

NL ARMS

Netherlands Annual Review of
Military Studies 2010

Advances in Military Navigation Technology

Patrick Oonincx
Anton Scheele
[eds.]

NL-ARMS is published under the auspices of the Dean of the Faculty of Military Sciences of the Netherlands Defence Academy (NLDA).

For more information about NL-ARMS and/or additional copies contact the editors, or the Faculty Research Office of the Faculty of Military Sciences of the NLDA, at the address below:

Faculty of Military Sciences of the NLDA

Faculty Research Office

P.O. Box 90.002

4800 PA Breda

phone: +31 76 527 32 16

fax: +31 76 527 33 22

email: FH.OudeRikmanspoel@NLDA.nl

NL-ARMS

- 1997 The Bosnian Experience
JLM Soeters, JH Rovers (eds.)
- 1998 The Commander's Responsibility in Difficult Circumstances
ALW Vogelaar, KF Muusse, JH Rovers (eds.)
- 1999 Information Operations
JMJ Bosch, HAM Luijff, AR Mollema (eds.)
- 2000/01 Information in Context
HFM Kirkels, MV Metselaar, GCA Steenbakkens (eds.)
- 2002 Civil-Military Cooperation: A Marriage of Reason
MTI Bollen, RV Janssens, HFM Kirkels, JLM Soeters (eds.)
- 2003 Officer Education - The road to Athens
HFM Kirkels, W Klinkert, R Moelker (eds.)
- 2004 Defense Logistics - Winning Supply Chain Networks
HFM Kirkels, W Ploos van Amstel (eds.)
- 2005/06 Terrorist and counterterrorist operations
MGD Rothman, RJM Beeres, HFM Kirkels, JLM Soeters (eds.)
- 2007 Defense Accounting Control & Economics. Evidence from the Netherlands
RJM Beeres, EJ de Bakker, HFM Kirkels (eds.)
- 2008 Sensors, Weapons, C4I and Operations Research
ThM Hupkens, H Monsuur (eds.)
- 2009 Complex operations: Studies on Lebanon (2006) and Afghanistan (2006-..)
MJ de Weger, FPB Osinga, HFM Kirkels (eds.)
- 2010 Advances in Military Navigation Technology
PJ Oonincx, CA Scheele (eds.)

Cover design and group photography: Peter J. de Vries, Multimedia NLDA

Printed and bound by: Ten Brink B.V., Meppel, NL

ISSN: 1387-8050

Contents

Introduction	5
<i>Patrick Oonincx & Anton Scheele</i>	
Global Navigation Satellite Systems: Status, Plans and Threats	9
<i>Christian Tiberius</i>	
The World of Software Receivers for Satellite Navigation	21
<i>Benoit Muth</i>	
Software-based Interferometric Positioning	41
<i>Benoit Muth, Patrick Oonincx & Christian Tiberius</i>	
Terrain Referenced Navigation Using a GPS Approach	55
<i>Patrick Oonincx & Daniela Vaman</i>	
Base Line Performance of the Low Cost LORADD SP Receiver	73
<i>Cees de Groot & Anton Scheele</i>	
Performance Improvement of the Low Cost LORADD SP Receiver	93
<i>Cees de Groot & Anton Scheele</i>	
Determining the Hydrodynamic Derivatives of a Basic Model of the REMUS AUV	103
<i>Dick Engelbracht & Paul Wolkenfelt</i>	
On Line Measurement of High Frequency Ship Motions of Channel Bound Ships	125
<i>Dick Engelbracht & Wim van Buuren</i>	
Cross Wind Loads on Ships and Complex Structures	149
<i>Max van Hilten†, Dick Engelbracht & Paul Wolkenfelt</i>	
Applying NEC to UAS Operations Using an Evolutionary Approach	171
<i>Jochum Tadema, Armand Goossens & Erik Theunissen</i>	
Self Separation Support for Unmanned Aircraft Systems	197
<i>Jochum Tadema & Erik Theunissen</i>	
Crew Resource Management Training and Research in a Military Organization	221
<i>Hilde van Ginkel</i>	
Training Tools for Research	241
<i>Rico Bloemberg</i>	
Contributing Authors	253

Introduction

Patrick Oonincx & Anton Scheele

Historical background

The moment man began travelling the earth man started to navigate. Thus navigation may be considered to be one of the oldest sciences. Until the beginning of the 20th century navigation has almost exclusively been the realm of the nautical community, because navigation on land did not pose real problems and aviation had not yet emerged. History shows a constant effort by scientists, artisans and practising navigators to develop methods and instruments to find time and hence position at sea and to measure direction and speed. One of the first documented experiments regarding the earth is the measurement of Eratosthenes of Syene who in the third century before Christ assessed the circumference of the earth. Knowledge of the shape and dimension of the earth is paramount for navigation.

Driven by economical incentives and as a direct consequence of some major shipping disasters the 17th and 18th centuries show impressive and creative attempts to solve the problem of finding the geographical longitude at sea culminating in clocks that function properly during a long period of time on a moving ship and instruments that can be used to measure angles on a rolling and pitching ship. Positioning, timing and ship behaviour are the first topics that appear in navigation. The first articles in this volume of NL ARMS therefore deal with positioning and timing and ship behaviour.

Positioning and timing

We begin with an introductory article on the status of present and future Global Navigation Satellite Systems (GNSS), followed by an article on

practical implementation of the Global Positioning System (GPS) in a so-called software receiver and an article on the use of this receiver to measure the direction and distance (the baseline) between two positions.

The next article describes how methods and techniques developed for GPS receivers can be used for Terrain Referenced Navigation (TRN) which is a completely different way of positioning. In TRN one estimates the position by matching an observation of the shape and nature of the terrain (or sea bed) with the shape and nature of this terrain stored in a database. Especially for military navigation (which does not only deal with travelling from one position to another, but also with picture compilation and weapon delivery) TRN is of importance because it provides a means of positioning which is independent of GNSS. Independency of GNSS is urgently required because GNSS's are very susceptible to (non)deliberate interference (jamming).

Ensuring the availability and continuity of navigation systems for normal navigation and for picture compilation and weapon delivery is the reason why we started research into another system for positioning and timing, eLoran. Two articles give an overview of the performance of a low cost eLoran receiver which may be used as a back up for GNSS.

Ship behaviour

Navigation using the systems mentioned above and for instance Inertial Navigation Systems (INS) can be improved by using an accurate mathematical model of the platform in conjunction with knowledge of the forces acting on it. The first article on ship behaviour describes the estimation of coefficients of such a mathematical model for the REMUS Autonomous Underwater Vehicle.

The next article discusses research that was aimed at accurately estimating the Under Keel Clearance (UKC) of a ship in real time. The UKC is the distance between the keel of the ship and the sea bed. Accurate real time estimation of the UKC is important to avoid grounding which may have tremendous economical, environmental and operational consequences. Alternatively it might extend the tidal window within which ships with a certain draught are allowed to enter port or within which a military operation in shallow waters is to take place.

The third article deals with the influence of wind on the behaviour of a ship. Knowledge of the effect of wind is especially of interest for ships with a large wind sensitive surface above the waterline like the Landing Platform Docks HNLMS Rotterdam and HNLMS Johan de Witt and the future Joint Logistic Support Ship (JSS).

Navigation of unmanned vehicles

December 1903 marked the beginning of the aviation era giving navigation a new dimension: besides the horizontal dimensions latitude and longitude and the time dimension the dimension height became of more interest. Initially aircraft were manned but more and more unmanned aircraft (UAV, Unmanned Aerial Vehicles) are being used.

In this volume of NL ARMS two papers discuss current research going on in the UAV lab at the Netherlands Defence Academy. The first paper deals with the integration of UAVs, UAV Control Stations and Command & Control in a larger network, enabling seamless access to the desired platform and payload on a time-share basis. Different levels of connectivity in a network are considered and the Network Enabled Capability levels will be related to functions and connectivity. Based on experimental results, an analysis of existing datalink standards, the feasibility and the potential of the evolutionary approach will be discussed.

The second paper on this topic deals with the concepts Self Separation and Conflict Avoidance for UAVs. It begins with a presentation of eight criteria that should be considered when defining requirements for a future separation assurance and collision avoidance system. Using these criteria, it is illustrated how the concept of conflict probing and the associated scalability enables a range of possible implementations, specifically matched to the available data, interfaces and displays.

Crew resource management

It used to be that knowledge and skills to navigate a ship or aircraft were mainly the privilege of a few people (sometimes only one person) on board. This one person did all the observations, decision making and actual controlling of the direction, speed and if applicable height of the platform. Advice from others, however competent, was not always appreciated. The present speed and complexity of operations and the huge amount of in-

formation with which the navigator is presented make such an attitude irresponsible and dangerous. The last few decades we therefore see Crew Resource Management (CRM) becoming an essential and integral part of navigation.

The first article on CRM focusses on the importance of CRM in military operations and the necessity to lay a scientific foundation for CRM training and coaching. Results of research into the influence of team, individual and situational characteristics on team interaction and team outcome are presented.

The last article of this volume of the NL ARMS addresses the problem of the relation between the number of navigation systems a navigator can use and his navigational performance. Does he or she perform better when more information is available? These kinds of questions are very hard to address in the real navigation world. Therefore we used the Full Mission Bridge Simulator (FMBS) of the NLDA making the FMBS not only a very valuable tool for education and training but also for research.

Summary and epilogue

In this introduction we sketched the contents of this NL ARMS on the basis of the evolution of navigation from a restricted field of application (nautical) and executed by a select group of navigators working in almost 'splendid isolation' to a wide field of application (nautical, aeronautical, CRM) executed by groups of people sharing different knowledge and skills working as a team. This evolution has not yet come to an end. Since the beginning of the satellite navigation era navigation tends to become more and more common property. Civil (TomTom) as well as military (Battle Field Management Systems) applications on the level of a single person have been introduced on land increasing the field of navigation enormously. It is the ambition of NLDA's Navigation Technology department to serve as a think-tank in the field of navigational education, research, training and policy for the Netherlands armed forces.

The editors like to thank Mrs. Deborah Trimpe Burger-Hogg for proof reading and making corrections with respect to the English language. Furthermore, we are grateful to Mr. Peter de Vries, Multimedia office NLDA, for reproducing the group picture of the authors and the cover page of this year's NL ARMS.

Global Navigation Satellite Systems: Status, Plans and Threats

Christian Tiberius

Introduction

In this contribution the current status and future development of four Global Navigation Satellite Systems (GNSS) are reviewed. They are the American Global Positioning System (GPS), the Russian Glonass, the European Galileo and the Chinese BeiDou Compass.

All four systems offer, or will offer, civil navigation services (publicly open), as well as restricted access navigation services, which means for GPS, Glonass and likely for Compass as well, military services. The European Galileo system will offer so-called Publicly Regulated Services (PRS) for use by the government.

The basic principle of satellite navigation — referred to as standalone positioning, or single point positioning — consists of measuring, with a receiver on, or close to the Earth, distances from three satellites to the receiver, each time by observing the travel-time of the radio signal, as transmitted by the satellite. Three distances, to objects or points with known position coordinates — the satellites — allow us to determine the position of the user receiver, geometrically, in three dimensions. To this end, the GPS ground segment, with stations around the Earth, of which the position coordinates have been accurately established, and using the radio signals transmitted by the satellites as well, determines the positions of the satellites, predicts them ahead, and uploads the information to the satellites, so that they can relay this information to users all around the world.

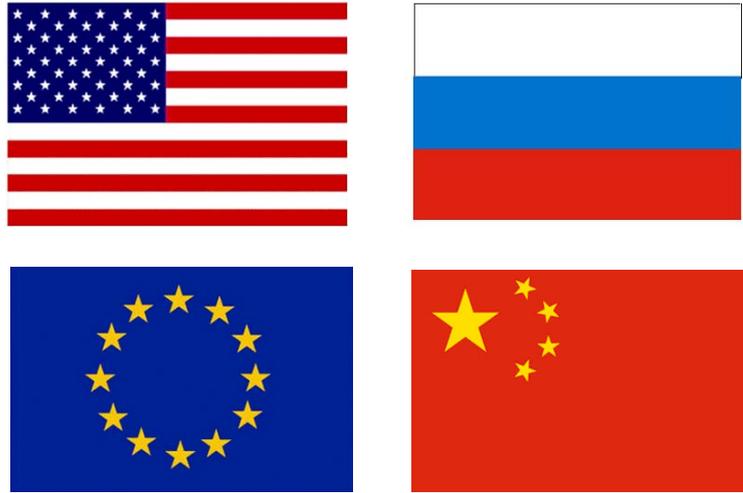


Figure 1: Nations currently operating and/or building a Global Navigation Satellite System (GNSS). The Global Positioning System (GPS) by the United States, Glonass by the Russian Federation, Galileo by Europe and BeiDou Compass by the PR of China.

One complication remains. The user receiver typically is not equipped with an accurate clock, like an atomic standard (which the satellites do have on board). When the user receiver clock is ahead, the measured distances are systematically too long, but all by the same amount. This problem is overcome by observing — simultaneously — at least a fourth satellite. With four distances, to satellites at known positions, the user receiver can determine its three-dimensional position, and its clock offset with respect to GPS system time as well. The resulting position accuracy lies in the order of 5–10 meter, under favourable circumstances, which is an open sky, and no tall obstacles, like buildings, blocking the view.

In this contribution it is outlined that in the present decade, global satellite navigation will mature, with diversity of system control, and a rich variety of signals, though by means of a small case study — demonstrating how, in an unlikely event, navigation service can go wrong — a warning is issued against over-reliance on available satellite navigation services.

Global Positioning System (GPS)

The Global Positioning System (GPS) has been developed by the US Department of Defense (DoD), and is operated by the US Air Force (USAF). The first satellite was launched in February 1978. Various generations of GPS satellites are Block I, II, IIA, IIR and currently IIR-M. From this year on, launches of block IIF satellites are scheduled, and third genera-

tion GPS block III satellites are scheduled from 2014 onwards. Last year, the US Government Accountability Office (GAO) issued a report on GPS, doubting whether the Air Force will be able to acquire new satellites in time to maintain GPS service without interruption. According to analyses by the GAO, the constellation may fall below the nominal number of 24 satellites, in the years to come.



Figure 2: GPS block IIF satellite, to be launched from 2010 on. This series of satellites offers triple frequency civil signals.

The ground-segment of GPS originally consisted of five stations (world-wide), and one at Cape Canaveral. In the meantime six ground-stations have been added. GPS is fully operational, currently with 31 satellites in orbit, instead of the nominal 24 constellation. The US Coast Guard is the primary point of contact for civilian users: www.navcen.uscg.gov.

Figure 3 (on top) shows the actual mission duration per satellite. The satellites are ordered, along the horizontal axis after launch date. The 7th and the 40th satellite were never put into orbit because of launch-failures. Satellites in green are still active today, as of February 2010, and some of them have been launched very recently, in March and August of 2009. The design lifetime of block II/IIA satellites is seven and a half years, and of block IIR satellites 10 years, with mean mission durations of six years and seven and a half years respectively. Only the first 11 satellites were block I satellites, the rest are all block II. It can be seen that in general GPS satellites largely outlive their mean mission duration and design life time. There are some 10 satellites exceeding even more than 15 years, which is double the mean mission duration!

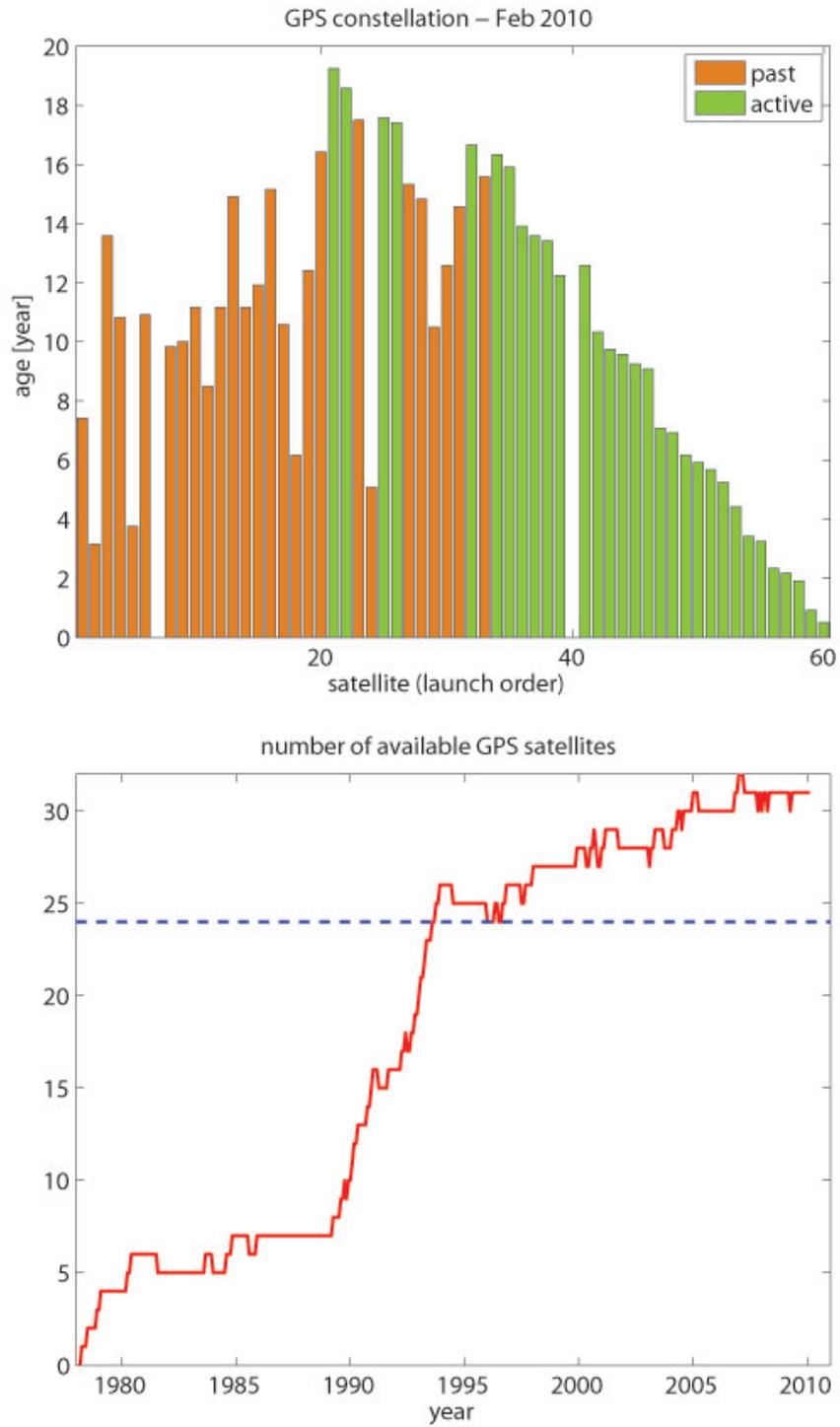


Figure 3: Top: active mission duration per satellite, green satellites are still active today. Bottom: the number of active GPS satellites as a function of time.

The graph at the bottom of Figure 3 shows the number of active GPS satellites in orbit as a function of time, ever since February 1978, up until today. The nominal GPS constellation consists of 24 satellites. Today there are 31 satellites, of which one is set unhealthy (PRN01/SVN49). So far, in a constellation with more than 24 satellites, typically newer satellites fly in tandem, side by side, with older satellites, by which we have, from the geometric point of view, effectively a 24 satellite constellation. Very recently (January 2010), the US Air Force has begun repositioning GPS satellites, as a transition to what they call a 24+3 constellation. This policy change was driven at least in part by the desire to improve satellite visibility for US and allied military operations in Afghanistan and Iraq, where mountainous terrain can hamper signal coverage for troops on the ground. The transition will take up to 24 months to implement fully.

Glonass

The GLObal NAVigation Satellite System (Glonass) has been built on the order of the Russian Ministry of Defence, and is maintained by the GUKOS. The first satellite was launched in October 1982. Since Autumn 2003, second generation satellites (Glonass-M) are being launched. A set of three Glonass-M satellites was launched, in one go, in December 2009. Third generation satellites (Glonass-K) are expected from this year onwards. They will offer civil signals on three frequencies.

The ground-segment of Glonass consists of four stations in Russia. Glonass has had a difficult time in the recent past, with only a few satellites in orbit. The system now seems to be approaching full operational capability. There are 20 satellites in orbit, of which two are in maintenance, as of February 2010. The point of contact for Glonass is the Russian Space Agency: www.glonass-ianc.rsa.ru.

Galileo

The Galileo system is being developed by the European Commission (EC) and the European Space Agency (ESA). Two prototype satellites have been launched so far, GIOVE-A in December 2005 and GIOVE-B in April 2008. The next step is to launch four satellites (later this year, and in 2011) for the In-Orbit Validation (IOV). A contract for 14 operational satellites was awarded in January 2010. These satellites should fly by 2014. The award of a subsequent 18 further satellites was left open. With the full constellation

Galileo should have 30 satellites in orbit (27, plus 3 active spares), but the final Full Operational Capability (FOC) date has not been specified yet.

The ground-segment of Galileo will consist of a worldwide network with about 40 stations. Useful websites on Galileo are www.gsa.europa.eu of the European GNSS Supervisory Authority, www.ec.europa.eu/transport/galileo of the European Commission, and www.esa.int/esaNA/galileo.html of ESA.

BeiDou Compass

The BeiDou (Compass) Navigation Satellite System is being developed by the Chinese government. The first Medium Earth Orbiting (MEO) satellites were launched in April 2007. The system has so far three satellites in orbit (the third one, a geo-stationary satellite was launched in January 2010). The system will consist of 27 satellites, supplemented by five geo-stationary satellites, and three inclined geo-synchronous satellites, making a total of 35 satellites. A 12 satellites-system with regional (Asia-Pacific) coverage is planned within a few years from now; the full system with global coverage for 2020. Recently a website has been launched www.beidou.gov.cn, so far only in Chinese language.

Vulnerability of satellite navigation

With four Global Navigation Satellite Systems fully operational by the end of the decade, users on Earth can enjoy signals, at multiple frequencies in the L-band of the Electro-Magnetic (EM) spectrum, from 1.1 to 1.6 GHz, from over 110 satellites. Then there should be, on average, about 30 satellites in view above a 10 degrees elevation anywhere on Earth.

The future of navigation is looking bright, very bright. The above developments provide diversity in system control. There is no single nation that can switch off, or disrupt all systems. There will be many satellites, and there will be substantial diversity in frequency, in terms of the radio-spectrum. This will mitigate risks of interference and jamming. Thereby, satellite navigation starts to play an important role in programs for civil aviation. For instance in Europe, EUROCONTROL, the European Organisation of the Safety of Air Navigation, recently presented its policy, implying a gradually increasing reliance on satellite navigation, based on a multi-constellation and multi-frequency GNSS. More specifically, for Communication, Navigation and Surveillance (CNS), in connection with Air

Traffic Management (ATM), there is a trend and evolution towards 4D trajectory management, for which satellite navigation is an important enabler.



Figure 4: In Safety-of-Life applications as aviation, in particular during critical phases of flight as approach and landing, satellite navigation shall provide a very high level of service. View of the Kaagbaan at Amsterdam Airport Schiphol.

To be used in aviation, in particular during critical phases of flight as approach and landing, satellite navigation shall provide a very high level of service. Correctness, within tight bounds (Alert Limits), of the position solution shall be guaranteed to extremely high levels of probability. In operating an aircraft, the risk of so-called Hazardously Misleading Information (HMI) due to the navigation system is typically budgeted at the 10^{-7} to 10^{-9} level. For Safety-of-Life (SoL) applications as aviation, we quote, as an example, the Galileo SoL core system performance requirements (without receiver contribution) with respect to integrity, as a risk of $2.0 \cdot 10^{-7}$ in any 150 seconds, with an Alert Limit (AL) of 12 m for the horizontal and 20 m for the vertical component.

GPS anomaly: case-study

From long term performance monitoring it is known that the GPS does very, very well. But no technical system is perfect. In this case-study we demonstrate that, though the events are very rare, GPS does not always



Figure 5: Though the US GPS has proved to function very reliably, it is *not* however without any failure or anomaly! An actual example will be shown in the next section.

meet the set position accuracy figures. Using a permanently installed receiver and antenna (Trimble 4700 with geodetic chokering antenna) at the TU Delft GNSS observatory, an anomalous event was detected 1 January 2004. The position coordinates of the installation are accurately known and the graph of Figure 6 shows the stand alone position solution (as a user would get), minus the accurately known reference. The position error is expressed in a local East, North and Up frame, to ease interpretation. Almost a 17 minutes time span is shown, from 18h17m to 18h33m (UTC), where the receiver measurement interval was 10 seconds. The position error behaves as expected up to epoch 180, with errors in the order of 5 to 10 meter, whereas in the last three minutes of the graph, the position errors suddenly start to grow to as much as half a kilometer! Around 19h, hence in just half an hour, the position error grows to 15 kilometer. And later the maximum position error became some 70 km without any warning. This shows how satellite navigation can go terribly wrong. Such an event is rare, but can not be completely excluded.

Through analysis it was found that satellite PRN23 was causing the trouble. As can be seen in the so-called skyplot of Figure 6, PRN23 was observed at that time at high elevation. It was almost directly overhead and thereby severely impacting the position solution. The satellite was healthy according to its navigation message. This anomaly impacted satellite navigation performance mainly in Europe and Russia.

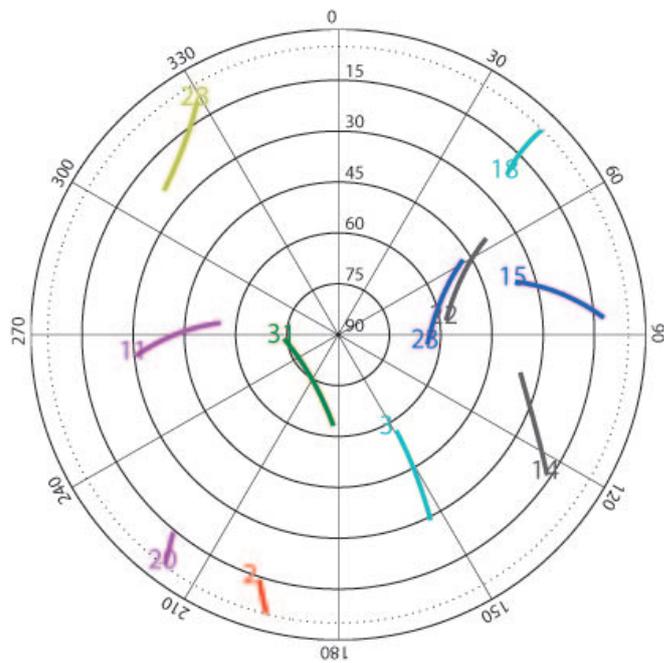
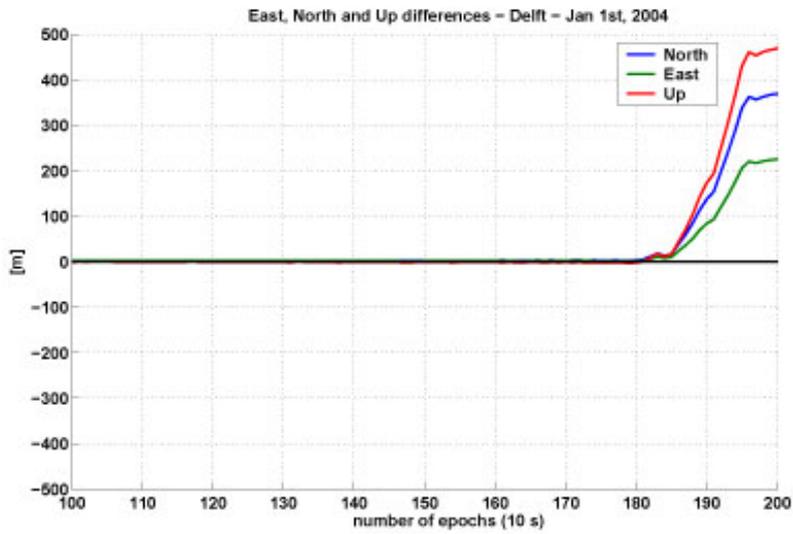


Figure 6: Top: position error expressed in local East, North and Up direction. Bottom: skyplot of GPS satellites, showing azimuth and elevation angles to the satellites, as locally observed in Delft, The Netherlands.

The GPS Support Centre later stated and explained that the anomaly was due to a failed atomic frequency standard (AFS) on SVN/PRN23. A lack of hard failure indications in the satellite telemetry coupled with satellite visibility limitations on the Master Control Station's (MCS) L-band monitor station network made this anomaly difficult to characterize and resulted in the transmission of Hazardously Misleading Information between approximately 18h33m and 21h28m (UTC).

The satellite was not really to blame for this, as this one, a block IIA satellite, was launched on 26 November 1990, and hence at the time of the event, already 13 years old. The satellite was, at that time, out of view from the GPS ground segment stations, so no alarming or corrective action could be undertaken. One month later, this satellite was officially decommissioned from service.

During this event, the position accuracy delivered was definitely not meeting what one usually expects from GPS. In particular when one is not aware of the anomaly and the resulting large bias in position, this could lead to dangerous situations; the GPS receiver says you are at a particular location, but in reality you really are somewhere else.

Actually the performance standards, which the US government commits to provide to civil GPS users, regard an instantaneous so-called Signal-in-Space (SIS) User Range Error (URE) exceeding 30 meter as Hazardously Misleading Information (HMI) when the satellite is set healthy and when the so-called User Range Accuracy (URA) is below a certain threshold. The probability of HMI in the performance standards equals 0.002, resulting from 3 assumed failures per year, lasting no longer than 6 hours each. A probability of 0.2% is still bigger than absolute zero, and it always will be! This means that, though satellite navigation by means of GPS is very, very reliable, nominal performance is *not guaranteed* 100% of the time.

Concluding remarks

What can we do to prevent such failures and anomalies in GNSS? We could deploy an additional network of ground stations in the region of interest, and have the satellite navigation service monitored (and relay alerts and warnings to users in the region). This is referred to as augmentation, and is a means of achieving sufficient integrity at system level. System aug-

mentation still does not cover anomalies and errors in signal propagation and in user equipment.

Using a redundant set of measurements, we could check integrity. In the context of satellite navigation, this is referred to as Receiver Autonomous Integrity Monitoring (RAIM). With, for instance, distances measured to 7 satellites, and only 4 unknown parameters to determine (3 position coordinates and the receiver clock offset), one has three measurements ‘too many’. This surplus of measurements can be used to have measurements (partly) checking each other. For instance an unexpectedly large error in only one of them can be detected ‘by the others’. Formally this is carried out through statistical hypothesis testing. Doing so on the data of the casestudy immediately leads to PRN23 being detected and identified as the faulty satellite. In an automated way, measurements from satellite PRN23 can be disregarded in the final solution, and the user is provided with a bias-free position solution. The navigation system ‘repairs’ itself. Adding more systems and more measurements (by using multiple signals and frequencies) is generally beneficial to measurement redundancy, and thus to monitoring performance.

