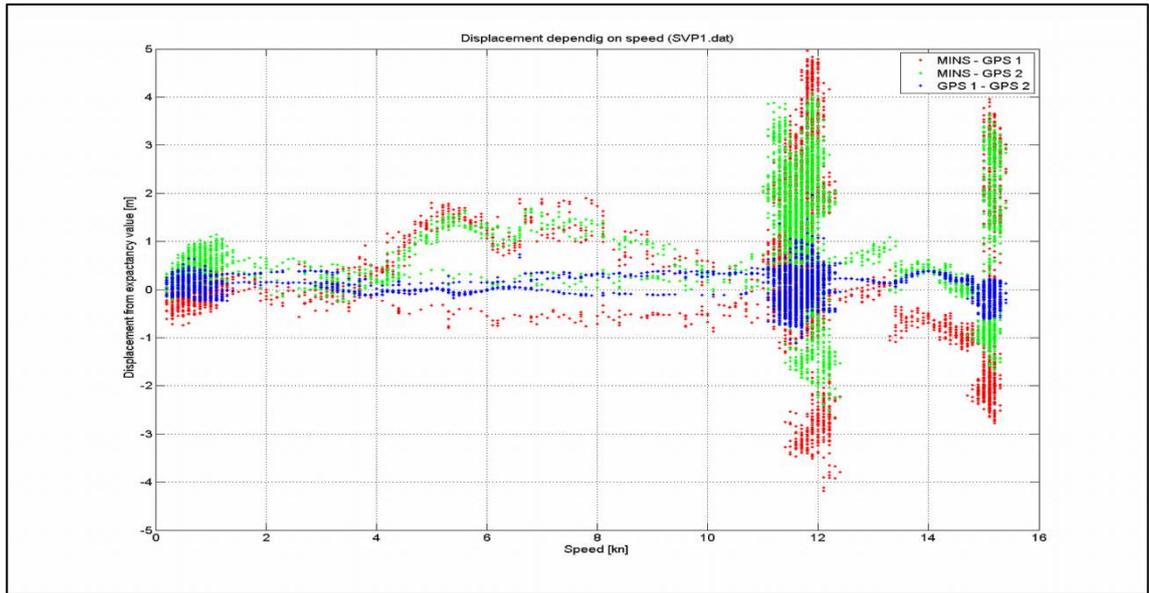


Figure 4: the correlation between speed and distance displacement.

The green and the red colors present the big displacements between the MINS and both GPS receivers, while it is clear that the displacements in the distance between the GPS receivers (presented in blue) are much less correlated with the ships speed.

From the previous figures it is clear that the MINS is producing some errors as the deviation of displacement of the distance between the MINS and the GPS antennas is higher than the deviation occurs in the displacement of the distance between the GPS antennas.

Furthermore, another behavior has been observed where the correlation with the ships speed is also distinguished. In Figure 5, a plot has been made between different speed of the ship and the difference between the calculated positions and the actual built in position of the MINS, and as we could see the error is increasing when the ship is speeding up (Schenke, 2006).

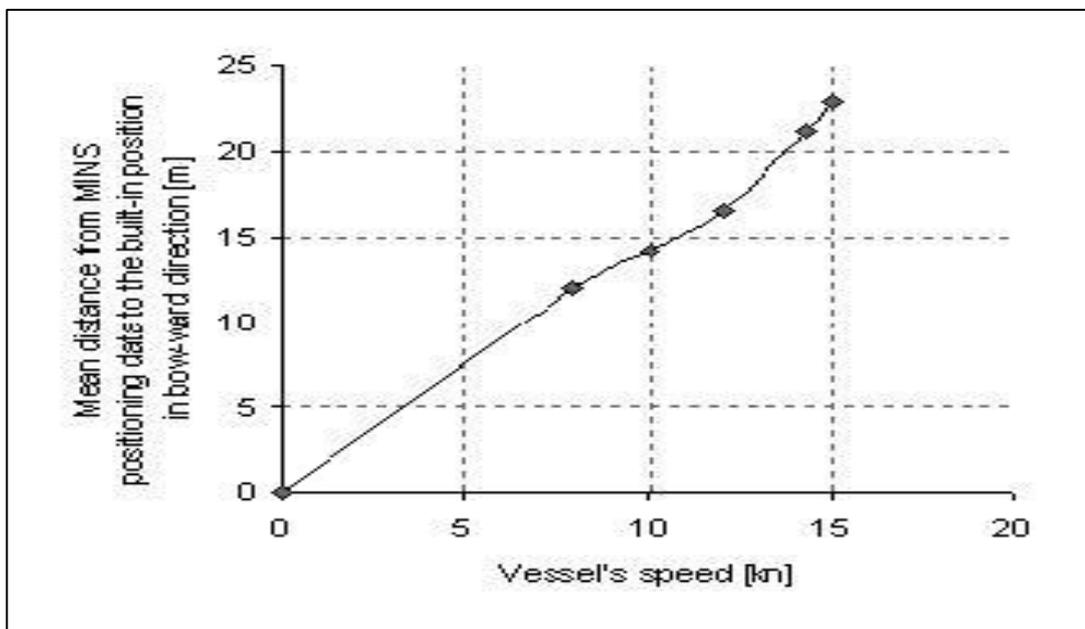


Figure 5 the difference from the built in position of the MINS and the recorded position

This difference between the calculated position and the actual position of the MINS is increasing toward the bow of the ship, and at maximum speed of 15 knots, this error is up to 23 meters toward the bow. This behavior had been observed in most of the positions produced by the MINS, and this indicates a systematic characteristic of this kind of error.

An interesting situation had been noticed in the MINS calculated positions when the ship is sailing northward or southward on meridians that are multiple of 6° . The longitude position jumps around 6 degrees to the east or west directions when the ship is sailing close to those meridians. Figure 6 shows the unusual jump in the MINS longitude when it was heading north and south close to the meridian -66° .

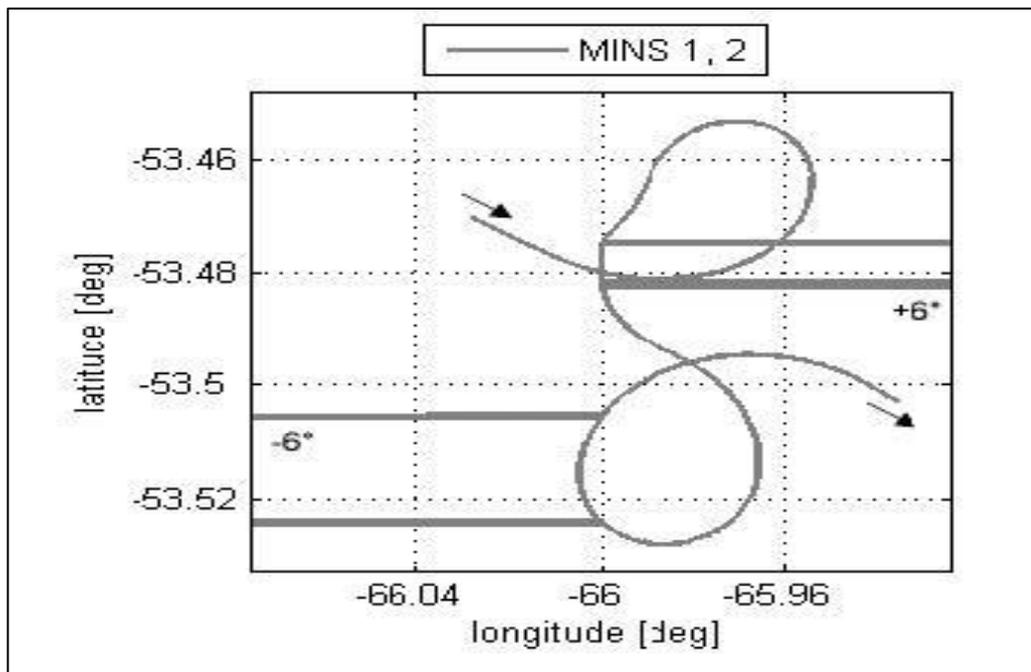


Figure 6 Jump in the longitude position of the MINS

Reviewing these errors and the interesting errors recognized in previous investigations give us a good reason for providing adequate analysis of the navigation information and for assessing it before implementing it in the GIS environment PANGAEA.

1.3 Thesis objectives and methodology

This thesis primarily demonstrates the need for fault detection, and exemplifies the need for proper design of a navigation filter, which tests the information provided by different sensors and defines the outliers which occur in the navigation track of the research vessel R/V Polarstern. A statistical analysis of the ship's track will be demonstrated using many parameters and measurements from different sensors in order to detect the outliers in the data.

The final product of this work will be a software application that provides evaluated, assessed and continuous ship positions with date and time stamp as well as a quality indication about the goodness of the position. The quality information will give any user the reliability of the position at any time of the expedition. Moreover the information about the heading of the vessel as well as the roll and pitch values will be provided each second as well.

From now on the name Master track will refer to the final output file provided by the software application. The software will provide output files that will be compatible with PANGAEA specifications for archiving and storing the Master track.

The master track will be obtained mainly from the MINS system after applying many methods and tests on its products. These tests are based on statistical approaches, and it is built to compare different values from different sensors, in order to detect the outliers in the navigation track. We should emphasize that this thesis will focus on the outliers and the rough artifacts in the navigation track. This means that the methods do not enhance the accuracy of any individual sensor, and it will not focus on the detection of neither systematic nor random errors in the measurements, but it will rather implement some comparison and statistical assessments in order to define and fix the rough errors in the data.

Finally the software will have the possibility to work in an offline mode for post processing the old expeditions, and also it will be implemented in an online mode for the assessment of the real time data onboard R/V Polarstern. The efficiency of the software will be that it serves the general aim of Alfred Wegener Institute, by providing a useful tool to assess the old navigation tracks and produce the Master track that complies with the PANGAEA project. Furthermore, the online assessment is a great tool for saving time and efforts of the post processing in the future.

1.4 Outline

In the first chapter we have described the efficiency of the geographical information system in the marine research field, and we have seen the importance of following standards of archiving scientific measurements. Moreover we have described the plan of Alfred Wegener Institute to localize its navigation tracks in PANGAEA and the need to evaluate these tracks. Also we have seen some of the associated outliers in the navigation tracks of the research vessel R/V Polarstern, and we have introduced the general objectives of this thesis to develop assessment software to provide the evaluation of the navigation tracks of R/V Polarstern, in offline and online mode, and to provide a corrected track as well.

In the next chapter we will clarify the situation onboard R/V R/V Polarstern explicitly and the responsible sensors will be explained in terms of possible error sources, and also the inputs of our software for both online and offline modes will be described.

Chapter2

Description of sensors and data (State of the art)

2.1 The description of the situation onboard R/V Polarstern

Detecting and fixing faults in any system have a very broad range in the scientific world, and achieving a specific goal regarding this approach requires a full understanding of the components of the system as well as the interactions between these components. As the task of this thesis is to evaluate the navigation track of R/V Polarstern, and detect the outliers in the old data, we have to know how the navigation system of the ship is working and what the error budgets of the components are.

At first we have to understand the different kinds of errors that could occur in a scientific measurement. In general we can distinguish between three types of errors: Systematic errors, Random errors and Gross errors. In the first type the error follows certain physical or mathematical rules and can be caused by the instrument, the measurement's environment or a human factor. The difference with the random error is that the second type does not happen in a systematic way, but it rather occur randomly, and could be due to the instrument or the measurement routines. The gross error could happen due to human mistakes, wrong measurement methods or blunders and malfunctioning in the used instrument and this type does not follow certain rules as well (Fan, 2010). And as mentioned earlier, our efforts will be focused on the rough outliers in the data, and this thesis will not cover the detection of systematic and random errors in the navigation information.

2.1.1 Positioning

The R/V Polarstern is equipped with many different navigation systems for different purposes. Table 1 shows these systems in more details.

Table 1: The navigation systems and GPS receivers onboard R/V Polarstern

	Modell description	manufacturer	Number	frequency	Integrated filtered position
GPS receivers	Z12	Ashtech	1	Dual	---
	MX-400	Leica	1	Single	---
	MS750	Trimble	2	Dual	---
	MX521	Leica	2	Single	---
Navigation distribution systems	MINS	Raytheon-Anschütz	2	---	Yes
	NACOS	STN-Atlas	2	---	Yes
	PODAS	Werum	1	---	---

Different GPS and navigation systems are used for different purposes onboard R/V Polarstern. The Ashtech receiver is used for time synchronization of systems onboard, while the Leica receiver works with differential mode by receiving correction signals from reference stations controlled by Sky-fix-decoder system. In addition, MX521 receivers are not in use and they only provide a backup for the MINS system. The reminder two Trimble receivers provide precise positioning with dual frequency with accuracy up to ± 10 -50 meters (Iffland, 2004).

However, we know that biases and errors affect all GPS measurements: pseudo-range, carrier phase or Doppler, and their combined magnitude affect the positioning result. These Biases are the systematic errors that produce the deviation between the true measurements and observed measurements. The sources of biases may have physical bases, such as the atmosphere, but may also enter at the data processing stage.

Using a differential GPS is not appropriate for R/V Polarstern where the expeditions take place in the Polar Regions, and there are limit chances for receiving corrections from reference stations. Thus the usual routine to obtain the position is done by using the two Trimble receivers that provide a precise point positioning (PPP) with dual frequency. We do not forget that R/V Polarstern is on service since decades and the positioning accuracy has been improved with time. The DGPS system has been stopped working due to the fact that the dual frequency PPP service provided by the Trimble receivers is reaching accuracy within ± 10 meters. Figure 7 shows an example about the accuracy of the GPS receiver (Trimble-1). This scatter plot represents the ship's positions when it was birthing in the port. Each color represents the recorded positions

in 30 minutes. The latitude/longitude grid lines are shown and the area of each square is equal to about 4.4 meters. It is shown that after five hours of measurements there were small shifts of the positions within the 30 minutes, besides a jumping error every 30 minutes intervals.

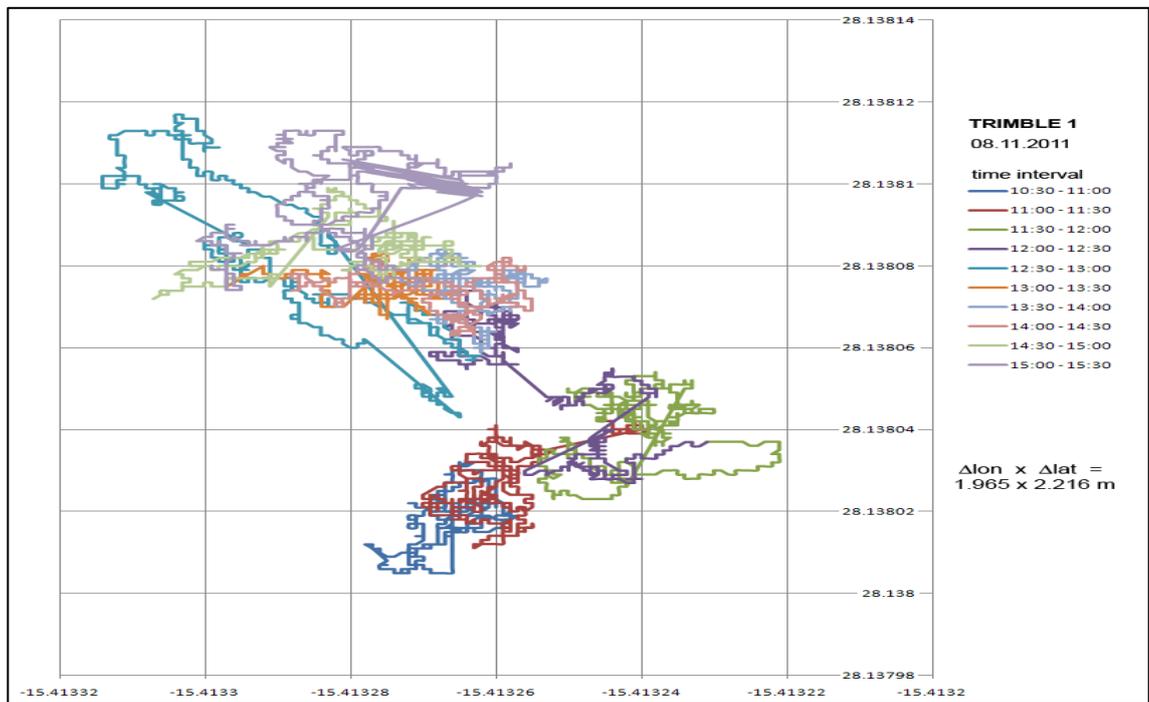


Figure 7: observed error in the positioning of Trimble-1 receiver onboard R/V Polarstern

Onboard the ship there are two MINS systems (MINS-1 and MINS-2) and they are integrated with the GPS receivers (TRIMBLE-1 and TRIMBLE-2). The MINS platform supplies the scientific groups with all navigation data like the position, the speed and the attitude of the ship with its components roll, pitch and heading (El Naggar, 2007).

After the shut-down of the Selective Availability by the US government, the Trimble-1 GPS receiver and the MINS-1 are assigned as permanent position sensors, and the other Trimble receiver and the MINS-2 are backup sensors. However, switching between these systems is available from the bridge of the ship and an automatic mode is also available. The data gathered with the Trimble receivers will be the base of our algorithms in order to detect the outliers in the navigation data as well as generate the Master track.

However, the navigation system NACOS-55-3 has been developed by STN-Atlas GmbH with a current accuracy ± 5 meters. It is supplied with navigation data by the laser based Marine Inertial Navigation System “MINS”. This system is mainly used for supplementing the positions of the ship used by the navigation crew, and it is not part of the scientific measurement, and this separation had been put following some regulations for safety considerations.

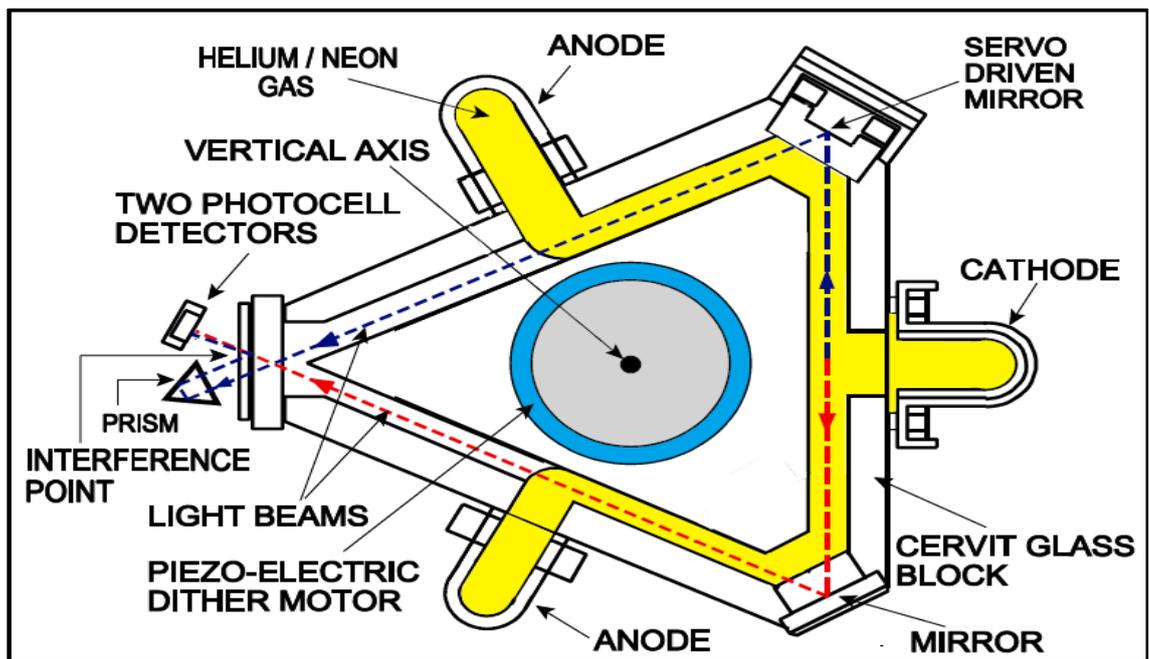


Figure 9 Ring Laser Gyroscope diagram

The Marine initial can indeed determine angles, angular velocities and accelerations very accurately, but it has no absolute positions. The MINS therefore calculates from the original location and position of a system, the time and the acceleration acting on the current position and location. The accuracy of the system is listed in the next table.