Measurement redundancy can also be enhanced by adding other sensors. For robust and reliable navigation, sensors should be complementary. Different sensors should have different principles of operation and characteristics and different failure modes (with satellite navigation one could think of vision-based positioning and of inertial navigation). In a close integration, all measurements are brought together and processed integrally, thereby exploiting all available measurement redundancy to the maximum. All data are fused to provide one integral position solution. In this sense, adding a back-up system or sensor is just a rudimentary form of sensor integration, as one effectively uses the back-up, only in case of the primary sensor failing.

Further Reading

In this article the current status and development of GNSS has been discussed. Of course, besides this overview, many interesting articles and textbooks exist on the topic for further reading. Here, we would like to mention a few.

A decent, up-to-date textbook on satellite navigation is the book by Misra & Enge [1], that deals with several topics to do with GPS, like GPS signals, GPS receivers and measurements. For the Glonass system a presentation by Kovenko [2], at the 2009 European Navigation Conference last year, gives a good overview of the current status and plans. Concerning aviation issues, the Eurocontrol policy on GNSS is well described in [3]. Furthermore, integrity with (satellite based augmented) satellite navigation, in the context of (civil) aviation, is discussed in [4].

References

- [1] P.Misra and P. Enge, *Global Positioning System - Signals, Measurements, and Performance*. Lincoln, (Massachusetts): Ganga-Jamuna Press, 2006.
- [2] V. Kovenko, "Satellite system Glonass - status and plans," *Proc. ENC-GNSS 2009*. Naples (Italy), 2009.
- [3] F. Salabert, A. Hendriks, R. Rawlings and R. Farnworth, "EUROCONTROL Policy on GNSS," *Proc. ION-GNSS 2008*, Savannah (Georgia), pp. 1109-1115, 2008.
- [4] J. Oliveira and C. Tiberius, "Quality Control in SBAS: Protection Levels and Reliability Levels," *Journal of Navigation*, Vol. 62, No. 3, pp. 509-522, 2009.

The World of Software Receivers for Satellite Navigation

Benoit Muth

Introduction

Software-Defined Receivers (SDR) are radio receivers that contain as few hardware components as possible. In such a receiver, the signal processing block then starts as close as possible after the antenna, and is carried out with scripts running on a general-purpose processor contrary to traditional hardware radio receivers.

The application of this concept to receivers for Global Navigation Satellite Systems (GNSS) signals was initiated in the latter part of the nineties. Several implementations have been developed all over the world since, at a time where existing GNSS constellations are being modernised and new constellations being constructed. Software receivers not only offer several advantages in terms of processing compared to the traditional GNSS receivers architectures, they also find various and sometimes unexpected applications. This paper contributes with an update on the different GNSS software receivers, a couple of years after the first state of the art [1].

The US GPS is the only fully-operating GNSS since the early nineties. This system has been undergoing several modernisations: new generations of increasingly performant satellites are being launched, and signals offering higher ranging accuracy are being transmitted. Regional augmentation systems to GPS have been developed, such as the European system EGNOS, opening the way to the development of complementary global navigation systems such as Galileo.

The existing GNSS systems, like GPS, are each increasing the number

of transmitted signals, with modernised signals in the existing bands and signals in the new frequency bands. This new situation allows for more navigation observables and different combinations of navigation observables, which enables the easier correction of, for example, atmospheric effects and yields more accurate position and velocity estimates. These improvements might be taken further by combining observations from different GNSS systems, as long as these GNSS systems are compatible with each other and interoperable.

For an overview of existing GNSS systems and the future of these systems, we refer to the preceding contribution *Global Navigation Satellite Systems: status, plans and threats* in this year's NL-Arms by Christian Tiberius.

Software Receivers

Software receivers are digital implementations of receivers for radiofrequency signals, that ideally contain as few hardware components as possible. The digital signal processing (DSP) of the stream of discrete-time samples starts right after the antenna, and is achieved by software scripts running on a general-purpose processor. The hardware in a software receiver would then just be made of an antenna attached to an analogue to digital converter (ADC) delivering signal samples to a DSP block. A software receiver does not transmit any signals, to the contrary of a software radio, which is a digital implementation of a transceiver, that not only receives but also transmits signals.

Software-defined receivers

The processing of signals sampled following Nyquist's theorem, an approach called direct digitisation, is not possible or too expensive with the capabilities of *actual* ADCs and DSP boards, and might not become very much easier in the coming years. Indeed, the frequency range and dynamic range of the ADC are usually too limited, and the current available processing powers are too limited, at least for real-time operation. In practice, these limitations are then bypassed by adding analogue components after the antenna of the ideal software receiver architecture, making so-called software-defined receivers. The difference between software and software-defined receivers is illustrated for GNSS in Figure 4. Traditional analogue radio architectures, i.e. radio front-ends, are then still needed to cope with

the limitations of the ideal software receiver and efficiently implement a software-defined receiver.

The most spread architecture is the heterodyne architecture that down converts the received signal to a (lower) intermediate frequency (IF), in one or several stages with one or several sets of oscillators, mixers and filters, one set for each down conversion stage. In order to reduce the power consumption of the receiver and the price of analogue components, some receivers implement the homodyne architecture, also called direct conversion architecture, which achieves the down conversion in one step of the RF received signal to baseband or quasi-baseband. This approach is particularly fruitful for the reception of wideband signals or of the combination of several narrowband signals (for the application to GNSS see [2]).

Features of the software receiver

The software receiver approach might first bring cost reduction in the design of radio receivers, since for instance, the displacement of the ADC towards the antenna enables the suppression of analogue components and digital filters in the front-end. Second, the software receiver approach brings flexibility in the design of radio receivers: the single-purpose narrowband components can be replaced by wideband components allowing for the fusion of reception capabilities for different radio systems. The software approach then enables the reception with one device of signals with different modulations and encryptions, or the integration of navigation functions in one electronic device with other non-radio systems.

More important, the software approach enables the re-programmability of the hardware, consequently the updating and improvement of the radio receiver capabilities. Eventually, if the processing power offered by market PCs in the early decennia was not sufficient to have software receivers with real-time processing capabilities, Moore's law ruling the speed of processors has made it now possible for demanding applications, such as GNSS receiver processing to run on ordinary PCs.

GNSS receivers

In this section we concentrate on software receivers for GNSS satellite signals, like the signals involved in GPS.

Purpose of the receiver

GNSS satellites currently transmit so called direct sequence spread spectrum signals, using Code Division Multiple Access (CDMA), i.e. transmission of multiple signals on one frequency using different codes for the signals. This imposes on GNSS receivers to be receivers for CDMA signals, that are either BPSK or BOC modulated [3,4]. Since ideal software receiver currently cannot be built, GNSS receivers depicted, as in Figure 1, implement the classical heterodyne radio architecture to collect the RF signal with an analogue front-end that is located between the antenna and the ADC. In the Digital Signal Processing (DSP) block, usually implemented in hardware through Application Specific Integrated Circuits (ASICs), the received satellite signals are synchronised and demodulated, achieving the simultaneous production of ranging observables and extraction of the navigation information transmitted by the satellites. Once this data has been processed, the receiver can compute its own Position, Velocity and Time (PVT).

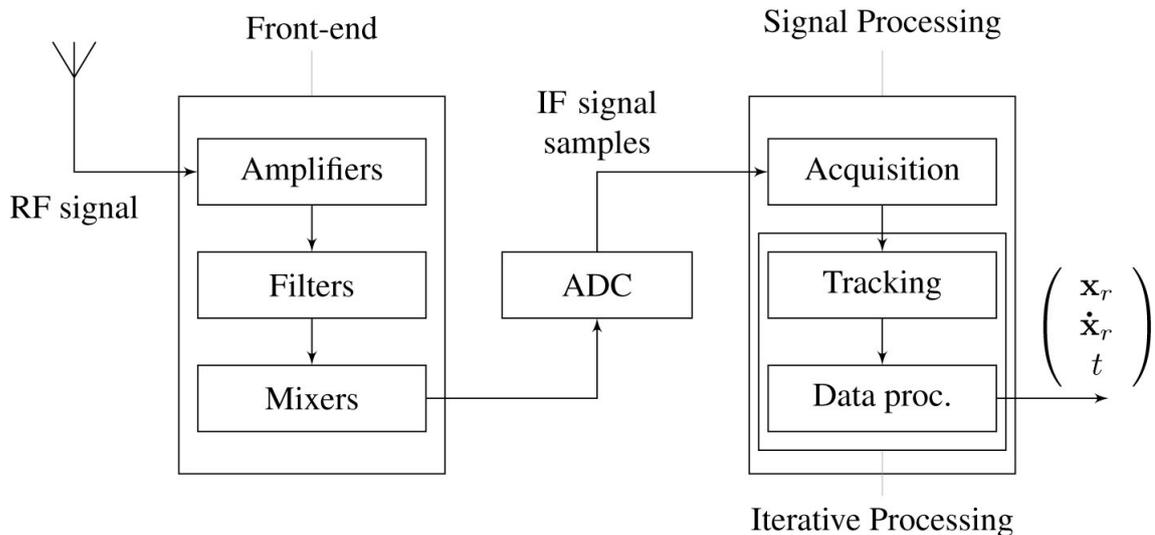


Figure 1: Simplified architecture of a GNSS receiver.

Architecture of the software receiver

The classical receiver implements the heterodyne architecture, that pre-processes the received signal. This means, as depicted in Figure 1 that the received signal is (bandpass) filtered, amplified to an acceptable level,

down-converted to a lower, intermediate, frequency (IF) to make it processable and digitised, i.e. transformed into a stream of discrete-time samples. The receiver then proceeds to signal acquisition, that is, the detection of the received signals and a coarse estimation of the code phase and Doppler frequency synchronisation parameters. The signal processing is then split into different channels corresponding to the detected satellites. In each channel, the received signal is synchronised so as to remove the code and carrier from the signal, i.e. extract the navigation data, and compute navigation observables, pseudorange, carrier phase, accumulated Doppler. In the end, the extracted navigation data is decoded and the receiver PVT estimates $[\mathbf{x}_r^T \dot{\mathbf{x}}_r^T \delta t^T]^T$ can be computed. In Figure 1 the box embedding the tracking and data processing blocks represents the continuous processing part inside the receiver.

A global model of the GPS receiver is presented in [3-7]. To still take advantage of the fastest technologies, one of the trends in GNSS receivers is to consider software routines run by a Field Programmable Gate Array (FPGA), or by a Reduced Instruction Set Computer (RISC) microprocessor instead of a PC. A detailed quantitative analysis of the design, the implementation of such processor architectures for GNSS SDR receivers is presented in [8], including an exhaustive study of existing Application Specific Instruction Processors (ASIP) based architectures.

Receiver starting options

The direct environment and history of a GNSS receiver have such impact that when switched on, the receiver faces different conditions, respectively called cold, warm and hot start conditions. A hot start happens after a relatively short interruption of the reception, from a passage in a tunnel until a shortage of less than two hours. A warm start occurs when the GPS receiver is switched on more than four hours after being switched off at the same place. In this case the receiver does not have any valid ephemeris, although it does have a valid almanac. To the contrary, a cold start happens when no valid timing or location information is available to the receiver, which has not been used for several days, or has travelled far from the last location known to it. These cold start conditions are the only type of conditions that matter to characterise the receiver performance.

The receiver's RF front-end

As illustrated in Figure 1, the RF front-end most of the time implements the heterodyne radio architecture, that filters the spectral band of interest, mixes it to an IF, filters and amplifies the down converted signal that is next transformed by an ADC into a stream of discrete-time samples [3,4,6,7,9-12]. These pre-processing operations in the analogue front-end shall perform identically independently from the reception conditions at the antenna, i.e. independently of the received signal strength and the value of the signal carrier frequency, independently also of potential unintentional interferences from external sources and independently of the self-noise from the front-end. As mentioned above, given technological limitations, the direct sampling of the received signal at RF is not possible, while the sampling at IF or baseband is. Besides the number of quantisation bits, the sampling is characterised by the value of the sampling frequency f_s which in turn is linked to the choice of the sampling strategy [9]. Most front-ends deliver signal streams containing one phase of the received signal. Depending on the signal structure and the receiver specifications, one might want the phase and quadrature components of the received signal.

Signal processing in GNSS receivers

The signal processing in a GNSS receiver is basically made of three blocks: the acquisition block, the tracking block and the data processing block. These three blocks achieve the synchronisation and demodulation of received satellite signals by means of signal processing algorithms, the production of ranging observables and the extraction of the navigation information. Both acquisition and tracking stages are based on the measure of the similarity, correlation, between a received satellite signal and a signal template. For the acquisition, this measure is a two-dimensional quantity depicted in Figure 2. From the tracking outputs the navigation message can be decoded and a position, velocity and time (PVT) solution can be produced.

The acquisition part

While the RF front-end processes one single stream of data, the DSP block separates the different satellite signals. The first task of the DSP, the acquisition, is then to detect which satellites are in view of the receiver and for each satellite, to yield coarse estimates of the synchronisation parameters, the signal travel-time and carrier frequency. This estimation-detection pro-

cedure is possible thanks to the properties of the GNSS signal spreading sequences. On the whole, the acquisition is by far the most demanding part of the whole receiver chain in terms of computational cost. This cost is highly variable according to the acquisition method, the initial status of the receiver and the environment of the receiver. Acquisition techniques are based on the correlation of the received signal with a signal replica. Numerically, this correlation measure is computed on a predefined, discrete search grid. The accumulation of the correlation values in a coherent and/or non-coherent way, makes up for a simple but efficient means to enhance the sensitivity of the receiver. Challenging acquisition scenarios include the processing of weak signals and signals with multi-trajectory propagation, multipath, that can be found in urban canyon or indoor environments.

Several factors impact the signal acquisition performance. Satellite-receiver dynamics and characteristics of the receiver, such as the sampling frequency or IF frequency, impact the resolution of the search grid. The implementation of the correlation operation, the integration length, the search strategy are other useful settings of the parameter estimation and the signal detection. The precision (bias and accuracy) of the synchronisation parameter estimates, the sensitivity of the algorithm and the mean acquisition time qualify the performance of the acquisition algorithm. The choice of the acquisition method and the setting of its parameters usually results from a trade-off between computational load / processing speed of the DSP and sensitivity of the algorithm.

The tracking part

The next signal processing step, called tracking, aims at refining and keeping track of the synchronisation parameters, the code phase and carrier frequency, based on the coarse parameter estimates, obtained during the acquisition. The tracking then achieves for the received signal the continuous estimation of the synchronisation parameters and of the GNSS signal parameters : the code phase (travel time), the Doppler frequency shift, accumulated Doppler shift and carrier phase. Of course the accuracy of these parameter estimates has an impact on the precision of the pseudorange measurement, i.e. the distance from the receiver to the satellite. In practice, the signal tracking is divided into two functions, the code tracking and the carrier tracking, that adjust the code phase and carrier phase (fre-

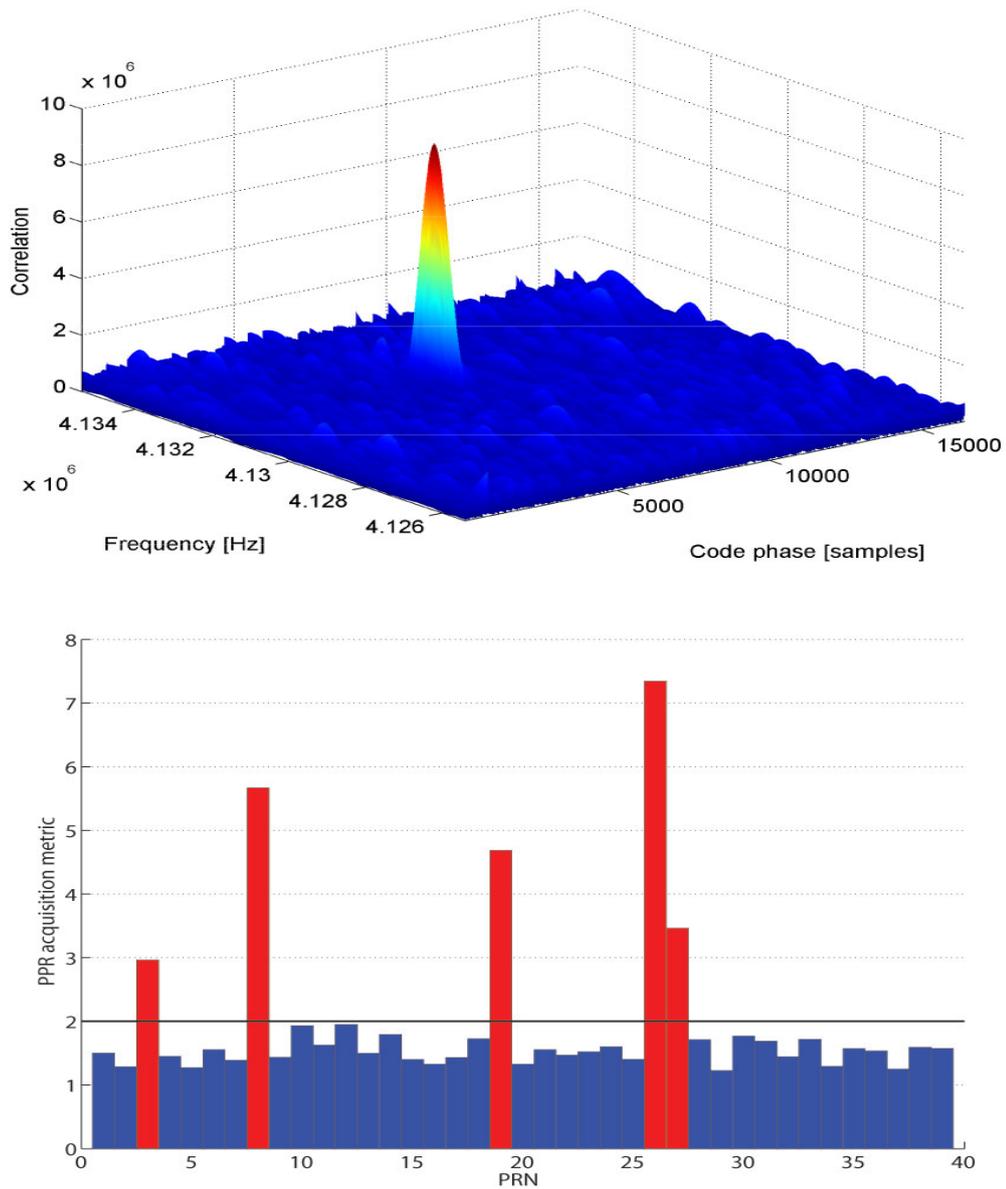


Figure 2: Two-dimensional correlation function of the received signal with the replica for PRN 3 (top) and Acquisition metrics (detection power) of all GPS satellites (bottom): red satellites are detected, blue satellites are considered as absent. The peak on the left plot indicates the presence of the satellite signal. The coordinates of the peak indicate the time-offset and frequency offset of the received signal with regard to the nominal frequency due to the satellite receiver dynamics.

quency) of a replica to the code and carrier phases of the received signal. This tracking is implemented in traditional DSP architectures, with closed-loops called Delay-Lock-Loops (DLL) and Phase-Lock-Loops (PLL). Both code and carrier tracking loops are made from a detector/discriminator, a

filter and an oscillator updating the parameter tracked in the loop.

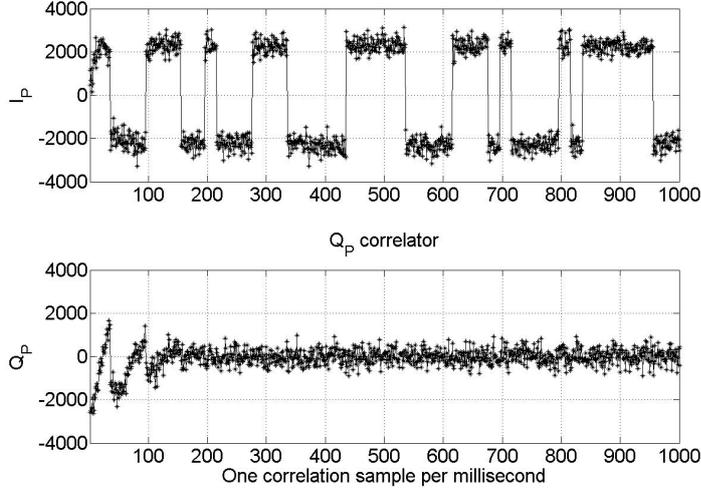


Figure 3: Tracking outputs for satellite signal of space vehicle (SV) 3 recorded in an experiment in the surroundings of Den Helder. In the figure: demodulated phase and quadrature arms (90 degrees shift) showing that the navigation data appear in the prompt-phase arm I_P (in phase).

A lock detector is used to check whether these tracking loops are locked or have lost lock. In this last case, the acquisition must be carried out again for the considered satellite. Otherwise if the tracking is successful, the signal stream at the output of the tracking delivers the navigation data carried by each satellite signal. Next to the navigation data the tracking block yields, for each visible satellite ranging measurements, the code pseudorange, possibly the accumulated Doppler and the carrier phase pseudorange, enabling the computation of a position-velocity-time (PVT) solution. Obtaining these measurements is very well known when the receiver is static and not subject to interference. On the other hand, current receiver challenges include the determination of correlation losses due to multipath or interference.

The main factors impacting the tracking performance are the signal modulation, the chip (one code element) length, the type of discriminator, the early-late correlator spacings, the sensitivity of the tracking loops to the received signal strength, the satellite-receiver dynamics, interferences or multipath. On the other hand, the tracking performance can be characterised mainly by the (non-linear) dynamic and stability properties of the PLL and DLL.

Data processing in the receiver

Once the signal has been demodulated, the data processing first consists of synchronizing and decoding the 50 Hz navigation data. The decoding is achieved in several steps, since the navigation data is structured for GPS, in words containing 30 bits, 10 words forming a (300-bit) subframe, 5 subframes making a (1500-bit) page. Consequently, if a receiver does not have its position, a long enough signal must be read to be sure to obtain the complete navigation data. Once the navigation data has been decoded, the receiver has the knowledge of the satellite's position, of clock corrections terms, can synchronise with the GPS time and can proceed to the computation of pseudoranges and the estimation of its position, velocity and time [3,13].

Software receivers for GNSS

As for other radio frequency receivers, GNSS receivers become all-digital receivers. The software receiver context then not only offers to GNSS receivers advantages in terms of flexibility and reconfigurability, but also enable new applications. Technologically the ideal GNSS software receiver is not feasible, and an analogue front-end is necessary in practice to build a GNSS software-defined receiver, as depicted in Figure 4.

Software-defined receivers first make up for a good framework for developing the GNSS receiver DSP [14,15] and improving the positioning accuracy by integrating other systems to the GNSS receiver platform. Considering the signal processing only, the SDR flexibility is especially needed when designing a new receiver. One can indeed more easily make a multi-frequency receiver, such as a GPS L1 / L2 receiver, with multiple signals on these carrier frequencies such as the C/A code and P(Y) code on L1. One should also implement more easily a multi-constellation receiver, such as a GPS / Glonass receiver. Considering the PVT accuracy, the GNSS receiver can be integrated with, for example, inertial systems such as INS, or other navigation systems such as LORAN. Secondly, considering the reception of one signal from one GNSS system, specific scenarios, such as interference, multipath or weak signals conditions reception can be studied and the fine tuning of the implemented algorithms is made easier. Thirdly, in terms of integration, the software receiver approach not only enables the integration of several navigation functions on one platform to improve

the navigation solution but also, by integration, the sharing of one single DSP platform, e.g. a cellphone, with other applications such as cellular communications or embedded applications [16].

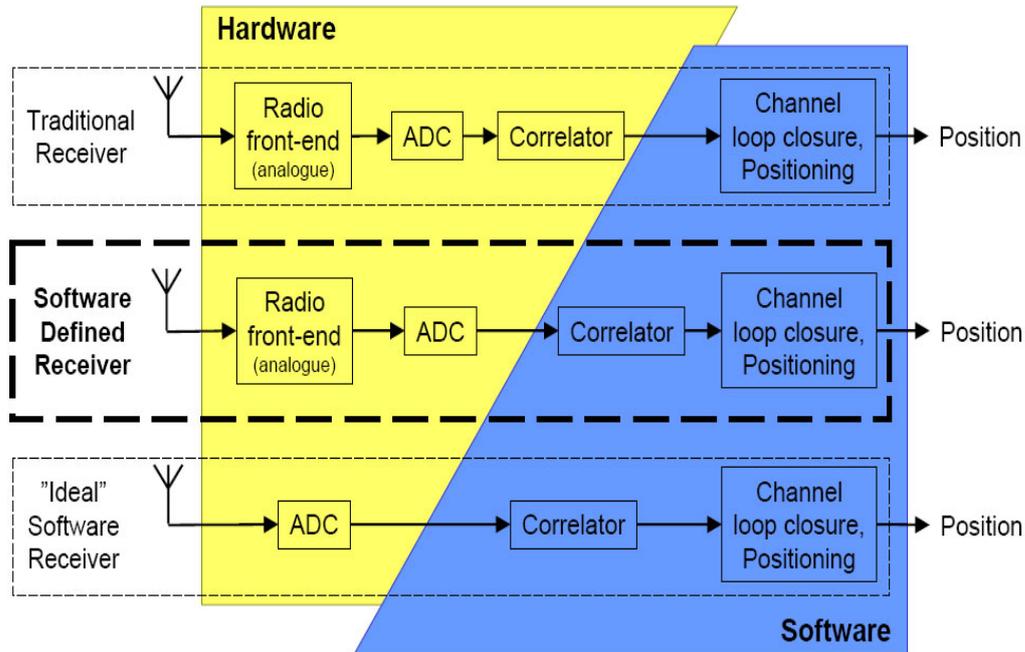


Figure 4: Different implementations for GNSS receivers: software receiver, software-defined receiver and hardware receiver, from [14].

Brief history of GNSS software receivers

The first front-end for GNSS software receivers has been presented in [10], together with the theory of bandpass sampling and other sampling strategies. The first GPS L1 C/A software receiver was then borne. The first realtime software receiver for GPS L1 C/A signals was presented in [17]. Several overview articles have been devoted to GNSS software receivers, the first of which is in the author’s opinion the most comprehensive [1]. Other articles with different depth have followed : a list of resources on software receivers in 2005 was established in [14]. Further technical hardware and software-related issues inherent to GNSS software receivers are to be found in [18,19]. Overview articles including safety concerns in civil aviation or military contexts, such as reliability of the position delivered by software receivers, are to be found in [1,20].

The first book on GPS and Galileo software receivers appeared in [4] together with software receiver scripts and data records. These data records



Figure 5: Low-cost antenna and (SiGe) software receiver front-end for pre-processing the received satellite signals. This front-end connects to a PC on which the recording and processing of the data signals take place. Images are from Sparkfun.

were collected with the front-end depicted in Figure 5. Further, [21] presents improvements brought to the software scripts and results that include (1) frequency spectrum monitoring plots, (2) multipath mitigation and tracking results, (3) FPGA and real-time implementation results, (4) implementation of multiple correlators for new GPS and Galileo signals and (5) the bump-jumping technique for removing ambiguities in side-peaks of the auto-correlation function of the new BOC signals.

Software receivers open new perspectives for the improvement of the accuracy of the pseudorange measurements and then of the positioning accuracy with standalone receivers. Indeed, modifications to the receivers such as an increase in the number of correlators, mitigation of physical channel propagation effects acquisition and tracking of weak signals, wipe-off of the navigation data thanks to assisted GNSS, or implementation of the vector delay-lock-loop (VDLL) become possible with a software-defined receiver. Technological trends such as the consideration of Graphical Processing Units, multi-core processing technologies shall supply FPGAs for handling large computational burdens [15,22]. Currently, particular needs concern the availability of hardware for down-converting the received signal, and digitising the variety of RF spectral bands. One example of multifrequency software receiver is illustrated in [23], where GPS L1, L2C and L5 signals are synchronously collected thanks to a triple-frequency front-end. In this case, since the L2 and L5 signals do not bear any navigation message, pseudorange measurements can still be obtained in the software-defined

receiver SDR framework by sharing the information from the L1 navigation message.

New applications by software receivers

Besides the new possibilities for standalone processing offered by software receivers in terms of programming flexibility and integration with other systems, software receivers enable the testing and validation of new algorithms. It becomes possible to add in software new functions to the receiver that work independently from the signal processing subsequent to the PVT estimation : such as time-domain histograms for BOC signals run lengths [24] or verification of the integrity of the receiver position [3]. Besides these applications, software-defined receivers have soon become useful to implement signal quality monitoring functionalities at different scales. GPS L1, L2 and L5 signals are being collected triple-frequency thanks to a front-end with very high sampling rate, for monitoring anomalies and exceptional events based on the raw data signal samples only [25].

GIOVE-A and GIOVE-B signals are being monitored using Chilbolton's astronomic telescopes, allowing for the assessment of transmitter non-idealities in a software receiver environment [26-28]. However the first objective of the application of the software-defined receiver approach to GNSS was the search for more computational efficiency for the traditional acquisition and tracking tasks. The implementation of new acquisition techniques such as the FFT-based parallel-search acquisition [29] was one of the first steps in this direction, giving birth to several new ideas for the purpose of signal acquisition [30,31] in the frequency-domain and [32,33] in the time-frequency domain.

As GNSS constellations are multiplying and their utilization is becoming widespread, computational efficiency inside the receiver processing and positioning accuracy remain major challenges for standard receivers. Other challenges arise for aviation or military users needing trust, integrity, on the provided receiver position. Indeed, the receiver has to cope with threats to the receiver accuracy such as unintentional or intentional interference, jamming. These phenomena have been studied for several years and are known. New challenges arise with the protection of receivers against spoofing attacks. In such cases, a signal resembling the broadcast satellite signal, but different, is sent towards the GNSS receiver to cause it to produce a

wrong position. Additional checks, signal authentication, are implemented for the receiver to know it produces not only a reliable position, but also in an adequate way and with reliable signals. Signal authentication is then necessary to have the receiver acknowledge the signal processing required for producing ranging measurements and further estimate a position. Signal authentication or anti-spoofing within software receivers, is a more recent field of investigations, where techniques using signals encrypted by the service provider, such as the US GPS P(Y) code, make it possible to the expert user to check the authenticity of the received signal. The first GPS L1 C/A code signal spoofer has been developed using a software radio [34], while the first anti-spoofing technique considers the joint processing of signals collected at different locations, close one to the other [35], to access the military P(Y) code sequence for authentication.

A second application involving the processing of signals received at two different stations is the measurement of the three-dimensional baseline vector between the stations [36,37]. The next contribution in this year's NL ARMS will discuss this application in more detail.

In a similar way to the interferometric positioning, GNSS-reflectometry (GNSS-R) implements the cross-correlation of two received GNSS signals in post-processing. GNSS-R consists in the recuperation of signals reflected by the Earth and the determination of properties of the reflected surface. Indeed, two antennas being connected to a receiver above the surface to study, the receiver collects for each visible satellite, one direct signal from the satellite and one signal reflected by the surface, typically an icy surface, the sea, or the ocean. Reflectometry experiments have been carried out where the receiver was placed on a satellite [38]. Compared to the case of the interferometric positioning, the cross-correlation energy of the direct and reflected signals is more spread and exhibits a specific shape that depends mainly on the reflected surface, and the parameters to retrieve [39-41]. Possible applications of the GNSS reflectometry principally lie in remote, Earth-sensing applications. For instance one can retrieve information on the ocean wind and waves (scatterometry), on the ocean mean height (altimetry) or information on sea ice, land surface topography or near surface soil moisture. These applications are possible thanks to the post-processing capabilities of software receivers.

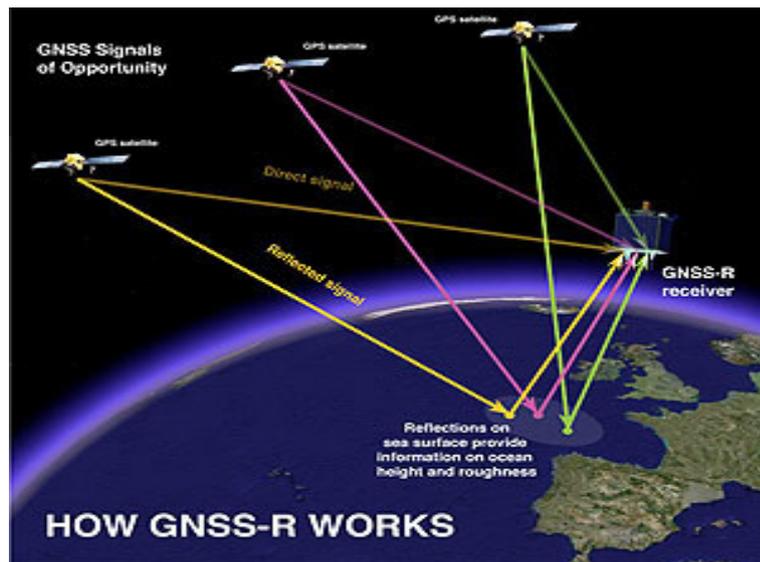


Figure 6: GNSS reflectometry consists in recuperating GNSS signals reflected from a surface. This picture from the University of Southampton depicts the application of GNSS-R to scatterometry or altimetry purposes above the North-Atlantic ocean.

Conclusions

Software receivers result from the breakthroughs in analogue components technology and processing technology. Consideration of GNSS and GNSS receivers have been becoming widespread since the early nineties. As a consequence the late nineties have seen the birth of software receivers for GNSS as research and development tools for traditional GNSS receiver signal processing, processing benchmarks of new signals and new constellations, development platforms for new implementations of old concepts and simply, new applications. GNSS software-defined receivers are tools that enable receivers to be quickly reconfigured and adapted to different environments, which matters in critical environments.

Simple experiments such as the interferometry are already possible with low-cost front-ends, yielding accurate single-difference positioning results. GNSS are tools that enable the study of signal processing immunity to unintentional and intentional interferences or worse, resistance to signal spoofing. Indeed signal authentication algorithms exist, to guarantee the integrity of the signal processing, not only of the estimated position. Eventually, as processing capabilities emerge, software-receivers enable applications that are highly-demanding in terms of computational load, such as GNSS reflectometry.

References

- [1] P.-L. Normark, “Software Based GPS : The SDR concept can be applied to receivers for Global Navigation Satellite Systems, giving improved flexibility and reducing development and production costs,” *Swedish Journal of Military Technology*, vol. 4, pp. 16–21, 2003.
- [2] R. Weiler, P. Blunt, P. Jales, M. Unwin and S. Hodgart, “ Performance of an L1 / E5 GNSS Receiver using a Direct Conversion Front-End Architecture,” *Proc. of the International Technical Meeting of the ION*, pp. 1478–1489, 2008.
- [3] P. Misra and P. Enge, *Global Positioning System - Signals, Measurements, and Performance*. Lincoln, (Massachusetts): Ganga-Jamuna Press, 2006.
- [4] D. Akos, K. Borre, N. Bertelsen, P. Rinder and S. Jensen, *A single-frequency software-defined GNSS receiver*, Birkhäuser, 2006.
- [5] J. O. Winkel, “Modeling and Simulating GNSS Signal Structures and Receivers”, Ph.D. dissertation, Universität der Bundeswehr München, 2003.
- [6] B. Eissfeller, “Ein dynamisches Fehlermodell für GPS Autokorrelationsempfänger,” Universität der Bundeswehr, München, 1997.
- [7] B. W. Parkinson and J. J. Spilker, *Global Positioning System: Theory and Applications*. Washington: AIAA, 1996.
- [8] G. Kappen, V. Pieper, L. Kurz and T. Noll, “Implementation and Analysis of an SDR Processor for GNSS Software Correlators,” *Proc. 21th ION GNSS Conference*, 2008.
- [9] D. M. Akos and J. B. Y. Tsui, “Design and Implementation of a Direct Digitization GPS Receiver Front End” *IEEE Trans. on Microwave Theory and Techniques*, vol. 44, pp. 2334–2339, December 1996.
- [10] D. M. Akos, “A Software Radio Approach To Global Navigation Satellite System Receiver Design,” Ph.D. dissertation, Ohio University, August 1997.

- [11] D. M. Akos, M. Stockmaster and J. B. Y. Tsui, “Direct Bandpass Sampling of Multiple Distinct RF Signals,” *IEEE Trans. on Communications*, vol. 47, pp. 983–988, July 1999.
- [12] The Danish GPS Center, “The SoftGPS Project”, web pages provided <http://gps.aau.dk/softgps/>,” 2010.
- [13] E. Kaplan, *Understanding GPS - Principles and applications*. Boston: Artech House, 2005.
- [14] R. Babu, “Web-based resources on software GPS receivers,” *GPS Solutions*, vol. 9, pp. 240–242, 2005.
- [15] M. Hollreiser, “Galileo Receiver Design and Performance,” Lecture Notes Galileo Course, Navitec, ESTEC, 2006.
- [16] D. M. Akos, “The role of Global Navigation Satellite System (GNSS) software radios in embedded systems,” *GPS solutions*, vol. 7, p. 14, 2003.
- [17] B. M. Ledvina, M. L. Psiaki, S. P. Powell and P. M. Kintner, “Real-Time Software Receiver Tracking of GPS L2 Civilian Signals using a Hardware Simulator,” *Proc. 18th ION GNSS Conference*, pp. 1598 – 1610, 2005.
- [18] C. Kelley and D. Baker, “OpenSource GPS, A Hardware/Software Platform for Learning GPS: Part I, Hardware,” *GPS World*, 2006.
- [19] C. Kelley and D. Baker, “OpenSource GPS, A Hardware/Software Platform for Learning GPS: Part II, Software,” *GPS World*, 2006.
- [20] J. H. Won, T. Pany and G. W. Hein, “GNSS Software Defined Radio: Real Receiver or Just a Tool for Experts ?,” *Inside GNSS*, pp. 48–56, 2006.
- [21] E. Vinanden and D. M. Akos, “Improvements to A Software-Defined GPS and Galileo Receiver: Single-Frequency Approach,” *Proc. ION International Technical Meeting*, pp. 2230–2236, 2007.
- [22] M. Hollreiser, “Galileo Receivers - Challenges and Performance,” *Proc. 34th European Microwave Conference*, Amsterdam, pp. 321–324, 2004.

- [23] C. O’Driscoll, D. Borio, M. Petovello, T. Williams and G. Lachapelle, “The Soft Approach: A Recipe for a Multi-System Multi-Frequency GNSS Receiver,” *Inside GNSS*, pp. 46–51, September 2009.
- [24] B. Muth, P. Oonincx and C. Tiberius, “BOC(m,n) signals: a Time-Domain Histogram Pattern,” *European Journal Applied Signal Processing*, vol. 2007, April 2007.
- [25] S. Gunawardena, “Development of a Transform-Domain Instrumentation Global Positioning System Receiver for Signal Quality and Anomalous Event Monitoring,” Ph.D. dissertation, Russ College of Engineering and Technology of Ohio University, June 2007.
- [26] O. Montenbruck, C. Gunther, S. Graf, M. Garcia-Fernandez, J. Furthner and H. Kuhlén, “GIOVE-A initial signal analysis,” *GPS Solutions*, vol. 10, pp. 146–153, March 2006.
- [27] M. Rapisarda, P. Angeletti and E. Casini, “A Simulation Framework for the Assessment of Navigation Payload Non-Idealities,” *Proc. Navitec Conference*, 2008.
- [28] M. Soellner, C. Kurzhals, W. Kogler, S. Erker, S. Thaelert, M. Meurer, M. Malik and M. Rapisarda, “One Year In Orbit : GIOVE-B CBOC E1 Signal Quality Assessment,” *GPS Solutions*, pp. 28–38, 2009.
- [29] D.J.R. Van Nee and A.J.R.M. Coenen, “New Fast Code-Acquisition Technique Using FFT,” *Electronic letters*, vol. 27, pp. 158–160, 1991.
- [30] D. Lin and J. B. Y. Tsui, “Comparison of Acquisition Methods for Software GPS Receiver,” *Proc. ION GPS Conference*, pp. 2385–2390, 2000.
- [31] J. S. Hong, J. W. Lee, G.-I. Jee, Y. J. Lee, J.W. Kim and C. G. Park, “GPS Signal Processing Algorithm for Software GPS Receiver,” *Proc. ION GPS Conference*, pp. 2338–2345, 2000.
- [32] P. Oonincx and B. Muth, “A Fractional Fourier Based Correlator for Detecting Joint Time and Frequency Offsets,” *Proc. 15th International Conference on Digital Signal Processing*, Cardiff, pp. 623–626, 2007.

- [33] P. Oonincx, H. Postma and B. Muth, “Generalized parallel-search strategies using the Fractional Fourier Transform,” *Proc. European Navigation Conference (ENC-GNSS) 2008*, Toulouse, France, 2008.
- [34] T. Humphreys, B. Ledvina, M. Psiaki, B. O’ Hanlon and P. Kintner Jr, “Assessing the Spoofing Threat: Development of a Portable GPS Civilian Spoofer,” *Proc. ION International Technical Meeting*, pp. 2314– 2325, 2008.
- [35] S. Lo, D. De Lorenzo, P. Enge, D. Akos and P. Bradley, “Signal Authentication: A Secure Civil GNSS for Today,” *Inside GNSS*, pp. 30–39, 2009.
- [36] B. Muth, “Software-based GNSS L1 Interferometric Positioning,” *Proc. 21th ION GNSS Conference*, Savannah, 2008.
- [37] B. Muth, P. Oonincx and C. Tiberius, “Differential Observables for Software-Based GPS Interferometry,” *Proc. 17th European Signal Processing Conference*, Glasgow, pp. 2166–2170, 2009.
- [38] P. Jales, R. Weiler, C. Underwood and M. Unwin, “First Spaceborne Demonstration of Galileo Signals for GNSS Reflectometry,” *Proc. ION International Technical Meeting*, 2008.
- [39] T. Elfouhaily, D. Thompson and L. Linstrom, “Delay-Doppler analysis of bistatically reflected signals from the ocean surface: theory and application,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 560–573, 2002.
- [40] S. Gleason and M. Unwin, “Development and testing of a remote sensing instrument using GNSS reflectometry concepts,” *Proc. IEEE International Geoscience and Remote Sensing Symposium IGARSS’03*, vol. 7, pp. 4332– 4334, 2003.
- [41] S. Esterhuizen and D. Akos, “The Design, Construction, and Testing of a Modular GPS Bistatic Radar Software Receiver for Small Platforms,” Notes of the Workshop on GNSS-R, 2007.

Software-based Interferometric Positioning

Benoit Muth, Patrick Oonincx & Christian Tiberius

Introduction

This article discusses the revival of the interferometric positioning experiment with GPS signals that aims at measuring the three-dimensional vector between two stations. Such an experiment was carried out in the late seventies, the early days of the GPS system, almost simultaneously by two competing teams that had means, astronomic telescopes, without any common measure with those at stake in our setup. From a hardware point of view, the innovative aspect of this paper lies in the implementation of the experiment with software receivers. GNSS software receivers are attractive tools given the low price of the corresponding front-ends and the flexibility of receiver processing implementations in software.

From a positioning point of view, several reasons account for the re-discovery of this experiment. First, a differential positioning technique between two stations, such as relative positioning with GPS signals, originates from the field of interferometry. Secondly, beyond relative positioning with code measurements only, the use of precise carrier phase measurements for GNSS positioning also originates from the interferometry framework. The low price of the software receivers front-ends and the increased precision of the differential positioning techniques are the attractive features of the experiment.

Interferometry

Principles of interferometry

The interferometry framework consists here in setting two identical software receivers consisting each of an isotropic antenna, a radio front-end

collecting signals in the L1 band and a laptop as depicted in Figure 1, and located at both ends of an unknown baseline to observe simultaneously and independently several radio sources, i.e. several satellites. The unknown three-dimensional vector between the two antennas is called the interferometer baseline vector [1].



Figure 1: Signals were recorded simultaneously 9 December 2008, at two stations equipped with software receivers and located in the surroundings Den Helder, The Netherlands. Left: map (Google) indicating the baseline location, where the yellow pins indicate the two stations. Right: software receiver recording set-up at the Stoomweg station where the small antenna is positioned on top of the tripod.

When measuring the phase differences of the signals received at the two antennas from a specific source, the direction of the baseline vector can be determined relative to the direction of the source. If the baseline length can be ignored compared to the distance from the source to the baseline, the received signal is considered as a plane wave arriving to the antennas from the same direction, the signal paths are parallel and the product of the Time-Difference-Of-Arrival (TDOA) of the two signals by the speed of light equals the projection of the baseline vector on the direction of the source as in Figure 2. However for the interferometry with GNSS signals, this plane wave approximation cannot be applied to any baseline, given the distance of the satellites to the Earth. In the exact case the GNSS signal has a spherical propagation and the deviation from true plane waves is proportional to the square of the baseline length divided by the satellite's distances, which is taken into account in the linearized observation equations [1].

Interferometry is a technique that consists of studying the interferences created by the superposition of two waves. This technique has been applied

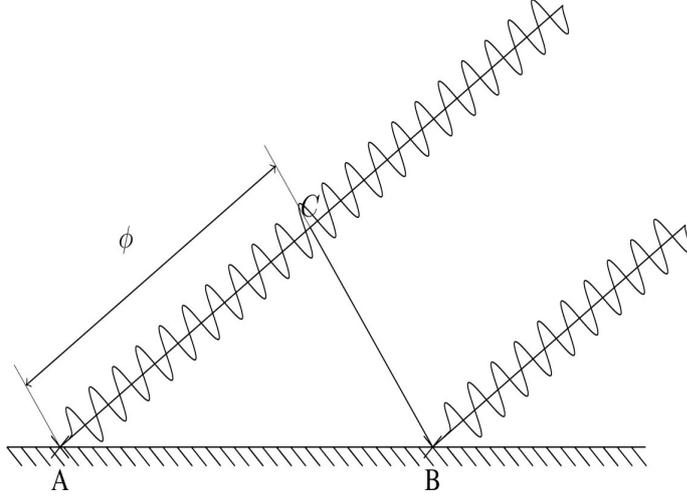


Figure 2: Interferometric configuration considering the plane wave approximation for the signal received from one source at the two stations A and B.

extensively in a variety of fields, e.g. astronomy, optics or plasma physics, also for positioning, first with signals from astronomic sources, second with GNSS signals [2]. Generally speaking, the TDOA of any two signals at the receiver can be measured with two types of observables: phase and group-delay observables. The simplified model of the phase observable ϕ is

$$\phi(t) = \frac{2\pi f}{c} \mathbf{B}(\mathbf{t}) \cdot \mathbf{s} + \phi_{media}(t) + \phi_{instru}(t) + 2\pi A \quad (1)$$

where the phase ϕ is in radians, f is the signal frequency and c the speed of light. In a chosen reference system \mathbf{B} is the baseline vector, \mathbf{s} the unit Line Of Sight vector from the reference point to the source and therefore $\mathbf{B}(\mathbf{t}) \cdot \mathbf{s}$ is the projection of the baseline in the direction of the source. Next, $\phi_{media}(t)$ and $\phi_{instru}(t)$ are imprecisions of the observable respectively due to random short-term changes of the media, e.g. multipath, and of the instrumentation, e.g. clock errors. A is the integer number of carrier cycles called the ambiguity. The derivative of the phase ϕ with respect to the angular frequency $2\pi f$ is the interferometric group-delay

$$\tau(t) = \frac{1}{c} \mathbf{B}(\mathbf{t}) \cdot \mathbf{s} + \tau_{media}(t) + \tau_{instru}(t) \quad (2)$$

in seconds. These phase and group delay observables contain the whole baseline vector information and correspond in the case of GPS to the single-difference carrier phase and code observables respectively. Interferometric positioning, i.e. the computation of a baseline with single-differencing techniques has been achieved first with analogue hardware processing GPS

signals [2,3]. Next, the evolution of receiver technologies gave birth to radio front-ends specifically designed for GNSS software receivers. Means to collect experimental GNSS data could then be acquired for a few hundred euro. The opportunity to acquire such front-ends has been another motivation for reproducing the interferometric experiment in a Software Defined Receiver framework [4], where not only the TDOA but also the differential Doppler between the received signals were computed yielding a two-dimensional correlation map.

Interferometric positioning is a semi-codeless technique, since the structure of the spreading sequences is not needed to produce the observables. This property enables on one hand the use of satellites whose code structure is unknown to a particular receiver, e.g. Galileo in an early stage or the Chinese GNSS Compass. On the other hand, the spreading codes are convenient to assign the peaks in the integrated correlation to specific PRNs, determine the satellite geometry and compute the baseline. This is why raw code phase and Doppler estimates are computed in a standalone mode for all satellites in each data stream, before being differenced and matched to the interferometric code and Doppler observables. The baseline vector can be computed using group delay or phase observables that are best modeled into observation equations including terms due to various phenomena such as atmospheric effects for very precise applications and/or for very long baselines. In our case, these differential atmospheric effects can be neglected considering the baseline length. Accurate reference coordinates for one station lead to useable single-difference observation equations.

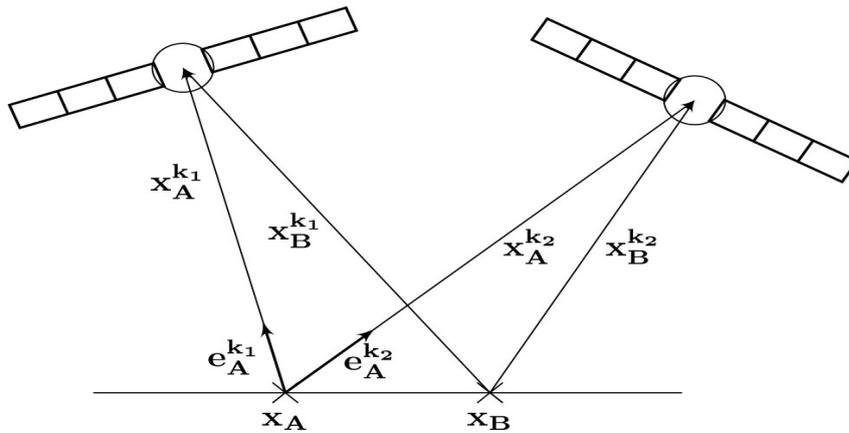


Figure 3: Configuration of satellites and receivers for GNSS interferometric positioning and GNSS relative positioning.

For one observation instant at stations A and B recording signals from the same satellite k , the single-difference equation for code observables reads

$$E [\Delta P_{AB}^k] = [-\mathbf{e}_B^k \quad c] \begin{bmatrix} \Delta \mathbf{x}_{AB}^k \\ \Delta \delta t_{AB} \end{bmatrix} \quad (3)$$

where ΔP_{AB}^k is the vector of measured group delays, \mathbf{e}_B^k is the unit LOS vector from station B to satellite k , $\Delta \delta t_{AB}$ the single difference receiver clock error. The satellite clock error is absent from this equation where all terms correspond to differences between the receivers, or between the reference receiver and the satellite. The single-difference equation for phase observations Δp_{AB}^k contains the additional phase ambiguity \mathbf{A}_{AB}^k .

$$E [\Delta p_{AB}^k] = [-\mathbf{e}_B^k \quad c \quad \lambda \mathbf{I}] \begin{bmatrix} \Delta \mathbf{x}_{AB}^k \\ \Delta \delta t_{AB} \\ \mathbf{A}_{AB}^k \end{bmatrix} \quad (4)$$

where Δp_{AB}^k is the vector of measured phase difference, \mathbf{e}_B^k is the unit LOS vector from station B to satellite k , $\Delta \delta t_{AB}$ the single difference receiver clock error and \mathbf{A}_{AB}^k the single difference ambiguity [5]. These ambiguities can be left floating, though the float solution obtained with less than one minute of phase observations would be less precise than the baseline obtained with group-delays. In other words, with equipment enabling such short recording durations only, fixing the ambiguities is necessary to take advantage of the precision of the carrier phase measurements.

Interferometric observables

Received signal

In our study the GPS L1 raw data have been recorded with the Sparkfun GN3S front-end depicted at page 31 of this NL ARMS. that includes a magnetic patch antenna and a SiGe SE4110L GPS chip [6]. The analog part of the receiver downconverts the signal to an intermediate frequency f_{IF} of 4.1304 MHz, the analog-to-digital converter (ADC) produces samples at f_s equal to 16.3676 MHz that are delivered to a laptop through a USB connection and a driver only is needed to have a functioning receiver. However the recording duration is limited and the stability of our front-end exemplars is a cause for concern, as the signal's spectra have been observed at the time of the experiment, shifted and unbalanced, both to a severe extent. The sampled received signal x is the sum of all K satellite signals

arriving in A, having each its own PRN code, amplitude, Doppler frequency and Time-of-Arrival, plus a white Gaussian noise process n modeling the noise from the environment and the front-end.

Cross-correlation

The cross-correlation function (CCF) of the two received signals yields a correlation map featuring peaks representing localized versions of the individual CCF of all received satellite signals present in the two data streams. These peaks can be best described by the auto-correlation function (ACF) of the transmitted GPS signal, which in turn can be approximated to the ACF of its one millisecond periodic PRN spreading code. Assuming the received signals have the same Doppler, their differential code phase can be found by maximising the sampled CCF of the discrete signals collected at the two stations

$$\hat{c}_{AB}[m] = \frac{1}{N_p} \sum_{n=0}^{N_p-|m|-1} s_A[n]s_B[n - |m|] \quad m \geq 0 \quad (5)$$

where N_p is the length of signals s_A and s_B . However the two receivers being far enough from each other and having unsynchronized, drifting, clocks, the signals received from one satellite most of the time have different Doppler frequencies at the two stations. The cross-correlation of the received signals is then maximal only for the differential code offset and the differential Doppler frequency. The sampled CCF $c_{AB}[m, f]$ is computed iteratively after introducing bins corresponding to frequency shifts on the signal. The baseline length together with the sampling frequency impact the length of the correlation interval containing the received signal's energy. These quantities, as well as the total length of the experiment, influence the computational cost of the correlation operation. We chose an irregular frequency grid having more bins close to the IF frequency than far from it. Denoting \odot as the circular correlation operator using the DFT [6], the correlation c of one millisecond of data collected at A and B writes

$$c(\tau_{AB}, f_{AB}^d) = s_A(t, f) \odot s_B(t + \tau_{AB}, f + f_{AB}^d) \quad (6)$$

The CCF of two noisy GPS signals has a SNR far lower than the CCF of one received GPS signal and a clean replica. Consequently to detect the satellite's energy, we chose to non-coherently integrate the received signal's correlation over 1000 code periods. The cross-correlation of two frequency-shifted copies of the signals then yields a two-dimensional correlation map

containing all the TDOA and differential Doppler information of the signals received at the two stations.

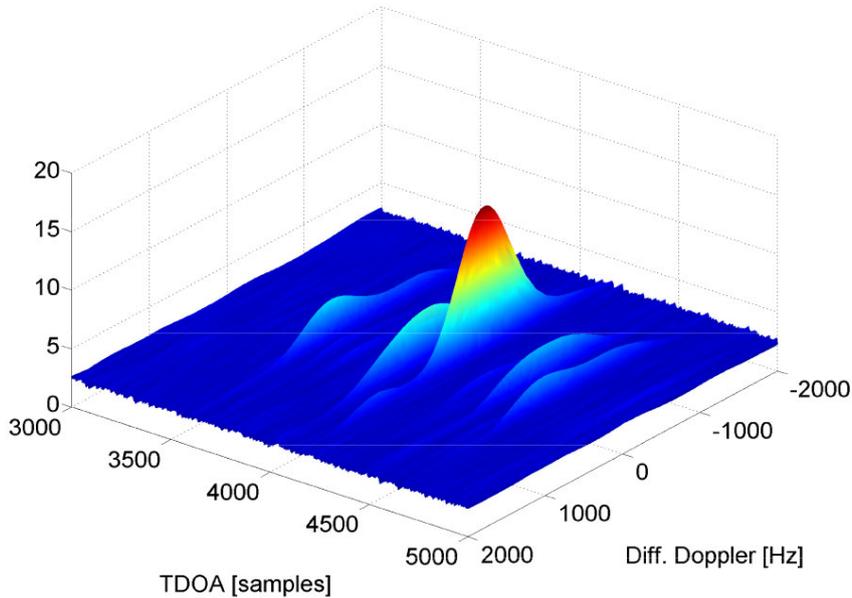


Figure 4: Region of a correlation map containing the received signals energy, obtained for a one second observation duration. Horizontally the map is parametrised by the TDOA and the differential Doppler, while the color-coded height corresponds to the correlation value. This figure shows peaks localised in time and frequency corresponding to the cross-correlation of the signals simultaneously received at the two stations. Outside the peaks region the correlation surface can be considered as flat.

Measurements

The resolution of the differential measurements is limited by the dimensions of the correlation map, which is in turn conditioned by the computational cost of the correlation operations and the available processing memory. On one hand the sampling frequency drives the time resolution of the code phase observables. On the other hand the frequency resolution of the Doppler observables is constrained by the integration time as for the acquisition in a classical standalone receiver. The observables can then be refined in the time-domain by interpolating the correlation values in the detected frequency bin, resulting in TDOA estimates with better resolution. While TDOA measurements can be extracted from the correlation map, differential phase observables are more difficult to obtain and not considered here, since our hardware does not allow the production of these measurements.

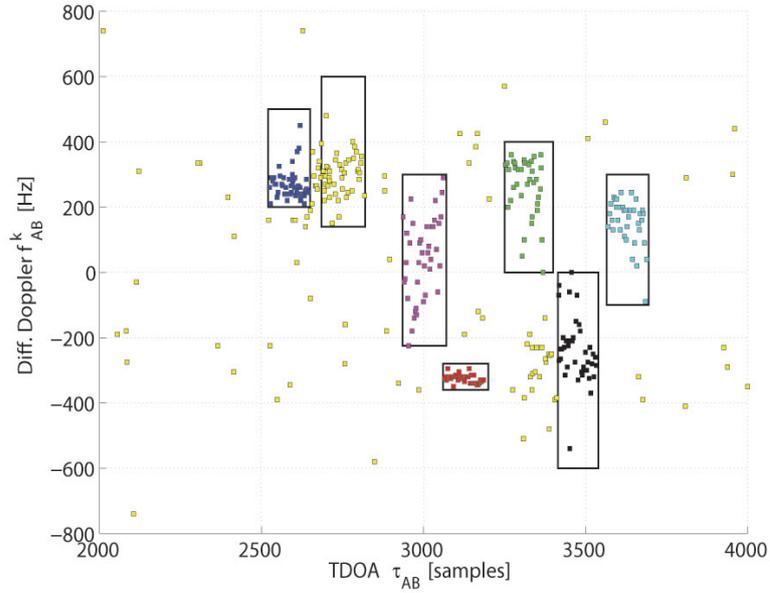


Figure 5: Differential code-Doppler observables obtained from 37 observation durations of one second each. Several color-coded clusters of points corresponding to different satellites can then be identified.

Identification of satellites

The map of observables contains data clusters corresponding to visible but unknown satellites we identify thanks to their differential Doppler. Indeed the code-Doppler couples of all satellites in each data record can be computed separately with an open-loop processing that uses non-coherent integration and produced code phase and Doppler observables for each data stream. This open-loop processing is preceded by standard code and carrier acquisition steps, with increasingly fine frequency estimation stages. The sequential processing then uses integration to increase the sensitivity of the reception and yields at each recursion estimates for the code phase $\hat{\tau}^k[n]$ and Doppler $\hat{f}_k^d[n]$. One correction is applied to compensate for the non-integer number of samples per code period and for the Doppler shift on the code sequence giving a new code phase $\hat{\tau}^{k*}[n]$. The strongest satellites in each data record were kept and those obviously unable to lock discarded. The satellite signal strengths are depicted in Figure 6.

Collecting the code phase and Doppler information for all satellites yields maps for both signals, which are differenced and brought back to

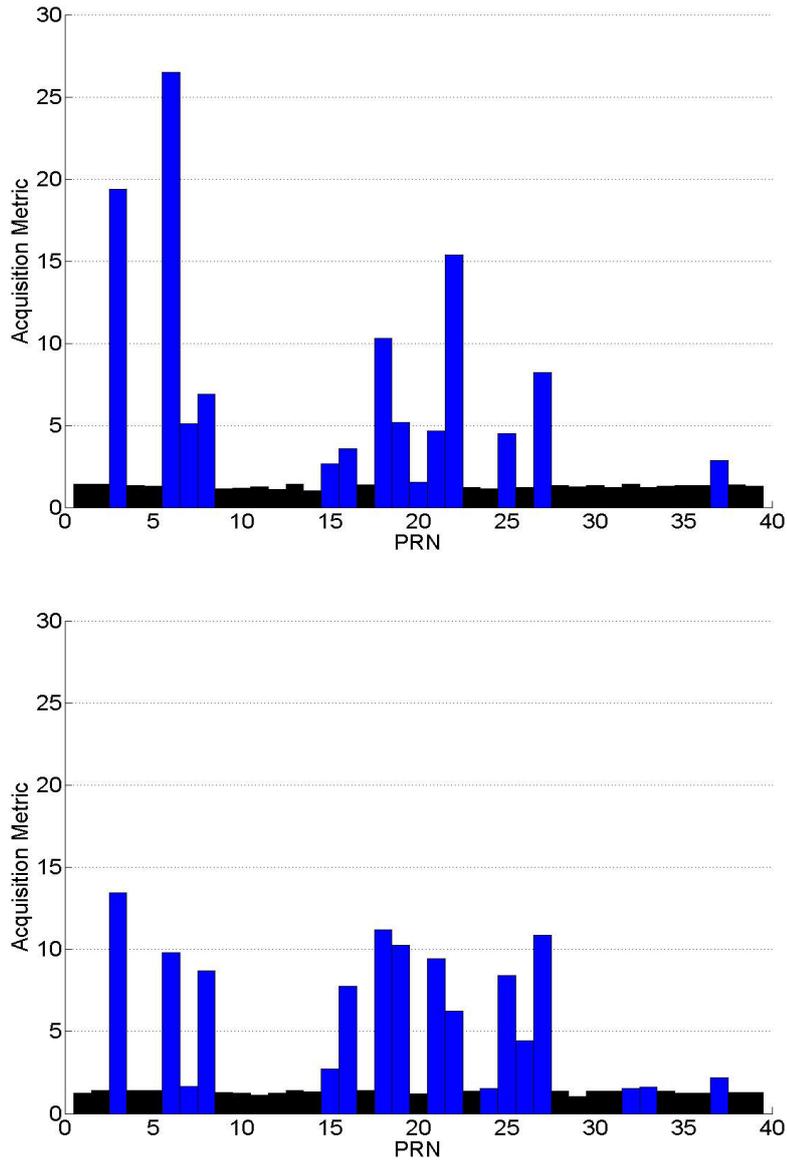


Figure 6: Acquisition metrics using non-coherent integration of ten milliseconds long signals recorded at the two stations, after correcting the initial start offset. The bottom plot shows a significant degradation in the received signal strength compared to the upper station. We link this phenomena to an anomaly in this specific exemplar of the front-end.

a one code period. These differential observables have a known PRN and then constitute a reference for those from the direct interferometry. The matching between the maps of observables from the interferometry and from the standalone processing, then the standalone observables, need only be precise enough to discriminate different satellites within a map.