Table 2: The performance of the MINS as defined by the manufacturer.

Sensor	Error
Heading	< 3 arc min sec
Roll	< 1.4 arc min
Pitch	< 1.4 arc min
Position	< 0.1 nm (with SPS GPS)
Angular rate	< 0.046 °/s

With this level of accuracy the MINS provides the ship with position where relative accuracy in the centimeter level is approached. This is due to the fact that inside its ICU many filtering algorithms (mainly Kalman Filter) are applied to control the raw data and to implement smoothing techniques before the final navigation data are distributed.

2.1.3 D-Ship and input data files

After a short description about the various sensors and systems onboard, we have seen that the permanent system to provide all users and scientists with navigation data is the integrated MINS system. Now we have to know how the data are distributed, archived and provided.

Onboard R/V Polarstern, data of a number of measuring systems and sensors, including the navigation data are stored in the DSHIP data acquisition system. This system which is built by Werum GmbH provides high technologies for recording, processing, visualizing, distributing and archiving marine scientific data. Recording data is available from different instruments via standard interfaces such as NMEA0183/2000, IEEE 488 and via networks as well. The navigation data are sampled every second as non-validated raw-data in physical units. Moreover, scientists have direct access to the raw-data archive of DSHIP. In many cases, this service - offered by DSHIP - already satisfies the users, who are interested only in certain data of a distinct voyage.

Downloading data had been done via the internet using the data retrieving utilities. A pre-defined format was chosen to download the data. The template is called PODAS, and this provides a text file with 21 columns representing 21 channels.

Figure 10 shows the way to download the data from the D-Ship web page. A good feature is the possibility to choose the beginning and ending of the records which gives flexibility in data handling. Also the time interval is 1 second with date and time stamp for each line, where each line represents a complete record at a time.

DAVIS | SHIP

**List of
DAVIS-Ship
- FS Polarstern -
parameters**

Export V3.0

Inpitted information about the extraction file:

Time/date of export		Result	
Start date/time	18.10.2005 00:00:00	File name	051018_podas
End date/time	18.10.2005 23:59:59	User name	aaljuhne
Duration [s]	86399	Mail address	aaljuhne@awi.de
Interval	1 [s]	Get mail	<input type="checkbox"/>
Duration of specified Extraction: (ca.) 3 h			
File format		Error/invalid value pattern	
Separator	Blank	Error value numeric	#
End of record marker	New line	Error value alphanumeric	Blank
Date / time format	YYYY/MM/DD*HH:mm:ss	Error value position	#
* = Separator		Include invalid values	<input type="checkbox"/>
Decimal symbol	.	Skip invalid lines	<input type="checkbox"/>
Header row	<input type="checkbox"/>	Fit to format	<input checked="" type="checkbox"/>
Max. data file size	0 MByte <input checked="" type="checkbox"/> unlimited	Save order	<input type="checkbox"/> as SavePodasNewBathyF

Figure 10: The configuration of the downloaded input files in D-Ship.

Within the download facility a complete cruise could be downloaded per file, and the assessment will produce a Master track of the complete cruise in one file as well.

The important parameters that have been used for the analysis of the navigation track are listed in Table 3, where other parameters were not important for our analysis but they could be used for further development of the software.

Table 3: The parameters in PODAS file that have been used in the software

PARAMETER	EXAMPLE
Date	2011/07/01
Time	00:00:00
MINS Latitude	78.830510
MINS Longitude	-1.110668
MINS Heading	198.8
MINS Roll	-0.1
MINS Pitch	0.6
Trimble-1 Latitude	78.830332
Trimble-1 Longitude	-1.111247
Trimble-2 Latitude	78.830333
Trimble-2 Longitude	-1.110372

However, we should emphasize that the focus was on the cruises that had used the Trimble/MINS system as a navigation system. In older cruises different systems had been used for providing the navigation information of the ship, and therefore the implementation of the software should be configured with those systems in order to have the desired results.

For the online implementation our input data will be created as an NMEA stream and the needed parameters will be grabbed from the network via TCP/IP protocol. Therefore the same algorithms could be applied in both the offline and the online implementation after some appropriate configurations. Connection with the D-Ship server is normally done via server/client network where D-Ship is a client that receives the required information from the user (the server in this case) and sends the desired information back to the user.

Up to here we see that with this system our offline and online approaches are possible because we can download the archived data from previous expeditions and with the possibility to choose specific templates as well to use it for our analysis. In parallel the implementation of the software for online analysis is also possible via various connection possibilities with the server onboard.

2.1.4 How the components work and interact

Up to this end we described the sensors involved in our analysis and generation of the Master track of R/V Polarstern, and we have determined the important parameters in order to approach our goal. Before going through the methods and the algorithms that have been applied in this thesis, it is good to understand the functionality and the process of the navigation track that we are going to evaluate and assess in order to produce the final Master track for the vessel. The complete process and interaction between the sensors and the system can be clarified in Figure 11.

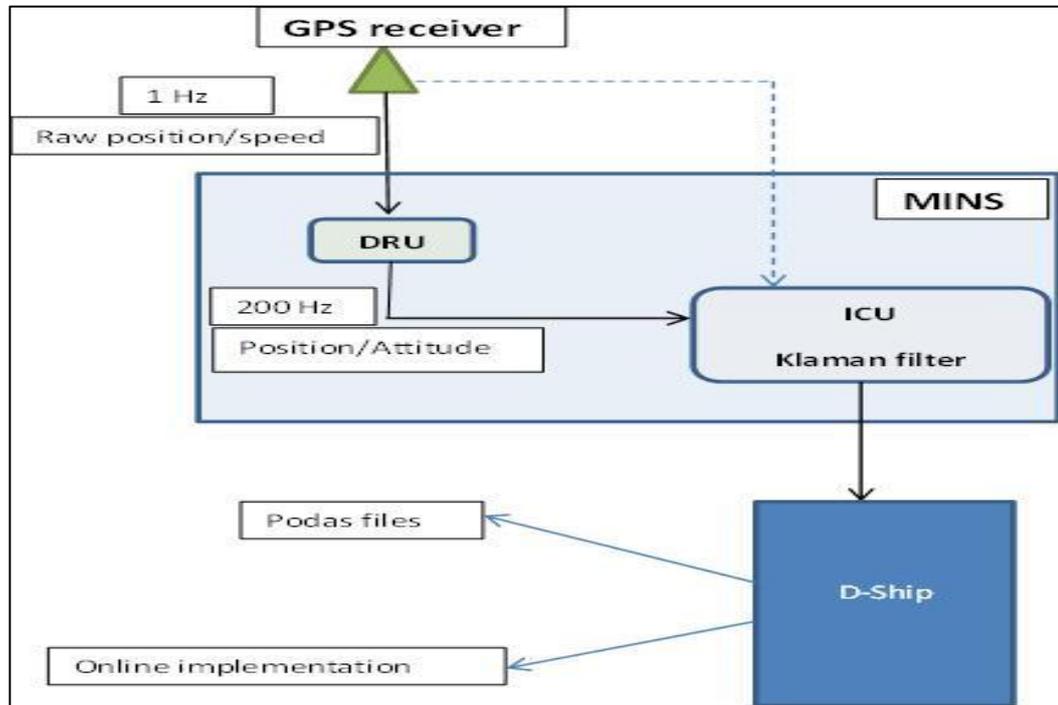


Figure 11: Flow chart of the interaction of all systems.

The position is calculated by the GPS receiver with a frequency of 1Hz (one second) and it is transformed to the MINS position. Inside the DRU unit, the updated position is calculated with a higher frequency (200 Hz) and this updated dead reckoning positions are processed in the ICU unit where Kalman filtering techniques are applied to have a smoothed results. Furthermore, the corrected and updated positions are sampled in D-Ship and therefore the sampling interval is 1 seconds for either the offline and the online mode, and the final product of the software will be with the same sampling interval as well.

The flowchart strengthens the idea that the filtering techniques implemented in the MINS are the source of the errors produced in the data, because the bridging possibility shows that the error is done after the implementation of the filtering techniques by the MINS.

Figure 12 shows the local ship frame and the locations of the sensors that will be involved in our assessments. The MINS is located very close to the center of gravity. The MINS position will be the reference from which the positions of the Master track will be centered to. The local reference system has been defined in the alignment survey of R/V Polarstern that have been accomplished by “OVERATH & SAND SHIP SURVEYORS” (SURVEYORS, 2010).

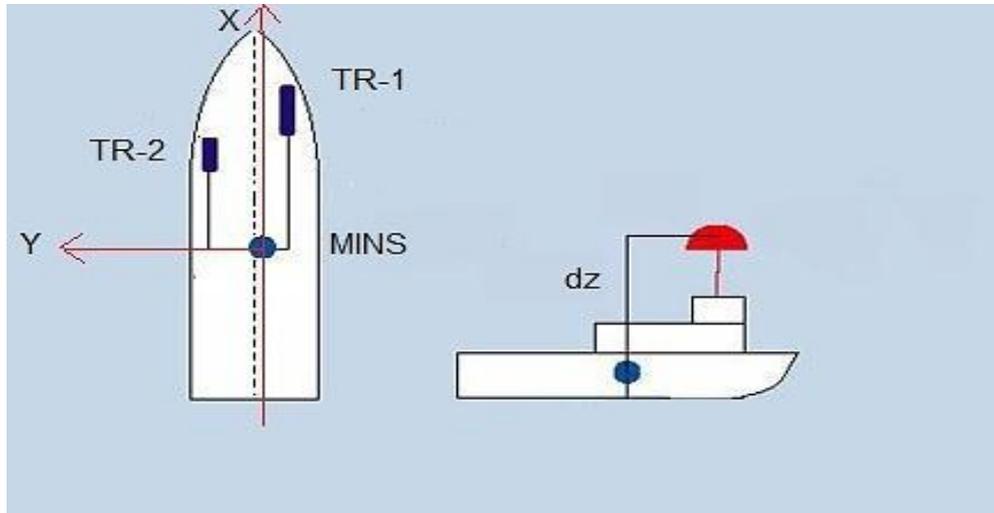


Figure 12: Installation of the sensors onboard R/V Polarstern

There are different local systems used onboard R/V Polarstern and the right hand side system shown in the previous figure is defined in the alignment of the ship, but a left hand system is used also in this work when we talk about the lever arms in the next chapter.

The D-Ship system provides only two dimensional positions with latitude and longitude projected in the World Geographical System WGS84. However the depth information is also available but it is not allowed for public users to use it because of some regulations that strict the accessibility to these data in the economic zones. For this reason the assessment of the navigation information will produce a 2-D position as well.

2.2 Outline

In this chapter the main navigation systems installed onboard R/V Polarstern had been introduced with the descriptions of their main functionalities. Moreover the MINS system was also described. we have also described all the sensors and systems that involve in the distribution and provision of the Navigation information, and a flow chart has introduced the interaction between all these components. Moreover we have introduced the local frame of the ship and the location of different sensors onboard.

In the next chapter the algorithms for outlier detection and correction will be described in details and how the offline and the online mode of the software work. We will also see the different characteristics of the visible control of the program.

Chapter3

Methods for data analysis

Before going through any kinds of techniques used in data analysis, it is useful to provide some information about the raw materials that we are dealing with. This includes:

- 1- Information about the data itself (the archived files-the NMEA stream from the network onboard R/V Polarstern) and this has been covered in the previous chapter. (See Figure 10).
- 2- Information about how the navigation data are produced and distributed and how different sensors interact during these processes. We have also provided this information with details in the previous chapters (see Figure 11).
- 3- Information about errors and examples of outlying data associated with the measurements and the possible correlations between different variables that could help as to achieve our objectives and this have been introduced in the first chapter.

After defining this information, the next step will be easier to achieve; that is to use some analysis techniques to interpret the data based on our information and to produce the desired results. Thus, the analysis is not only picking up an appropriate technique, but the pre-knowledge should be provided and judgments should be taken when interpreting the data based on the pre-knowledge that we have.

As we have seen in the previous chapter, six parameters will be gathered from the server and interpreted in our software either second by second in a real-time process or in post processing dealing with files. The NMEA stream will be produced to contain these parameters:

- The position from the GPS Trimble receivers (Trimble-1 and Trimble-2).
- The position from MINS.
- The attitude parameters from the MINS (heading, roll and pitch).

Our objective is to produce the Master track that consists of a single evaluated position of the ship every second. This track had been chosen to be centered in the MINS location for two reasons. First, the MINS location is nearly at the center of gravity of the ship, which is the most stable point regarding the motion of the ship, and this in parallel is the best representative point for the Master track. Second, the MINS implements very effective techniques for filtering the raw data calculated by the GPS receivers (Kalman Filtering) and the integration of MINS/GPS enhances the relative accuracy of the position. Therefore, having a preprocessed data in hands is much better for having smoothed Master track than dealing with raw data.

Because of this fact the analysis of the acquisition of R/V Polarstern will be mainly based on analyzing the MINS position, identifying the outliers associated with it, interpreting the outlying data and finally providing these evaluated data as a Master Track for the ship.

During this analysis, the algorithms will also use other measurements from GPS receivers in many comparison steps, therefore the analysis will be extended to include all the other variables used in the algorithm. This is also due to the fact that the final product will not only have the assessed position, but also other parameters such as the motion parameters as well as the quality indication.

In many data analysis tasks where data have been recorded and sampled, detecting the outliers and interpreting them is an important challenge as these outliers could significantly affect the results of any scientific measurements. The methods for outlier detection could be divided into two major parts: the univariate methods and the multivariate methods. The former methods deal with the outliers that occur within a single independent variable, while the latter interacts with different dependent variables in the data.

In a statistical point of view, we may consider that all the measurements and variables that are deployed in the dynamic system of the ship could be interpreted independently as individual variables using univariate methods for assessment. For example the roll of the ship could be assessed and evaluated as a single individual variable varying with time. Later on, the detection would be based on comparison procedures in order to judge if the measurement is an outlier or not.

The algorithm is using a decision based filtering technique which means that if the tests consider a value as an outlier it will be replaced with another appropriate value, otherwise it is kept in the data. This is very useful filtering technique in our case because we want to preserve the position if it is not an outlier and we also concern on replacing a “bad” position with another good one instead of deleting it. This will keep a continuous acquisition with a position every second.

Hence the algorithm is providing a detection method as well as a replacing method the implementation of the algorithm requires a specification of the following attributes:

- 1- Startup stage.
- 2- Pre-handling of the inputs.
- 3- Threshold selection for detecting different outlier types.
- 4- The replacement method for replacing the detected outlier with a corrected one.
- 5- Further smoothing steps.

We will change the order of the explanations of these attributes for adequate understanding. Moreover, these attributes have been achieved based on two kinds of assessments techniques. These techniques are:

- The single epoch sliding window.
- The sequence data sliding window.

Within the single epoch window tests, only single input is required for achieving the desired result of this test, while in the other hand some tests need sequences of data inputs to perform adequately, and this is achieved with the sequence data sliding window.

Figure 13 illustrates the simple idea behind the sliding windows for one variable during the analysis. When a single input data is transferred from the network to the software, it is first examined in a single process (single epoch sliding window indicated by the solid boxes) to do some preparations that are needed for further treatments.

Afterwards, the value is stored together with the resulted information from this test, and will wait for series of values to be examined. In further step other tests will be applied on this sequence of preprocessed inputs for further analysis.

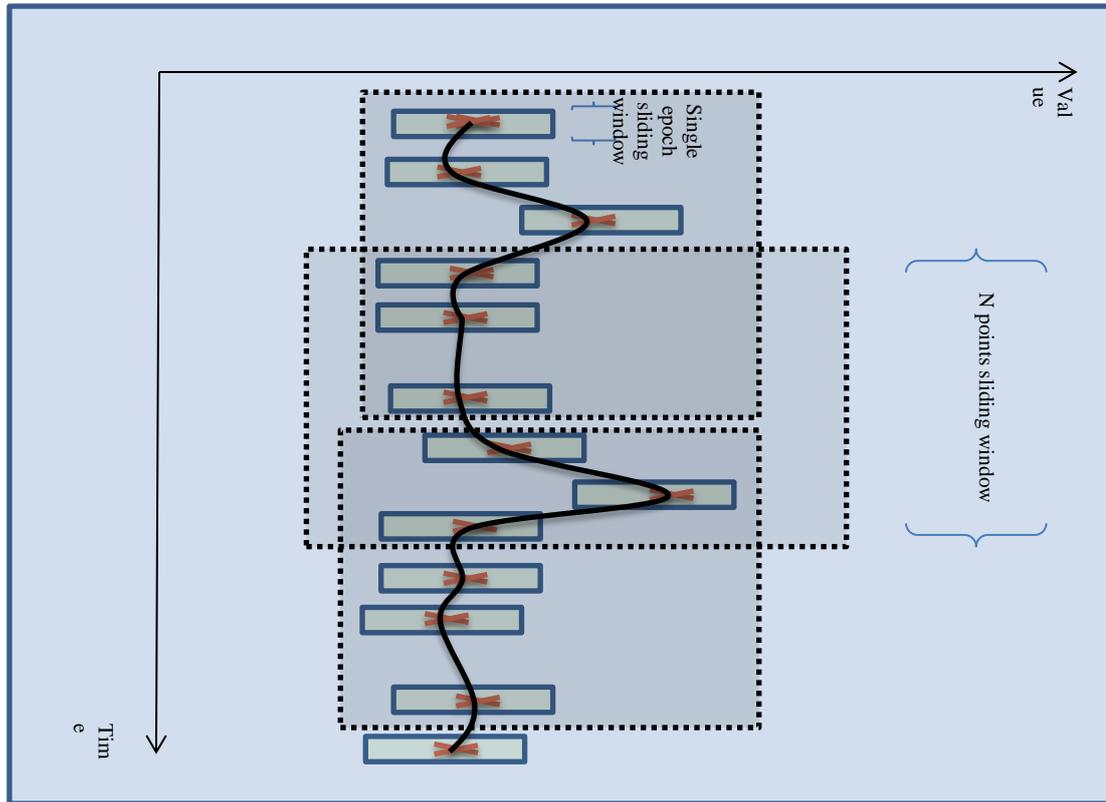


Figure 13 Single and sequence sliding windows used in the filtering algorithm

An overloading technique had been chosen for the sequence sliding window tests (presented as dashed boxes) in order to have sufficient analysis. This means that when a sequence is processed, the next one will consist of inputs from the previous sequence and new inputs as well. The shift of the window had been chosen with one element shifting. So if a first sequence of data consists of (N) number of points from $\{P_{k-6}, P_{k-5}, \dots, P_k\}$ where (P) is the measured point and (k) is the index of the point, the next sequence window will consist of $\{P_{k-5}, P_{k-4}, \dots, P_k, P_{k+1}\}$. The size of the window will be discussed in details later in the next section.

3.1 Plausibility and domain tests

The input data for the software can be categorized into four major inputs. These inputs are:

- **Existed/un-existed data:** The grabbed data from the stream could be received incorrectly for some reasons (due to network disturbances or malfunctioning of the measuring sensor for example) and this un-existed data is represented in the D-Ship server with the (#) character. For example a missing value of the roll would be simply represented by (###. #) if it does not exist.
- **Valid/invalid data:** In rare cases an extreme blunder value could be received from the stream which has been disturbed during the transferring phase from the sensor to the D-Ship system or from D-ship to the user, and this value is lying outside the logical range that it must belong to. For example due to a failure in the transferring phase a latitude value could be received out of the range $\pm 90^\circ$, and this will be considered as an invalid input.
- **A good data:** This successfully passes the outlier test and is considered non outlying value.
- **An outlier:** this fails in the outlier tests and is considered as an outlier which needs a correction process.

After grabbing the NMEA stream, and correctly reading the string line in the software, the first step in the algorithm is to find the missing data and to check the validity of it. This is achieved by the Plausibility and domain tests. A conversion of the missing values is implemented for better handling of the inputs. Thus, any missing or invalid data presented in the stream will be converted to an identical value that represents the missing and invalid data. For example, a gap in the latitude of the MINS could occur giving a missing value of string with eight decimals (#####, #####) and this is converted to a numerical value (9999.0 was chosen) that will represents the missing value in the stream from now on.

These plausibility and domain tests that check the validity and the existence of the input data do not require any connections between different variables, therefore they are implemented in the single epoch sliding window in which a single value is entering the window and these tests examine it and produce the results that go to the next step.

The result of this pre-handling step is not more than the information about the input value whether it is existed and valid or not. Based on this information a decision will be taken afterwards in the correction process.