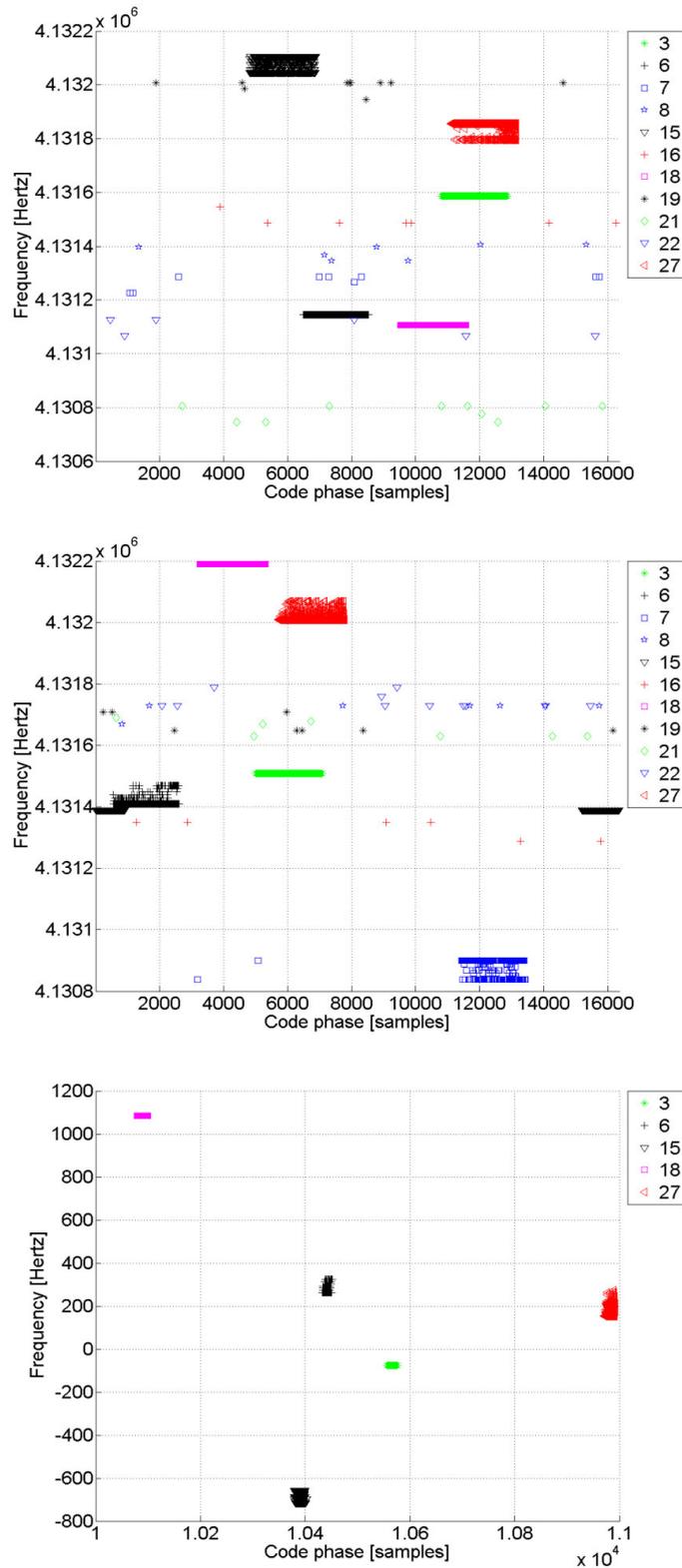


Figure 7: Code-Doppler results from the standalone processing at the two stations (top and middle) and differenced results for the strongest satellites (bottom) that is used as a reference for identifying the interferometric observables.

Correction of the clocks

Each receiver is based on several clocks: the oscillator in the front-end and the laptop hardware clock, from which a signal time-tag is produced. Both laptop clocks were synchronised using a LAN connection to an Internet time server shortly before the experiment, and the recordings are triggered using the job scheduler on each laptop. Now, the different clocks are drifting causing difficult synchronisation of the records' start. The standalone closed-loop PLL and DLL processing at each receiver delivers prompt correlator outputs containing the navigation data. The correlator output sequences having a 1 kHz sampling rate can be cross-correlated, yielding a coarse estimate of the offset between the records' start. Long enough sequences should be taken to ensure common data bits and the estimation is repeated for several satellites adding reliability to the estimated value.

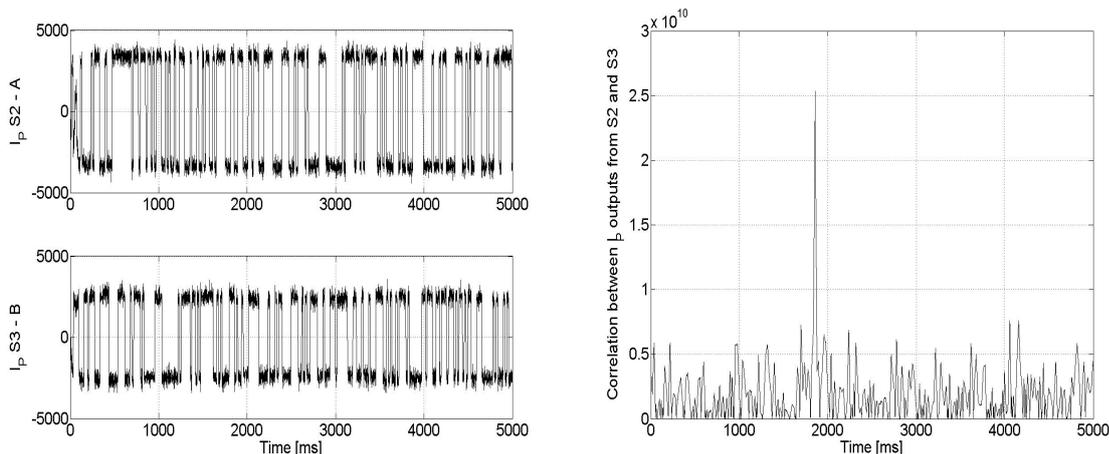


Figure 8: In-phase prompt correlation outputs at the two stations (left) and coarse estimated time-delay between the start of the records, resolution of one code period

Next, the offset is computed for all recordings in the measurement campaign and depicted in Figure 9, where it changes considerably in time, thus showing an anomaly in the analogue part of one of the software receivers.

Conclusions

In this paper we presented a methodology to produce and identify differential code measurements used for interferometric positioning with GNSS signals. Having dealt with the synchronisation between the start of the

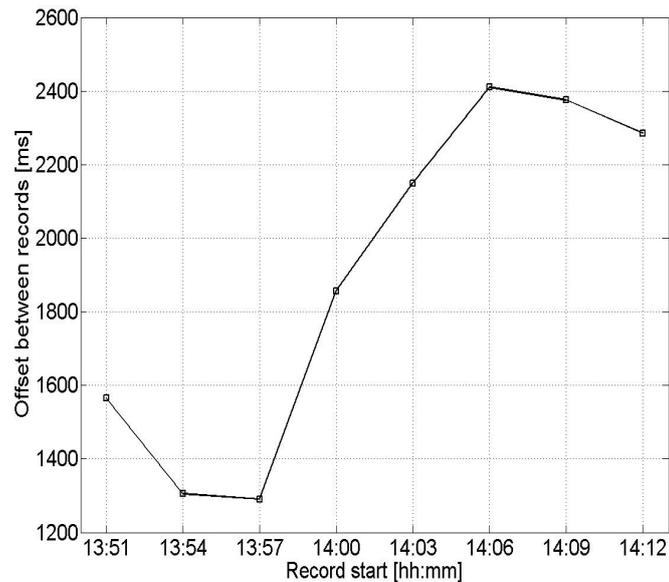


Figure 9: Differential clock offset for all signals recorded during the measurement campaign.

records, the processing of the raw data collected at two stations yields integrated cross-correlation maps. Processing both data streams separately enables the identification of the received satellite signals in the correlation maps. The whole experiment is then semi-codeless. The production of code measurements is well understood while the collection and processing of phase measurements, theoretically leading to a better baseline precision, remains challenging especially with the front-ends in our possession. Indeed, while the interferometric concept is straightforward and powerful, while the data collection is easy, the expected overall performance cannot reach a geodetic level given the low price of our equipment.

References

- [1] G. Seeber, *Satellite Geodesy*, 2nd edition. Berlin/New York: de Gruyter, 2003.
- [2] C. C. Counselman and I. I. Shapiro, “Miniature interferometer terminals for earth surveying,” *Journal of Geodesy*, vol. 53, pp. 139-163, 1979.
- [3] P. F. MacDoran, “Satellite Emission Radio Interferometric Earth

- Surveying Series - GPS Geodetic System,” *Bull. Geodesy*, pp. 117-138, 1979.
- [4] B. Muth, P. Oonincx and C. Tiberius, “GNSS Software-based Interferometry,” *European Journal of Navigation*, vol. 7, pp. 10-16, August 2009.
- [5] P.Misra and P. Enge, *Global Positioning System - Signals, Measurements, and Performance*, 2nd edition. Lincoln, (Massachusetts): Ganga-Jamuna Press, 2006.
- [6] D. Akos, K. Borre, N. Bertelsen, P.Rinder and S. Jensen, *A single-frequency software-defined GNSS receiver*, Birkhäuser, 2007.

Terrain Referenced Navigation Using a GPS Approach

Patrick Oonincx & Daniela Vaman

Introduction

To navigate, a position estimation system is needed. Inertial Navigation System (INS), Satellite based navigation such as the Global Positioning System (GPS) and Terrain Referenced Navigation (TRN) represent three different types of technology to obtain position estimates.

An INS uses measurements of acceleration to compute relative velocity and displacement. Accelerations and rotations are measured along three perpendicular directions. This type of technology was already being used in the early sixties for the guidance of Intercontinental Ballistic Missiles (ICBMs). Figure 1 (left) shows a picture of the INS used in the Titan II missile. Today's state of the art in acceleration sensors allows the realization of a solid state INS with a weight below one gram. Figure 1 (right) shows an example of a solid-state accelerometer integrated on a miniature circuit board.

Already, more than 35 years ago, the first TRN system, using a radar altimeter and a terrain elevation database for position estimation, was tested on a Hound-Dog missile. The first GPS satellite was launched more than 30 years ago.

All three types of navigation technology are used in military systems. For example, the AGM-86B cruise missile uses a terrain contour-matching guidance system, combined with an INS. A later model, the AGM-86C/D uses an onboard GPS coupled with its INS. With GPS, the vast consumer-electronics market has been a strong driving force behind innovations in

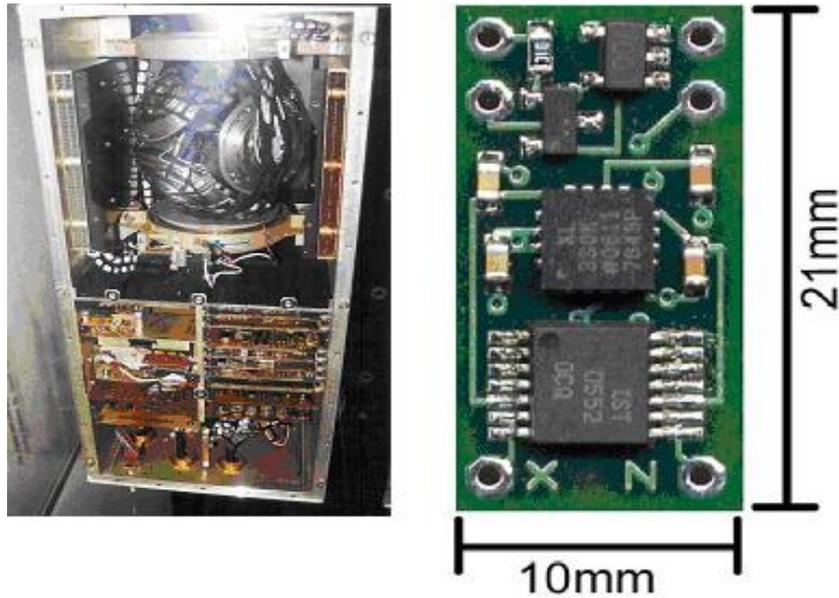


Figure 1: INS used in a Titan II Intercontinental Ballistic Missile (left) and a solid state 3-axis accelerometer with I/O circuitry (right).

the area of GPS-based applications, and has thus provided the business case for the continuous development of receiver technology. However, for Terrain Referenced Navigation (TRN), such a vast consumer market has never existed. With the advent of GPS, some of the originally intended applications of TRN even ceased to exist. However, given today's state-of-the-art jamming technology it is fair to assume that the uninterrupted availability of GPS in future conflicts will be impacted.

Due to the dissimilarity with GPS and INS, an integration of TRN can increase the availability of accurate position estimates in situations where GPS is temporarily unavailable and the reduction of accuracy of the INS with an increase in time limits the operational capability. Although not interesting for consumer-type applications, certain military operations can benefit from such a capability.

The developments in the area of digital signal processing, combined with the continuing increase in computing performance have enabled the realization of coin-size GPS receivers. Developments in the area of acceleration sensors have provided the possibility to build solid-state INSs which weigh less than a single gram. The developments in the area of data storage enable the creation of hand-held devices containing world-wide hi-resolution elevation maps.

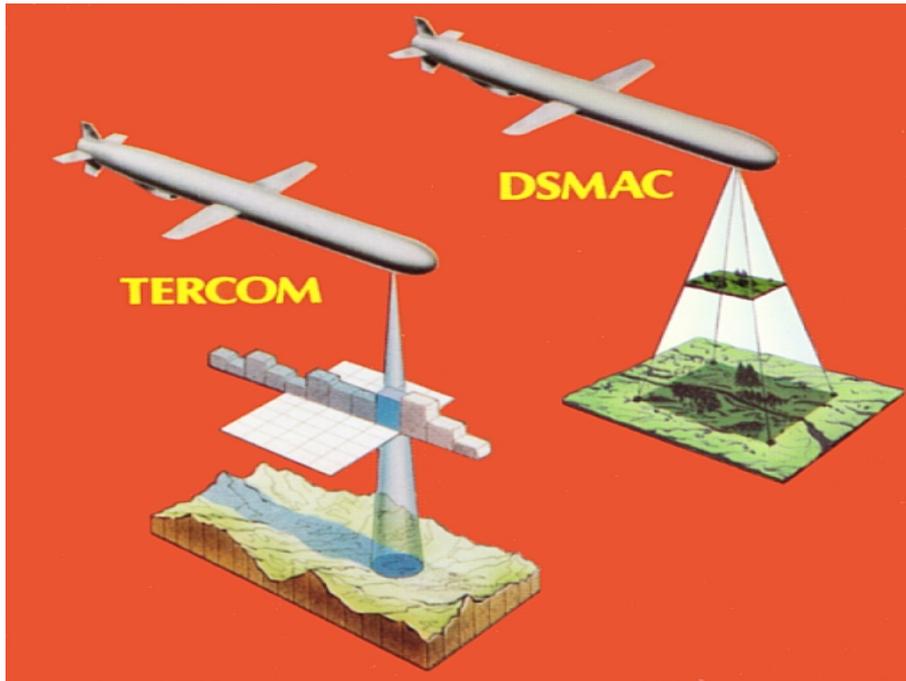


Figure 2: Two types of terrain related navigation for missile guidance: TERCOM generates position fixes by sampling terrain heights and correlating the elevation profile with ground truth information; DSMAC correlates stored images with snapshots of the terrain beneath the missile to generate a position fix.

With the availability of the technical components needed to realize a system that integrates GPS, TRN and INS, the challenge now lies on the development of the functionality which utilizes the available measurements to estimate the location. Application specific performance requirements regarding availability, accuracy, integrity, update-rate and susceptibility to intentional interference, in combination with specific possibilities for trade-offs, will cause the optimum combination to be determined by the intended application(s).

Terrain Referenced navigation vs. GPS

To be able to define both hardware and processing requirements for a given application, a better insight into the possibilities and limitations is needed. The goal of the research conducted within the NLDA's Navigation Technology department is to contribute to the creation of a design framework which can be used to match technology and processing concepts to a particular set of system performance requirements. Here the focus lies on exploring the potential of digital signal processing concepts for signal acquisition and tracking in the area of TRN.

TRN techniques provide the user's position by using a stored digital map of terrain elevation data together with real time measurements of a platform's height above the ground (altitude/depth). The basic rationale of the system is to find the best match between the features of the environment and the stored database, in the same way that a person would compare a landmark with a map, see Figure 2. The methods for obtaining position from the measurements can be divided into two categories: sequential (each measurement is processed separately) and batch processing (a series of terrain height measurements, known as transect, are processed together). The TRN system we present in this paper belongs to the second category.

In order to have a clear match between the measurements and the terrain map, it is necessary to have a good indication of the velocity and the orientation of the platform. When dealing with measured errors in the speed and course parameters we found similarities with the problems associated with the tracking of arrival times and Doppler offsets in a GPS receiver.

To estimate the user's position, velocity and time in a GPS receiver a series of different processes must be accomplished, namely capture the signals transmitted by the satellites, identify the satellites in view, measure the signal's transit time and Doppler shift and decode the navigation message to determine the position, velocity and clock parameters. The transit time and Doppler shift are changing parameters in time, because of the movement of the satellites. However, once determined on a coarse scale it is possible to keep track of these parameters in an elegant way, because these parameters will only change slightly during each measurement. In a TRN system course and speed of a platform, e.g. a submarine, are also expected not to change heavily during a dense set of measurements.

Furthermore, with GPS the presence of a Doppler shift can cause the estimation of the right transit time of a satellite to deteriorate, while in the case of TRN an error in the estimate of the velocity deteriorates the correlation of the measurements and the terrain database. Similar to the correlation with multiple codes (satellites) in case of GPS, in TRN an uncertainty in the track of a platform will require the correlation with samples from a range of directions. These issues also apply to the tracking phase.

In the next section we will revisit GPS signal processing techniques, that will be used in our proposed TRN tracking algorithm.

GPS revisited

This section focuses on the digital processing stages that take place in a GPS receiver. After a short description of the GPS signal, following [1,2], we will recall the concepts necessary to understand the similitude between GPS signal processing and TRN, as briefly discussed before.

Each GPS satellite transmits messages continuously on two radio frequencies in the UHF band, referred to as link L1 and L2. Each GPS signal comprises three components:

- carrier (refers to the RF sinusoidal signal with frequency $f_{L1} = 1575.42$ MHz and $f_{L2} = 1227.60$ MHz),
- ranging code (refers to a family of pseudo random noise (PRN) sequences: each satellite transmits a unique (civil) coarse-acquisition C/A code on L1 and a precise (military) P code on both L1 and L2),
- navigation data (refers to the coded message consisting of data on the satellite's status, ephemeris and clock bias parameters).

For simplicity we will restrict ourselves here to signal models using only the C/A code on the L1 frequency. This comes down to

$$s(t) = \sqrt{2P} D(t) x(t) \cos(2\pi f_1 t + \theta_1),$$

with P the average power, $D(t)$ the navigation message and $x(t)$ the C/A code. The code $x(t)$ is of length 1023 bits and repeats itself every millisecond. The message $D(t)$ is of length 12.5 minutes and is transmitted at 50 bits per second. Both x and D are built up by a series of block function (chips) of amplitude 1.

In a GPS receiver, the incoming signal must be multiplied with a locally generated C/A code (corresponding to the specific satellite) and mixed with a locally generated carrier wave. Two parameters are needed: the (Doppler) frequency and the code phase of the incoming signal. The motion of the transmitter satellite relative to the GPS receiver causes the Doppler

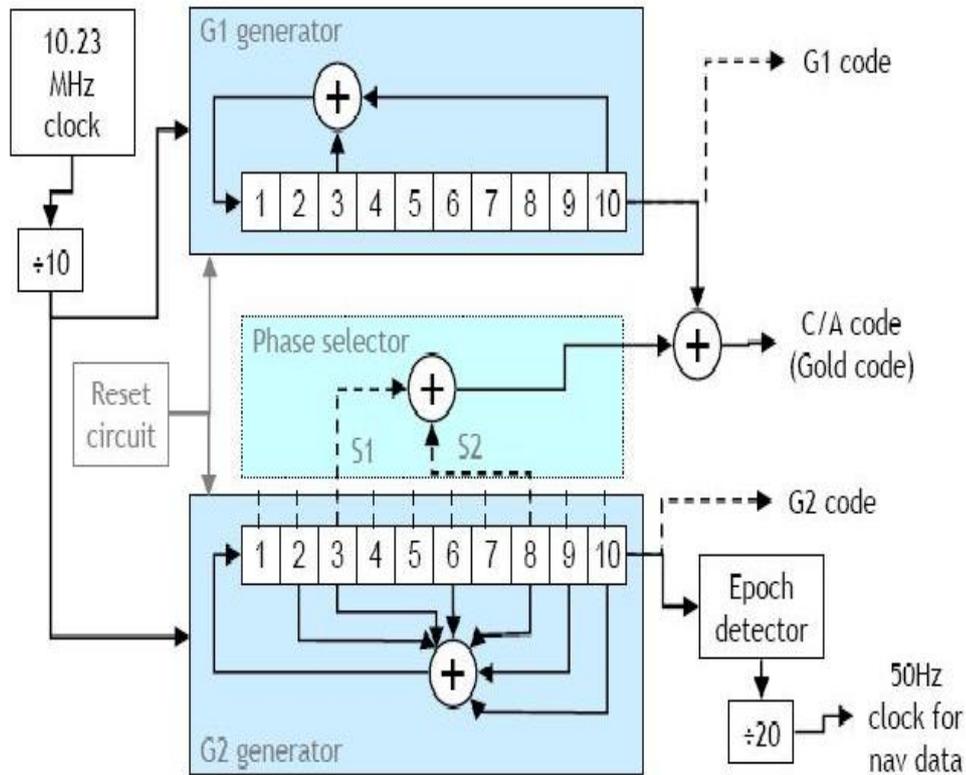


Figure 3: C/A code generator combining two maximum-length sequences into a Gold sequence.

shift: a difference between the frequency of the received signal and the frequency at the source. Due to the transit path, the point in the data block where the C/A code begins is also unknown to the receiver.

Frequency and code offsets are unknown and must be estimated. This is done in a two-step process. First, the receiver attempts to acquire the signal, by finding coarse values for the parameters. The acquisition module performs a two dimensional search on all possible shifts in code phase and frequency. A fast algorithm for this search is known as parallel search and is based on a paper by Coenen and Van Nee [3]. Once the parameters are estimated, they will be refined and kept track of as they change continuously over time. The tracking module is implemented with feedback control loops, containing two parts. Code tracking continuously adjusts the replica code to keep it aligned with the code in the incoming signal. Carrier frequency/phase tracking generates a sinusoidal signal to match the frequency and phase of the incoming signal. The used methods are based on the GPS signal properties.

Our attention has been mainly oriented towards the C/A codes $x(t)$ and the way their properties are keyed in the GPS digital signal processing stages. This code belongs to a unique family of sequences referred to as the Gold codes [4]. A Gold code is the sum of two so-called maximum-length sequences, see Figure 3, of the same length and is always balanced, i.e. the numbers of zeros and ones differs by only one. The GPS C/A code contains 512 ones and 511 zeros. The C/A code is characterized by important correlation properties, that play a key role in the acquisition stage in a GPS receiver. There the correlation function $R^{k,l}(\tau)$ is computed, given by

$$R^{k,l}(\tau) = \frac{1}{T_c} \int_0^{T_c} x_k(t) x_l(t - \tau) dt,$$

with k, l denoting satellite identification numbers and T_c the code length on which we integrate. In case of the Gold code we have the following properties

$$\begin{aligned} R^{k,k}(0) &= 1, \\ |R^{k,l}(t)| &\leq 65/1023 \approx 0.064, \quad t \neq 0, \end{aligned}$$

for any k, l including $k = l$. Obviously, in case of a Gold code, the correlation function is a strong tool, since it almost vanishes for every combination of satellites and time transits except for the appropriate one.

For acquiring a signal, the receiver generates a replica of the known C/A code and attempts to align it with the incoming code by sliding the replica in time and computing the correlation. The correlation function exhibits a sharp peak when the code replica is aligned with the code received from the satellite, providing the code phase estimate.

To keep track of transit times Delay Lock Loops (DLL) are used, based on an early-late discriminator, see [2]. The DLL is based on a two-correlator structure. Each correlator is set with a small time offset relative to the promptly received signal code timing phase, producing early and late signals, as seen in Figure 4.

The three correlation outputs (early, late and prompt) are compared to see which one provides the highest correlation. Figure 5 shows two examples of this approach. At the left the incoming code was shifted in time, resulting in a maximum correlation for the late replica, while at

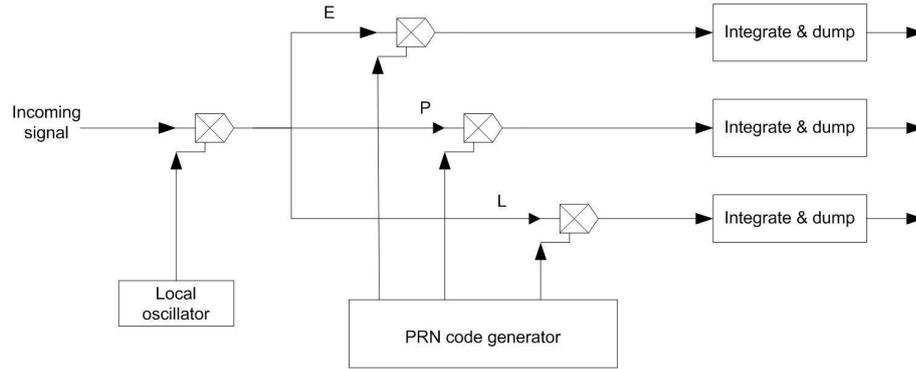


Figure 4: A basic GPS DLL diagram: the incoming signal is multiplied with a generated carrier wave and a generated code replica. In three branches, the Early, Late and Prompt correlation results are Integrated and Dumped.

the right, the incoming code remained unchanged in time, resulting in a maximum correlation for the prompt replica.

The early-late tracking loop performs well for GPS signals, due to the type of code that is used. For general signals (codes) there is no guarantee that this tracking loop performs equivalently well. However, in the case of terrain elevation data we expect this type of feedback loop to be of interest to keep track of course and speed of a platform, for at least certain regions of a terrain. For tracking the Doppler frequency offset a similar feedback loop approach is used, called the Phase Lock Loop (PLL). This tracking loop is not discussed here, since it will not be used in our proposed TRN algorithm.

The TRN approach

In the TRN system that we deal with, a sequence of samples is collected by the platform, formed from several depth measurements along a line. The database of terrain maps is searched for the ground truth profiles that best match with the recorded profile and this indicates the present position. The estimation process consists of correlating the measured elevation profile with database extracted elevation profiles.

Depth values from positions between the known elevation points are often required. These values are extracted by interpolating the map. From several interpolation techniques available, the bicubic method has been chosen for implementation in the algorithm. The bicubic interpolation [5], fits a bicubic surface through existing data points. The new interpolated

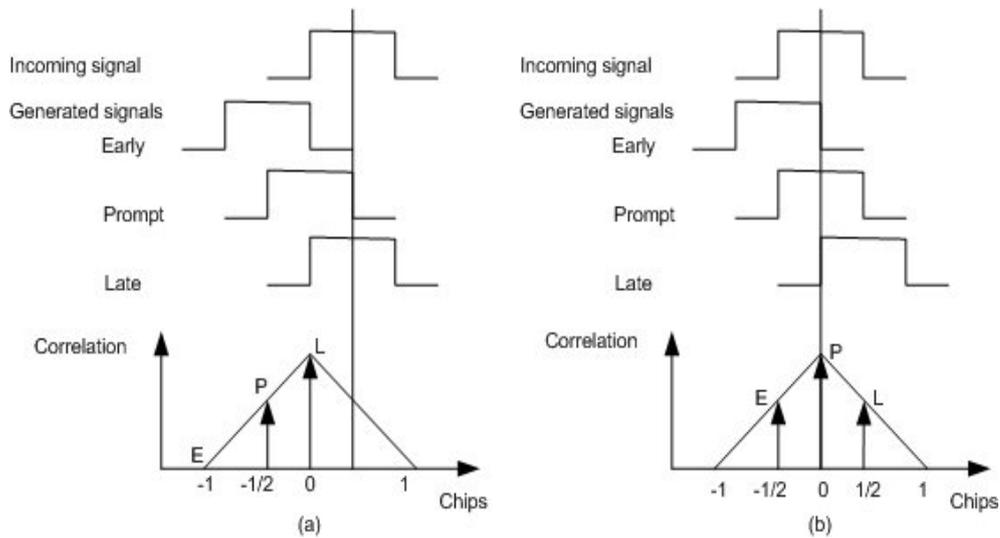


Figure 5: Code tracking: three local codes are generated and correlated with the incoming signal. (a)The late replica has the highest correlation; code sequence must be delayed. (b)The prompt code has the highest correlation; code sequence is properly tracked.

value is computed by weighting sixteen known surrounding elevation points in the database. The method is known for producing smooth surfaces. Extracted depth values will deviate from the true depth, but interpolation errors can be minimized if the velocity and course of the vehicle are accurately determined. An error in the course will determine shifts in the direction of the extracted tracks compared to the real one. An error in the speed will cause a modification in the length of the extracted tracks compared to the real one.

In order to expect accurate results from the TRN system, velocity and course of the platform should be acquired and monitored. As mentioned before, this issue is comparable to the topic of tracking code onset times and Doppler offsets in a GPS receiver. In accordance with the GPS receiver approach, a TRN correlation algorithm in two stages was implemented. During the ‘acquisition module’ a two dimensional search of all possible shifts in speed and course is performed, calculating coarse estimates for the two parameters. Next, the ‘tracking module’ begins by adding new depth measurements and deleting old ones. The position is progressively updated. The values of the speed and the course are refined and kept track of as they change during the platform’s movement.

Two tracking loops will be required. A block diagram of the implemented TRN system has been depicted in Figure 6 to illustrate the global

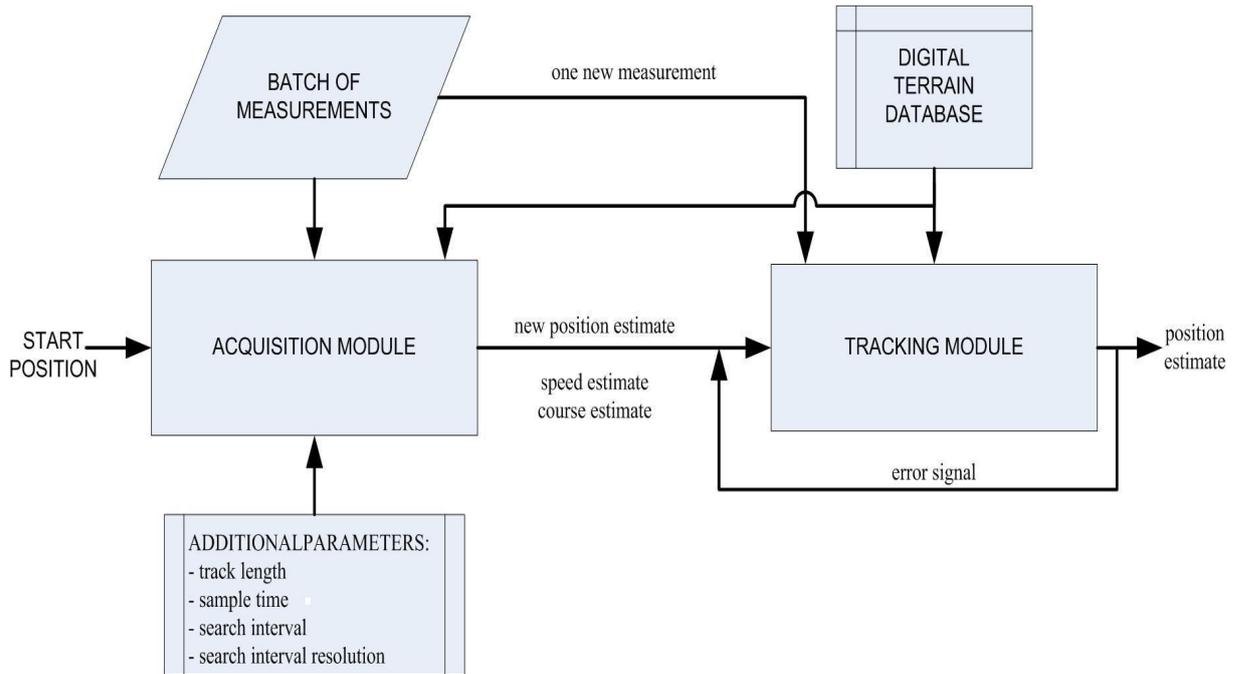


Figure 6: Block diagram of the proposed TRN algorithm.

operating structure. Here, we give a brief description of the functionality of each block in the diagram.

Digital terrain database

Shuttle Radar Topography Mission (SRTM) digital elevation models have been used as maps. These types of files contain only binary data corresponding to the height, with no georeferenced or other related information. The resolution of the map is 3 arc seconds, corresponding to 90m between elevation points. The SRTM elevation models cover the continental surface of Earth. It was preferred to have a real-world database, as we lacked access to accurate sea bed charts. However, a desert area has been used.

Batch of measurements

The depth measurements which are made by the platform come straight out of the reference database. They are extracted using the ‘correct’ values for speed, course and position. If the coordinates do not concur with the grid points, interpolation techniques are used. This profile is used in the error-free simulations. However, to get closer to a real approach, depth measurements errors must be added to this profile.

Acquisition block

The focus of the research was to investigate whether the platform’s position

can be tracked in time similar to a GPS approach, by estimating the velocity and the orientation of the vehicle. Based on these considerations, we assume that the start position has been accurately determined. Meanwhile the speed and the course are due to be estimated. The acquisition module is built up as a search process of these two parameters. A long batch of measurements gives a terrain elevation profile which is compared to several extracted profiles. Serial search acquisition is performed: for each value of the speed, different values for the course are taken and a profile using these values is extracted and correlated with the measurements. When all possibilities have been verified, a correlation matrix is built and plotted, and the best estimates are picked up and delivered to the tracking module, together with the position estimate.

Tracking block

The number of measurements in the tracking phase is fixed. When the next depth measurement is done by the platform, the profile propagates one measurement ahead, discarding the oldest one. The tracking loop has been built in the same manner for both speed and course estimation. The GPS early-late tracking concept has been used as a basic principle. A diagram of the basic TRN tracking loop has been depicted in Figure 7.

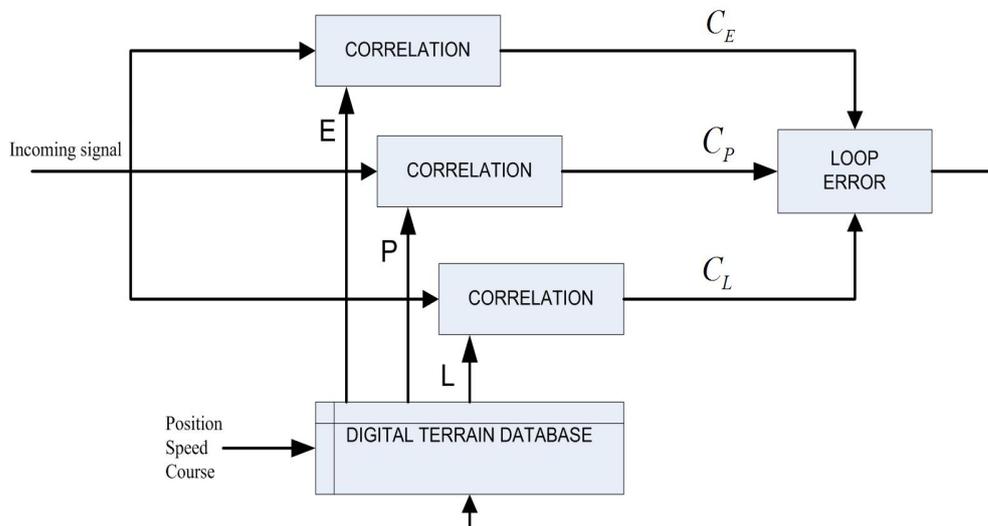


Figure 7: A basic TRN tracking loop diagram.

The incoming signal consists of the batch of measurements. We generate two local replicas, by extracting profiles from the database using two small offsets relative to the estimated speed and course. A prompt replica of the profile will be extracted, using the exact estimated values of the parame-

ters. The correlation between these tracks and the measurement profile is computed. The output is a numerical value indicating a matching property of the specific signal replica with the incoming signal. An error signal is provided from the combination of the correlation outputs.

Simulations and Preliminary Results

The ‘correlation’ method used in our approach does not compute correlation in a strict statistical sense, but uses a matching function that shows how well a measured bottom profile matches the map. For such matching functions we consider a series of m measurements: d_i will denote an incoming measured depth and $h_{i,k}$ the height in the map in the k th location, or simply written h_i , with $i = 1, \dots, m$. The different tested methods are introduced below:

- The Pearson product-moment correlation coefficient. This coefficient measures the linear dependence between two variables, see[6]. Based on samples of paired data, the coefficient is calculated as:

$$C_{d,h} = \frac{\sum_{i=1}^m (d_i - \bar{d})(h_i - \bar{h})}{\sqrt{\sum_{i=1}^m (d_i - \bar{d})^2} \cdot \sqrt{\sum_{i=1}^m (h_i - \bar{h})^2}},$$

with \bar{d} and \bar{h} the mean of the data d and h respectively.

- The average distance. In this case the absolute distance between the measured depth and the map depth in the candidate points is computed and then averaged over m :

$$C_{d,h} = \frac{1}{m} \sum_{i=1}^m |d_i - h_i|.$$

- The squared difference. The differences between the depth measurement and the correlation depths are calculated and then these values are squared and summed and divided by the number of measurements. The standard deviation of the differences is added to reduce the effect of the systematic errors. For non-correlating tracks, the deviation will be quite large. The formula for this correlation measure is given by

$$C_{d,h} = \frac{\sum_{i=1}^m (d_i - h_i)^2}{m} \cdot \sigma(d - h),$$

with $\sigma(d - h)$ the standard deviation of the difference data $d - h$.

Correlation between a measured profile and several extracted tracks has been calculated using, in turn, the three formulas. For simulations the following set up has been used:

1. The measured profile is error-less and is extracted from the database;
2. The extracted profiles are selected by varying, in turn, the error in the velocity and orientation;
3. A set of different scenarios were used.

Figure 8 shows an example of the difference in the three different types of correlators for different errors in the speed (the functions have been rescaled or reversed, for the sake of comparison). For the errors in course a similar figure can be depicted. The average distance function is imple-

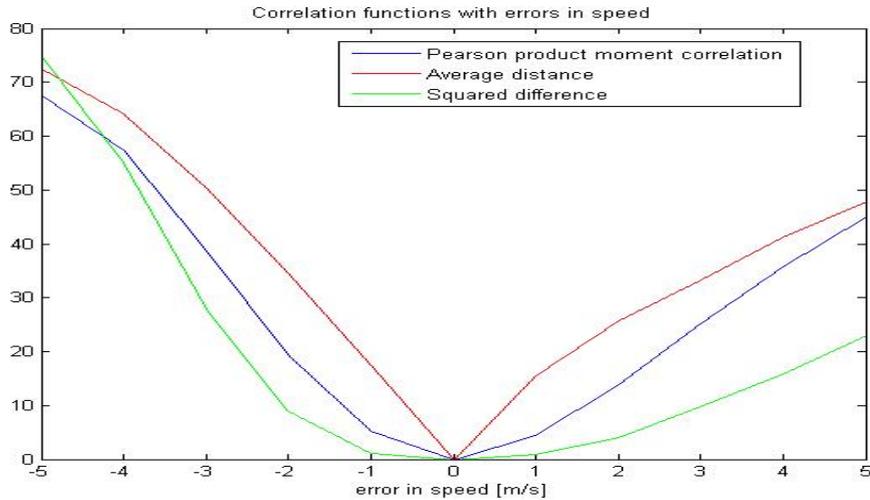


Figure 8: Correlation functions dependent on speed parameters.

mented in our TRN algorithm. It has been chosen because it is the most discriminating in the area of interest (convex, ‘V’-shaped). Choosing this correlating function will result in more notable differences for small errors of the parameters.

A key driver on the performance of the overall system is the terrain and its characteristics: the uniqueness of the ground profiles within the

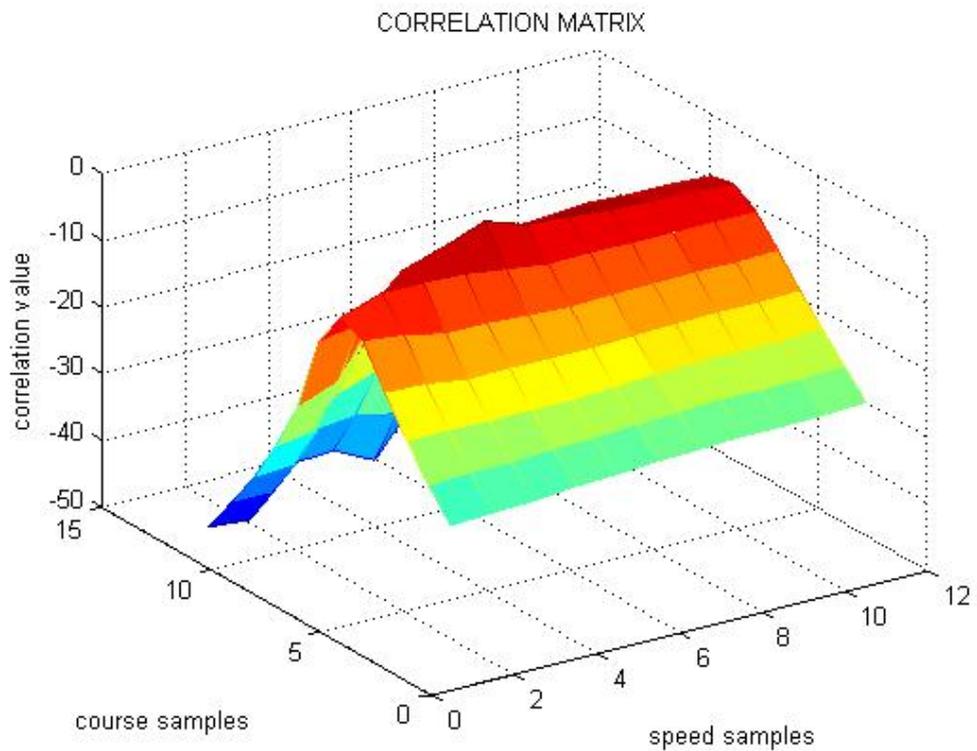
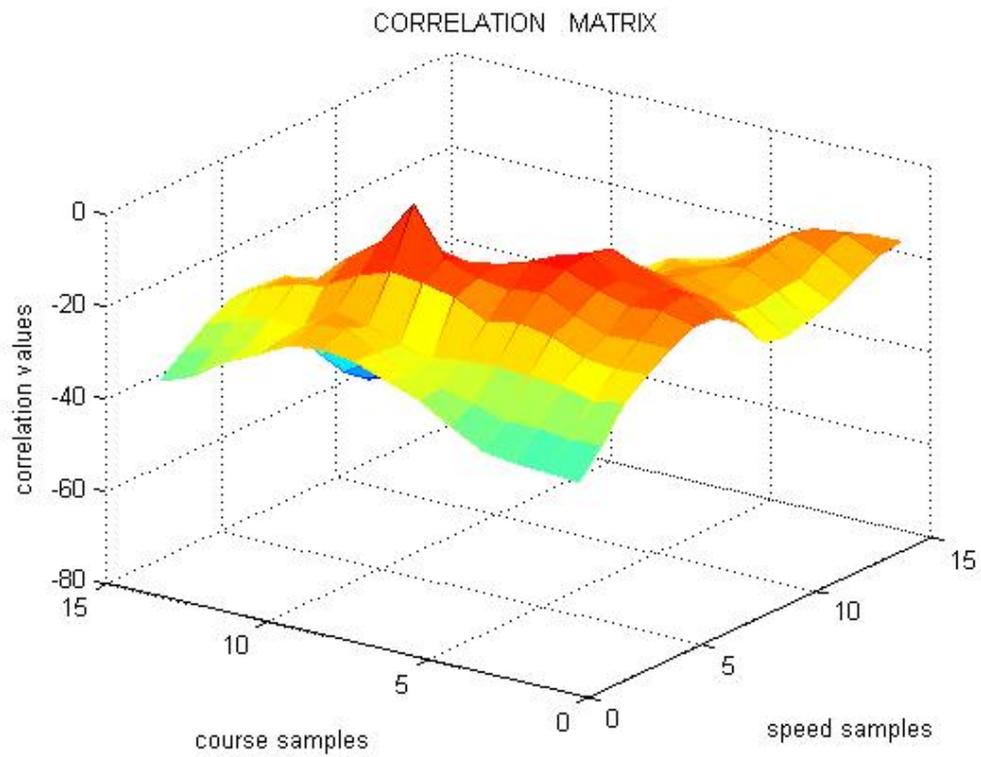


Figure 9: Displayed correlation matrices showing a clear visible peak (up) and a less visible peak (below)

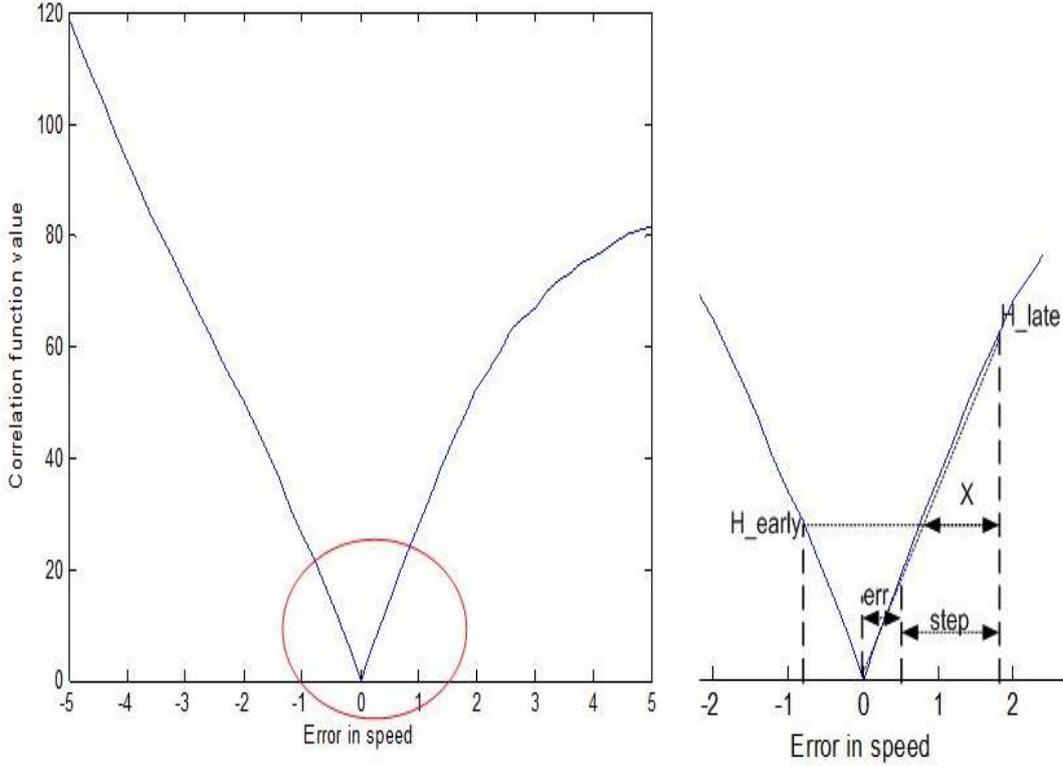


Figure 10: Average distance correlation function and error correction.

searched map area. The cross correlation between tracks may have a dynamic range, due to the unpredictability of terrains. The correlation plot will show a significant peak (auto-correlation peak) when the tracks are matched. However, additional peaks are likely to appear, leading to position ambiguity. Figure 9 shows two different situations: (Top) the auto-correlation peak can clearly be seen; (Bottom) the auto-correlation peak is less obvious and an additional peak can be noticed.

The existence of a prominent auto-correlation peak has been the premise of adapting the GPS ‘early-late’ tracking concept to TRN. Both tracking loops have been built up following the same rationale. Due to this remark, we restrict ourselves to the loop that follows the changes in speed. The early and late versions are separated by a distance we called step. Assuming the symmetry of the function and linearizing it, a correspondence between the error and the ‘X’ segment is assumed, as shown in Figure 10.

The error signal is calculated using:

$$X = \frac{\text{step} \cdot (H_{\text{late}} - H_{\text{early}})}{H_{\text{late}}},$$

with X , step, H_{late} and H_{early} as in Figure 10.

The performed tests have given good results, as the correction is determining a convergence for the estimated speed. However, the terrain shape has a big influence on the number of iterations needed in order to obtain a good estimation, as can be seen from Figure 11 (for speed, with similar results existing for course).

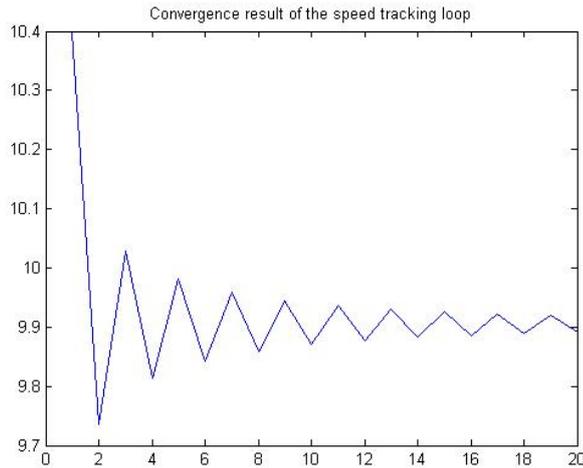


Figure 11: Convergence results of the tracking loop for speed.

Conclusions and Future Work

In this article a GPS inspired approach has been developed for TRN systems. We investigated whether the acquisition and tracking GPS concept has the potential to be applied in the TRN field with good results. The performance of the proposed system is analyzed using MATLAB simulations. The results obtained so far have not confirmed this assumption. Good estimates of the velocity and course of the vehicle are obtained. However, we are aware that we can not conclude that the testing is representative for terrains in general.

At most we can say it is representative for our own database. The terrain characteristics are a key driver in the performance of the overall system. Measurement and database resolution limitations can make it difficult to distinguish features nearby, while environments often contain repeating features.

The approach presented in this article has mainly been based on [7].

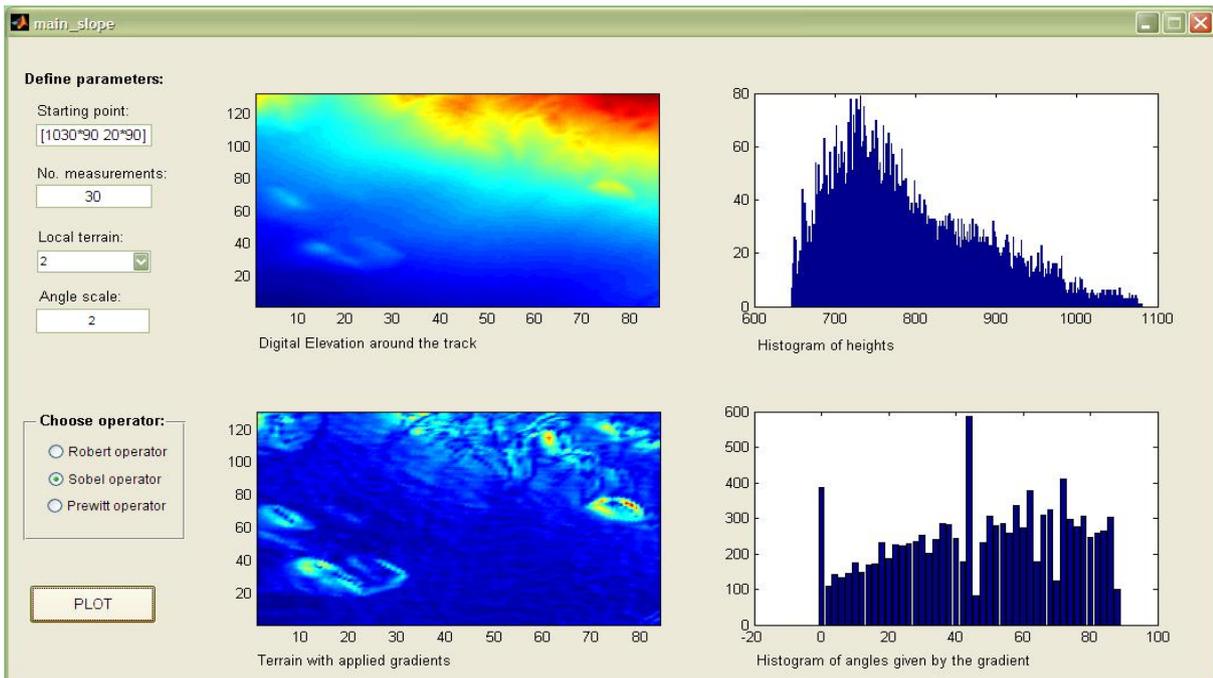
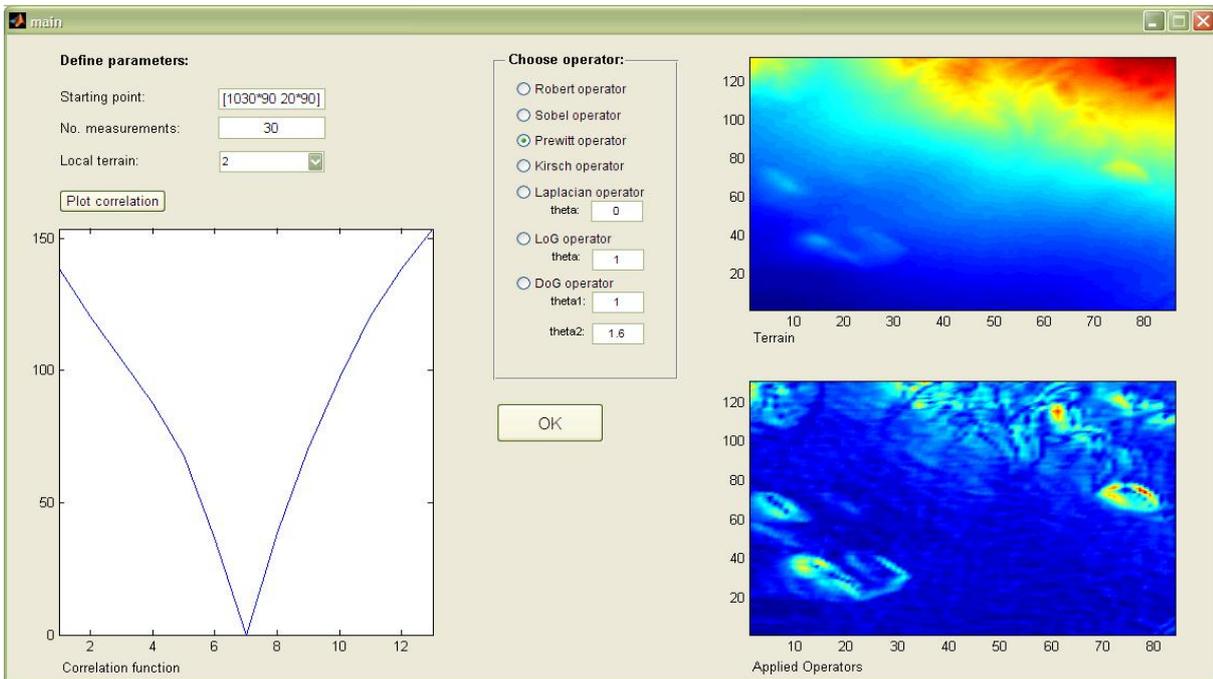


Figure 12: Screen shots of the newly derived simulation tool for studying terrain parameters and the relation to correlation results in more detail. In the upper picture: Transformation of the terrain using the Prewitt operator with corresponding correlation result. In the lower picture: Transformation of terrain using the Sobel operator with an analysis of heights and gradient angles.

Currently, we study the proposed algorithm in a more rigorous way by taking terrain parameters more into account. A new simulation tool has been built, see Figure 12 in order to analyze the relation between terrain conditions (depth, roughness) and the algorithm's performance. Primarily, a classification of the terrains that form the database of terrains we use is elaborated on. In a latter stage, the parameters found from the terrain description will be used to quantify the accuracy and availability that can be achieved by the proposed TRN system. Results from this work will be presented in an upcoming paper [8].

References

- [1] D. Akos, K. Borre, N. Bertelsen, P.Rinder and S. Jensen, *A single-frequency software-defined GNSS receiver*, Birkhäuser, 2007.
- [2] P. Misra and P. Enge, *Global Positioning System - Signals, Measurements, and Performance*, 2nd edition. Lincoln, (Massachusetts): Ganga-Jamuna Press, 2006.
- [3] D.J.R. Van Nee and A.J.R.M. Coenen, "New Fast Code-Acquisition Technique Using FFT," *Electronic letters*, vol. 27, pp. 158–160, 1991.
- [4] R. Gold, "Optimal binary sequences for spread spectrum multiplex", *IEEE transactions on Information Theory*, 619-621, 1967.
- [5] Application note: on bicubic interpolation topic.
<http://www.mathworks.com/access/helpdesk/help/toolbox/vipblks/ug/f12622.html>
- [6] J.L. Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *The American Statistician*, vol. 42(1), pp. 59–66, 1988.
- [7] D.Vaman and P.J. Oonincx, "Exploring a GPS Inspired Acquisition and Tracking Concept for Terrain Referenced Navigation," *Proc. Technical Meeting ION*, San Diego, 2010.
- [8] D. Vaman and P.J. Oonincx, "Performance Analysis of a GPS Inspired Terrain Referenced Navigation Algorithm," *Proc. European Navigation Conference GNSS 2010*, Braunschweig, 2010 (to appear).

Base Line Performance of the Low Cost LORADD SP Receiver

Cees de Groot & Anton Scheele

Introduction

Global Navigation Satellite Systems (GNSS) currently dominate the Positioning, Navigation and Timing (PNT) community and there are no signs that this will change in the near future. But for many years the weaknesses of GNSS have been well known and besides research into GNSS a lot of research into back up systems like Inertial Navigation Systems (INS) and eLoran has been done.

For military operations high availability of PNT data is of utmost importance. We estimate that the probability that GNSS is not available for a long period of time and in conjunction with that the risk that a military operation fails due to a lack of continuity of service is very small. In view of this an expensive back up system is not justified in our opinion. Therefore we began research into a low cost eLoran receiver. To this end we purchased the LORADD SP eLoran receiver made by the Dutch firm Reelektronika. The main data produced by this receiver are position, time and heading.

Besides low cost a back up system should at least meet the requirements for positioning and heading as stated in Tables 1 and 2. The requirements have been taken from several civilian documents since the Netherlands Armed Forces currently do not have an official policy on PNT¹.

¹The most recent version of a PNT policy for the Royal Netherlands Navy dates back to the mid nineties of the twentieth century. Whether the Royal Dutch Army and Airforce have a official PNT policy is not known.

Table 1: Position accuracy requirements according to different sources.

Source	Accuracy (95%) [m]
FRP2008 Table 4-3	8-20
ERNP Table 69	10
IMO Resolution A953(23)	10

Table 2: Heading accuracy requirements according to IMO Resolution MSC116(73).

Error type	Max. error [°]	Rate of Turn [°/s]
Static	1.0	0
Dynamic	1.5	0
Follow up	0.5	0-10

In this article the positioning and heading performance of the LORADD SP receiver will be investigated and tested against the requirements. Because it is a preliminary research into eLoran at the NLDA the research is also meant to give ‘hands on’ experience in this field and to generate research topics.