3.2 Outliers detection and identification

In This step a decision should be taken about a single value if it is an outlier or not. This decision follows a simple approach that can be clarified in the next example:

Let us assume a single position from the MINS. If this measured position is an outlier, this yields:

$$Y_i = x_i + \sigma$$

Where (x_i) is the true position and (σ) is the error associated with the measured position that is represented by (Y_i).

One of the basic techniques used in statistics to define a rough error is simply setting a threshold (T) and compare the measured value with this threshold. If the value falls within this pre-defined threshold, it will then be considered a good measurement otherwise it will be listed in the outlying data category.

Basically we can choose the simplest approach for selecting the threshold based on our knowledge about the typical variation of the data. In the last example we can treat the MINS position as a single value and try to evaluate it without getting any assessment from other devices as a first step. One test could be the speed test, where the distance can be computed each second. We know that the maximum speed of R/V Polarstern is about 16 knots, and this means a maximum speed of 7.7 meters per second. (1 knot equals 0.51444 m/s). The threshold that could be defined for the speed test is a range between 0 and 7.7 m/s, and any point exceeds this range will be suspicious to be an outlier. This threshold needs more justifications to cover what could be logical values of the speed. In any filter when a threshold technique is used the sharpness of the filter is sensitive to the threshold selection. This philosophy has two edges. A question should be asked about what the filter is built to achieve. What is more important? To clear a point that is not an outlier or to remain a point that could be an outlier?

This kind of approach is important when the filter technique is providing a deletion solution to the suspicious points and whether this will affect the data or not.

In our filter we are providing a replacement of the outlier data with a correction method that will be discussed in the next section, but we still need to define the sensitivity of our detection algorithm. We mainly concern about the rough errors and therefore the filter could be directed more toward keeping the data and providing relatively bigger ranges when examining it. A maximum speed of 10 m/s had been taken to allow some extra variation over the identical maximum speed, and this extra variation is added to cover all the reasons that could increase the speed of the vessel in the natural situations. These reasons could be the effect of the current, the wind or the motion of the vessel. Another reason could be recognized for such addition of the maximum allowance of the speed if we take a second look on Figure 5. We can recognize a linear correlation between the vessel's speed and the systematic increment of the error produced by MINS. In the figure we recognize that at speed of 5 m/s the MINS is calculating a position with a shift of around 7 meters toward the bow of the vessel, and at speed of 10 the shift is around 14 meters. Taking into account a maximum acceleration of the vessel from any initial speed and with full capacity of the engine could not exceed 1 m/s^2 , thus the maximum error in the speed caused by MINS due to this systematic error would be calculated from this linear relation as follows:

$$\epsilon = 7 \times 1 \div 5 \approx 1.3 \text{ m/s}$$

Where (ϵ) is the maximum error per second that could be associated with the MINS performance because of the systematic error described previously. Moreover if we consider the wind and currents that could affect the speed of the vessel a maximum speed of 2 m/s could be added for the cases when the ship is sailing with sea current and wind that may add 4 knots/hour. Adding these variations to the maximum speed of the ship will bring us to a range between 0 and 10 m/s which is the appropriate threshold for testing the speed of the ship.

However observing the data and visualizing the track of the MINS give us another reason for this soft handling and non-sharpness filtering of the examined data. This is due to the fact that the MINS is considered as a very accurate scientific device that provides a very accurate positioning, thus it is more likely to have a good position than to have a bad position from the MINS.

Nevertheless, Figure 14 shows that even if the speed test could detect the point in the upper case where the initial speed was relatively high, and the distance does not lie within the pre-defined threshold, but in the lower case when the ship is moving in a slow speed and the outlier occurs but the distance was still valid compared to the threshold, the point will be considered as a good point. Therefore another test is required to cover this situation, and here we recognize that the change of speed (the acceleration) is the parameter that could detect the outlying point independent from the initial speed.

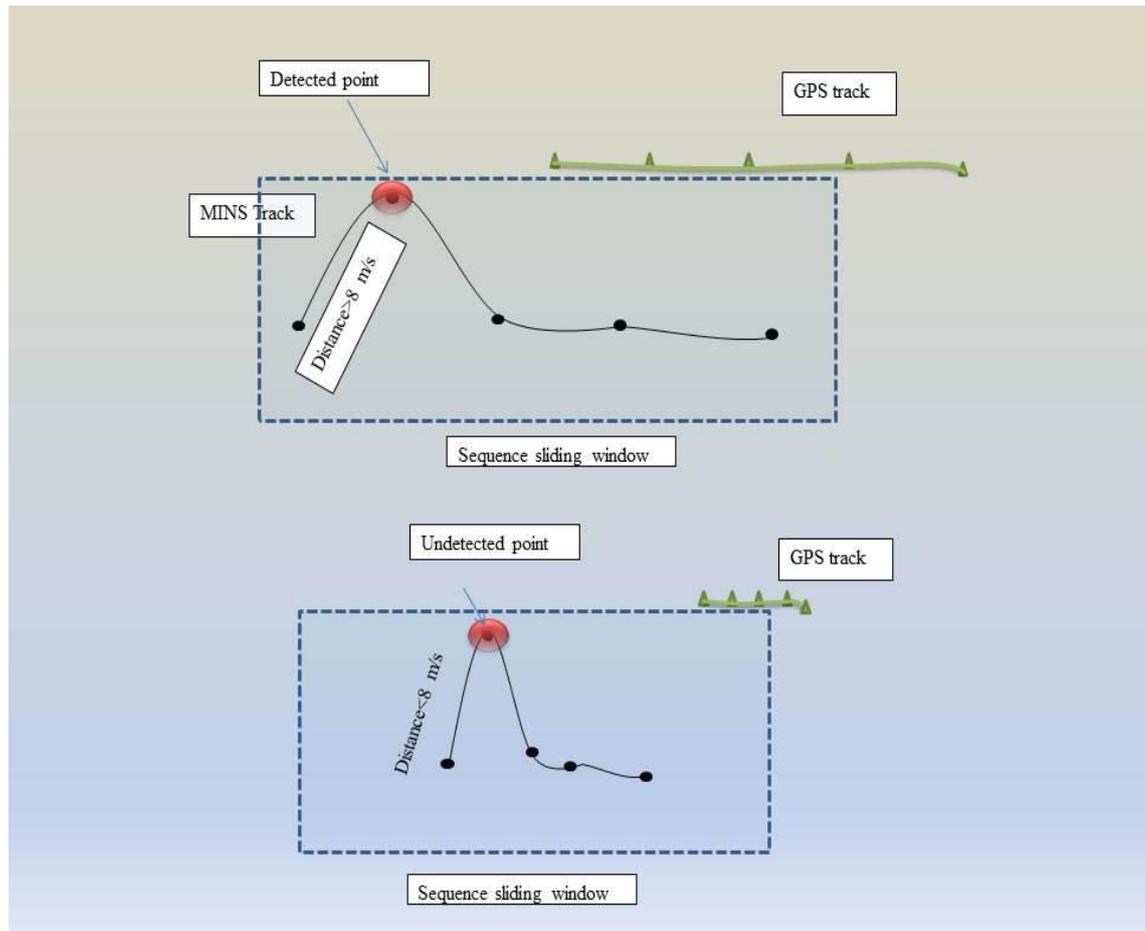


Figure 14: The speed and acceleration tests for the MINS position

The threshold of the acceleration check had been defined between the minimum and the maximum change of speed that could occur in R/V Polarstern per second. A range between (0) and 1 m/s^2 is defined to test the acceleration of the vessel. These absolute values had been chosen with rule of thumb as logical values knowing that the acceleration of the ship could not exceed this range in natural situations. Within these tests (the speed and the acceleration tests) a sequence of positions is required to compute a sequence of speeds and accelerations. The acceleration check had been used also to check the GPS positions of both antennas before providing any correction decision. This is necessary to check whether one of the GPS antennas or both of them are producing also a jump in the position when the MINS is producing the error. In this case, the jump of the position in one of the GPS antennas indicates that the MINS is following a symmetric behavior of the GPS. It also indicates that we cannot consider the GPS position in the correction procedure in this case.

However, these tests depend on the MINS position only as individual measurements, and they provide a good identification for spike outliers that could occur in the MINS position and any sudden jump in the position could be detected. But these detections reach a certain level where another type of outliers may occur and will need different tests to be applied in order to be detected.

There is another type of outlier that had been observed (see Figure 2) where sequence of positions is shifted, and this needs other tests to identify them. Figure 15 shows an example about the mentioned situation.

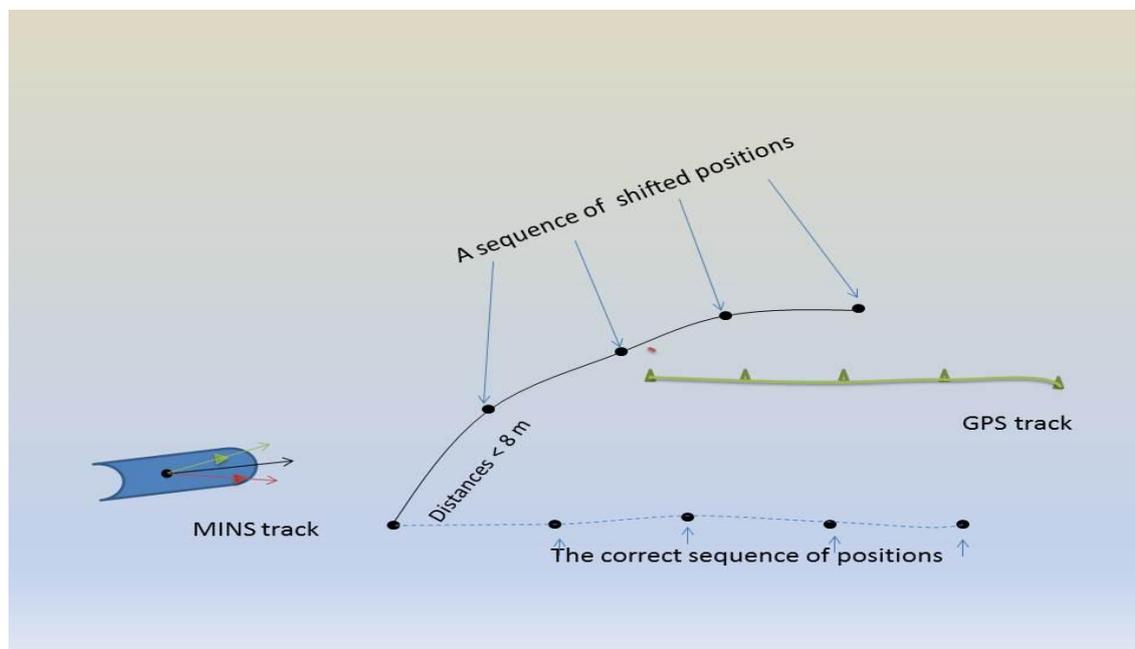


Figure 15: The shift error type occurs in MINS position.

When small deviations in the MINS positions occur like in the previous figure neither the speed test nor the acceleration test will detect this shifting because no sudden change in the distance or in the speed happened. This kind of errors could also be seen in Figure 2 where huge shift in the MINS position could be recognized. This example needs different tests to be applied, which needs other parameters to be involved in sequence of comparison procedures.

On board the ship, we have two GPS receivers that compute the position (Trimble-1, Trimble-2) and we will use these measurements to compare it with our single position in the last example.

The distance computation between two points had been studied for years. The flat Earth distance between the two points in the Cartesian coordinates could be computed using the Pythagoras theorem:

$$Distance = \sqrt{(x1 - x2)^2 + (y1 - y2)^2}$$

Where ($x1, x2$) are the latitudes of the first and the second point, and ($y1, y2$) are the longitudes of these points respectively.

Another method to calculate the distance is the Haversine Formula which calculates the distance on a spherical Earth. The formula for any two points on a sphere is:

$$haversine\left(\frac{d}{r}\right) = haversine(\varphi_2 - \varphi_1) + \cos(\varphi_1) \cos(\varphi_2) haversine(\omega_2 - \omega_1)$$

Where:

$$haversine(\theta) = \sin(\theta/2)^2 = \frac{1 - \cos(\theta)}{2}$$

(d) is the distance between two points along a great circle.

(r) is the radius of the sphere (which is the Earth's radius here).

(φ_1) and (φ_2) are the latitudes of point 1 and 2 respectively.

(ω_1) and (ω_2) are the longitudes of point 1 and 2 respectively.

And the ($\frac{d}{r}$) term is always in radian.

Moreover in geodesy the distance over ellipsoid is calculated using the inverse geodetic calculation in which the shortest distance between two points on the ellipsoid “the geodesic” as well as the azimuth and the inverse azimuth is calculated. Another useful calculation is the direct geodetic calculations where the known inputs are the position of the first point as well as the azimuth and the ellipsoidal distance to the second point, and the output is simply the position and the reverse azimuth of the second point. These calculations are illustrated in Figure 16.

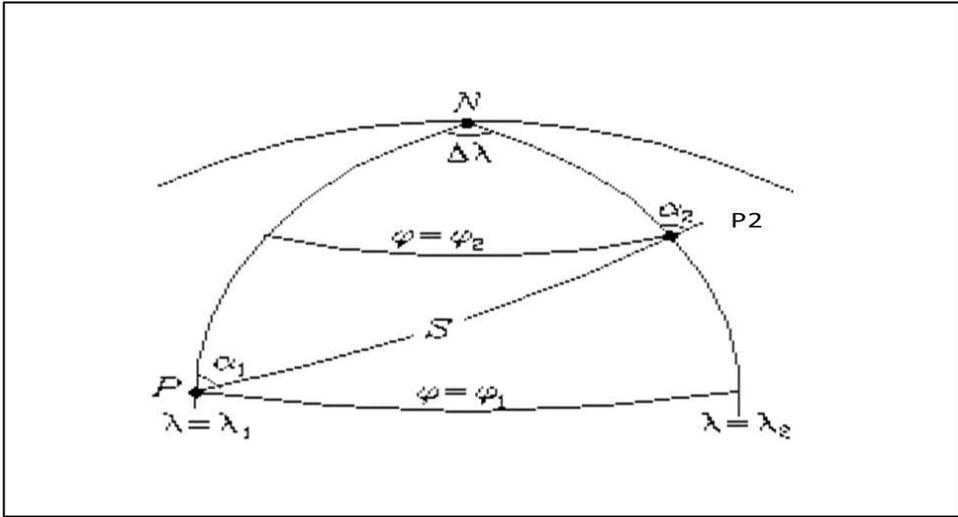


Figure 16: The parameters used in the geodetic calculations (Ghilani, 2005).

Where (S) is the ellipsoidal distance and (α_1, α_2) are the azimuth and the inverse azimuth from point P to P2 respectively.

The geodetic calculations give a better approximation of the distance than the previous two methods. However, even if the computation of the distance between two points is accepted with any of the previous methods as the error of computation is negligible for short distances, but we are going to use the Inverse geodetic problem to compute the distances as well as the azimuth because these two results could both be parameters for our fault detection methods. The computations of the geodesic and the azimuth are implemented using the Mid Latitude Formulas founded by C.F. Gauss, which is described in details in (Walter, 1964) and (IMO-IMA 4th Course on Nautical Cartography, 2003).

All of the previous computations are calculated in two dimensions (using latitudes and longitudes of the points). This includes the ellipsoidal calculations used in this thesis. Figure 17 shows an illustration of the computed parameters that will be used in the upcoming tests.

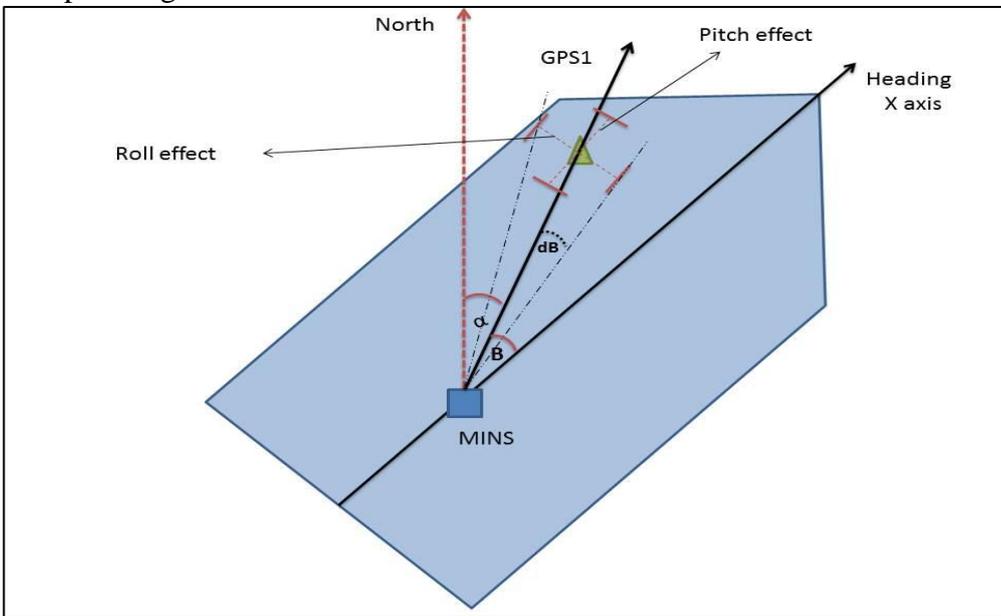


Figure 17: Angle computations and the thresholds consideration.

The azimuth between the two devices will be used in the detection algorithm. As we can see from the figure, the new parameter that had been used for the detection of the outlier is the angle (B) which represents the angle between the heading of the ship and the azimuth from the MINS to GPS1. In addition we should define a threshold that gives this angle some variability due to the motion of the vessel.

The threshold of the angle had been choosing after taking into account the roll and the pitch effects that could change the azimuth (α). Before calculating the amount (dB) shown in the figure above we should first calculate the angle (B) in a well-defined situation. The angle B had been calculated taking a static situation of the vessel when it was berthing at the harbor of Bremerhaven. Figure 18 shows the computation of the angle in this situation.

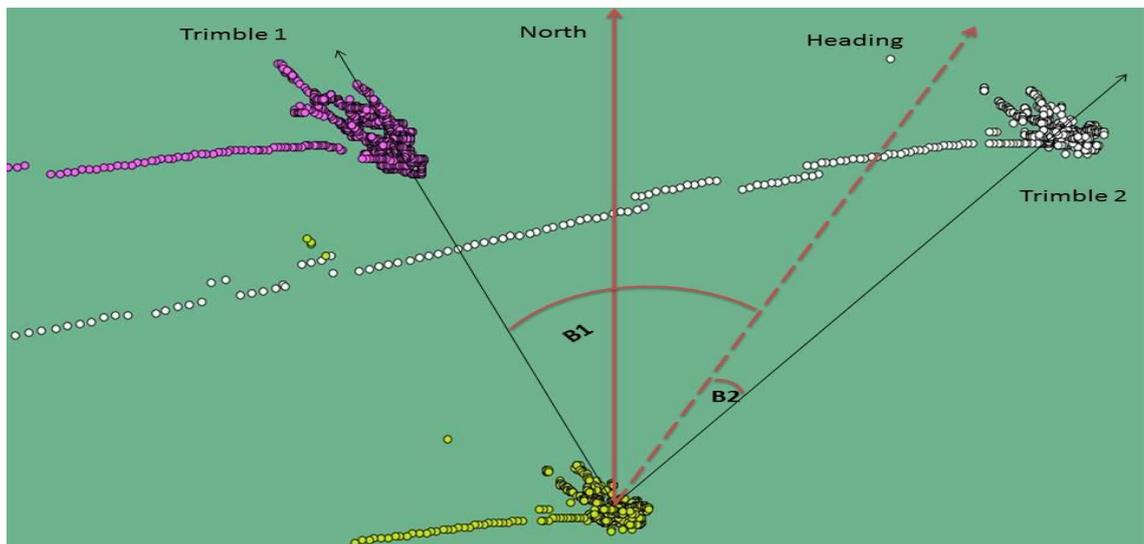


Figure 18: The computation of the angles B1 and B2 between the heading of the ship and the azimuth from the MINS to the GPS location.

Calculating the appropriate positions of the devices had been done using simple averaging of the positions over time. The averaged positions used to calculate the angles (B1) and (B2) are listed in the following table.

Table 4: The average positions that had been used in the computation of the angles for the angle test.