Loran-C and eLoran

Loran-C is a positioning system operating at 100 kHz that measures time differences (TDs’) between carefully shaped pulses that are transmitted by synchronised transmitters; these TDs’ can be transformed into the position of the receiver antenna. Enhanced Loran (eLoran) is an improvement of Loran-C that measures the time of arrival (TOA) of a pulse and using the time of departure (TOD) of this pulse calculates the travel time τ the pulse needed to propagate from the transmitter to the receiver antenna. τ can be transformed into the distance d_{TR} between transmitter and receiver antenna by multiplying with the speed of light c :

$$\tau = TOA - TOD \Rightarrow d_{TR} = c \cdot \tau$$

We can write d_{TR} , which is the distance along the geodetic line, as a function of the known latitude and longitude of the transmitter, the unknown latitude and longitude of the receiver antenna and the unknown clock error of the receiver yielding a relation -the observation equation- with one measurement (TOA) and three unknowns which can be solved when measurements of at least three transmitters are available.

When used in combination with an H-field antenna the receiver is capa-

ble of measuring the direction from the receiver to the transmitter relative to the reference direction of the antenna. Based on the known position the true bearing towards the transmitter can be calculated. At any epoch t the relation between heading $H_{loran}(t)$, calculated bearing $B_{calc}(t)$ and measured relative bearing $RB_{obs}(t)$ reads (see Figure 1):

$$H_{loran}(t) = B_{calc}(t) - RB_{obs}(t) \quad (1)$$

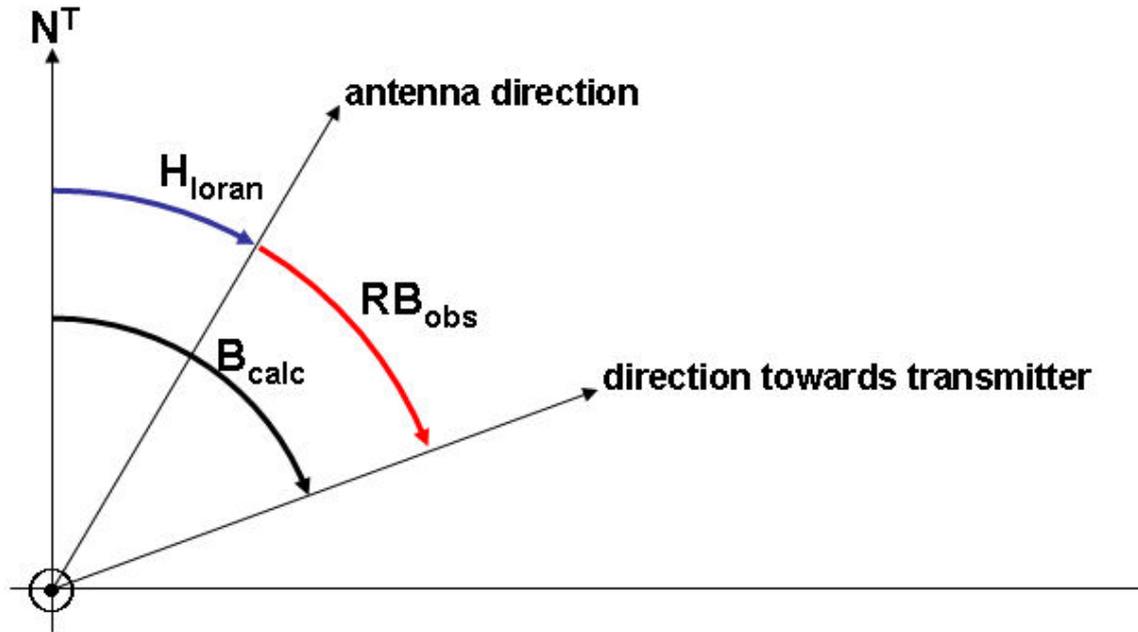


Figure 1: Relation between heading, observed relative bearing RB_{obs} and bearing B_{calc} relative to true north N^T .

According to Table B-3 of [5] the positioning accuracy of Loran is 460 m (95%); the heading accuracy is not stated. The accuracy of eLoran is not (yet) stated. For a comprehensive description of (e)Loran see [1].

The LORADD SP receiver

For our research we used the LORADD SP Integrated GPS/eLoran receiver of the Dutch firm Reelektronika. It can be used either with an E-field antenna or with an H-field antenna. It is a so-called all in view receiver which means that it tracks the signals of all transmitters it can receive.

Although it can track many signals only the ones with a certain (unknown to us) signal-to-noise ratio are used for the calculation of the position. For positioning at Den Helder Roads the receiver only used the signals of transmitters in Sylt (Germany), Lessay (France) and Anthorn (England). The receiver integrates measurements over a period of 5 seconds before processing yielding one eLoran position every 5 seconds.

Besides eLoran measurements, the receiver is capable of measuring GPS ranges. Using Eurofix range and range rate corrections these measurements produce DGPS positions. (D)GPS and eLoran can be used stand alone or integrated. When both (D)GPS and eLoran are available (D)GPS can be used to calibrate eLoran giving an improved performance of eLoran when (D)GPS is temporarily unavailable due to for instance interference or jamming.

When used in combination with an H-field antenna the receiver can produce heading. It is not clear whether the receiver uses the strongest signal for the heading, or the signal from the nearest transmitter or depending on their signal-to-noise ratio a combination of measurements from all transmitters in view. Heading measurements are integrated over one second intervals before processing.

For a comprehensive description of the LORADD SP receiver see [2].

Preliminary remarks and assumptions

Due to the fact that it is unknown which and how many measurements are used for the heading, the performance of the heading output might change when the relative geometry of the transmitters change i.e. when the measurements are made in another area. Therefore all measurements have been made at Den Helder Roads and vicinity only (see Figure 2). As a consequence the results will only be valid for this area.

The heading output of the receiver is calculated using Formula 1 for which the true bearing $B_{calc}(t)$ to the transmitter has to be calculated. The distance to the nearest transmitter in combination with the expected accuracy of Loran (460 m) are such that an error in $B_{calc}(t)$ due to a random shift in position may be ignored. The measured headings have only been corrected for relative bearing dependent errors which are inherent to the

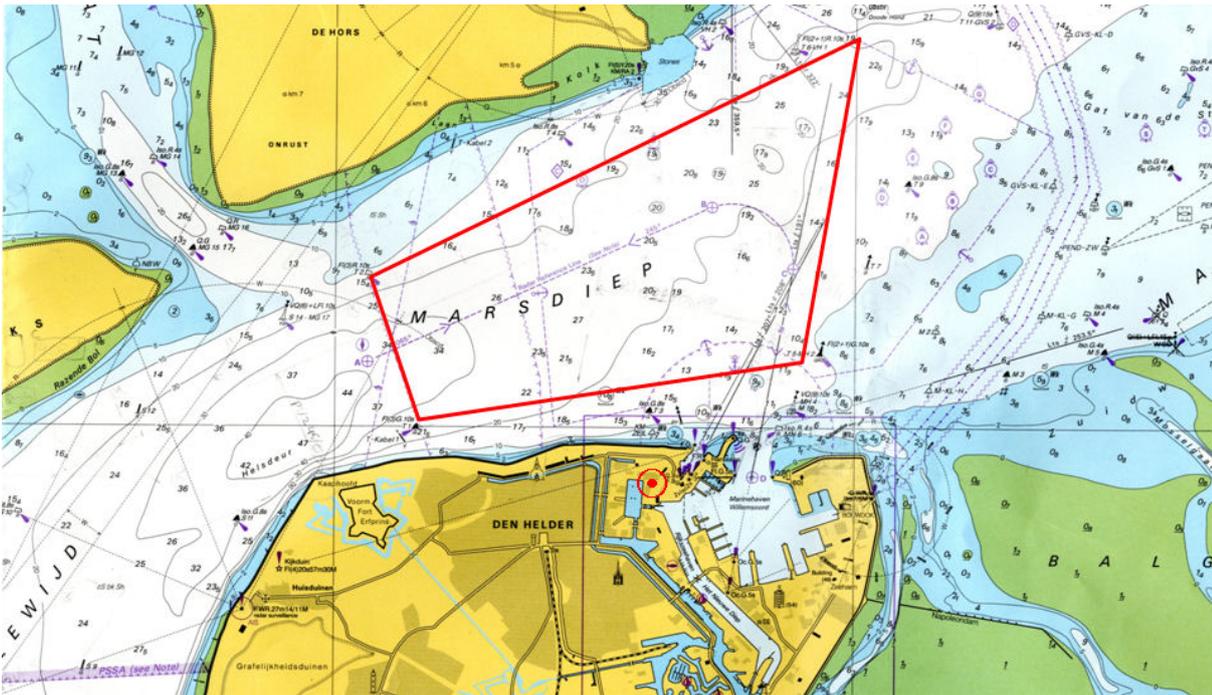


Figure 2: Copy of a portion of Netherlands Nautical Chart 1546, ‘North Sea Netherlands, Zeegat van Texel and Den Helder Roads’. The antenna position for the static measurements is marked by the red circle. The red lines indicate the track sailed for the dynamic measurements.

H-field antenna.

To assess the accuracy of the position of Loran we started the research using measurements without any corrections except the corrections known as Primary Factor (PF) and Secondary Factor (SF). This yielded a base line accuracy to be used as a reference to assess the effect of future improvements (see the next article).

All static measurements have been done on the roof of the Klooster, one of the buildings of the Netherlands Defence Academy (NLDA) in Den Helder. All dynamic measurements have been done on a small harbour tug (see Figures 3 and 4 respectively). The fieldwork has been done by Jelle Woudstra, Hessel de Groot and John den Ouden, students at the NLDA.

Position measurements

Static positions

To assess the accuracy of the position in the static case we measured DGPS ranges over a period of 48 hours and calculated the position of the com-

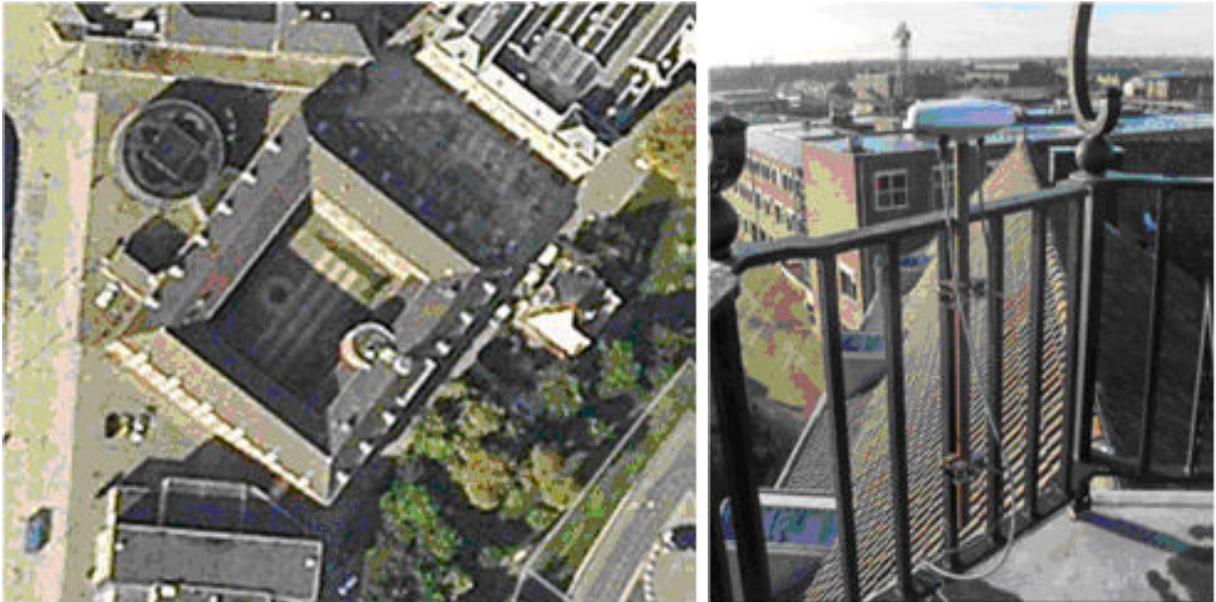


Figure 3: Left: Aerial view of the Klooster, one of the buildings of the NLDA at Den Helder. Right: Close up of the antenna location on top of the Klooster.

bined GPS/eLoran antenna on the roof of the Klooster. This position ($52^{\circ}57'43.248''\text{N}$, $004^{\circ}46'26.668''\text{E}$) has been used as a reference position throughout.

To gain a first impression of the accuracy eLoran signals of Sylt, Lessay and Anthorn were observed and logged at 5 second intervals from 13 to 20 february 2009. In Figure 5 the positions are plotted relative to the reference position. The mean eLoran position is in (205.8, -159.7) which is 260 m in the direction 128° relative to the reference position. The covariance matrix



Figure 4: Harbour tug Wierbalg used for dynamic eLoran measurements.

of the observed positions in $[m^2]$ reads:

$$\begin{pmatrix} 20.4 & -18.3 \\ -18.3 & 24.3 \end{pmatrix}$$

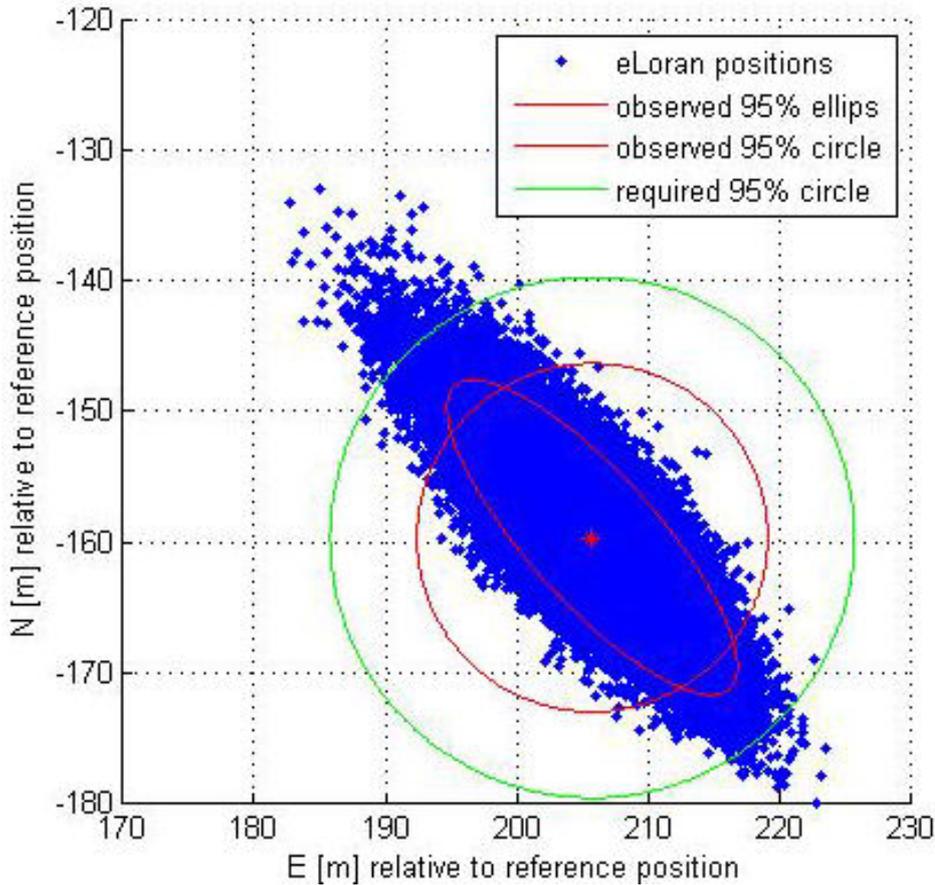


Figure 5: Scatter plot of eLoran positions observed at the Klooster between 13 and 20 february 2009.

Although there is a large bias, the spread of the eLoran positions around the mean value is rather limited: the 95% ellipse has a semi major axis in the direction 138° - 318° with a length of 15.6 m; the semi minor axis has a length of 4.9 m. The radius of the 95% circle equals 13.4 m. This indicates, as expected from literature, that the absolute accuracy of Loran is poor while the precision is good.

The direction of the semi major axis nearly coincides with the direction towards Anthorn (295°). The antenna position is very close to the baseline between Sylt and Lessay. Based on that we may conclude that

the orientation of the ellips and the bias are mainly caused by the fact that there is only one transmitter (Anthorn) producing one position line in the NE-SW direction while there are four position lines in the NW-SE direction because Sylt and Lessay are both dual rated.

As can be seen in Figure 5 the center of the ellipse which is the mean position over the observation period is not in the center of the scatter plot, but slightly to the south east. This might indicate that the mean position changes over time. Therefore we calculated the mean positions and mean R95 over periods of one hour. This indicated a change in position error between 253 and 270 m and a change in R95 between 6 and 16 m (see Figure 6. Note however that the position error minus 250 m is plotted.). Since there seems to be a relation between these changes and day and night we also calculated the mean position error and R95 during the day (07.00 -17.00 UTC) and night (17.00 - 07.00 UTC). In Figure 7 these values are indicated as asterisks.

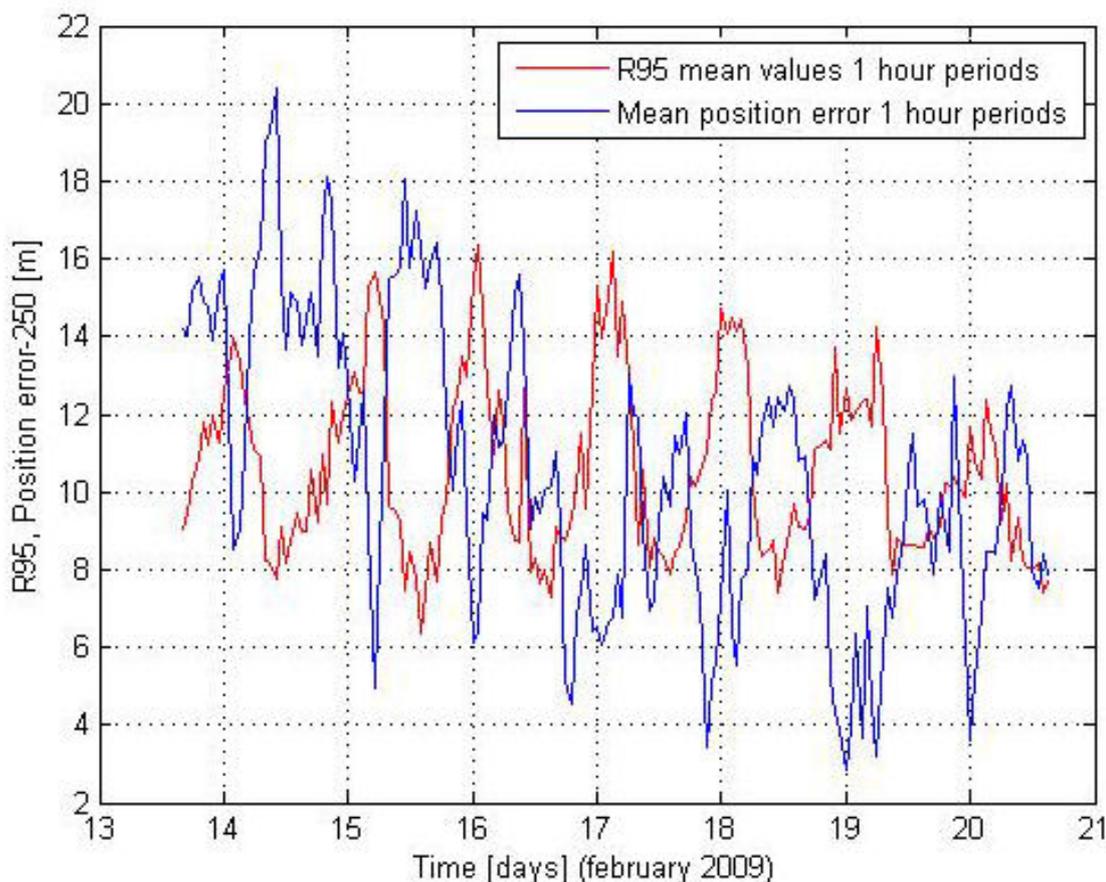


Figure 6: Position errors and values for R95 averaged over one hour periods.

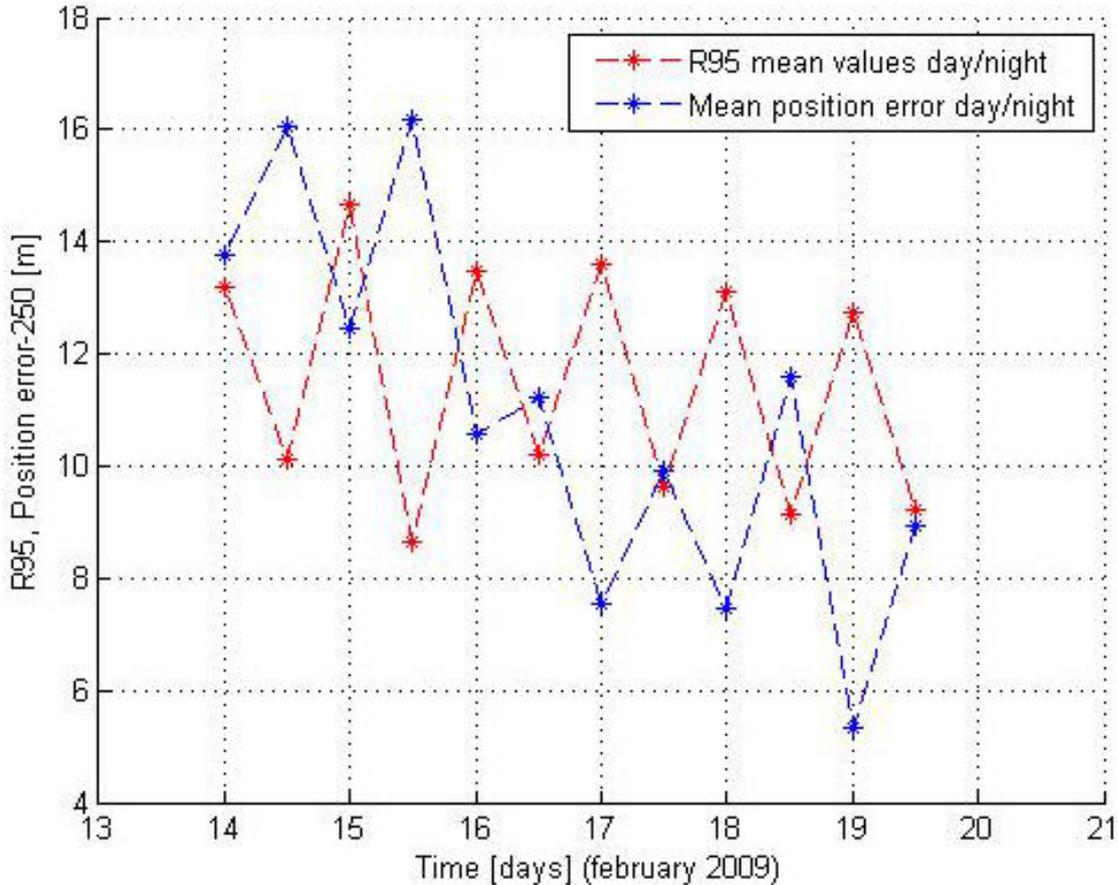


Figure 7: Position errors and values for R95 averaged during the day and night.

At Den Helder Roads the base line accuracy of eLoran positions in the static case is well within the 460 m as stated in Table B-3 of [5] but does not fulfill the requirements of Table 1. However, since the radius of the observed 95% circle is smaller than the radius of the required 95% circle (in its most relaxed version) the precision does meet the requirements.

From literature it is known that H-field antennae produce relative bearing dependent errors in position [1,3]. In order to assess the influence of the antenna direction on the bias and 95%-ellipse we measured positions in 8 different directions during 15 minutes (180 positions) in each direction. These measurements revealed that the distance between the reference position and the eLoran position varied in the direction of the semi major axis with plus or minus 40 m relative to a mean eLoran position as a function of the antenna direction (see Figure 8). An influence of antenna direction on the precision (size and orientation of the ellipse) was not observed.

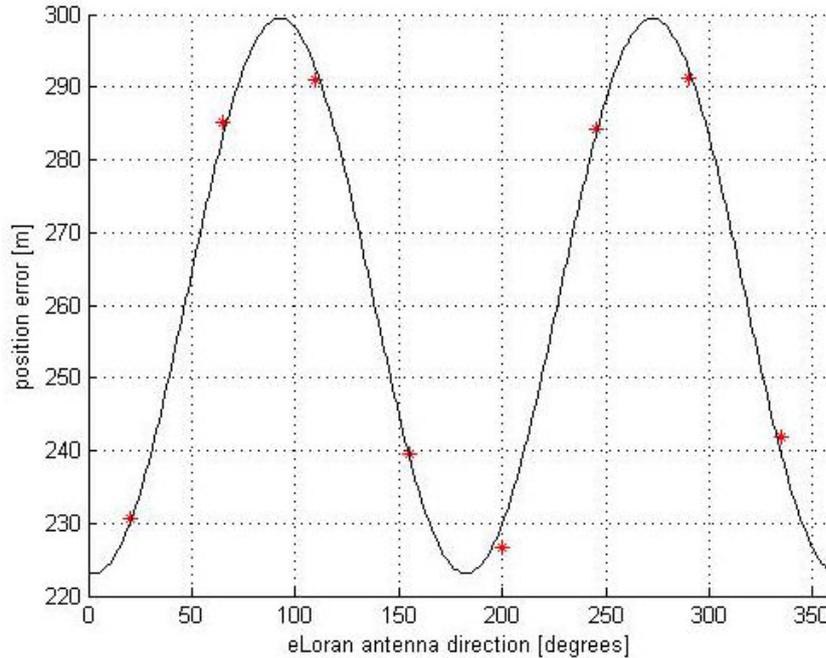


Figure 8: Position error as a function of antenna direction.

Dynamic positions

To get an impression of the degradation of the eLoran position in dynamic circumstances we compared calibrated eLoran positions with positions of an independent Trimble agGPS332 DGPS receiver, which served as the reference (ground truth). The positions were logged sailing the same circuit in Den Helder Roads twice.

Figure 9 gives the scatter plot of these positions. Since eLoran was continually calibrated with the LORADD DGPS we expected the accuracy to be in the order of the accuracy of DGPS (5 m (95%)). But although the 260 m bias has been removed by the receiver the precision has decreased in comparison to the static case. The radius of the 95% circle has increased to 28 m. When we look at the scatter plot of Figure 9 the increase of the R95 is probably caused by the relative bearing dependent error because several clusters of positions are apparent.

The accuracy of eLoran positions that are continually calibrated does not meet the requirements of Table 1.

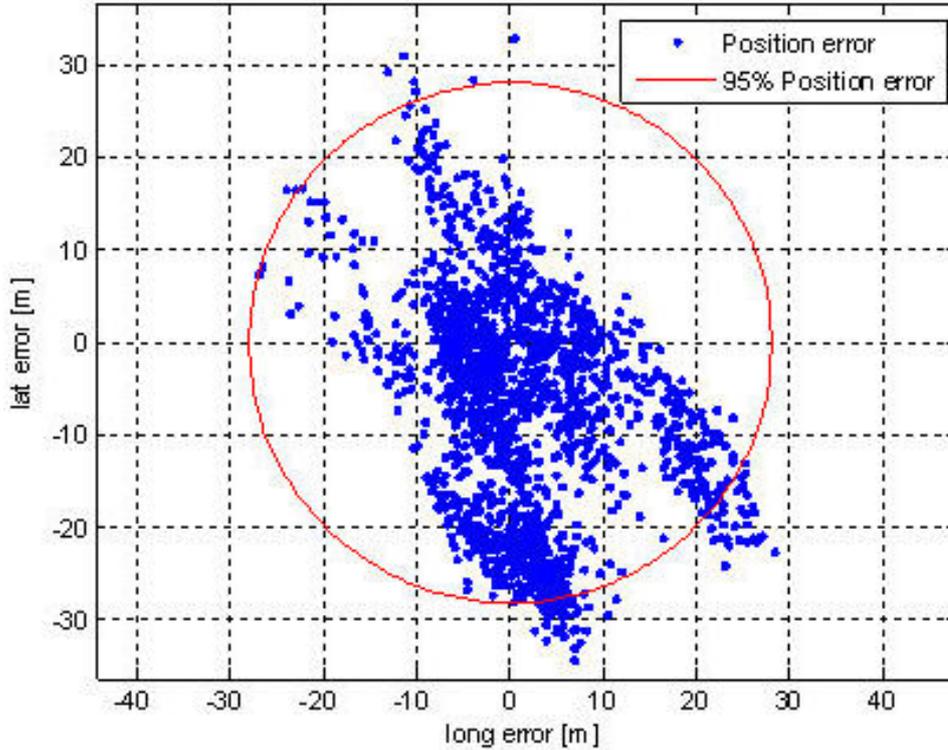


Figure 9: Observed errors between DGPS and calibrated eLoran sailing a circuit in Den Helder roads. Due to the relative bearing dependent error position errors are in clusters. Note the increase of the R95 compared to Figure 5.

Heading measurements

As well as the range measurements, the heading measurements also suffer from relative bearing dependent errors. To correct for these errors we constructed a correction function which gives the correction to be applied to the observed eLoran heading as a function of the relative bearing towards Sylt. Since the main source of these errors (re-radiation by objects in the vicinity and/or by the own vessel) was different on the roof of the Klooster and on board of the tug we constructed a function for the static measurements and one for the dynamic measurements.