Position	MINS		Trimble 1 GPS		Trimble 2 GPS	
	Latitude	Longitude	Latitude	Longitude	Latitude	Longitude
	53.566815	8.555035	53.566986	8.555222	53.566988	8.554939

The heading of the ship was 19.3°, and the azimuth from the MINS to both GPS antennas had been calculated using the inverse geodetic calculation leading to the following results:

$$\alpha_1 = 33.0639^\circ.$$

$$\alpha_2 = 341.62^\circ.$$

Where (α_1) is the azimuth from the MINS to the Trimble-1 antenna, and (α_2) is the azimuth from the MINS to the Trimble-2 antenna.

The desired angles (B1) and (B2) are calculated by the simple relations:

$$B1 = \text{azimuth} - \text{heading} = 33.0639 - 19.3 = 13.7639^\circ.$$

$$B2 = (360 - \text{azimuth} + \text{heading}) = (360 - 341.62) + 19.3 = 37.68^\circ.$$

After finding the expected angles we should define the threshold in which the calculated angles should be compared with. The variability for these angles comes from what is called the lever arm effect. Due to the geometric distance between the MINS and the GPS antennas, the attitude parameters play a major role in changing the angle (B) and this should be taken into account when defining the threshold for the test.

On board R/V Polarstern, The geometric distances between these devices have been accurately calculated during the alignment survey of the vessel R/V Polarstern, and these distances were measured according to the local ship reference system (See Figure 12). The next table shows the lever arms between these devices in meter unit.

Table 5: The lever arms of different devices in the ship's reference system

Sensor	X [m]	Y [m]	Z [m]
MINS	0	0	0
GPS Trimble 1	22.777	-5.460	21.525
GPS Trimble 2	17.303	12.408	21.536

The implementation of the motion effects is called the lever arm corrections. In Figure 17, the dashed red lines represent the variation of the GPS position due to the roll and the pitch of the ship. These corrections are important to determine the threshold of the angle test as well to correct the lever arms between the sensors that will be essential in the transformation procedures when replacing a bad point as we will see in the next section.

The computation of the lever arm corrections are applied by finding the rotation matrix around the axes of the ship's body frame. A detailed derivation could be reviewed in (Rowe, 1996). The summation of the resulted rotations around each axis is written as follows:

$$x = \cos(P) \times dx + \sin(P) \times dz$$

$$y = \sin(P) \times \sin(R) \times dx + \cos(R) \times dy - \sin(R) \times \cos(P) \times dz$$

$$z = -\sin(P) \times \cos(R) \times dx + \sin(R) \times dy + \cos(R) \times \cos(P) \times dz$$

Where (x, y, z) represent the corrected arms between the MINS and the desired GPS antenna.

(dx, dy, dz) are the original lever arms shown in (Table 5).

(R, P) are the roll and pitch values.

However, we should emphasize that different systems are applied onboard R/V Polarstern and for this purpose the lever arms used in this calculations were taken from the left hand system of the ship. in addition, and due to the lack of the yaw information this had been set to zero in order to get the above equations as we do not take into account the rotation around the (z) axis.

Positive roll are defined in this system to the starboard side of the vessel and positive pitch is when the motion is to the bow of the vessel. As we recognize from Figure 17 which represent the situation of the Trimble-2 antenna, the lever arms between both sensors (see Table 5) had been corrected using the previous formulas by considering two situations that are shown in Figure 19.

- 1- When a roll of (10°) and a pitch of (-10°) are applied this leads to a new lever arms that are calculated using the previous formulas. The resulted arms lead to minimum angle $(B - dB \approx 31^\circ)$.
- 2- The opposite motions had been taken for the maximum (B) leads to a maximum angle $(B + dB \approx 40^\circ)$.

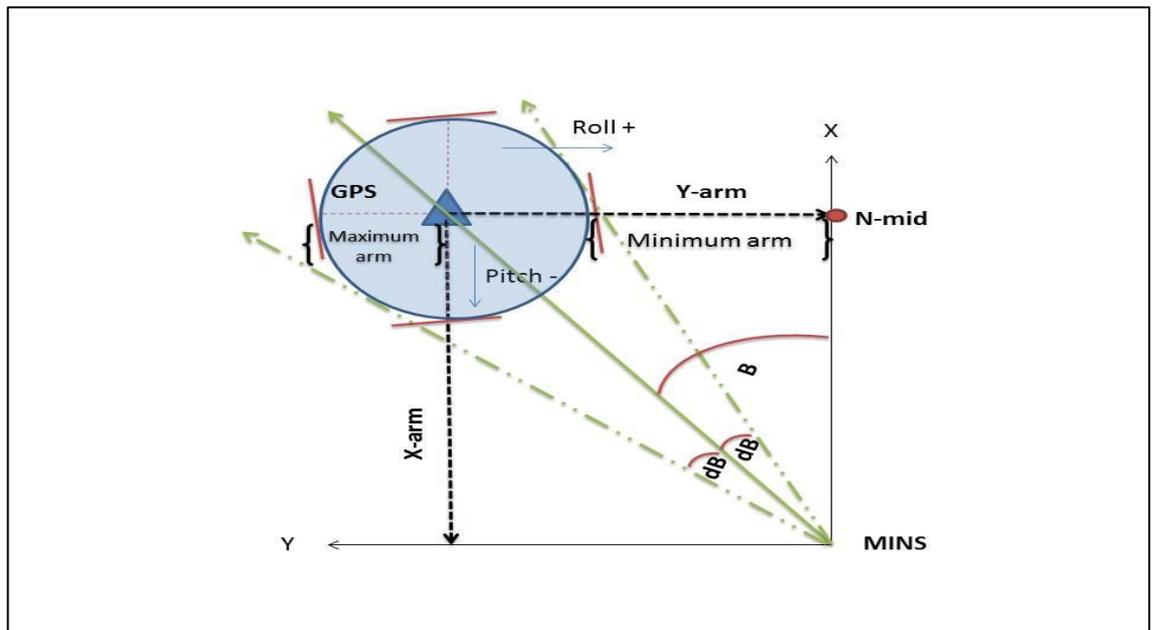


Figure 19: The threshold calculation for the angle test of Trimble 2 antenna.

The same approach had been applied for computing the threshold for the angle with Trimble 1 antenna and the threshold had been set between (4°) and (18°) . Now after defining the thresholds for the angle test, the detection method could be understood better looking at Figure 20.

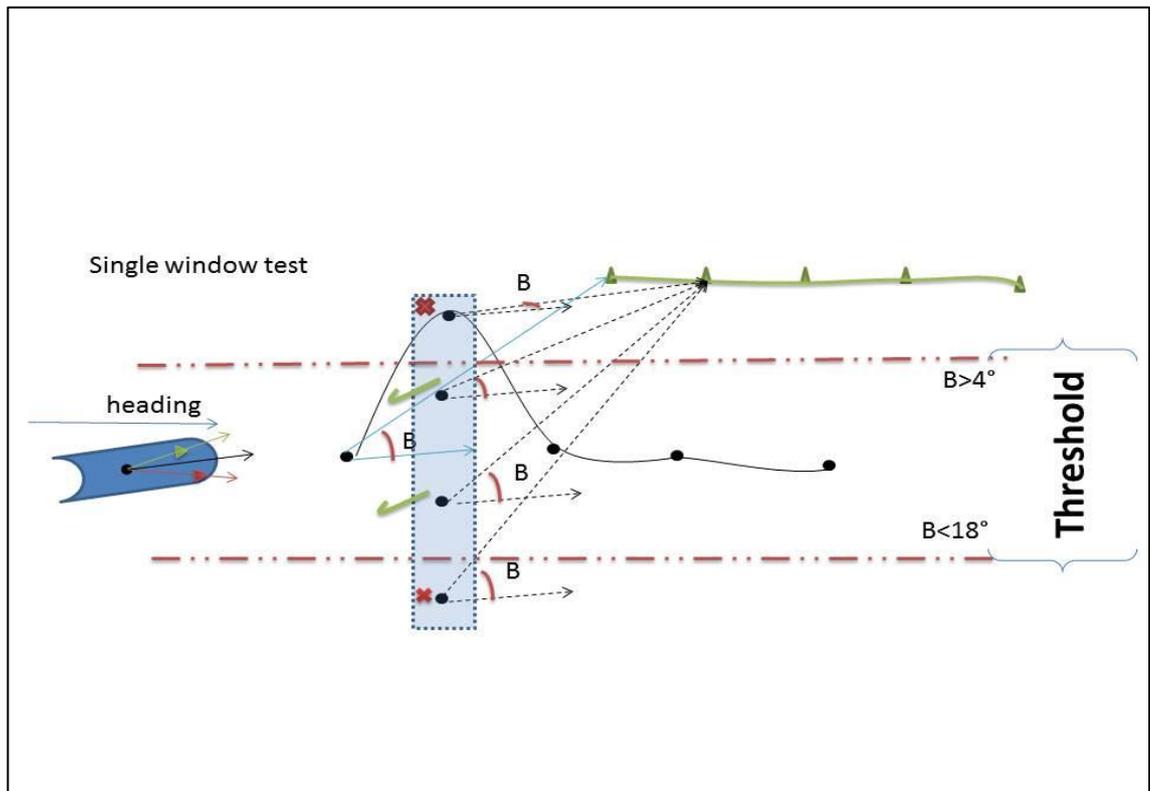


Figure 20: single epoch window test for the angle test.

The threshold detects any outlying position outside the pre-defined range of the angle (B). This test could also detect any shifting or sequence of erroneous positions from the MINS.

3.3 Outlier correction

After defining all the methods associated with the detection of the outliers, it's time to provide solutions for replacing the outlying point with another appropriate one. The correction methods based on three kinds of solutions, all of them will be generated when an outlier is detected in a first step. Then, these solutions will be evaluated to take a decision which of them is providing better accuracy. The generation of these solutions is divided into two categories:

- **The transformation solutions**

In this procedure, two positions are transformed from both GPS receivers (Tremple-1, Tremple-2). The transformation is based on the direct geodetic problem. Let us assume a known position of Tremple-1 is provided represented by (φ_1, λ_1) for latitude and longitude respectively. We need the azimuth and the ellipsoidal distance in order to be able to calculate the MINS position, but we only have one input which is the position. Indirect transformation based on our knowledge about the lever arms between these devices are used for completing this procedure. As seen in the Figure 19, at a first step the position is transformed from the GPS to the point (n-mid) and we can determine the azimuth based on the information of the heading of the ship. In this figure, the heading is 19.3° , this yield to an azimuth from GPS1 to the point (n-mid) equals:

$$azimuth = heading + 90^\circ.$$

Moreover, the ellipsoidal distance will be approximated to be equal to the lever arm on

the (y) axis in the ship's local coordinate frame. After having all the required inputs for the direct geodetic calculations, now we are able to calculate the position of the point (n-mid). This process is implemented for second time to transform the position from the point (n-mid) to the MINS where the azimuth is simply the opposite direction of the heading (199.3°) and the ellipsoidal distance will be the lever arm on the (x) axis in the ship's frame.

The lever arms should be corrected in order to have accurate transformations. Therefore, we need the information about the attitude of the vessel when we are generating the transformation solutions. Because of this the attitude parameters should be corrected first before any corrections to the positions are applied. The correction of the attitude parameters is done using the extrapolation solution.

- **The extrapolation solution**

This solution is applied when a missing data or invalid value is detected in one of the attitude parameters or in the position of the devices. When an outlier is detected in a sequence of positions gathered from the MINS, and no solution from the transformation procedure is available, the sliding window test is providing a solution for the new erroneous position entering the window. This procedure produces the solution after applying a polynomial regression on the data inside the sliding window except the last point (the outlier). For example if we have a new sequence of positions from the MINS, and the last position of the sequence is an outlier, missing or invalid, a regression model is built to fit all the previous positions in the sequence (that had been already treated).

For a simple review of the polynomial regression fit, if we define data set with 2D positions that consists of latitude and longitude as follow

$$X_i, Y_i (i = 1, \dots, n)$$

Then we will treat each of the variables as an individual variable varying with time, thus we have two time series data to find a regression model for, that are: the latitude (X_i) and the time (T_i), and the longitude (Y_i) with time intervals (T_i) also. Therefore we will have now the two time series sequences:

$$X_i, T_i (i = 1, \dots, n)$$

$$Y_i, T_i (i = 1, \dots, n)$$

We will clarify the process of one variable as this process is similar with all the variables that apply the extrapolation process for fixing the error.

Generalizing from a straight line (i.e., first degree polynomial) to a K degree polynomial

$$T = +a_1x + \dots + a_kx^k$$

The residual is calculated by the square of the summation of the differences between the measured values and the estimated values provided by the model. This yield:

$$R^2 \equiv \sum_{i=1}^n [T_i - (a_0 + a_1x + +a_kx^k)]^2$$

If we take the partial derivatives then we have:

$$\begin{aligned}\frac{\partial(R^2)}{\partial a_0} &= -2 \sum_{i=0}^n [T_i - (a_0 + a_1 x + \dots + a_k x^k)] = 0 \\ \frac{\partial(R^2)}{\partial a_1} &= -2 \sum_{i=0}^n [T_i - (a_0 + a_1 x + \dots + a_k x^k)] x = 0 \\ \frac{\partial(R^2)}{\partial a_k} &= -2 \sum_{i=0}^n [T_i - (a_0 + a_1 x + \dots + a_k x^k)] x^k = 0\end{aligned}$$

These lead to the equations

$$a_0 n + a_1 \sum_{i=1}^n x_i + \dots + a_k \sum_{i=1}^n x_i^k = \sum_{i=1}^n T_i$$

And

$$a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 + \dots + a_k \sum_{i=1}^n x_i^{k+1} = \sum_{i=1}^n x_i T_i$$

And

$$a_0 \sum_{i=1}^n x_i^k + a_1 \sum_{i=1}^n x_i^{k+1} + \dots + a_k \sum_{i=1}^n x_i^{2k} = \sum_{i=1}^n x_i^k T_i$$

Or, in matrix form

$$\begin{bmatrix} n & \sum_{i=1}^n x_i & \dots & \sum_{i=1}^n x_i^k \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \dots & \sum_{i=1}^n x_i^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_i^k & \sum_{i=1}^n x_i^{k+1} & \dots & \sum_{i=1}^n x_i^{2k} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n T_i \\ \sum_{i=1}^n x_i T_i \\ \vdots \\ \sum_{i=1}^n x_i^k T_i \end{bmatrix}$$

This is a Vandermonde Matrix that arises in the polynomial least square fitting. We can also obtain it by writing:

$$\begin{bmatrix} 1 & x_1 & \dots & x_1^k \\ 1 & x_2 & \dots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{bmatrix}$$

Pre multiplying both sides by the transpose of the first matrix then gives

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^k & x_2^k & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^k & x_2^k & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{bmatrix}$$

So we came to the same equation. As before, given (n) points (X, T) and fitting with polynomial coefficients (a_0, a_1, \dots, a_k) gives

$$\begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^k \\ 1 & x_2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix}$$

In matrix notation, the equation for a polynomial fit is given by

$$\mathbf{T} = \mathbf{Xa}$$

This can be solved by pre multiplying by the transpose (\mathbf{X}^T) ,

$$\mathbf{X}^T \mathbf{T} = \mathbf{X}^T \mathbf{Xa}$$

This matrix equation can be solved numerically, or can be inverted directly if it is well formed, to yield the solution vector

$$\mathbf{a} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{T}$$

This procedure is applied on the latitude of the MINS to find the regression model that provides the best fit of the latitudes as a time series varying with time. After having this function new latitude is calculated. So if we have 10 points and the last is an outlier for example, then a regression models for latitude and longitude are built for the first 9 points and an extrapolation is done to fine the 10th values of latitude and longitude.

If we consider that we have three solutions for the position of the MINS when an outlier is detected, a decision should be taken to choose the better solution among all that provides more accurate position. This decision will be specified in the next section where finalization of the algorithms is produced.

- **Decision making, smoothing and finalization:**

Up to this point a sequence of positions from the MINS are evaluated for detecting any possible outlier. A correction vector is generated through the correction process. Whenever an error is presented, the vector solutions are provided and wait for the decision which will determine the replacement point. The smoothing process takes place at this stage for triple purposes:

- 1- It will provide a final smoothing approach for the MINS position to eliminate the small variability in the MINS track. This will be done using a second order polynomial smoothing regression to replace the MINS position with the estimated point calculated by the regression function. This smoothing procedure will enhance the solutions provided for the outliers. The second degree of the polynomial had

been chosen for this step for better smoothing in the first hand. Also the window size for evaluation the sequence of points was chosen with 10 points, and for this size the track of the ship could not exceed the polynomial curve within 10 seconds only, and therefore the second order of the smoothing fit is appropriate.

- 2- It will make the decision of which solutions is more adequate for the outlier point. When a point is identified as an outlier the solution from the GPS antenna (trimble-1) is provided to replace the outlying point. It had been chosen first due to the fact that in lots of observations of the plotted tracks of the GPS-1 and 2, more stable tracks were much more likely to be produced by Trimble-1.

After applying the solution it should be evaluated first. The reason of that is simply illustrated in the next figure. The figure shows an example about a specific situation. In the upper right case we recognize erroneous positioning produced by the Trimble-2 where stable positions are provided by the Trimble-1 GPS. In contrast the opposite situation is happening in the lower case where the better positioning is produced by the Trimble-2 antenna.

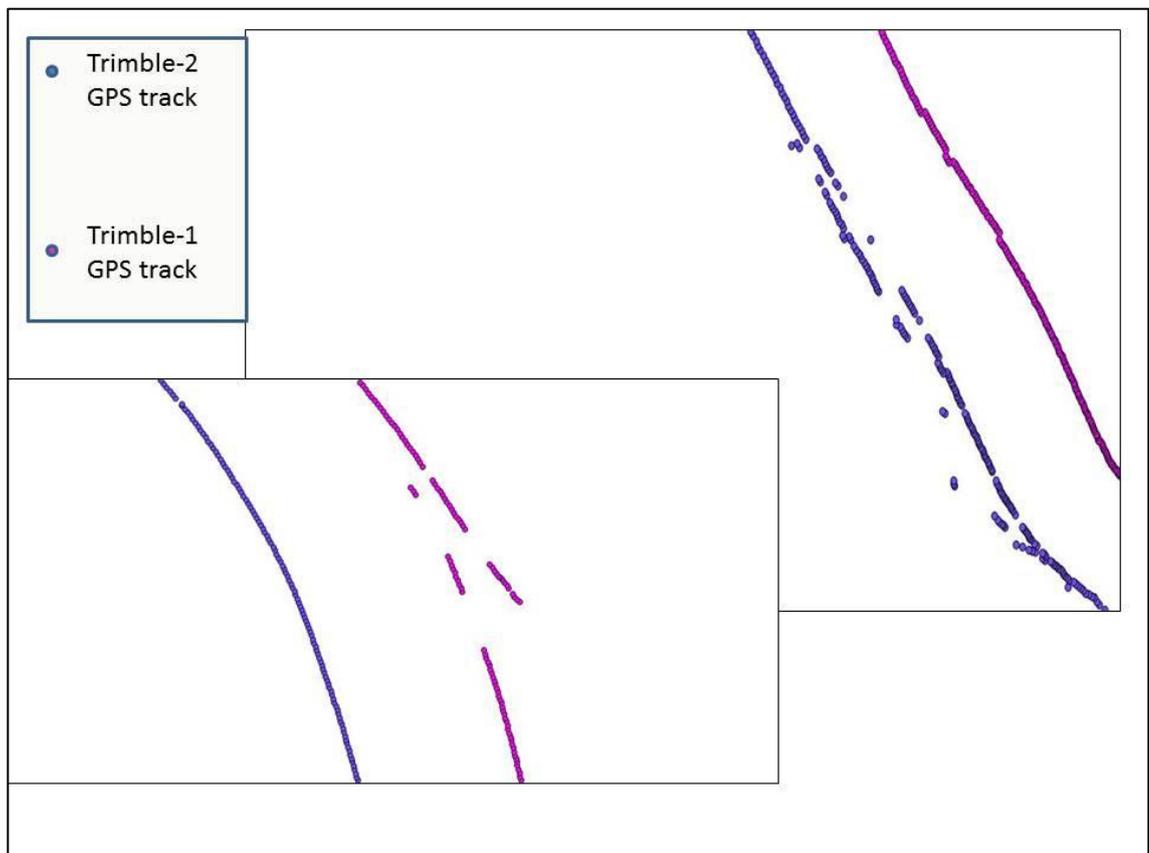


Figure 21: Bad Position produced by Trimble-(1-2) GPS antennas.

Therefore, when a transformation procedure is applied to replace the bad position from the MINS, the method should evaluate the solutions first in order to choose the better solution. This is done as follows:

When the first solution is provided a regression function is built including the solution point. The maximum error of the function is computed that represents the maximum deviation of a point from the regression function. If the maximum error

exceeds a pre-defined value, the solution will not be considered and the second solution is replaced. The whole process is repeated to evaluate the new maximum error produced.

At the end of this process the point with the minimum error is chosen to be the replacement solution of the outlying point, and the smoothed positions are outputted as Master track values.

- 3- The last objective achieved in this final stage is producing a quality indication about each single position of the master track. The average error produced by the final smoothing regression function had been chosen as an indicator about the quality of the whole sequence of the points generating it.

The smoothing regression is implemented on the latitudes and longitudes separately as individual time series as we have seen earlier. Thus, two average errors will be produced from both the latitude regression model and the longitude regression model. Because both variables are presented in a degree unit, this implies that the average errors produced will be in degree units as well. This is not very good representative especially when the final product will be represented in a GIS environment, and the metric unit is better for such purpose. Thus, a conversion to the metric unit is produced using the following approximation:

$$\sigma_{latitude} = average\ error_{lat\ regression} \times 111111.111$$

$$\sigma_{longitude} = average\ error_{lon\ regression} * 111111.111 \times \cos(\varphi_{average})$$

$$total\ quality = \sqrt{\sigma_{latitude}^2 + \sigma_{longitude}^2}$$

Where, the number (111111.111) represents the average length of one degree of the latitude. However we should take into account the convergence of the longitudes towards the poles therefore we multiply with the cosine of the mean latitude of the sequence points that we are calculating the quality for.

Until here we have covered all the methods used in the filtering algorithm, and we have explained each process with all of its functionalities. The next step is to put all of these functions in action and to connect them appropriately to produce the desired output. The next flow chart in Figure 22 illustrates the main processes of the algorithm that had been built in the software with the connections of the functions.

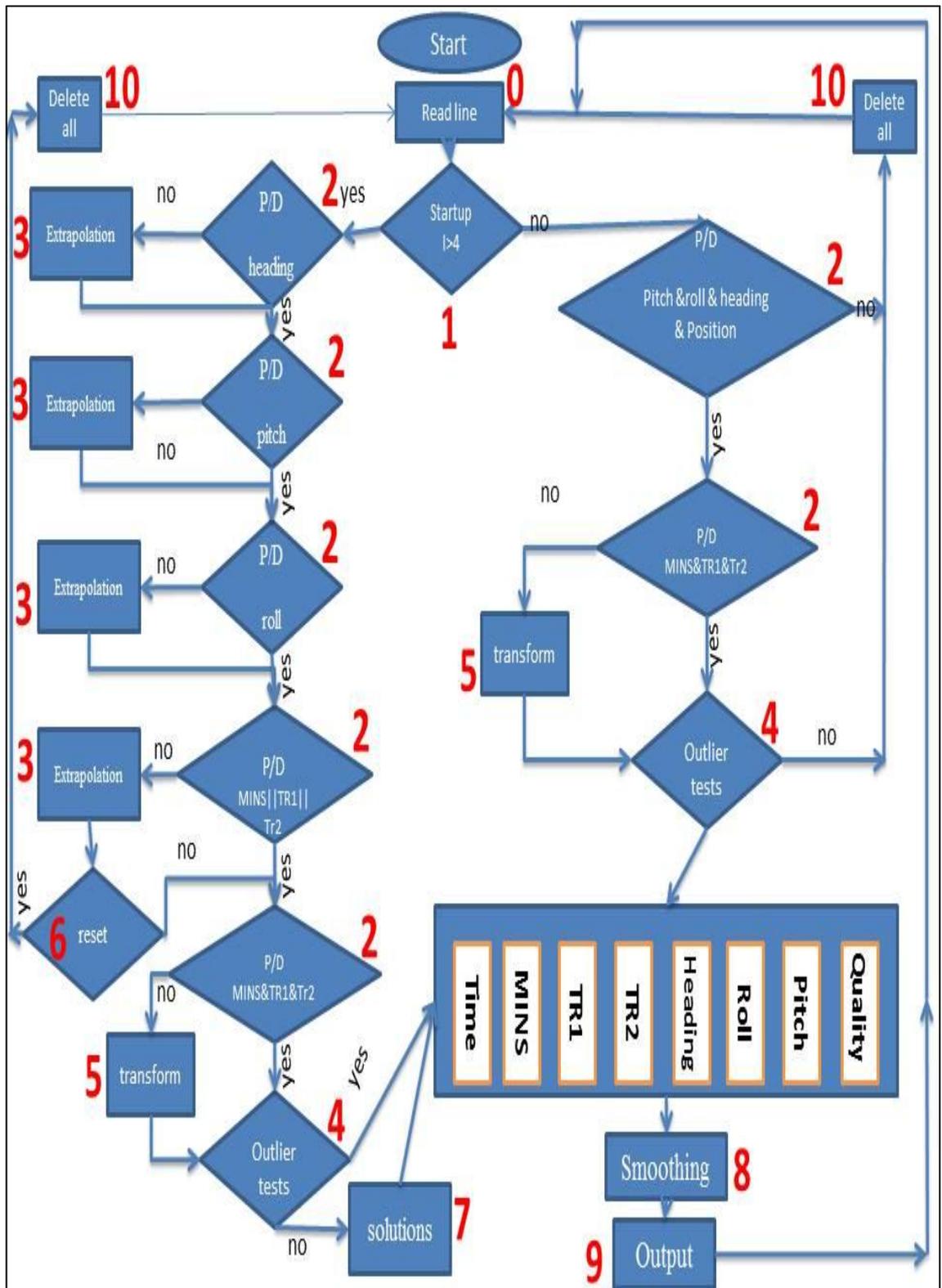


Figure 22: Flow chart shows all the functions used in the algorithms.

When the application starts to receive the records from the stream or from reading the lines in the input file (using the function read line indicated with number 0), a counter is generating indices for the lines for better handling the inputs, and each examined record will be stored in a global vectors that hold the results of each process.

A startup stage examines the first lines of the process (this is indicated with number 1 in red). This stage is necessary because the algorithm is using a sequence sliding window that evaluates a set of inputs together. With the startup stage we ensure that the algorithm starts the process based on good records. The difference between the startup stage and the normal processing of the lines is in the response to the outlying points. In the startup stage the aggressive response is taken and points are deleted, while in the normal processing stage the response is taken with correction intuitions. In the startup stage the data are examined with the domain and plausibility tests, and with the outlier tests, but no correction is taken but deletion of the outlier is taken.

The number of records that should be provided in the startup stage was chosen to be 4 good records in a row. This means that the software will end the startup stage and move to the normal processing stage only after receiving four records that successfully pass the outlier tests. However this could delete some results at the beginning of the analysis if this condition is not met.

For example if at the beginning of the analysis two lines successfully passed the tests of the startup stage, they will be stored in the global vectors. If the third line failed to pass the tests the function [Delete all], which is indicated with number 10, will be activated to delete all the stored variables in the vectors as well as the indices of the lines, and the startup process will be repeated. Normally the cruise starts at the harbor, and at the beginning there would be records at the harbor and at this time of the cruise deleting some results does not affect any scientific measurements.

After the startup stage the algorithm will examine each record with the domain and the plausibility tests, and a correction will be applied for the heading, roll and pitch using the extrapolation process (indicated with the number 3). This process will use the information stored in the global vectors from previous records to produce the extrapolated value and to replace the invalid data of the heading, roll or pitch with it.

Examining the position information will take place afterwards using three routines.

- 1- The position information from all devices is examined with the domain and plausibility tests to check whether all positions are existed and valid. If all of them are not, then the same extrapolation process used for the heading and the attitude parameters will be applied.
- 2- If some of the positions are existed and valid and some are not, a transformation process will be applied to generate the missing positions (the transformation is indicated with number 5 in the flowchart).
- 3- Afterwards the positions are examined with the outlier detection methods described earlier for checking the errors (this is indicated with number 4).

Moreover a reset condition had been built automatically to reset the algorithm of the software when twenty complete records had been received as missing data. This indicates that the systems onboard the ship are shut down (when it arrives a port for example), or it could be a reason of problem in the distribution system onboard the vessel. In other words the reset function works perfectly with the association with the startup stage, and whenever a specified number of missing records is received in a row the reset condition is met and the software restarts the startup stage where receiving a good results again is the condition for outputting the data (the reset condition is indicated with number 6).

Furthermore, the solutions of the outliers are generated using the transformation

function and the extrapolation function. In the first one the information of the lever arms are taken into account and the position is transformed from the GPS receiver to the MINS location (the solutions are indicated with number 7).

At the end of this step the global vectors are fed with the results of the previous tests, and will wait for certain amount of records to be filled with, in order to proceed to the next step. A number of ten records were chosen to further proceed in the algorithm. This number simply indicates the length of the sliding window test. It has been chosen with a rule of thumb that the maximum curve that the ship may follow in 10 seconds is expected to be represented with a second order polynomial regression fit, and this is exactly the applied regression used in the smoothing process afterwards.

Finally we should emphasize that as the inputs of the application could be either a file that consists of the records of the expeditions, or a stream gathered from the network, and because the time of this work did not help to test the application onboard the ship, the online assessment approach had been achieved virtually with the aid of a simulator that works as a server and send the contents of a chosen file to the application second by second to simulate the situation onboard R/V Polarstern. The connection between the simulator and the application had been adopted as a server-client network where the application is acting as a client that receives the information from the server (the simulator).

Figure 23 shows the simulator and its connection functionalities, where lines are sent to the specified IP address of the software with time interval that could also be specified.

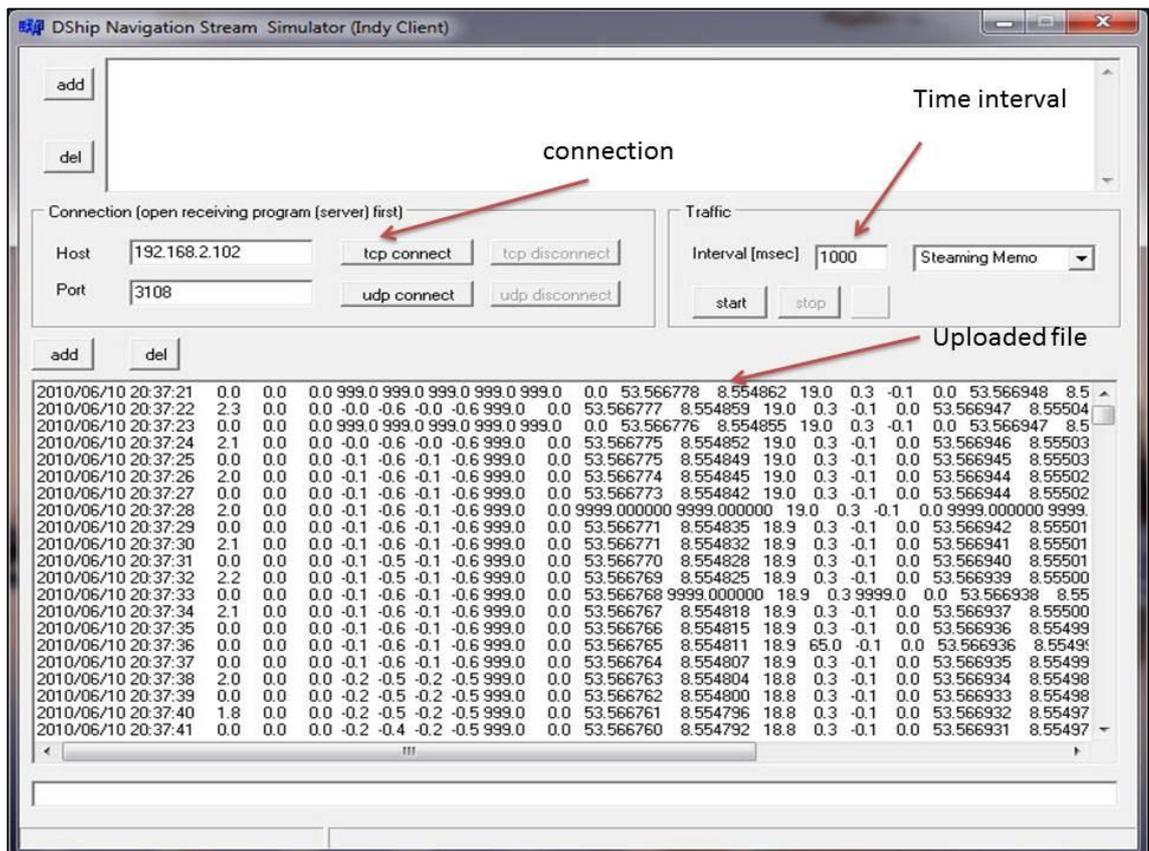


Figure 23: the simulator of D-Ship server.

3.4 Master track generation and PANGAEA standards

Meeting the standards of PANGAEA is a key objective of this work, and the final product will be implemented directly into the system as a new dataset for the R/V R/V Polarstern. This product is simply our final output file or the Master track file.

The Master track file will be generated with our software application for the previous cruises of R/V Polarstern as well as for the new expeditions. In this file seven columns are provided for the user and listed in Table 6.

Table 6: the header of the columns in the Master track file

Master track column information						
Date/Time	Latitude	Longitude	Heading	Roll	Pitch	Quality

Moreover, PANGAEA is using the Digital Object Identifier system DOI, which provides a technical infrastructure for the registration and use of datasets on digital networks. In addition GEOCODES are used in the system for geo-reference the data in space and time on Earth. A list of pre-defined GEOCODES is available, and they are mandatory to be included in any dataset that should be stored in the system. These mandatory GEOCODES are:

- 1- Date and time: they are presented as one column in the dataset.
- 2- Latitude.
- 3- Longitude.

All GEOCODES are defined as float values (except the date/time) and thus they could be defined with any chosen precision. However the Master track file includes in its first three columns the mandatory information that meets PANGAEA specification. The precision of the geographical coordinates was set to 7 digits.

Furthermore, any dataset in PANGAEA should have header information that consists of the metadata about this dataset. The information of this header are categorized in four tabs:

- Basics tab.
- Config tab.
- Details tab.
- Web tab.

More details of these tabs could be found in (Grobe, 2006). However, header information for the Master track file had been built, and the contents of the header file are shown in Figure 24.

```
PANGAEA:
/* DATA DESCRIPTION:
Author:
Title:
Reference:
Export Filename:
Event:
PI:
Parameter:
DataSet Comment:

Other version:
Project:
Topologic Type:
Status:
Login:
*/
Event label
```

Figure 24: header information for the Master track file with metadata description.

This header provides metadata description for the user with all the connections and links to the references.

However, due to the huge size of the produced Master track file which may limit the ability of providing the contents of the Master track as tables shown in PANGAEA, another version of the dataset had been produced which is the generalized Master track file.

The generalization of the Master track file was produced using Ramer-Douglas-Peucker algorithm. The purpose of this algorithm is as follows: When a curve is generated with a big number of points, the algorithm finds a similar curve with fewer points. Figure 25 shows the principle of the applied algorithm.

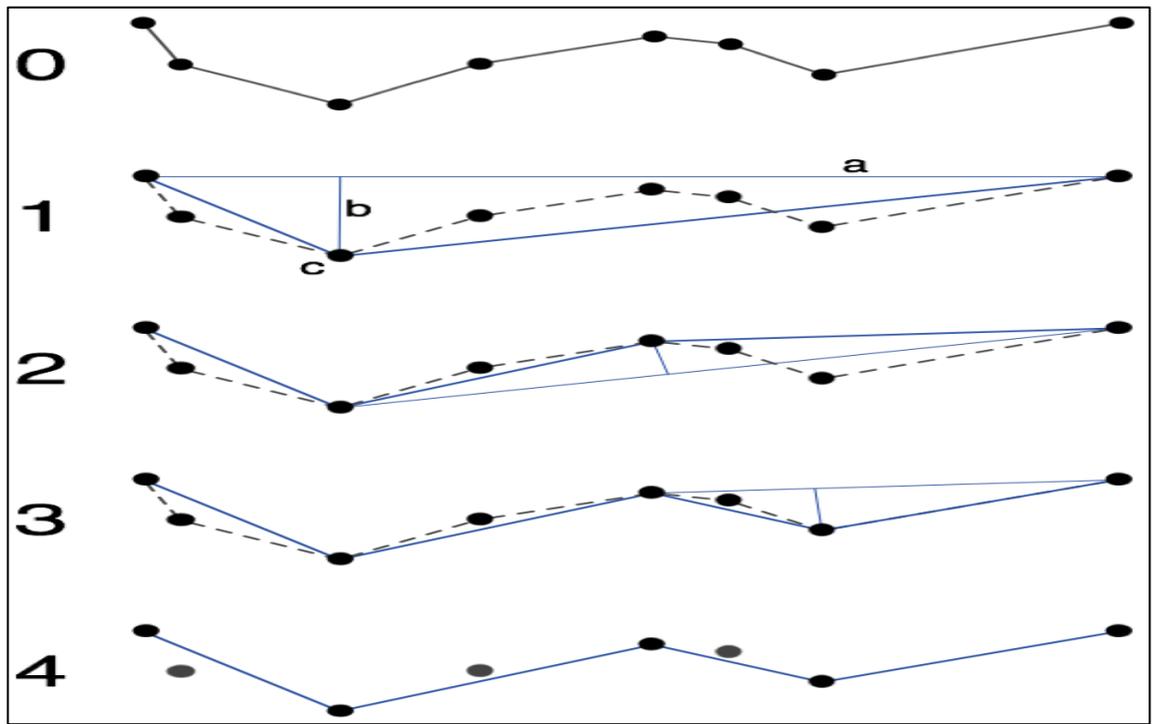


Figure 25: The principle of Ramer–Douglas–Peucker algorithm. (Peucker, 2012).

As seen in the figure, the algorithm is applying an iterative process where the start and the end points are kept in the new curve. After connecting these two lines it examines the distance from the resulted line with the furthest point from the line, if this distance is bigger than a prior value, the examined point is kept in the data, and the process repeated taking into account the start, the end and the new points. New lines are produced and new points are examined using the prior value until the process ends. We recognize that out of 8 points the algorithm successfully reduced the amount of the points to 5 point producing a similar curve at the end.

3.5 outline

In this chapter the methods of the analysis used in this work had been explained with details. The functions used in the algorithms for detecting the outliers are clarified as well as the methods for correcting the outliers are explained.

Furthermore, the specifications of PANGAEA were presented and the necessary information that should be provided in the final output was produced. We also covered the full process of the algorithm with a flowchart that explains the connections between different functions and processes that are presented in this work.

In the next chapter the programming codes and interface of the software will be explained with all the possibilities for users. Finally the results and the discussions of the final outputs will be presented in chapter five.

Chapter4

Software description and implementation

In this chapter we will cover the main functions and classes used in the software as well as the interface and the different options that are provided for the user.

This software was developed using C++ programming language under Windows 7 operating system. Borland C++ builder6 environment was used for programming the application. It provides very useful tools for graphical user interface and portable applications. The software was built as a standalone application with an executable file (win.exe) that could be used as a portable software, thus there is no need to download all the codes and the include files with the application.

Although the application is developed for Windows system, the classes and functions that have been used were developed under ANSI standard, and therefore it can be compiled in any C++ environment that uses the ANSI standard for C++.

However, separate classes were used for the interface and the analysis algorithms. The next table shows these classes with some of the functions that have been used in the application.

Table 7: the classes that have been used in the software.

Class name	functions used in the application
cEllipsoid	Used for the computation of the geodetic calculations.
TNak3Record	Used for the interpretation of the inputs from D-Ship
AnalysisFunctions	Used for the testing the inputs and for outlier detection and correction.
cDatumTime	Used for the interpretation of the time of the records.

The main class that has been used for the analysis of the data as seen in the table is the *AnalysisFunctions* class. The header file of the class is shown in the next two figures.