For the construction of the correction function we compared $H_{loran}(t)$ with $H_{phins}(t)$, the heading yielded by PHINS for headings between 000° and 360° with steps of 30° . The correction $\Delta H(t)$ in $[\circ]$ at epoch t as a function of the heading reads (see Figure 10):

$$\Delta H(t) = H_{phins}(t) - H_{loran}(t) \quad (2)$$

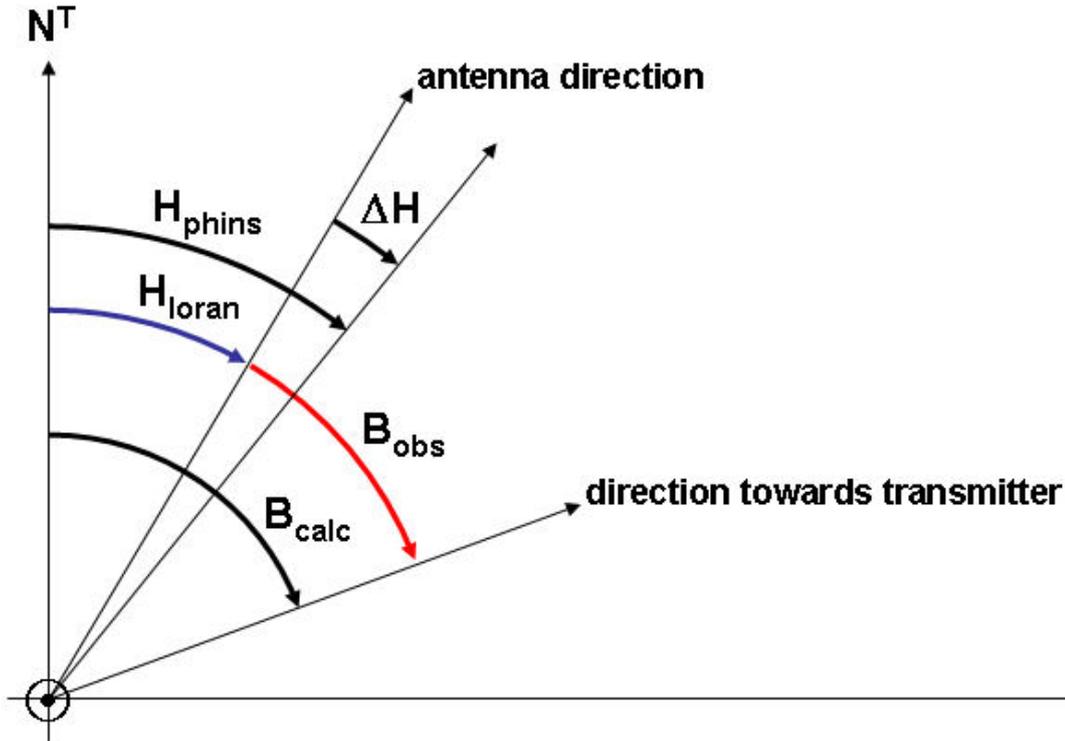


Figure 10: Relation between heading correction ΔH , observed eLoran heading H_{loran} and observed reference heading H_{phins} . PHINS is an accurate inertial navigation system that was used throughout the research to produce an accurate reference heading.

$\Delta H(t)$ consists of a systematic part which is mainly due to a difference in alignment between the eLoran antenna and the PHINS and a relative bearing dependent part.

In order to construct a function that is independent of time and in order to minimize the influence of random errors the measurements have been repeated several times. From literature [1,4] it is known that the correction function is a cosine with two periods in 360° and may well be written as a Fourier series.

Static headings

Using a least squares adjustment the next correction function has been derived for the static case (see Figure 11):

$$\begin{aligned} \Delta H(B) = & 1.63 + 0.53 \sin(B) + 0.24 \cos(B) \\ & + 0.77 \sin(2B) + 1.67 \cos(2B) \\ & - 0.01 \sin(3B) - 0.15 \cos(3B) \end{aligned}$$

In this function B is the relative bearing towards Sylt, which is the transmitter that is nearest and normally has the highest signal-to-noise ratio in the area Den Helder Roads. Terms of an order higher than three turned out to be negligible. The adjustment was based on 34 sets of measurements in four days in 12 directions. Each set consisted of about 600 measurements of $\Delta H(t)$.

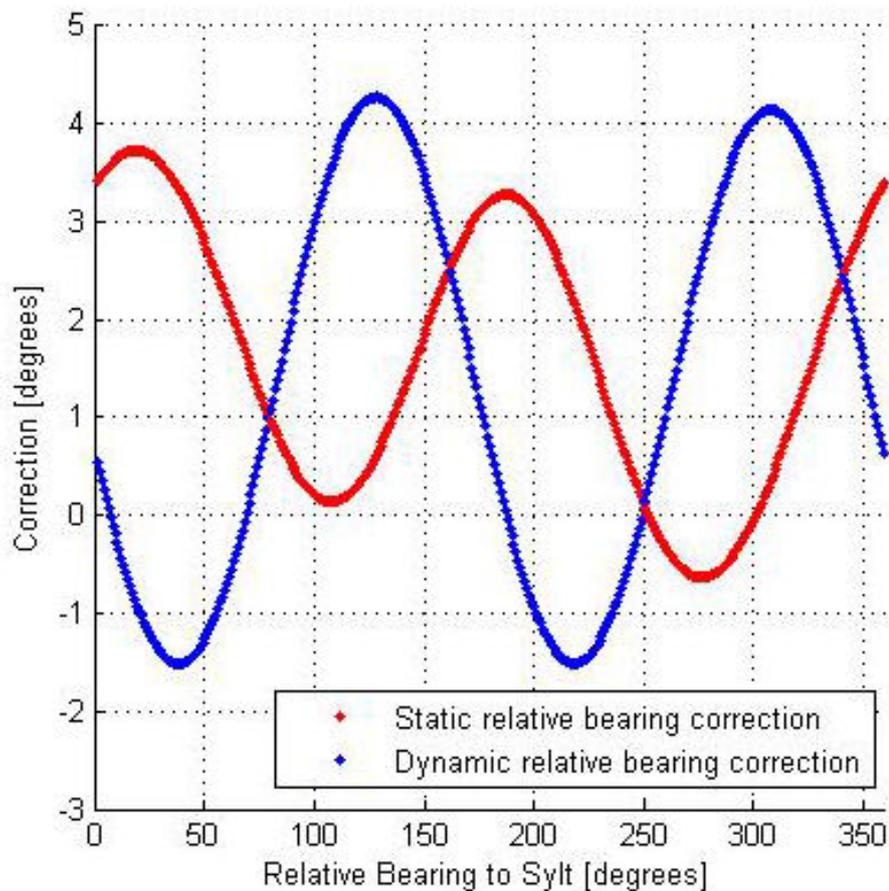


Figure 11: Correction to be applied to eLoran heading readings as a function of the relative bearing towards Sylt.

The mean value and standard deviation of each dataset was calculated after correction. This yielded mean heading corrections between -0.68° and $+0.59^\circ$ and standard deviations between 0.16° and 0.39° (see Figure 12). The overall mean value turned out to be 0.01° with a standard deviation of 0.46° [9].

In IMO Resolution MSC116(73) the static error is defined as the “error

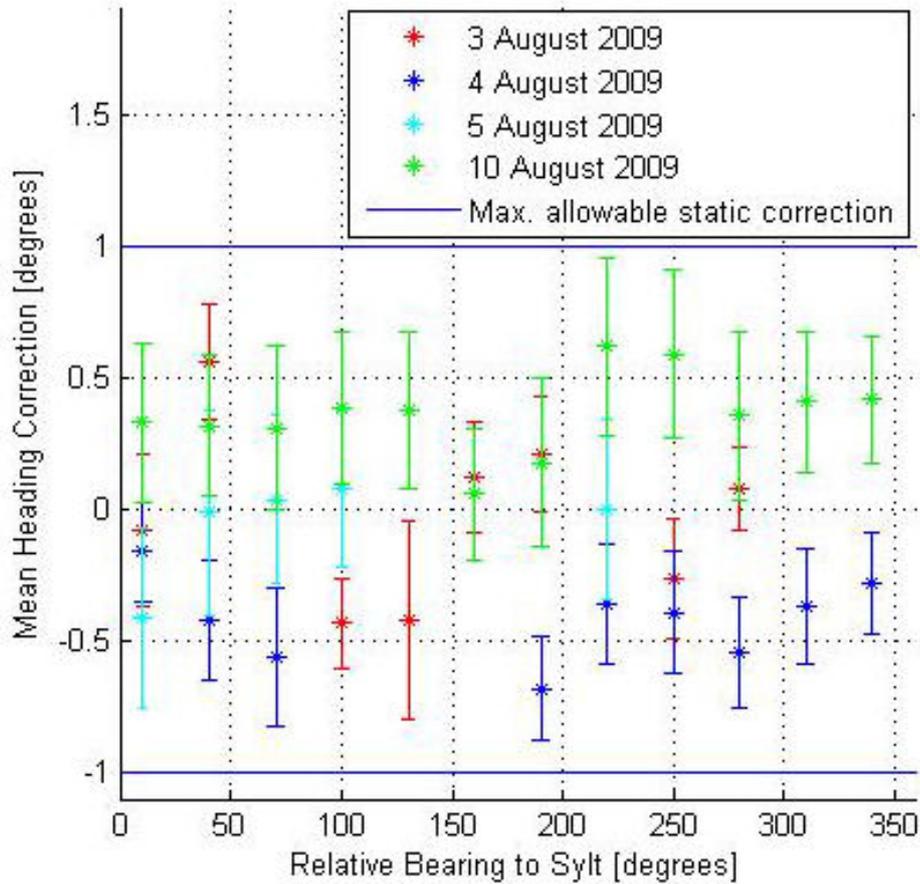


Figure 12: Mean heading corrections and standard deviations as a function of the relative bearing to Sylt.

which is caused by any reason and which stays unchanged in value during the operation of the system”. According to the requirements the maximum static error is plus or minus one degree. When we assume that a static error is the same as a systematic error, which is a plausible assumption in view of the definition, the LORADD eLoran compass fulfills this requirement the mean value of all individual sets and the overall mean value being smaller than one degree.

However because the definition is not quite clear the one degree maximum requirement could also mean that the probability that the static error exceeds one degree is zero². Assuming a normal distribution we then could argue that the one degree is equal to three times the standard deviation. In that case the compass might not fulfill the requirement since the standard

²Of course a probability will never be zero, but for a normally distributed stochastic variable it is known that the probability that the variable exceeds a value three times its standard deviation is only 0.3%. For practical reasons this probability can therefore be taken as zero.

deviation of a number of sets and the overall standard deviation exceed 0.33° .

Dynamic headings on a steady course

To derive the coefficients for the correction function to be used on the tug a circuit on the Marsdiep has been sailed three times in two days yielding some 20,000 measurements in all directions although not evenly distributed. The correction function for the dynamic measurements reads (see Figure 11):

$$\begin{aligned}\Delta H(B) = & 1.35 + 0.05 \sin(B) - 0.04 \cos(B) \\ & - 2.78 \sin(2B) - 0.68 \cos(2B)\end{aligned}$$

Again B is the relative bearing towards Sylt. In this case terms of order three and higher could be neglected. Since the re-radiation is only due to the tug itself the function resembles more the theoretical one with dominant terms of order two.

The mean value and standard deviation of the three corrected sets varied between -0.02° and 0.13° and 0.29° and 0.40° respectively.

IMO Resolution MSC116(73) defines the dynamic error as an “error which is caused by dynamic influences such as vibration, roll, pitch or linear acceleration.”.

For the requirement of dynamic errors the resolution states: “The dynamic error amplitude should be less than ± 1.5 degree. The dynamic error frequency should be less than 0.033 Hz equivalent to a period not shorter than 30 s if the amplitude of the dynamic error exceeds ± 0.5 degrees.”. The resolution does not state what the frequency is allowed to be in case the amplitude of the dynamic error is smaller than 0.5° . Since amplitudes are normally always positive the “ \pm ” might indicate that ‘about’ 0.5° is meant.

Also in this case it is not clear whether the 1.5° requirement is the systematic part of the dynamic error (on which a time varying or random part is superimposed) or a (nearly) three sigma value. No matter which one is meant, in both cases the compass fulfills the requirement.

Dynamic headings on a changing course

According to IMO Resolution MSC116(73) the follow up error is an “error which is caused by the delay between the existence of a value to be sensed and the availability of the corresponding signal or data stream at the output of the system. This error is e.g. the difference between the real heading of turning vessel and the available information at the output of the system.”.

Denoting the delay by k [s], the next relation between rate of turn $R(t)$ [$^{\circ}/s$] and follow up correction $\Delta H(t)$ is to be expected:

$$\Delta H(t) = kR(t)$$

In order to assess the follow up correction of the eLoran compass circle tests have been done with different rates of turn turning port as well as starboard. During each test the heading measured with PHINS at a certain epoch and the eLoran heading at that same time were measured and subtracted (see Formula 2). For each circle test the mean value of the follow up correction and its standard deviation were calculated. The results are plotted in Figure 13. As can be seen in Figure 13 the relation between the rate of turn and the follow up correction tends to be linear as expected. Using the calculated corrections the value for k was estimated to be 0.2. But it is very remarkable and unexpected that there is a bias of two degrees between the theoretical expected relation and the relation derived from the measurements. This bias is probably due to a misalignment of the antenna and/or PHINS.

Given the calculated corrections, the eLoran compass does not fulfill the requirements at all. However assuming the value for k is correct and the bias is due to errors in the measurements or in the measurement set up the compass might fulfill the requirements for rates of turn between zero and 2.5 degrees per second. For higher rates of turn the compass does not meet the requirements.

Discussion

The base line accuracy of eLoran does not meet the requirements. The scatterplots of the static as well as the dynamic positions show ellipses pointing towards Anthorn. This indicates that the ellipticity is mainly caused by random errors in the measurements of this transmitter. The

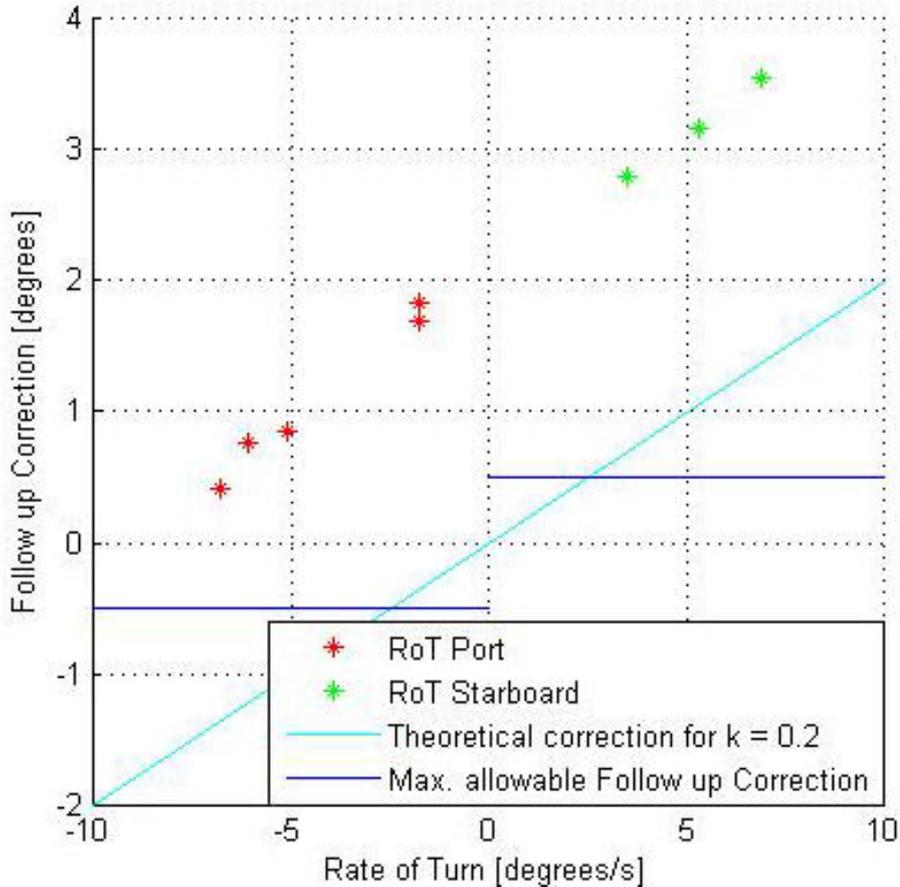


Figure 13: Follow up correction to be applied to the eLoran heading as a function of the rate of turn.

size of the ellipses of the dynamic positions was larger than that of the static positions. The reason for this is not clear from the measurements so far.

In the static case the mean positions show a bias that is directed towards/from Anthorn and changes along that direction as a function of antenna direction and as a function of the time of the day. Due to the fact that in the area Den Helder Roads the transmitters at Sylt and Lessay are nearly opposite each other all common errors in the signals of these transmitters will cancel yielding a better position accuracy in the direction Sylt-Lessay. Also in the dynamic case the mean positions show a bias, but the direction of the position shift is not towards/from Anthorn. When these biases turn out to be reasonably constant over a longer period of time we might transform them into (relative bearing and position dependent) corrections to remove them and improve the accuracy. Therefore a (plat-

form dependent) correction function for positions should be constructed analogous to the correction function for the headings.

Depending on the interpretation of the definitions in [6] the LORADD eLoran compass does or does not meet the requirements in the static and the dynamic case. According to the LORADD SP fact sheet the compass performs “better than one degree under normal conditions”. Although no definition is given of the term normal conditions the measurement results tend to confirm this statement. One should keep in mind that these results only hold when no re-radiation is present.

The performance of the LORADD eLoran heading when the platform has a rate of turn cannot yet be assessed. The trend in the graph of Figure 13 is in accordance with the theoretically expected relation, but the values of the corrections are not.

It is questionable whether the compass will be able to meet the requirements for the follow up error at all. Suppose a rate of turn of ten degrees per second to starboard and suppose the PHINS and eLoran heading at the beginning of the one second integration interval both are 000° . At the end of the integration interval the PHINS heading will then be 010° , but the integrated eLoran heading will be 005° . This yields a follow up error of at least 5° (one order worse than the requirement of 0.5°). Therefore the compass will probably not meet the requirements unless a correction is applied somewhere in the processing of the measurements.

Although H-field antennas have many advantages over E-field antennas [1] the main disadvantage is the fact that relative bearing dependent errors show up in the position as well as in the heading measurements. Therefore the use of an E-field antenna might be preferred especially in the case of maritime use of eLoran.

Conclusions

In this article the positioning and heading performance of the LORADD SP receiver have been investigated and tested against the requirements. The positioning and heading accuracy are summarized in Table 3.

Tested against the requirements the main conclusion is that as far as positions are concerned eLoran in its baseline configuration cannot serve as

Table 3: Summary of position and heading accuracy.

	Mean value	Accuracy (95%)
Position	260 m	12.4 m
Heading Static	0°	0.9°
Dynamic	0°	0.8°
Follow up	as yet unknown	as yet unknown

a back up for GNSS. However the measurement results are very promising in that way that the performance might be improved by removing the bias in the position and by applying a relative bearing dependent position correction.

The LORADD receiver can serve as a back up for a compass when no re-radiation is present.

Besides testing the performance of the receiver another research goal was to gain ‘hands on’ experience in the field of eLoran and to generate research topics. Although the LORADD receiver does not meet the requirements the knowledge of the possibilities of the receiver has increased. Besides this the results give rise to a number of research topics to improve performance.

References

- [1] W.J. Pelgrum, “New Potential of Low-Frequency Radionavigation in the 21st Century,” Ph.D. dissertation, Delft University of Technology, Delft, 2006.
- [2] Reelektronika, *LORADD series Integrated GPS/eLoran receiver; Installation and operational manual Version 1.2*. 16 january 2009.
- [3] G. Johnson, K. Dykstra, R. Shalaev, P. Swaszek and R. Hartnett, “Airframe Effects on loran H-field Antenna Performance,” presented at Institute of Navigation Annual Meeting, Cambridge, MA, 27 - 29 June 2005.
- [4] G.J. Sonnenberg, *Radar and Electronic Navigation*, 6th edition. London: Butterworth & Co., 1988.
- [5] United States. US Department of Defense, Department of Homeland Security and Department of Transportation. *2008 Federal Radionavigation Plan*. Springfield, VA, 2008.

- [6] International Maritime Organization, *Resolution MSC116(73); Performance Standards for Marine Transmitting Heading Devices*. London, 2000.
- [7] J.M. Woudstra, “De grootte van Loran Additional Secondary Factors in Den Helder,” B.Sc. thesis, NLDA, Den Helder, 2009.
- [8] H. de Groot, “Performance Analysis of an Integrated GPS-eLoran Receiver,” M.Sc. thesis, NLDA, Den Helder, 2010.
- [9] J.W. den Ouden, “Heading Determination Using eLoran,” M.Sc. thesis, NLDA, Den Helder, 2010.

Performance Improvement of the Low Cost LORADD SP Receiver

Cees de Groot & Anton Scheele

Introduction

In the previous article we discussed the basic performance of the LORADD SP receiver. One of the conclusions was that “eLoran in its baseline configuration cannot serve as a back up for GNSS”. In this chapter we will explore the possibilities to improve the results. Therefore we will first argue why further research is necessary. Then we will analyse the observation equation of eLoran and concentrate on improvement on the level of observations. Finally we will outline improvements on the level of positions, for which no measurements have currently been made.

Rapid Environmental Assessment and positioning

A lot of research into improvement of eLoran has been done in the past few years. The results show that accuracies in the order of 10 m (95%) are possible. So why should we do research again? The reason for doing our own research is the fact that all research so far has been done with (much) more expensive receivers in dedicated areas with adequate infrastructure (for instance Harwich harbour). As argued in the previous article an expensive back up for GNSS is not justified although for military operations a high availability of PNT data is paramount. Therefore we do the research with a low cost receiver in an arbitrary area with no infrastructure which is the kind of area to be expected for military operations.

In view of Rapid Environmental Assessment (REA) [1] the military command might have a need to assess quickly the influence of the environment

(e.g. ASF) on positioning. The aim of our research is to develop a means to provide the military command with corrections which yield positions that are accurate enough to serve as a back up for GNSS for the duration of the military operation.

eLoran observation equation

The observed travel time τ' of a Loran pulse between a transmitter and the receiver antenna in terms of distance quantities may be written as:

$$d' = d + PF + SF + ASF + B + \Delta UTC + P + r$$

In this observation equation d' is the measured distance ($c \cdot \tau'$) and d the true (geodetic) distance between the transmitter and receiver antenna. PF and SF are the Primary and Secondary Factor respectively which are deterministic. PF is an error that occurs due to the fact that the receiver calculates distances using the speed of light in vacuum while the speed of light in the atmosphere should be used. SF is an error that occurs due to the fact that the signal travels over sea water with no perfect conductivity inducing a delay. PF and SF in meters are calculated using the next formulae¹ with d in meters:

$$\begin{aligned} PF(d) &= 3.38 \cdot 10^{-4} \cdot d \\ SF(d) &= -122.1654 + 6.4597 \cdot 10^{-4} \cdot d + 1.1594 * 10^7 / d \end{aligned} \quad (1)$$

ASF is the Additional Secondary Factor. It is an extra error (delay) that occurs when the path between transmitter and receiver antenna is not an all sea water path. ASF depends on the topography and conductivity of the path and can be calculated using a model or can be measured in carefully selected positions. B and ΔUTC are the clock errors of the receiver and the transmitter respectively. P is the processing delay in the receiver and r a random residual error. The term $B + \Delta UTC + P + r$ is assumed to consist of a systematic part Δd and a residual stochastic part e yielding the next relation for d' :

$$d' = d + PF + SF + ASF + \Delta d + e \quad (2)$$

The systematic part can be estimated from the measurements. The stochastic part can be canceled out by measuring over a sufficiently long period of time.

¹The original version of this formula d is in statute miles.

When we want to improve the performance of the receiver the best thing to do in first instance is to correct for the biggest error namely the ASF. An expression for ASF is found rewriting Formula 2:

$$ASF = d' - d_{true} - PF - SF - \Delta d - e$$

In this formula d_{true} is known to a high degree of accuracy, for instance using DGPS or a fixed position. For absolute ASFs to be estimated the accuracy of τ' ($= d'/c$) and Δd needs to be in the order of nanoseconds which requires a very accurate clock in the receiver. The LORADD SP receiver does not meet this accuracy. Therefore we will discuss alternative methods to estimate ASFs as accurately as possible in the next sections.

Estimating ASF's using Millington's method

The simplest way to get a (rough) estimate of the ASF is by using Millington's method. In this method the path between the transmitter and the receiver antenna is divided into portions with the same conductivity. For each portion the delay is estimated; the sum of delays gives the total delay.

For the propagation paths between the Klooster and the transmitters at Lessay, Anthorn and Sylt we calculated the land portion and the sea water portion. For Lessay and Anthorn the first 40 and 200 kilometers respectively are over land. We assumed the path between Sylt and the Klooster to be an all sea water path, although there is a small land path of about 5 kilometers on the most westerly part of the Frisian island Ameland. Furthermore we assumed the Waddenzee to have a conductivity of sea water. Using the land and sea water portions and the graphs in figures 13 and 14 of [2] we estimated values for SF+ASF for a ground conductivity (σ) of 0.001 and 0.01 mho/m. The ASFs follow by subtracting the SF calculated with Formula 1. The results are in Table 1.

According to table F-1 of [3] we estimate the true conductivity of the land portions to be in the order of 0.005 mho/m (pastoral land, medium hills and forestation).

Table 1: ASFs estimated using Millington's method.

Transmitter		Lessay	Anthorn	Sylt
Distance total	[km]	611	572	309
Distance land	[km]	40	200	0
$\sigma = 10^{-3}$	[mho/m]			
SF+ASF	[μ s]	1.67	2.67	0.74
SF	[μ s]	0.97	0.89	0.74
ASF	[μ s]	0.70	1.78	0.00
	[m]	210	534	0.00
$\sigma = 10^{-2}$	[mho/m]			
SF+ASF	[μ s]	1.17	1.58	0.74
SF	[μ s]	0.97	0.89	0.74
ASF	[μ s]	0.20	0.69	0.00
	[m]	60	207	0.00

Estimating ASFs from observed differential ASFs

The clock of the (low cost) LORADD receiver is not accurate enough to estimate Δd to a sufficient degree of accuracy. Therefore the receiver cannot calculate absolute ASF's. Instead the receiver calculates the so called dASF which is the difference between the ASF of a measurement and a reference ASF_{ref} . ASF_{ref} is the ASF of the measurement from the transmitter with the largest Signal-to-Noise ratio, mostly being the nearest transmitter. By differencing Δd cancels:

$$dASF = (ASF + \Delta d) - (ASF_{ref} + \Delta d)$$

To estimate ASFs using dASF's we measured dASFs of the three transmitters in chain 6731 and the dASFs of the transmitter at Ejde (chain 9007). The propagation path between Ejde and the Klooster is completely over sea water, so the ASF of Ejde should be zero. Given that Sylt is the reference this leads us to the next relation:

$$dASF_{Ejde} = ASF_{Ejde} - ASF_{Sylt} = -ASF_{Sylt}$$

Substituting this relation into the relation of the dASF of Lessay and Anthorn gives:

$$\begin{aligned} dASF_{Lessay} &= ASF_{Lessay} - ASF_{Sylt} \\ \Rightarrow ASF_{Lessay} &= dASF_{Lessay} - dASF_{Ejde} \end{aligned}$$

$$\begin{aligned} dASF_{Anthorn} &= ASF_{Anthorn} - ASF_{Sylt} \\ \Rightarrow ASF_{Anthorn} &= dASF_{Anthorn} - dASF_{Ejde} \end{aligned}$$

Figure 1 gives the results calculated from static measurements at the Klooster between 13 and 20 February 2009.

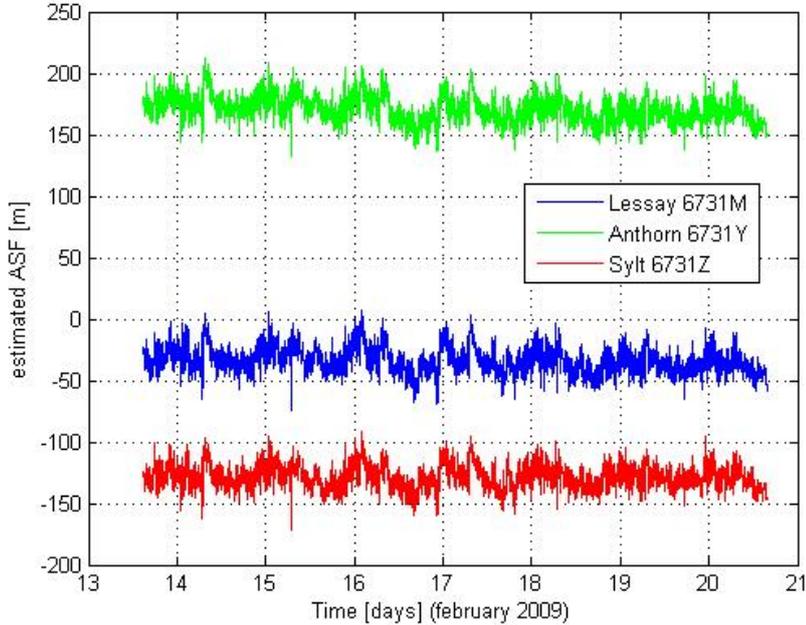


Figure 1: ASF's estimated using observed dASF's.

The mean values of the ASFs for Lessay, Anthorn and Sylt are mentioned in Table 2. The ASFs of Sylt and Lessay turn out to have a negative value. This is highly improbable. This might indicate that the ASF of Ejde does not equal zero although the path between Ejde and the Klooster is completely over sea water. Let us therefore assume again (as in the previous section) that the ASF of Sylt should be zero (or a small positive value). By adding 128 m to all ASF's we get the results of the second line of Table 2.

Table 2: Mean values of ASF's estimated using observed dASF's.

Transmitter		Lessay	Anthorn	Sylt
ASF ($ASF_{Ejde} = 0$)	[m]	-34	171	-128
ASF ($ASF_{Sylt} = 0$)	[m]	94	299	0

Estimating ASFs from eLoran positions and residuals

The LORADD receiver does not output values for Δd . In order to estimate ASFs we therefore first estimate values for $ASF + \Delta d$ according to the next

relation:

$$ASF + \Delta d = d' - d_{true} - PF - SF - e$$

Because the LORADD receiver does not output d' (or τ') we estimated d' from:

$$d' = d_{loran} + e$$

In this relation d_{loran} is the distance between the calculated eLoran position and the transmitter. Figure 2 depicts the relation between d' , d_{loran} , d_{true} , $ASF + \Delta d$ and e .

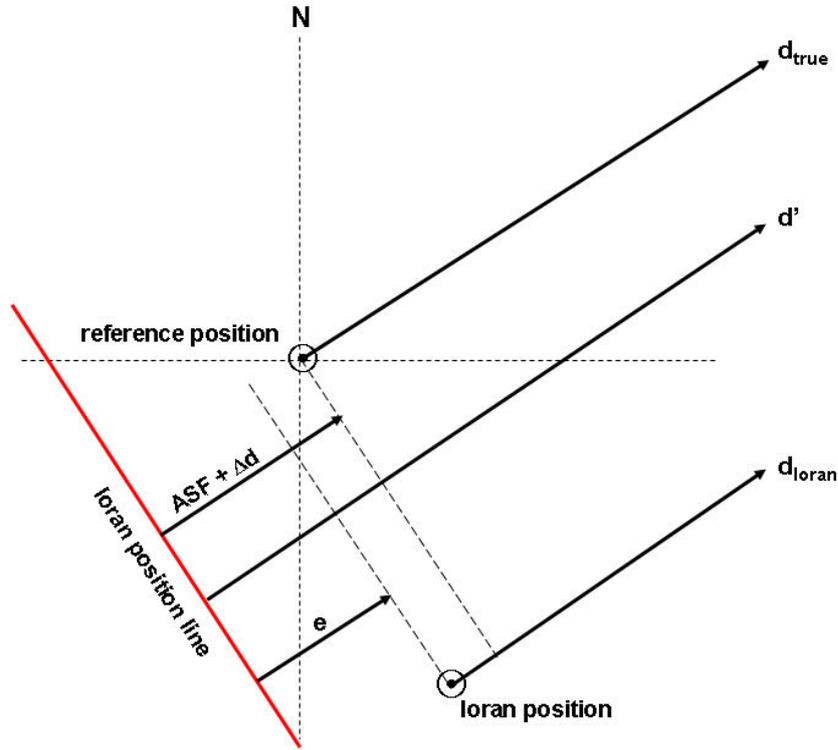


Figure 2: Relation between observed distance (d'), distance between transmitter and eLoran position (d_{loran}), distance between transmitter and true position (d_{true}), distance between observed eLoran line of position and true position ($ASF + \Delta d$) and residual e .