Figure 26 shows the declared functions of the class that we have discussed in the chapter three.

```

#ifndef AnalysisFunctionsH
#define AnalysisFunctionsH
#include "functions2.h"
#include <StdCtrls.hpp>

class AnalysisFunctions {

public:

    AnalysisFunctions(void);
    ~AnalysisFunctions(void);
    void clear(void);
    bool createStatistikFile(string fname);
    void deleteAll();

    //detection and solutions
    void read_lines(TNak3Record * nak);
    void transform_position_1(double& L1, double& B1, double& L2,
        double& B2, double& L3, double& B3,
        double heading, double roll, double pitch);
    void fill_MINS_fromTr1(double L1, double B1, double& L2,
        double& B2, double heading, double roll, double pitch);
    void fill_MINS_fromTr2(double L1, double B1, double& L2,
        double& B2, double heading, double roll, double pitch);
    void leverArmCorrection(double roll, double pitch, double dx,
        double dy, double dz, double& x2, double& y2, double& z2);
    double dist_check(double& L1, double& B1, double& L2, double& B2);
    bool validity_startup(double &L1, double &L2, double &L3,
        double &Lo1, double &Lo2, double &Lo3, double &heading,
        double &roll, double &pitch);
    bool outlier_startup(double L1, double L2, double L3, double lon1,
        double lon2, double lon3, double heading, double roll,
        double pitch);
    bool outlier_MINS_TR1(double latMins, double lonMins,
        double latTr1, double lonTr1, double heading, double roll,
        double pitch);
    bool outlier_MINS_TR2(double latMins, double lonMins,
        double latTr2, double lonTr2, double heading, double roll,
        double pitch);
    double max_err_check(double lat, double lon, int ix);
    bool speed_test(double L, double B, int ix);
    bool acc_test(double L, double B, int ix);
    //extrapolation routine
    double extra_heading(int ix);
    double extra_roll(int ix);
    double extra_pitch(int ix);
    double extra_latMins(int ix);
    double extra_lonMins(int ix);
    double extra_lonTr2(int ix);
    double extra_latTr2(int ix);
    double extra_latTr1(int ix);
    double extra_lonTr1(int ix);

    //output routine
    double poly_fit_check(vector <double>& latMins,
        vector <double>& lonMins, int ix);
    void poly_fit(vector <double>& latMins,
        vector <double>& lonMins,
        vector <double>& quality, int ix);
    void closeStatistikFile(void);

```

Figure 26: the header file of the Analysis functions class used in the application

Also Figure 27 shows the declared variables that are used in the application for different purposes such as the input and output routines and the statistical information for example.

```
vector <string> date_time_holder;
vector <double> latMins_holder;
vector <double> lonMins_holder;
vector <double> latTr1_holder;
vector <double> lonTr1_holder;
vector <double> latTr2_holder;
vector <double> lonTr2_holder;
vector <double> heading_holder;
vector <double> roll_holder;
vector <double> pitch_holder;
vector <double> quality_holder;
vector <string> source_holder;
vector <int> identification;
vector <double> lat_mins_original;
vector <double> lon_mins_original;
ofstream* aus;
cEllipsoid ell;
cDatumTime dt;
string str;
cPodasList* pds;
TNak3Record* nak;
ifstream ein;
string ausname;
int output_counter;
int ix;
int reset_counter;
double latmins_outlier_tr1,lonmins_outlier_tr1;
double latmins_outlier_tr2,lonmins_outlier_tr2;
double latmins_outlier_extra,lonmins_outlier_extra;
double max_err_tr1, max_err_tr2, max_err_extra;
//statistic calculation
int exist_MINS;
int bad_rec;
int exist_Tr1;
int exist_Tr2;
int exist_heading;
int exist_roll;
int exist_pitch;
int outlier_count;
int good_rec;
int total_rec;
int transformation_count;

protected:
};
```

Figure 27: The header file of the Analysis functions class that shows the different variables used in the software.

The functions described in the header file provide all the necessary implementations of the algorithm described in Figure 22.

Moreover, two libraries had been implemented in the software for special purposes. The first library is *NEWMAT 11 beta*. It is a library used for fast matrix operations and matrix algebra, and it is compatible with Borland environment that can be easily used for fast computations. Full details about the documentations and the installation and usage of the library could be found in (Davies, 2008).

Another library had been used in the application for the interpolation and extrapolation routines. This library is named *ALGLIB*, and it is a C++ library that consists of eleven packages for different computations that are useful for data analysis. The regression fitting routines as well as the extrapolation procedures were used from this library which also provides the useful error computations for time series datasets. Detailed information and documentations of the library are found in (Bochkanov, 1999).

Further, the software has two possibilities for working either in the online mode when the software is implemented onboard R/V Polarstern, or in the offline mode where the post processing of old files is desired.

4.1 Online mode

The online mode allows the user to connect to the network onboard R/V Polarstern to receive the NMEA stream and start analyzing the data. Depending on the specific expedition that the analysis will be applied for, the user should choose an output file in which the results will be copied to. The application had been built with orders. This means that in order to start the process in the online mode the user should specify the output file otherwise he will not be able to press the connection button. This will prevent the overwriting of older output files. The full path of the output will appear in the online window.

Furthermore, the log information are presented in a field that consists on the following information:

- The starting and the ending time of the process.
- The author who manages the processing procedure.
- Statistical information about the data like the number of the number of outliers that have been corrected.

This information could be produced whenever the process is ended, and could also be saved in a “.log” file that is chosen by the operator.

Figure 28 shows the online window in the software while processing stream data generated by the simulator of D-ship system.

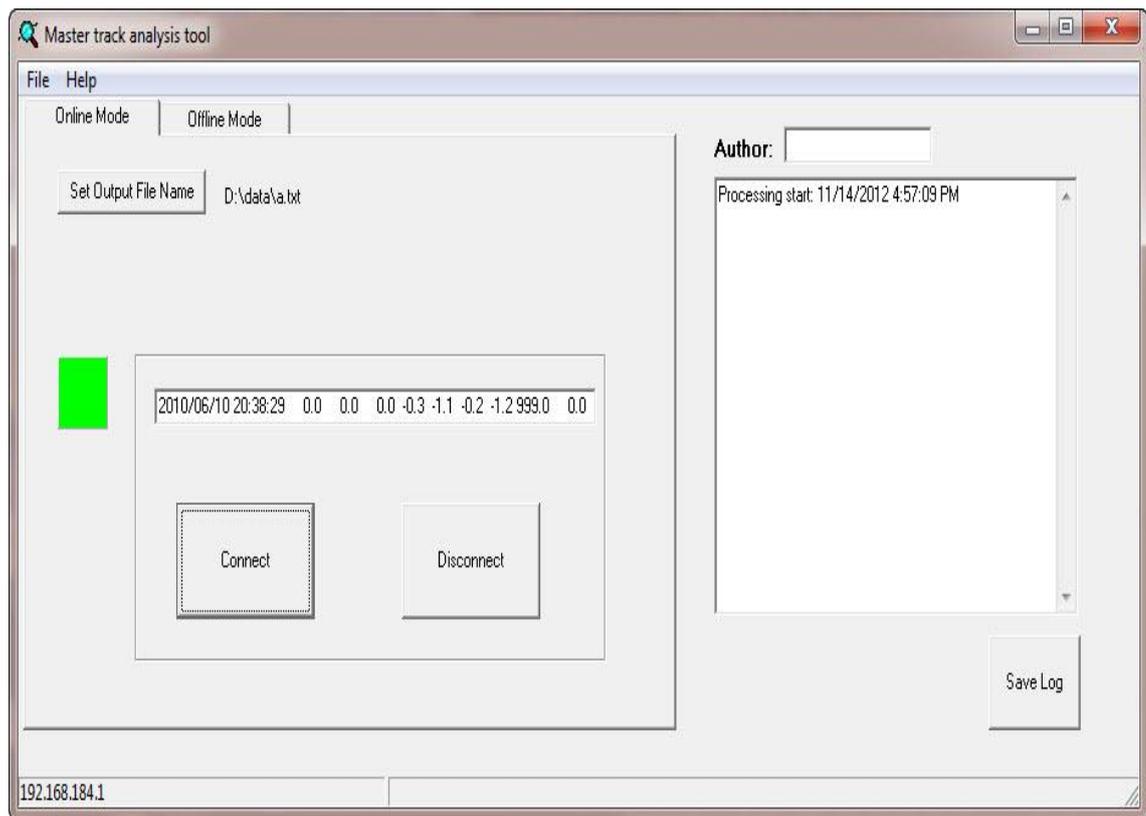


Figure 28: The online processing window while processing a dataset produced as stream using the D-ship simulator.

The visible control of the online processing window consists of a visible light which shows that the connection to the server is ok. It also consists of a text field that presents the received stream from the network.

4.2 Offline mode

The offline processing window has almost the same appearance of the online window. Here the user should also follow controlled orders in order to complete the process. For example, the user cannot press the analyze button before he goes through the steps of choosing the input file which needs to be analyzed, and the output file in which the results are to be copied to. Moreover, saving the statistical results is following the same procedure that is done in the online mode. The information about these statistical results is shown in a text field that can be saved to an output file.

However, due to the big size of the input files that could be processed and in order to let the user control the end of the process, the stop button allows the user to finish the analysis at any time if he is interested in analyzing specific amount of data inside the input file. The visibility control is added as a progress bar which shows the progress of the analysis.

In the next figure the offline mode was activated for analyzing an input file.

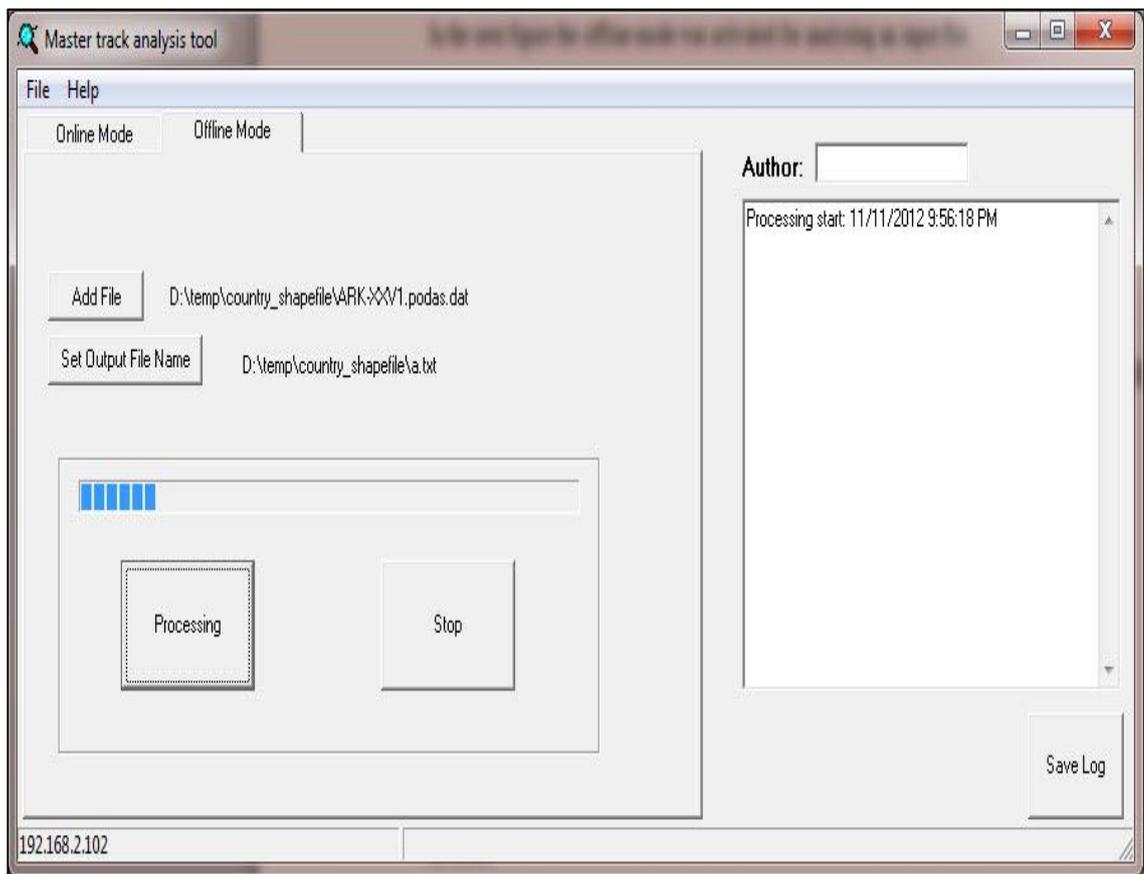


Figure 29: the offline window while processing an input file.

However we should emphasize that not all the expeditions of R/V Polarstern have the same file format, and our application is working only with the format “*PodasNewBathyFormat*” that can be chosen when downloading the files for processing. This format was replaced from 2003 and is available till now, while older formats are not in use but still available such as “*PoldatNewBathyFormat*”. However, the software needs some adjustments in order to be able to process those older formats.

4.3 Implementation onboard R/V Polarstern

This software is built for the implementation onboard R/V Polarstern, and at the time of this work the implementation onboard was not possible. However the software was testing using the simulator that simulates the D-Ship server onboard the vessel. It only needs some configurations regarding the connection methods with the server to work adequately. This will depend on the possible connections provided onboard the ship.

Moreover, the online implementation of the application can be applied using two methods. It can be used either as a standalone program that needs an operator for controlling the software and analyzing the data, or it could be implemented directly in the D-Ship server by implementing the algorithms inside the server to be part of its functionality. This shows the flexibility of the algorithm, where using classes with standard C++ language makes all functions and methods used in the software flexible to

be implemented directly into the server.

This software is a special tool for Master track generation of the R/V Polarstern. However the concepts of the algorithms used could be also applied for other research vessels to analyze the data and produce the Master track. However, this will need some configurations for the software and more efforts should be applied on making the application applicable for other vessels where completely different systems are used.

4.3 Outline

In this chapter we have clarified how the software was developed. We also showed different classes and libraries that have been used in the program. The header file of the Analysis class was introduced with all functions and variables.

Moreover, we have explained the possible implementation of the software and different windows in the interface and their functionalities. In the next chapter we will discuss the results and the product of the software and the implementation of the Master track into PANGAEA.

Chapter5

Results and discussions

In the final chapter we will show some results generated by the software and discuss the functionality of the algorithms, and the improvements and corrections that have been applied. Different datasets have been chosen for this objective and some handmade errors have been generated in some real datasets to see the efficiency of the applied methods.

For visualizing out results, we have chosen a GIS application that called “Quantum GIS”. This choice was due to the reason that firstly the application is free to use besides it provides powerful tools for visualizing spatial data information.

5.1 Quantitative results of the final product

The first dataset was part of the Expedition of the Research Vessel "R/V Polarstern" to the Arctic in 2010 (ARK-XXV/1) where R/V Polarstern had started from Bremerhaven port to Longyearbyen. This dataset had been chosen because it shows different behaviors of the associated devices used in our analysis. Figure 30 shows this part of the dataset that have been used.



Figure 30: Part of the expedition (ARK-XXV/1) used for results visualizations.

The geometry of the devices are shown in the next figure where we recognize that the distance between the devices make the position of the MINS is not between the GPS antennas.

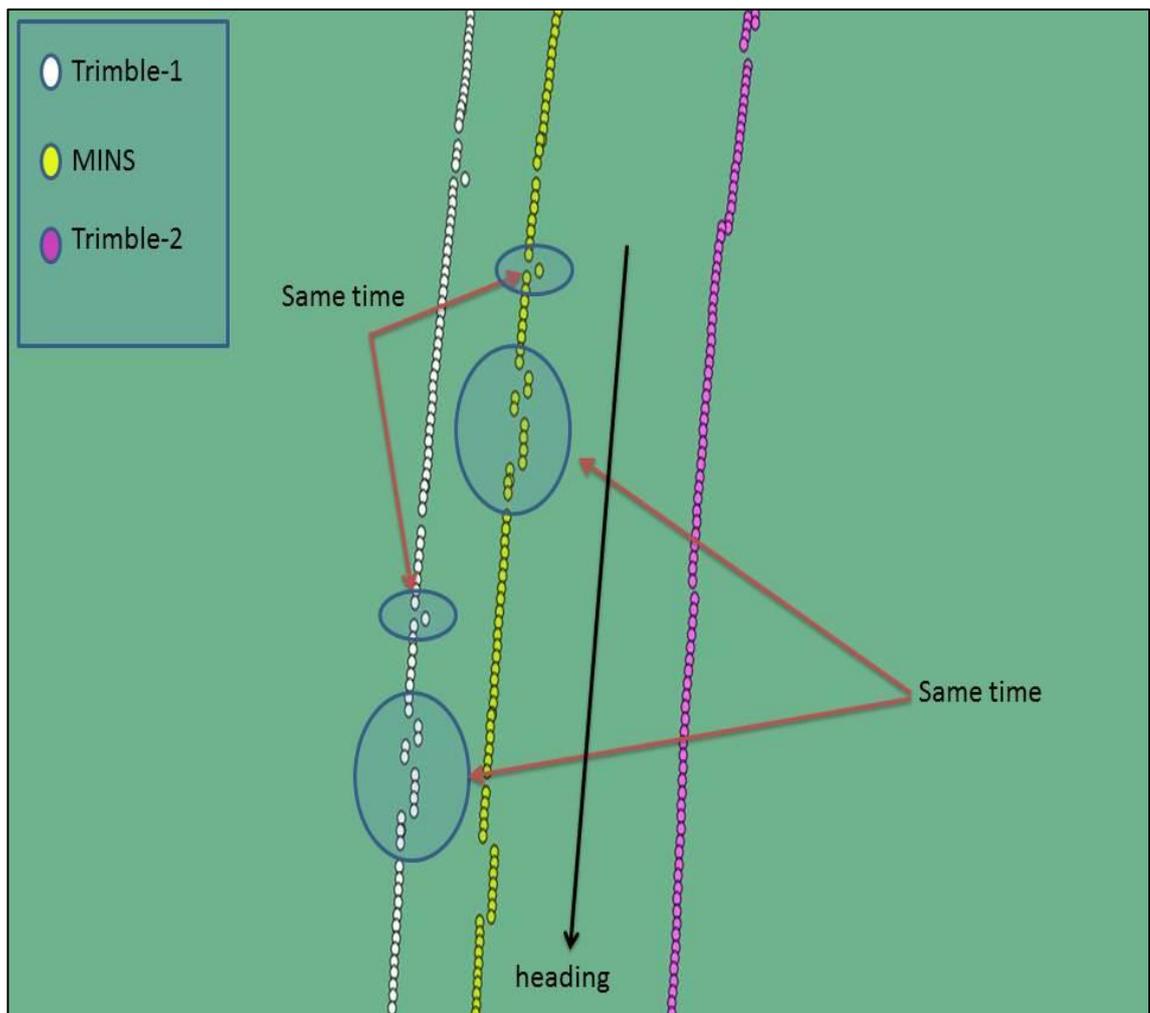


Figure 31: The effect of the geometric distance on visualizing the track of the devices.

We can see a symmetric behavior of the positions from MINS following the Trimble-1. In general it is observed that the MINS always follows the behavior of Trimble-1 antennas, which provides good accuracy. However in some cases this is not true as we will see in the next figures.

Figure 32 shows some errors produced by MINS.

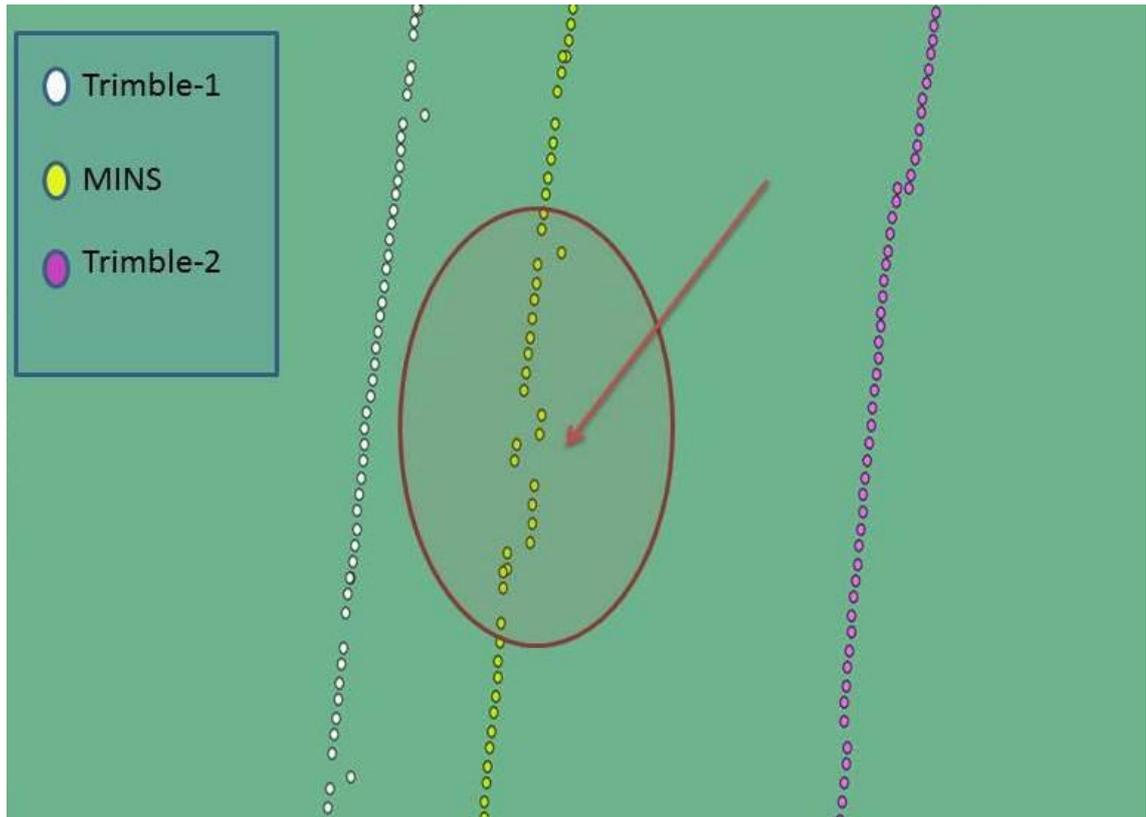


Figure 32: Error in the track produced by MINS.

We recognize small deviations made by the MINS. These small deviations may not be detected in the detection phase of the algorithm but the smoothing procedure will improve the position in this part as shown in the next figure.

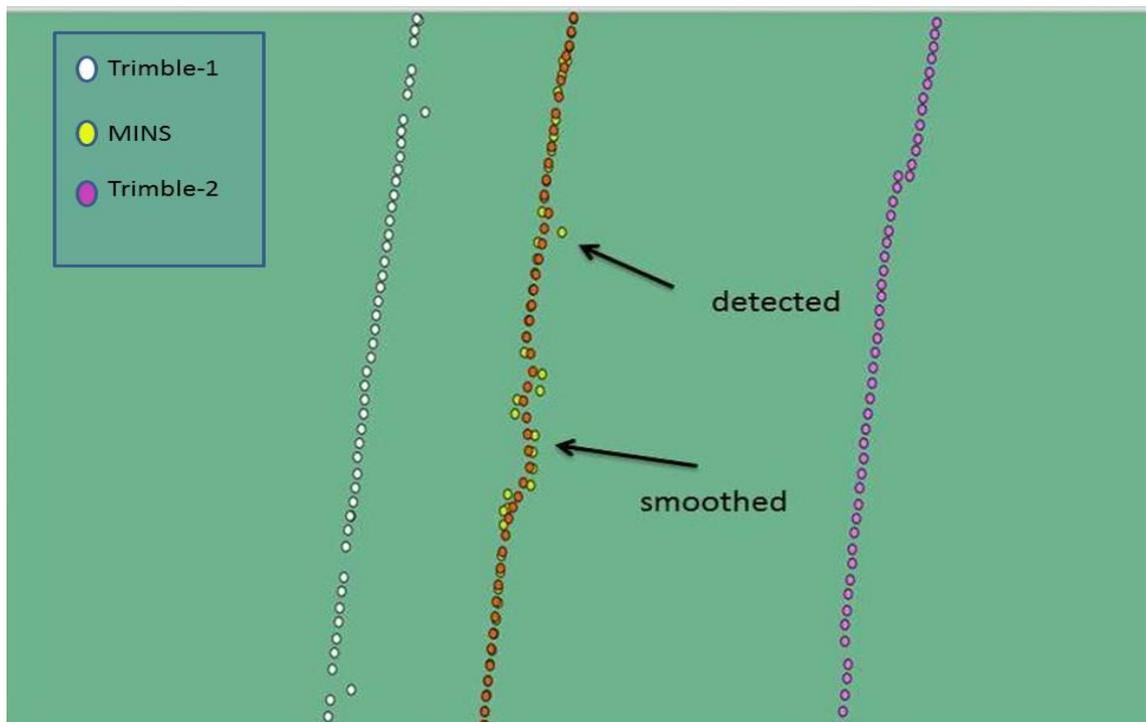


Figure 33: the smoothing routines of small deviations

Another interesting situation follows with the next figure where the MINS is following the behavior of the Trimble-1 GPS antenna which in addition produces some gabs in the positions.



Figure 34: error produced by Trimble antenna and a similar behavior followed by MINS.

The MINS is producing exactly the same track with the same gabs and errors produced by Trimble-1 GPS. Clearly the Trimble-2 GPS is providing better accuracy in this situation and thus the replacement of the outlier points will use its position to compute the new track. The resulting correction is shown in the next figure.

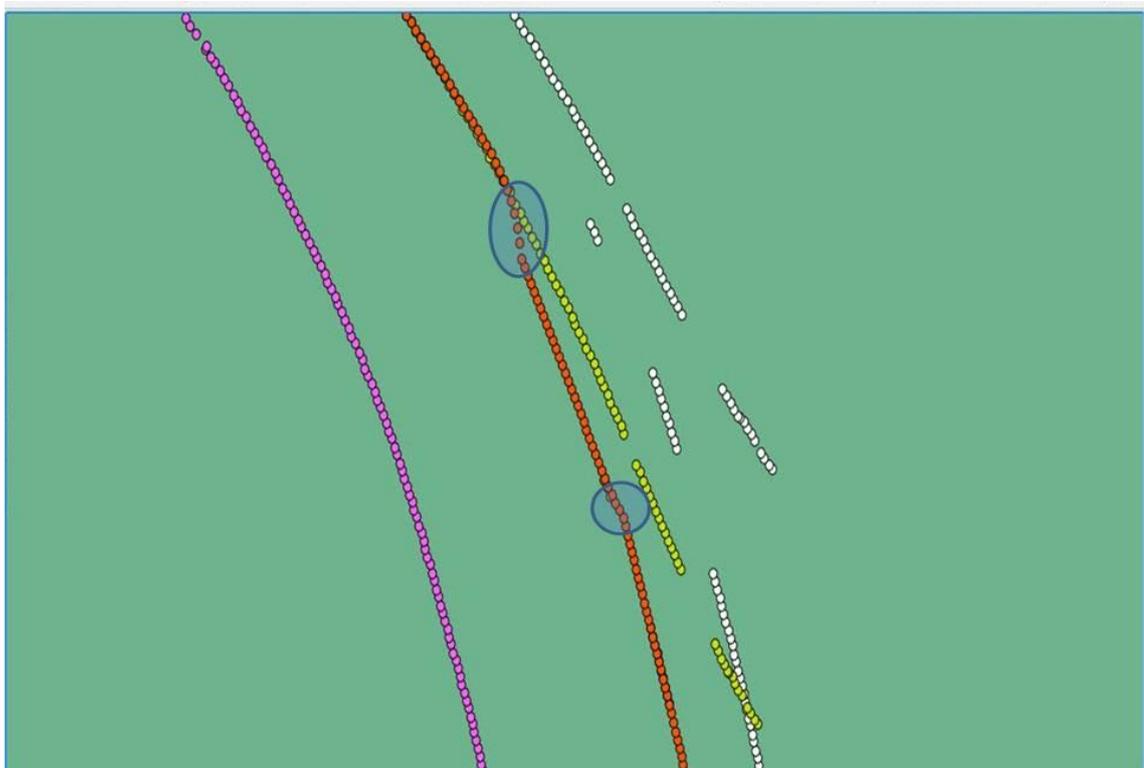


Figure 35: the correction applied to the MINS by the transformation.

The correction in the previous figure, which is represented in red, shows a much better positioning from the original MINS track in yellow. However at the start and the end of the correction part we see a small shift of the line, and this indicates that the transformation from the Trimble-2 antenna is not perfect, and maybe the lever arms provided for this transformation is not correct and it needs to be corrected by another alignment of the vessel.

Another behavior has been tested to see the improvement of the track when the MINS is producing a sequence of positions with shifting error. This behavior have been built by hand to see the solution as visualizing the datasets did not help too much in finding a similar situation. However this shift in the position is similar to what we have seen in the actual data in Figure 2. The next figure shows this situation but in smaller distance.

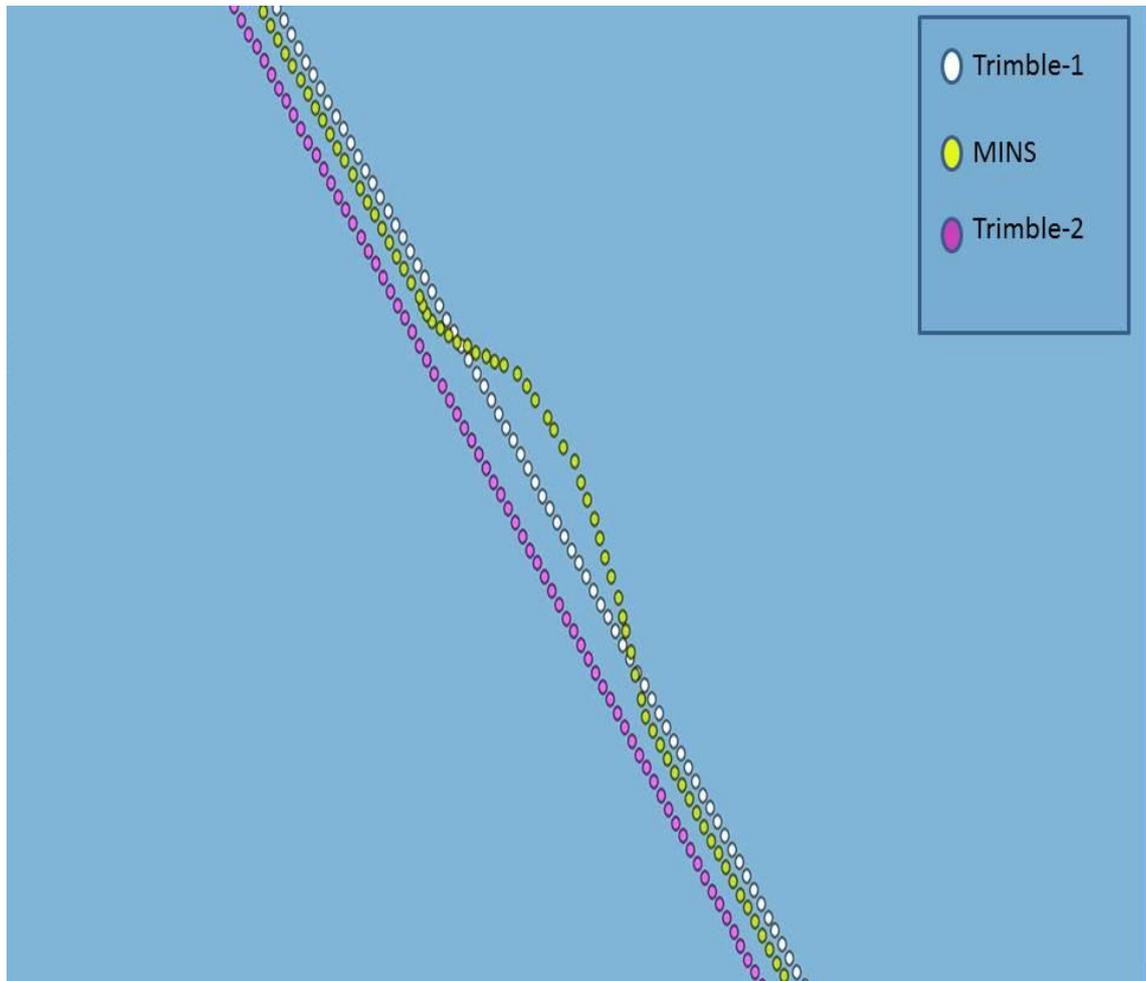


Figure 36: Shifted positions in the MINS track.

The solution had produced an ideal replacement of the error positions and this is shown in Figure 37. The red points fit exactly the original MINS positions before the handmade errors are added to the data, which means that the solutions are working appropriately.

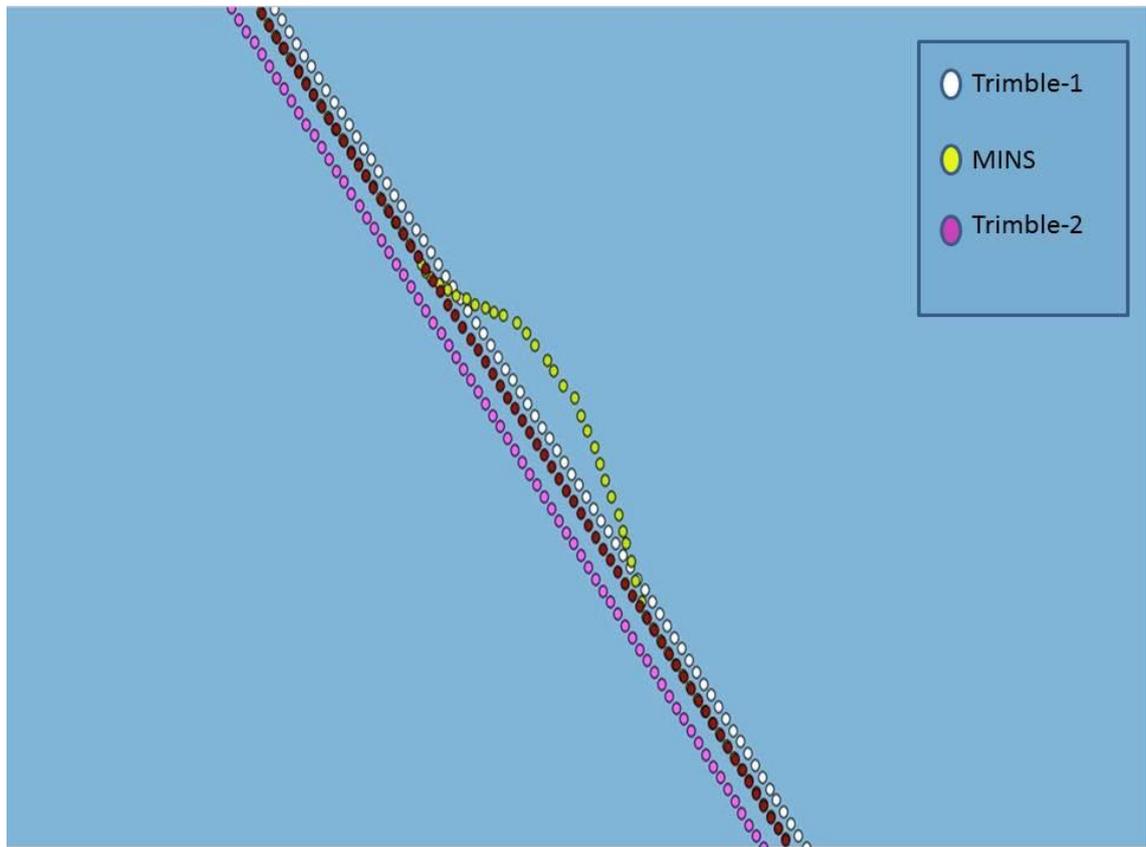


Figure 37: the correction of the shifting positions produced by the algorithms.

However the algorithms worked perfectly dealing with the rough outliers associated with the data, and the thresholds and conditions had been chosen to fit with this objective. Nevertheless the small outliers that are not detected in the detection procedures would affect the results with small deviation toward the outlier point, as a result of the smoothing procedure that still provides enhancements for these kinds of errors up to some levels.

In the next figure missing positions from the MINS were generated in the data. Around 2 minutes of contiguous positions were deleted from the dataset as shown in the figure.

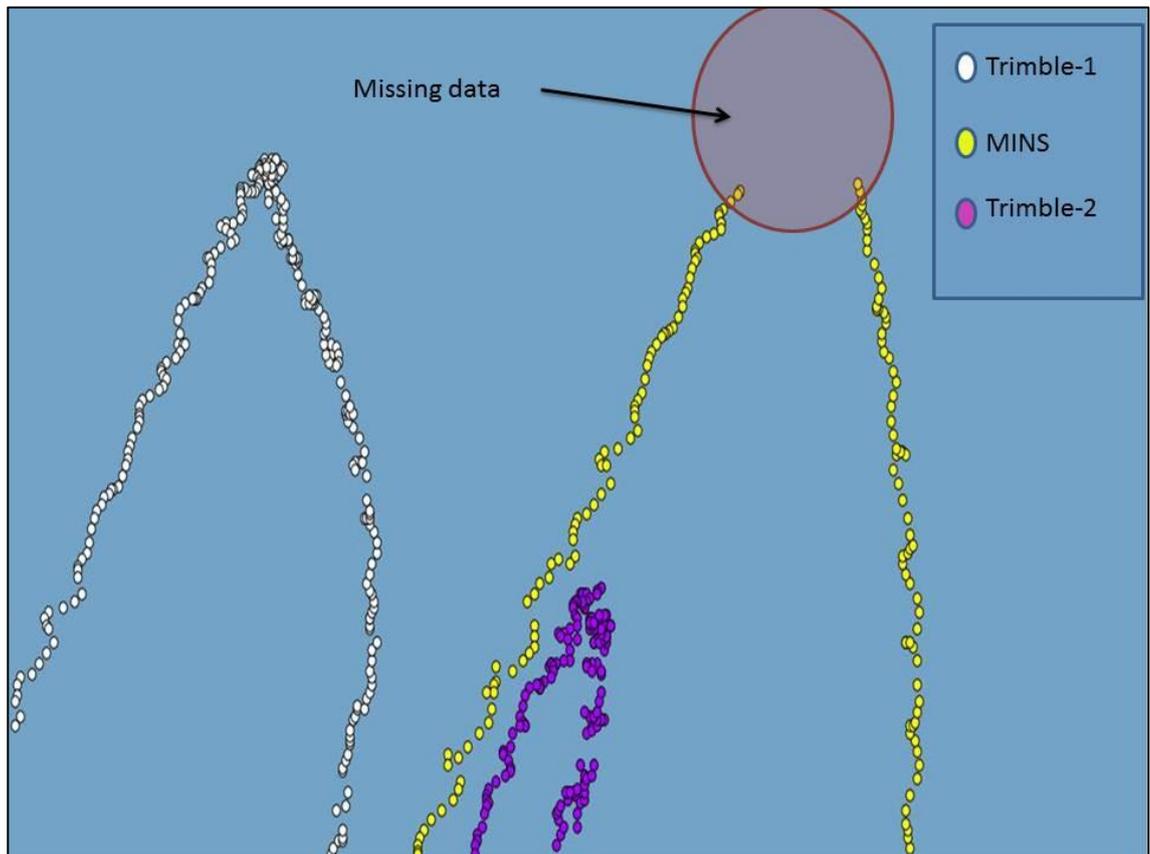


Figure 38: Missing data generated in the MINS positions.

The results of this section of data show good replacements of the missing points. However, it was intended to process these results when the ship was turning to see the efficiency of the solutions. Next figure shows the resulted data after the analysis is done.

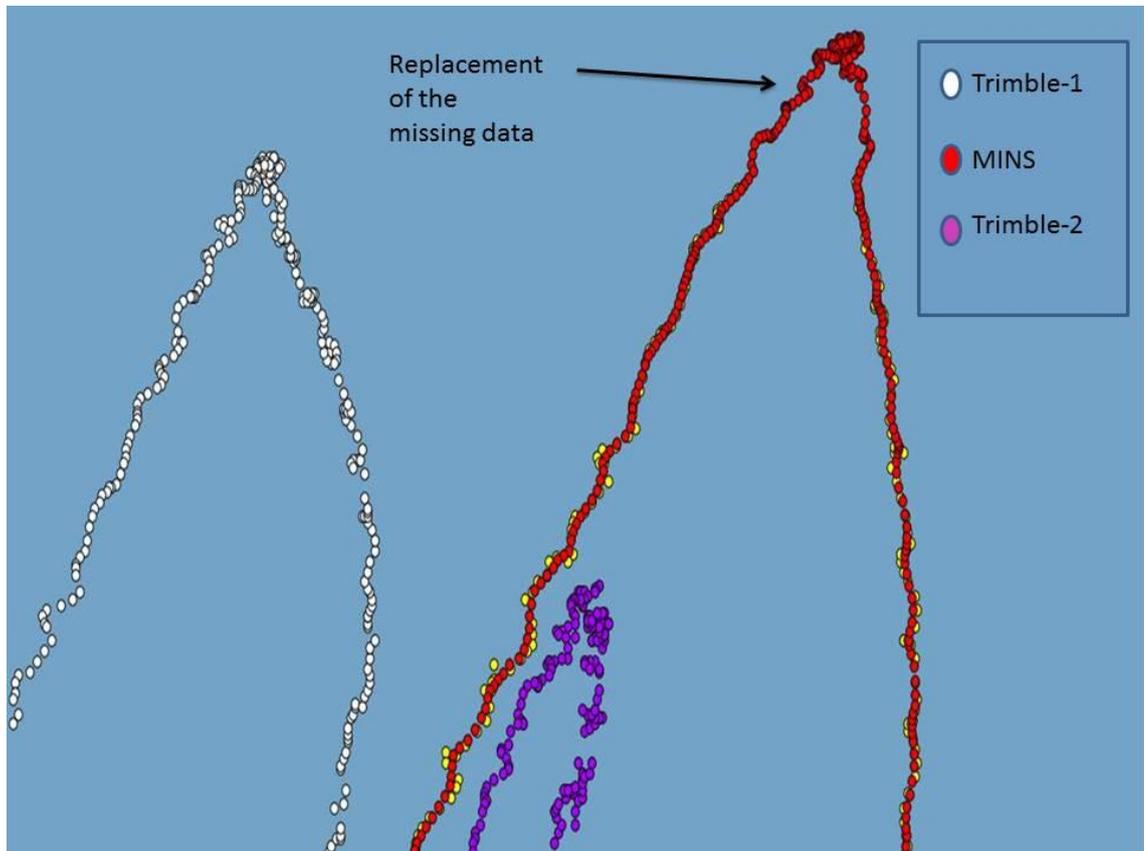


Figure 39: the replacement of the missing data produced by the application.

Until here we have discussed different datasets with the resulted Master track produced by the application. However, the statistical information is another output of the software that provides useful information about the data that have been analyzed. This information could be saved in a separate file where information is presented about the starting and the ending time of the process. Also the number of existed data provided by the devices is presented in the log file which is shown in the next figure. This information could be useful as statistical information about the performance of different devices used in the analysis,

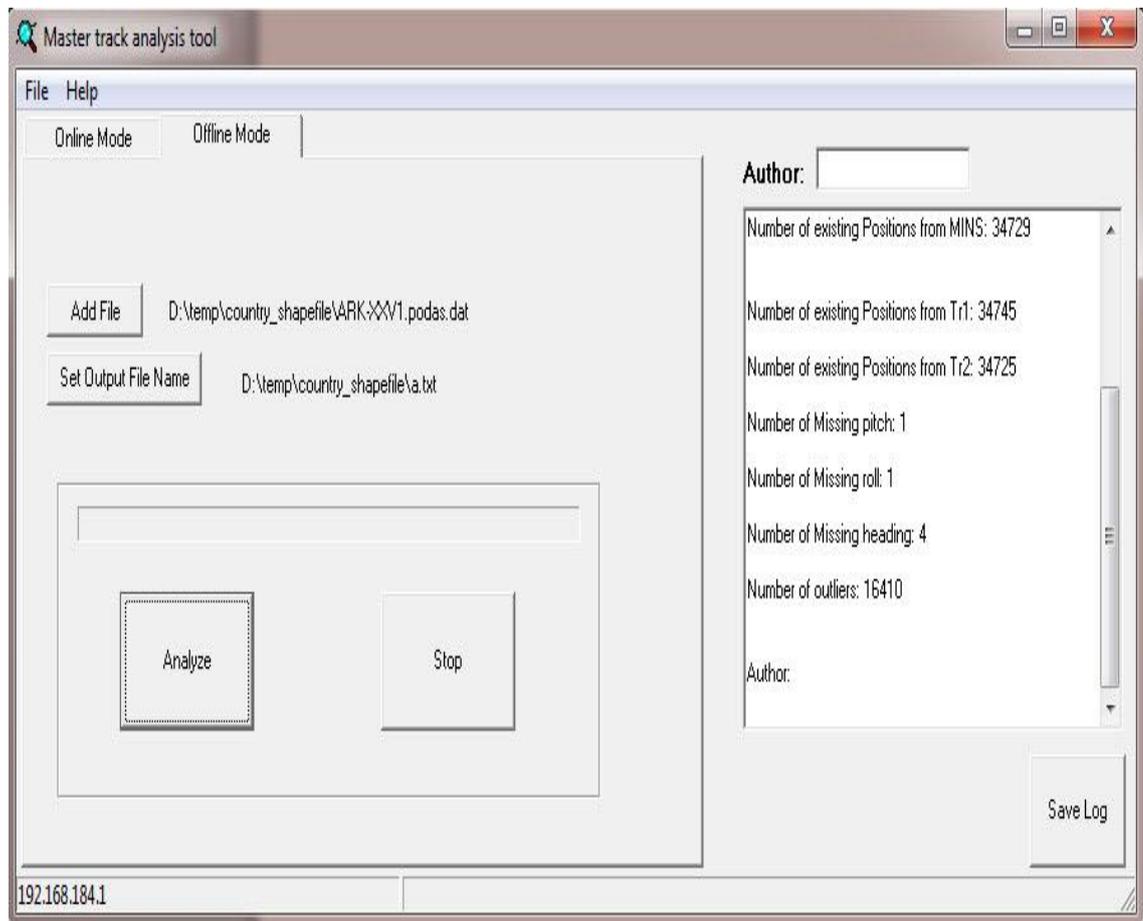


Figure 40: information that is presented in the log file at the end of the process,

5.2 meeting PANGAEA specifications

At the end of the work a sample dataset had been processed with the software and an output file had been generated. The sample dataset was the final destination in the Expedition of the Research Vessel "R/V Polarstern" to the Antarctic in 2010 (ANT-XXVI/4). The output file was compatible with the specifications of PANGAEA environment discussed earlier in chapter 3. The sample output file has been implemented into PANGAEA. Accessing to the dataset is available with the unique DOI (<http://doi.pangaea.de/10.1594/PANGAEA.793144>). The header information specified in Figure 24 is met and the output file had a header that specifies the author who produced the results as well as the processing date and time. The header also included the information about the columns presented in the output file as well.

Figure 41 shows the sample output file that had been implemented into the PANGAEA storage system.

Not logged in (log in or sign up)

PANGAEA*
Data Publisher for Earth & Environmental Science

Always quote citation when using data!

Data Description [Show Map](#) [Google Earth](#)

Citation: Aljuhne, Ammar; Krockner, Ralf; Dorschel, Boris (2012): Full resolution master track of POLARSTERN cruise PS75 (ANT-XXVI/4). Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Dataset #793144 (DOI registration in progress)

Reference(s): Körtzinger, Arne (2011): The Expedition of the Research Vessel Polarstern to the Antarctic in 2010 (ANT-XXVI/4). *Berichte zur Polar- und Meeresforschung = Reports on Polar and Marine Research*, 631, 73 pp, hdl:10013/epic.37888

Coverage: Median Latitude: 0.209630 * Median Longitude: -31.180485 * South-bound Latitude: -53.144700 * West-bound Longitude: -70.909100 * North-bound Latitude: 53.563960 * East-bound Longitude: 8.548130
Minimum Elevation: 20000.0 m * Maximum Elevation: -20000.0 m

Event(s): PS75/4-track * Latitude Start: -53.144700 * Longitude Start: -70.909100 * Latitude End: 53.563960 * Longitude End: 8.548130 * Date/Time Start: 2010-04-07T12:00:00 * Date/Time End: 2010-05-17T12:00:00 * Location: Atlantic, transit cruise * Campaign: ANT-XXVI/4 (PS75) * Basis: Polarstern * Device: Underway cruise track measurements

Other version: doi:10.1594/PANGAEA.793147

Comment: The dataset consists of seven columns: Date/time, latitude, longitude, heading, roll, pitch, standard deviation. Positions are calculated from Marine Inertial Navigation System (MINS), taking GPS coordinates and relative distances from 3D accelerometer into account. The coordinates are centered to MINS, which is nearly located in the pivot point of the ship. The parameters have been examined against rough errors and outliers. The master track is provided with interval of one second. See other version link for generalized version of this dataset.

License:  Creative Commons Attribution 3.0 Unported

Size: 37.0 MBytes

Download Data

[Download dataset](#)

[Contact](#)



Figure 41: The implementation of the Master track of the cruise (ANT-XXVI/4) into PANGAEA.

Due to the huge size of the Master track files as they have records with one second interval for long expeditions, this would limit the ability of visualizing the data, and therefore a generalized track had been used as another version of the data where a tremendous decreasing of the size had been achieved. The original Master track file had 3500000 records in total, and the final generalized track had 1044 points only. The generalized track had been implemented as another version of the Master track file and the next figure shows the other version of the Master track output which is also available with the DOI (<http://doi.pangaea.de/10.1594/PANGAEA.793147>). This generalized track could be visualized easily as seen in the figure and the records could be distributed as HTML, or could be downloaded as a tab delimited text file.

Not logged in (log in or sign up)

PANGAEA®
Data Publisher for Earth & Environmental Science

Always quote citation when using data! [Show Map](#) [Google Earth](#)

Data Description

Citation: Aljuhne, Ammar; Krockner, Ralf; Dorschel, Boris (2012): Generalized master track of POLARSTERN cruise PS75 (ANT-XXVII/4). *Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Dataset #793147 (DOI registration in progress)*

Reference(s): Körtzinger, Arne (2011): The Expedition of the Research Vessel Polarstern to the Antarctic in 2010 (ANT-XXVI/4). *Berichte zur Polar- und Meeresforschung = Reports on Polar and Marine Research*, 631, 73 pp, hdl:10013/epic.37888

Coverage: Median Latitude: 5.687521 * Median Longitude: -27.035859 * South-bound Latitude: -53.125798 * West-bound Longitude: -70.858334 * North-bound Latitude: 53.897340 * East-bound Longitude: 8.555033
Date/Time Start: 2010-04-07T00:00:01 * Date/Time End: 2010-05-17T05:19:50

Event(s): PS75/4-track * Latitude Start: -53.144700 * Longitude Start: -70.909100 * Latitude End: 53.563960 * Longitude End: 8.548130 * Date/Time Start: 2010-04-07T12:00:00 * Date/Time End: 2010-05-17T12:00:00 * Location: Atlantic, transit cruise * Campaign: ANT-XXVII/4 (PS75) * Basis: Polarstern * Device: Underway cruise track measurements

Other version: http://hs.pangaea.de/nav/mastertrack/PS75_4_generalized_master_track_shp.zip

Comment: Generalized from master track (doi:10.1594/PANGAEA.793144) executing Ramer-Douglas-Peucker algorithm (http://en.wikipedia.org/wiki/Ramer-Douglas-Peucker_algorithm) with maximum track deviation of plus/minus 50 meter. See other version link for ESRI shapefile of this dataset.

Parameter(s):

#	Name	Short Name	Unit	Principal Investigator	Method	Comment
1	DATE/TIME	Date/Time				Geocode
2	LATITUDE	Latitude				Geocode
3	LONGITUDE	Longitude				Geocode
4	Heading	Head	deg	Dorschel, Boris		

License: Creative Commons Attribution 3.0 Unported

Size: 1031 data points



Figure 42: The generalized track of the cruise (ANT-XXVII/4) and its implementation in PANGAEA.

5.3 assessment of the performance

The software was developed using classes and simple computation of different tests and procedures in order to provide the Master track of the vessel. The implementation of the application shows fast computation and performance.

In the offline mode the software was able to complete the analysis of huge files with adequate time. Approximately around 1000 records were analyzed per second which means that a complete cruise that consists of around 4000000 records could be processed with 4000 seconds or around 1 hour and 15 minutes.

However, the same computations are applied in the online process because the inputs in both modes are the same. This means that the application is providing a good analysis method that is compatible with the time of the online process.

Moreover, the visibility control of the software is a simple tools indicating that the software is working properly. In the offline mode the progress bar gives the indication that the process is running and shows also the approximate time of the ending time. In the other hand, the online mode has only a visible green color that indicates that the connection is working fine. However more visualization of the process would improve the visibility of the software. Thus improving the visualization tools of the application could be a new scope for further development. This could include for example a simple graphic visualization of the track. These improvements could slow the performance of the application. Thus using threads for parallel computation could improve the performance of the software when a graphic visualization is going to be used.

5.3 Outline

In the final chapter we have discussed different results of the application, and we have seen the improvements that have been produced to the original data by applying the correction algorithms provided in the software.

Smoothed results were presented when small deviations are shown in the data, and a replacement of missing values is done using the correction procedures of the application. Moreover the shift behavior of the MINS was discussed and the solution was presented as well. Moreover, the statistical information was also produced and presented in this chapter.

At the end the Master track and the generalized master track were implemented into PANGAEA, and the DOI of each file was presented as well.

In the next chapter we will conclude this work with a summary of what have been done in this thesis, and suggestions for further development of the software will be given.

Chapter6

Conclusions

The AWI had the idea of storing the navigation tracks of the R/V Polarstern in PANGAEA system. Some investigations showed some kind of erroneous behavior of the positioning systems onboard the vessel, and thus an analysis of these tracks had to be applied before storing them in PANGAEA to detect and correct the rough outliers associated with the data.

The objectives of this thesis were divided into two parts: on the one hand, this work focused on the analysis of the old tracks of R/V Polarstern and had provided a software application that analyzes the tracks and produces evaluated Master tracks of R/V Polarstern. On the other hand, this work has extended the objectives of analyzing old tracks of the vessel, and had provided online assessment for the under way tracks of the vessel using online processing.

The algorithms used in this thesis work similarly either in the online or the offline mode, where different tests examine the inputs and provide correction solutions for the outlying points. Moreover, smoothing routines were applied on the data using statistical methods such as regression polynomial fits. The navigation systems were studied before the analysis to understand the relations and the possible correlations between those systems.

Polynomial regression methods were used for this analysis, and some extrapolation procedures were applied for correcting the attitude parameters provided by the MINS. This brought us to another idea that could be considered in the further development of the software. We were using only two GPS antennas for our analysis, if a third GPS antenna is provided in the future, the correction of the roll and pitch could be applied by using the three antennas to compute the attitude of the ship and compare it with the values provided by the MINS. This method was not appropriate in our study because the ASHTECH receiver onboard the vessel was not working continuously during the older expeditions. Moreover, the Kalman Filtering techniques are another method that could be applied for same analysis, and it could be useful in the future to provide the analysis with this method to compare the results and see the advantages and disadvantages of both methods.

The application used a decision based filter that examines the data provided by the Marine Inertial Navigation System MINS to check of any rough errors. The corrections were applied on the outliers and the original data that are not erroneous were kept. Finally the Master track was provided and implemented successfully into PANGAEA.

However, this software was built specially for the R/V Polarstern. Further development could be done for globalizing the software and to make it applicable on other research vessels. This needs some configuration of the algorithms to correctly analyze different inputs from different navigation systems. Moreover, the software is compatible only with the files of a specific format. It also needs further development regarding the format of the input files.

The visualization of the resulted Master track could be a new scope for further development of this application, and providing visual picture of what is going on in real

time brings more credits to the analysis. This could slow down the performance of the software because the visualization of spatial data is a great consumer of the physical memory. Therefore using threads and parallel computations could be the solution of such difficulty.

In conclusion, the process of assessing the navigation track of a research vessel and providing evaluated information about the position of the ship during its expeditions has its own value because it touches the information that every scientist needs in marine science and improving the accuracy of this track reflects better results in all demands.

Bibliography

- www.pangaea.de*. (1998). Retrieved 2012, from www.pangaea.de/about/
- IMO-IMA 4th Course on Nautical Cartography. (2003). *IMO-IMA 4th Course on Nautical Cartography* (pp. 34-42). Trieste: IMO-IMA.
- AWI. (1980). <http://www.awi.de/en/infrastructure/ships/polarstern/>. Retrieved March 2012, from <http://www.awi.de/>:
<http://www.awi.de/en/infrastructure/ships/polarstern/>
- Bochkanov, S. (1999). *alglib open source*. Retrieved June 22, 2012, from Alglib numerical analysis library: www.alglib.net
- Bumke, K. (2011). *The Expedition of the Research Vessel "Polarstern" to the Antarctic in 2010 (ANT-XXVII/1), Berichte zur Polar- und Meeresforschung (Reports on Polar and Marine Research)*. Bremerhaven: Alfred Wegener Institute.
- Davies, R. B. (2008). *Newmat C++ matrix library*. Retrieved March 15, 2012, from Robert Davies: http://www.robertnz.net/nm_intro.htm
- El Naggar, S. F. (2007). *RV Polarstern handbook*. Bremerhaven: AWI.
- Fan, H. (2010). *Theory of Errors and Least Squares Adjustment*. Stockholm: Tekniska högskolan, 1997.
- FGDC. (2011, December 20). <http://www.fgdc.gov/>. Retrieved May 2012, from <http://www.fgdc.gov/>: http://www.fgdc.gov/index_html
- Ghilani, C. D. (2005, October 05). *Penn State Wilkes-Barre*. Retrieved August 05, 2012, from Geodetic modules: <http://surveying.wb.psu.edu/sur351/syllabus.htm>
- Grobe, H. D. (2006). *Archiving and distributing earth-science data with the PANGAEA information system*. Bremerhaven: Antarctica : contributions to global earth sciences ; Proceedings of the IX International Symposium of Antarctic Earth Sciences Potsdam, 2003 / Hrsg. Dieter Fütterer; Detlef Damaske; Georg Kleinschmidt, Hubert Miller, Franz Tessensohn; Springer, Berlin.
- Hannes Grobe, M. D. (2005, 11 01). *PangaWiki*. Retrieved September 13, 2012, from PangaWiki: http://wiki.pangaea.de/wiki/Main_Page
- Iffland, A. (2004). *Aufarbeitung und Visualisierung einer bathymetrischen Vermessung in Verbindung mit Seismogrammen der Sedimentechographie*. Bremerhaven: Alfred Wegener Institute.
- King, A. (1998). *Inertial Navigation – Forty Years of Evolution*. Coventry: GEC Review.
- Peterson Ray, H. A. (2008). *Profisional Pilot. The learn to fly website*. Retrieved August 13, 2012, from Aircraft systems and Electronics:
<http://selair.selkirk.ca/Training/systems/index.html>
- Peucker, D. D. (2012, September 28). *Ramer–Douglas–Peucker algorithm*. Retrieved

August 12, 2012, from Wikipedia:
http://en.wikipedia.org/wiki/Ramer%E2%80%93Douglas%E2%80%93Peucker_algorithm#cite_ref-0

Ron A. Cooper, T. J. (1983). *Data, Models, and Statistical Analysis*. Pittsburgh: Rowman & Littlefield.

Rowe, A. W. (1996). *High-Accuracy distributed sensor time-space-position information system for captive-carry field experiments*. California: Naval Postgraduate School.

Sarafidis Dimitrios, P. I. (2006). A tool for managing ISO 19115 compliant metadata for the spatial. *21st European Conference for ESRI Users* (pp. 1-2). Athens: ESRI.

Schäfer, A. (2011, 10 06). *Marine Daten Infrastruktur Deutschland*. Retrieved 07 11, 2012, from Marine Daten Infrastruktur Deutschland:
http://139.30.111.16/WebsiteMDI-DE/Veranstaltungen/Praesentationen/Workshop_I/20111013-MDI-DE-WS1-MaNIDA-Portal-Deutsche-Meeresforschung_Pfeiffenberger.pdf

Schenke, H. W. (2006). *The Expeditions ANTARKTIS-XXII/4 and ANTARKTIS-XXII/5 of the Research Vessel "Polarstern" in 2005, Berichte zur Polar- und Meeresforschung (Reports on Polar and Marine Research)*. Bremerhaven: AWI.

SURVEYORS, O. &. (2010). *SURVEY DOCUMENTATION*. Bremerhaven: OVERATH & SAND SHIP SURVEYORS.

Walter, G. (1964). *Geodätische Rechnungen und Abteilungen in der landesvermessung*. Stuttgart: K. Wittwer.

Appendix A

Contents of the CD

- Digital version of the thesis work
- Testing data folder
- The software application folder
- Quantum GIS installation folder
- Generalized track application folder
- D-Ship simulator folder