To calculate $ASF + \Delta d$ eLoran measurements were done from 13 February 2009 14:53:08 UTC until 20 February 2009 15:42:09 UTC with 1 minute intervals. Figure 3 shows the results graphically while Table 3 gives the mean values.

The next step is to try to eliminate Δd . Therefore we can follow two different independent ways:

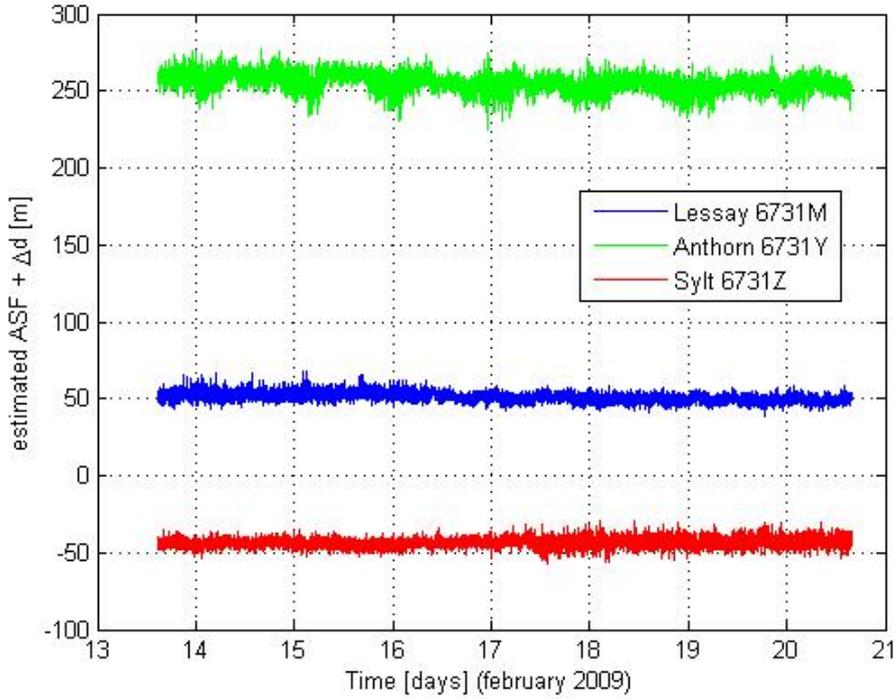


Figure 3: ASFs estimated from eLoran positions and residuals.

1. Based on the assumption that the ASF of Sylt should be zero we find a mean value of -43 m for Δd . The ASFs for Lessay and Anthorn can now be estimated to be 94 m and 298 m respectively (see Table 3).
2. The geometry of eLoran position lines in Den Helder Roads is such that position lines of Sylt and Lessay run parallel while the one position line of Anthorn is nearly perpendicular to them. This means that the distance between the reference position at the Klooster and the mean eLoran position may assumed to be equal to the ASF + Δd of Anthorn (see Figure 4). Using -43 as the value for Δd we conclude that the ASF for Anthorn equals 305 m ($= 261 + 43$) yielding an ASF of 6 m for Sylt and 100 m for Lessay.

Table 3: Mean values of ASF's estimated using eLoran positions and residuals.

Transmitter		Lessay	Anthorn	Sylt
ASF + Δd	[m]	51	255	-43
ASF (ASF _{Sylt} = 0 m)	[m]	94	298	0
ASF (ASF _{Anthorn} = 304 m)	[m]	100	304	6

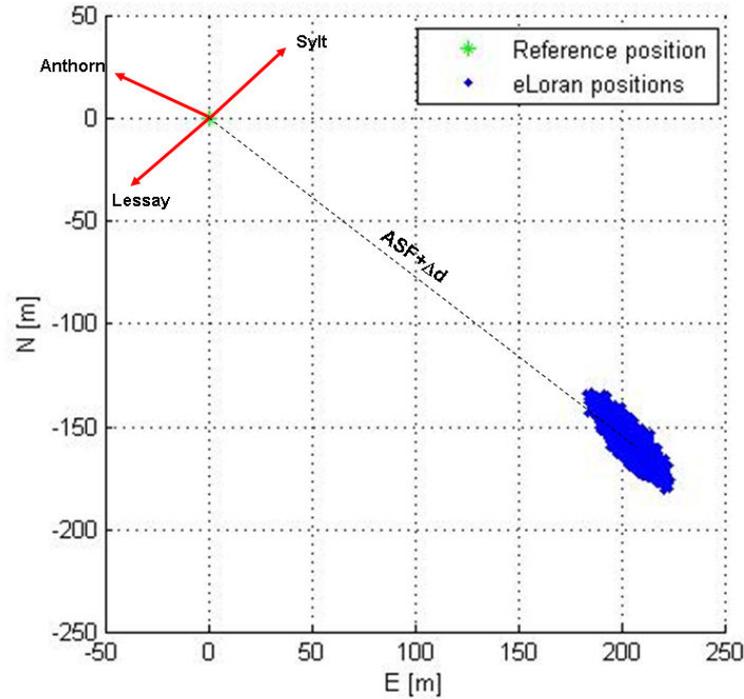


Figure 4: Geometry of eLoran position lines at the Klooster. The distance between the reference position and the mean eLoran position nearly equals $ASF + \Delta d$ of Anthorn (261 m).

Summary and discussion

The values for the ASFs of Lessay, Anthorn and Sylt for the position of the Klooster in Den Helder are summarized in Table 4. Outcomes with negative values for one or more ASFs have been omitted beforehand.

Table 4: Mean values of ASF's in m.

Method	Lessay	Anthorn	Sylt
Millington			
$\sigma = 10^{-3}$	210	534	0
$\sigma = 10^{-2}$	60	207	0
dASF's			
	94	299	0
eLoran positions and residuals			
$ASF_{Sylt} = 0$	94	298	0
$ASF_{Anthorn} = 304$	100	304	6

The resemblance between the ASFs calculated using dASFs and using eLoran positions and residuals is striking and deserves a closer look. When we calculate the correlation coefficient between both sets of ASFs we find

values of 0.44 and 0.60 for Lessay and Anthorn respectively. This indicates, neither method is completely independent, but the independency is such that both methods deserve supplementary research for validation.

The ASFs calculated using measurements are in the range of ASFs calculated using Millington's method.

Based on the results so far we assume that the ASF at the Klooster equals 94 m and 298 m for Lessay and Anthorn respectively.

In order to ascertain these values they should be entered into the LORADD receiver. Basically this is possible, but users themselves cannot enter an ASF map. From a point of view of the 'normal' user this is understandable to avoid errors which could affect the safety of navigation. From the point of view of research this is a drawback of the LORADD receiver. Also from a military point of view this is a drawback, because it reduces the flexibility and the independency of the military command.

Improving the performance on position level

Since the receiver does not offer the possibility for the user to enter ASF maps eLoran measurements cannot be corrected for ASF. Therefore for the time being we shifted the focus of the research from corrections on observation level to corrections on position level. The assumption is that the receiver always uses the same transmitters in a certain area. The mean bearing and distance between observed eLoran and DGPS positions can be used as a correction. No systematic field work has been done for this research so far.

Based on research mentioned in the previous article and based on small tests the next topics need our attention:

1. To avoid relative bearing dependent errors inherent to an H-field antenna measurements will be taken using an E-field antenna.
2. On a moving platform the LORADD receiver shows a difference (latency) between the eLoran position of a certain epoch and an (independent) DGPS position of that same epoch. Part of the research will concentrate on modelling this error.

The future of eLoran

On 28 October 2009 the President of the United States, Barack Obama, announced the termination of Loran-C and hence the termination of eLoran in the United States. According to [4] Obama stated:

“This system once made a lot of sense, before there were satellites to help us navigate. Now there’s GPS. And yet, year after year, this obsolete technology has continued to be funded even though it serves no government function and very few people are left who still actually use it.”

The US Coast Guard supports the president by stating [4]:

“In a submission to the Federal Register, the Coast Guard said Loran-C was not established as, nor was it intended to be, a backup for GPS. Other radio navigation systems, or operational procedures, can be used as backups for GPS navigation and other critical applications, the Coast Guard said.”

We will continue to do research into eLoran, because presently there is no adequate backup; since the European eLoran transmitters will be operational until 2022 we can continue the research. Over a period of 12 years there will likely be a GNSS incident serving as a wake up call to the PNT community.

References

- [1] J. Appelman, *One Theatre, One Forecast; Working Paper Maritiem Optreden nr. 2*. Den Helder: Maritiem Doctrine en Tactieken Centrum, 2009.
- [2] S.N. Samaddar, “The Theory of Loran-C Ground Wave Propagation - A Review,” *Navigation, Journal of the Institute of Navigation*, vol. 26, no. 3, Fall 1979.
- [3] United States, United States Coast Guard, “Appendix F; Millington’s Method,” *LORAN-C User’s Handbook*, COMDTPUB P16562.6
- [4] M.M. Ahlers, “World War II-era navigation system shutting down,” CNN.com, 8 february 2010.

Determining the Hydrodynamic Derivatives of a Basic Model of the REMUS AUV

Dick Engelbracht & Paul Wolkenfelt

Introduction

With the term manoeuvring we mean: to conduct a planned and controlled motion of the ship i.e. a horizontal translation and change of heading of the ship. In physical terms manoeuvring is a planned and controlled translation and rotation of the ship as a result of all forces and moments acting on the vessel. For a free moving body, three translations and three rotations are possible. They have to satisfy, and can be calculated from, Newton's second law of motion. A manoeuvring model is thus a system of equations of motion and the manoeuvring behaviour of a ship can be simulated by (numerically) solving this system of equations. Applicability of the model and its degree of correspondence with reality are important considerations in model building and is reflected in the number of equations and the number of terms in an equation. Any term or combination of terms represents a certain physical force or moment that affects the motion of the ship. Each term also contains a coefficient (or proportion constant). The essential step in producing a manoeuvring model is estimating these coefficients. In principle there are three options (i) model tests, (ii) empirical formulas based on model tests, (iii) full-scale sea trials.

Initially, simple models were used, such as that of Nomoto. In 1971 the Royal Netherlands Naval College (RNLNC) started the development of a night vision simulator. This simulator has been in operation since about 1974 and remained in use until the early nineties. The applied manoeuvring model describes three degrees of freedom and for one of them the Nomoto equation was used. A more elaborate model is that of Inoue [1]. Also

three degrees of freedom are described, the model contains approximately 23 terms including at least 23 coefficients to be determined. The terms used in the model are directed to their physical origins. This is certainly not the case with the Abkowitz model [2], a large number of coefficients have to be determined by the presence of a large number of terms resulting from a Taylor series expansion. Each term contributes to the forces or the moments acting on a ship but the physical meaning of an individual term is not always clear.

In the nineties the Delft University of Technology (TU Delft), faculty Maritime Technology (MT), department of hydromechanics, and the RNLNC, nautical sciences department, led by professor J.A Spaans, explored the possibility of determining the coefficients from full-scale sea trials [3]. As a starting point the model by Inoue was used. But, this model has some drawbacks if the values of the coefficients have to be calculated from full-scale sea trials [4].

Research question

The drawbacks have led to two research questions namely:

1. Is there a basic model that can simulate basic manoeuvres of ‘normal’ ships with ‘sufficient accuracy’ and which is valid and useful for a ‘wide range’ of speeds and rudder angles?
2. What is the optimal method to calculate the values of the coefficients of such a model, based on data derived from full-scale sea trials?

For surface ships both these research questions can be answered more or less in the affirmative.

Some time ago the navigation department of the NLDA was facing the question whether a mini-submarine, in this case of the type Autonomous Underwater Vehicle Remote Environmental Measurement Units (AUV REMUS), the same basic model can be used and whether in the same way the coefficients of the model can be estimated. The answer to these questions is given in this article.

Outline

This article does not discuss the mathematical and physical backgrounds that form the basis for composing a manoeuvring model. This can be found in literature [1-8]. In section two the basic manoeuvring model is described. Section three deals with the process to calculate the values of the coefficients of the model with data based on full-scale sea trials. Section four describes the AUV REMUS, and the process of identification and validation of the coefficients for the mathematical model of the REMUS. Section five ends with some conclusions.

The Manoeuvring Model

Conditions and restrictions of the model

The design of a basic manoeuvring model by which the coefficients are determined with the aid of full-scale trials is based on the following conditions:

- As few coefficients as possible, because on the sea the directions of motion can not be uncoupled from each other.
- No additional modification of the model for different types of prime movers, type of propeller etc.
- The model is suited only for basic manoeuvres, for some velocities and rudder angles, such as when the vessel is in transit mode.
- Wind forces and moments can be covered but are primarily considered as disturbances and ignored.

The model

After investigation, the following model is created:

$$\begin{aligned} \dot{u} &= c_{u1}rv + (c_{u2} + c_{u3}|u| + c_{u4}speed)(u - speed) \\ &\quad + c_{u5}(\delta^2 + \delta_d^2)(u \cdot speed)^{c_{u6}} \\ v &= c_{v1}r \\ \dot{r} &= c_{r1}(r^2)^{c_{r2}} \frac{1}{V_t^{c_{r3}}} + c_{r4}\delta(u \cdot speed)^{c_{r5}}, \end{aligned} \tag{1}$$

where:

u	longitudinal velocity [m/s]
v	lateral velocity [m/s]
r	rate of turn [rad/s]
V_t	total velocity $(u^2 + v^2)^{0.5}$ [m/s]
δ	rudder angle [rad]
δ_d	diving rudder angle [rad], towards the REMUS
speed	velocity of the vessel which corresponds to a certain telegraph position [m/s]

This model contains twelve coefficients to be determined. The input parameters are the rudder angle δ and speed. The variable *speed* is the final velocity of the ship that belongs to a position of the telegraph. To determine the coefficient with the aid of full-scale trials the variables u , \dot{u} , v , \dot{v} , r , \dot{r} , δ , δ_d and *speed* must be known.

The considerations which have led to this model, given the conditions, are the following:

X-equation

The first term represents the centrifugal force; this term also follows from the Euler equation. The resistance has the form of: $c_{u2}u + c_{u3}u|u|$. The propulsion is described as the difference between the set speed and actual longitudinal velocity u , adjusted by a coefficient C_{u4} . The advantage of the term (u -speed) is that in the end always a stationary speed is created equal to the set speed if v or r equals zero and δ and δ_d equal zero. The rudder term depends on the rudder position to the square and is supplemented by the current velocity u and the set speed.

Y-equation

According to Keizer, the transverse velocity v of the vessel can be described as a constant multiplied by the rate of turn of the ship. Extensions of the equation with terms in which appear the variables r and v are mostly ineffective, because the corresponding coefficients are not readily identifiable from full-scale trials. Furthermore, the transverse velocity is generally quite small.

N-equation

With c_{r2} equal to a half and c_{r3} and c_{r5} equal to zero, the equation of Nomoto remains. In the equation of Nomoto the coefficients are in fact not constant because the values of the coefficients remain dependent on

the set speed and rudder angle. With the extra variables and coefficients this must be overcome.

In principle the model can be extended with other forces.

Calculating the values of the coefficients based on full-scale sea trials

With the vessel the next trials must be sailed: natural stop trials, gradual acceleration trials, turning circle trials and zigzag trials. During the trials the values of the exact time, position, heading, position of the telegraph and rudder angle have to be collected and saved on a computer. Then the data has to be edited.

First of all the velocities, accelerations, both body-bound, rate of turn and angular accelerations should be calculated. By twice differentiating the measured earth-bound positions and rotating them to the body-bound coordinate system these variables can be calculated. A data file is created with the following variables: Time [s], x [m], y [m], u [m/s], v [m/s], \dot{u} [m/s²], \dot{v} [m/s²], ψ [rad], r [rad/s], \dot{r} [rad/s²], δ [rad], δ_d [rad].

The coefficients now can be calculated with the aid of two mathematical methods. First with the least square method (LSM) and secondly by solving the system of differential equations (SDE) of Formula 1 in a certain way.

The least square method

If we look at a natural stopping manoeuvre, the trial is as follows: first of all the vessel is ordered to sail on a straight line with a steady speed and the rudder angle, transverse speed and acceleration, rate of turn and angular acceleration all equal zero. Secondly the vessel is ordered to reduce the number of revolutions of the propeller to zero with all the above named variables try to stay at zero. Finally the trial is finished if the velocity of the vessel is very low or the vessel is stopped. From the system of differential equations (Formula 1) only the following equation remains:

$$\dot{u} = c_{u2} + c_{u3} |u|$$

In matrix notation for time step i from 1 to k :

$$\begin{pmatrix} \dot{u}_1 \\ \vdots \\ \dot{u}_i \\ \vdots \\ \dot{u}_k \end{pmatrix} = \begin{pmatrix} u_1 & u_1 \cdot |u_1| \\ \vdots & \vdots \\ u_i & u_i \cdot |u_i| \\ \vdots & \vdots \\ u_k & u_k \cdot |u_k| \end{pmatrix} \cdot \begin{pmatrix} c_{u2} \\ c_{u3} \end{pmatrix} \Leftrightarrow B = A \cdot C$$

Values of c_{u2} and c_{u3} are found by $C = (A^T A)^{-1} A^T B$.

Solving the system of differential equations

The computer program MATLAB features a subroutine called *fminsearch*. Fminsearch is a multidimensional unconstrained nonlinear minimization routine. Given the coefficients in the system of differential equations an initial value, the system can be solved. The values of the solved variables can be compared with the same measured values of these variables from the full-scale trials. An error ϵ can be defined as a difference between measured values and solved values so fminsearch can minimize this error by varying the coefficients in the system of differential equations. If the error is minimised a set of coefficients is determined.

The initial values can be calculated with the least square method or by estimation of the value of the relevant coefficient in another way. To get reasonable results, the total error is built up out of an error in position, two errors in the velocities, an error in the rate of turn and an error in the heading of the vessel.

Determining the coefficients or the identification process

As stated, three kinds of manoeuvres are conducted, namely: natural stopping manoeuvres, to determine c_{u2} en c_{u3} ; acceleration manoeuvres, to determine c_{u4} and turning circles manoeuvres, to determine the rest of the coefficients.

Four files are compiled. File A contains all the data belonging to the natural stopping manoeuvre, file B contains all the data of the acceleration manoeuvre, file C includes all the data of the turning circle manoeuvre and finally file D contains all the data of all the trials.

The process by which the coefficients are determined is a kind of iteration. In Table 1 the solution scheme can be seen.

Table 1: Solving scheme. To solve the coefficients 12 steps must be calculated. Bold face typed coefficients are endresults.

Step	Solving method, variable	File type	Initial values of coefficients	Final values of coefficients	Error based on
1	LSM, u	A		c_{u2}, c_{u3}	
2	SDE, u	A	step 1	c_{u2}, c_{u3}	$\epsilon_{pos} \cdot \epsilon_u$
3	SDE, u	B	0.0	c_{u4}	$\epsilon_{pos} \cdot \epsilon_u$
4	LSM	C		c_{u5}	
5	SDE, u	C	step 4 and 1.0	c_{u5}, c_{u6}	$\epsilon_{pos} \cdot \epsilon_u$
6	SDE, u	B	step 3	c_{u4}	$\epsilon_{pos} \cdot \epsilon_u$
7	SDE, u	C	step 5	c_{u5}, c_{u6}	$\epsilon_{pos} \cdot \epsilon_u$
8	LSM, v	C		c_{v1}	
9	SDE, v	C	step 8	c_{v1}	$\epsilon_{pos} \cdot \epsilon_v$
10	LSM, r	C		c_{r1}, c_{r4}	
11	SDE, r	C	step 10, 0.5, 0.0, 0.0	$c_{r1} \text{ t/m } c_{r5}$	$\epsilon_{pos} \cdot \epsilon_\psi \cdot \epsilon_r$
12	SDE, u, v, r	D	step 7, 9 and 11	$c_{u1}, c_{u5}, c_{u6}, c_{v1}, c_{r1} \text{ t/m } c_{r5}$	$\epsilon_{pos} \cdot \epsilon_u \cdot \epsilon_v$ $\epsilon_\psi \cdot \epsilon_r$

This process is more or less applied to a number of Royal Netherlands Navy ships (see Table 2).

Table 2: Ships of the Royal Netherlands Navy of which coefficients have been determined in the past.

Name	Period of measurement	Reference
Vlaardingen	June 1993	
Amsterdam	August 1996	
Zeefakkel	1990	[3]
	August 1997	
van Kinsbergen	August 2001	
Tydeman	July 2002	[9]
REMUS	2007, 2009	[10]

Some notes to the identification process

The calculation of the coefficients in steps is for a number of reasons. In the original approach to estimate the values of the coefficients [3], only the least squares method was used, this has three drawbacks.

First of all, the equations contain velocities and accelerations. They are calculated by applying a numerical differentiation formula to the measured positions and the headings of the vessel. This differentiation process

creates inaccuracies which adversely affect the values of coefficients to be determined. Secondly the least squares method can result in the character and or the value of a coefficient in a term being physically incorrect. Finally the model contains coefficients which are included as a power. They are difficult to calculate with a LSM.

By systematically varying the coefficients and solving the system of differential equations the above disadvantages do not arise, because the error which is to be minimized is based on positions, heading etc, and these variables are hardly affected by mathematical operations. Secondly by solving the system the coefficients must have a value. This value can be manipulated in character and in the size of the value during the process. For instance c_{u2} , c_{u3} , c_{u4} , c_{u5} , c_{v1} , c_{r1} and c_{r4} are necessarily negative or c_{u1} has to lie between pre-calculated values. Finally coefficients included as a power are automatically optimized.

Yet this process of solving the system of differential equations also has some drawbacks. The `fmin` search routine in MATLAB is based on the so called Nelder-Meade simplex method for function minimization:

- As more coefficients are to be estimated and the initial values lie far away from the final values, the necessary computer time becomes more than proportionately longer. Hence the recommendation to start with the estimated initial values as a result of the LSM.
- It is unclear whether this minimization routine is able to find the absolute minimum, experience shows that there are reasons to believe that it can also lead to a local minimum. This can be concluded from the fact that different initial values cause different final values.

The AUV REMUS

The AUV REMUS is torpedo-shaped, and can be equipped with different sensors to conduct measurements of its surroundings. Modelling is a very useful tool to either simulate the behaviour of the vehicle, or to improve the performance of the steering algorithms. For instance, when based on an accurate model, the usages of the vehicle fins become much more efficient. The simulated behaviour of the vehicle can also be used to predict the vehicles attitude and position, either to check the positioning data for major

errors, or to be used as a backup on the primary positioning system. In this latter case, a system could be developed, that calculates a future position estimate, based on the last known position and velocity, and on fin and thruster data. In a military context, this type of positioning system is very useful as a secondary system, because it does not depend on information from sources outside of the platform, such as satellites or beacons.

This study only looked at the manoeuvring behaviour of the REMUS in the **horizontal plane**. The behaviour in the vertical plane is treated in [11] and [12].

The REMUS vehicle

The Royal Netherlands Navy acquired the REMUS AUV for mine detection purposes. Autonomous means, in this context, that the vehicle is not connected to any object, or computer operator. It will carry out a pre-programmed mission, and then return to a rendezvous point. However, the REMUS is designed as an environmental sensing unit. This poses no problems however, because the REMUS can carry a variety of sensors. Its fuselage is made up from different compartments, each having its own functions. However, a couple of parts are essential for normal operation, and can not be altered. For instance, the battery compartment, and the tail are not to be altered. The tail contains, apart from the fins and the propeller, the computer hardware. The nose cone contains some navigational hardware. The other sensors can be altered to the user's preference. In the case of the Royal Netherlands Navy, an object detecting sonar can be installed. The added weight of the sensor modules are however limited. This is because the buoyancy must be positive. This will drive the vehicle to the surface when it is not propelled, for instance, in the case of a flat battery or when it is out of order for whatever reason. So, generally it can be said that the REMUS vehicle can serve many purposes as long as the sensors needed for the measurements fits the fuselage. For this research, no additional sensor will be used as the REMUS vehicle itself stores manoeuvring and position data.

To manoeuvre, the REMUS uses two sets of two fins and a single propeller. However, the command issued to these fins cannot be directly altered. Hydroid, the manufacturing company of the REMUS vehicle, equipped the REMUS with its own decision making program. The fin

angles are determined based on vehicle attitude, altitude, depth and mission. The propeller RPM can be directly altered, and is set in a mission file. Basically, the only control possibilities of the REMUS vehicle are to write a mission file. In this mission file a propeller RPM can be set. The vehicle can keep a given depth, or a given altitude above the sea bed. The mission is described using waypoints. The vehicle will determine its position, either based on the given positions of one or more beacons (with a pre-defined position and frequency), or by taking a GPS position. In this latter case, the vehicle will have to come to the surface. Based on this position information, its goal position (the waypoint in the mission file), and its current heading, the vehicle will calculate the required course alteration. In other words, this results in one error in the attitude that is calculated. The other errors are determined in a similar way. For instance, pitch and depth are calculated by the difference between the actual depth and the set depth. The roll should be zero, as the vehicle should be upright. This results in a given error in attitude. From this, the setpoint for the fins is calculated.

It should be noted, that a set of fin angle is needed to neutralise the roll, induced by the reaction force resulting from the revolutions of the propeller.

Dimensions

For the description of the vehicle specifications, three sources have been used. First of all, the website of the manufacturing company, which gives some basic specifications on the vehicle. Although a citation, some of the data has been deleted (shipping details for instance). Secondly, previous research is used by Prestero [11], and finally, some missing specifications are measured, estimated or calculated based on the vehicle owned by the RNLN.

Full-scale sea trials

As stated, to determine the coefficients of the model, some trials have to be conducted. The experiments took place somewhere on the Amstelmeer, which is a lake in the North-West of the Netherlands, near a village called De Haukes. During the trials a major difficulty is posed by the fact that the rudder can not directly be altered. Basically, the vehicle will sail either

from waypoint to waypoint, or on a given course. The rudder settings will be automatically adapted to follow either the given course or the course to the next waypoint. This fact made it very hard to conduct ‘zigzag trials’, because the turning points should be entered in advance as waypoints. They can not be based on turning speed, time or any other property. It has been chosen to omit this trial. This results in the fact, that turning the properties have to be fully derived from the turning circles.

The identification process

On the Amstelmeer some tracks are sailed and data is logged to a computer. For this study, fifteen pieces (experiments) are cut from these data. Four of them are natural stopping manoeuvres, five of them are acceleration manoeuvres and six of them are turning circle manoeuvres. The four files A to D are compiled.

Table 3 shows the whole identification process in a schematic way. Column one shows the number of successive steps. Column two gives the method used to determine the coefficients, LSM represents the least square method and SDE stands for solving differential equation. It also indicated which variables are solved. Column three is the type of file used to determine the coefficients. Column four shows the initial value of the coefficients and column five gives the final value of the coefficients. The last final values of the coefficients are shown in bold face letter type.

Results and analysis of the validation process

With the coefficients found, the fifteen trials can be simulated. Of three simulations, the results are shown. Figures 1 and 2 show the results of a natural stopping manoeuvre. In Figures 3 and 4 the results of an acceleration manoeuvre and in the Figures 5 and 6 the results of a turning circle manoeuvre can be seen.

The dotted green line represents the measured value of the trials; the solid blue line is the result of the simulation.

The X-equation

The natural stopping manoeuvre is used to determine the resistance of the vessel. Experience shows that in almost all cases good to excellent results are achieved. This is well illustrated in Figure 2, longitudinal velocity u

Table 3: The identification process.

Step	Process, variable	File type	Initial value	Final value
1	LSM 1, u	A		$c_{u2} = -0.0517$ $c_{u3} = -0.0896$
2	SDE 1, u	A	$c_{u2} = -0.0517$ $c_{u3} = -0.0896$	$\mathbf{c}_{u2} = -0.0162}$ $\mathbf{c}_{u3} = -0.1362}$
3	SDE 2, u	B	$c_{u4} = 0.0$	$c_{u4} = -0.0421$
4	LSM 2, u	C		$c_{u5} = 0.5399$ $c_{u6} = 1.0$
5	SDE 3, u	C	$c_{u5} = 0.5399$ $c_{u6} = 1.0$	$c_{u5} = -0.3620$ $c_{u6} = 1.1541$
6	SDE 2, u	B	$c_{u4} = -0.0421$	$\mathbf{c}_{u4} = -0.0391}$
7	SDE 3, u	C	$c_{u5} = -0.3620$ $c_{u6} = 1.1541$	$c_{u5} = -0.3619$ $c_{u6} = 1.1541$
8	LSM 3, v	C		$c_{v1} = -0.8766$
9	SDE 4, v	C	$c_{v1} = -0.8766$	$c_{v1} = -0.0759$
10	LSM 4, r	C		$c_{r1} = -0.5429$ $c_{r2} = 0.5$ $c_{r3} = 0.0$ $c_{r4} = -0.3094$ $c_{r5} = 0.0$
11	SDE 5, r	C	$c_{r1} = -0.5429$ $c_{r2} = 0.5$ $c_{r3} = 0.0$ $c_{r4} = -0.3094$ $c_{r5} = 0.0$	$c_{r1} = -176.8837$ $c_{r2} = 2.8177$ $c_{r3} = -2.8616$ $c_{r4} = -0.2983$ $c_{r5} = 2.0092$
12	SDE 6, u, v, r	D	$c_{u1} = 1.14$ $c_{u5} = -0.3619$ $c_{u6} = 1.1541$ $c_{v1} = -0.0759$ $c_{r1} = -176.8837$ $c_{r2} = 2.8177$ $c_{r3} = -2.8616$ $c_{r4} = -0.2983$ $c_{r5} = 2.0092$	$\mathbf{c}_{u1} = 0.7980}$ $\mathbf{c}_{u5} = -0.6639}$ $\mathbf{c}_{u6} = 0.8277}$ $\mathbf{c}_{v1} = -0.0459}$ $\mathbf{c}_{r1} = -7.3560}$ $\mathbf{c}_{r2} = 0.7812}$ $\mathbf{c}_{r3} = 0.8896}$ $\mathbf{c}_{r4} = -2.1965}$ $\mathbf{c}_{r5} = 0.5016}$

against time. Because the propeller is not modelled separately, the effect of the propeller on the resistance is ‘embedded’ into the coefficients c_{u2} and c_{u3} .

The position accuracy is in general less. This has to do with the fact that with an ever lowering velocity, the disturbances have a greater influence on the path the vessel sails. These disturbances are corrected during the trial

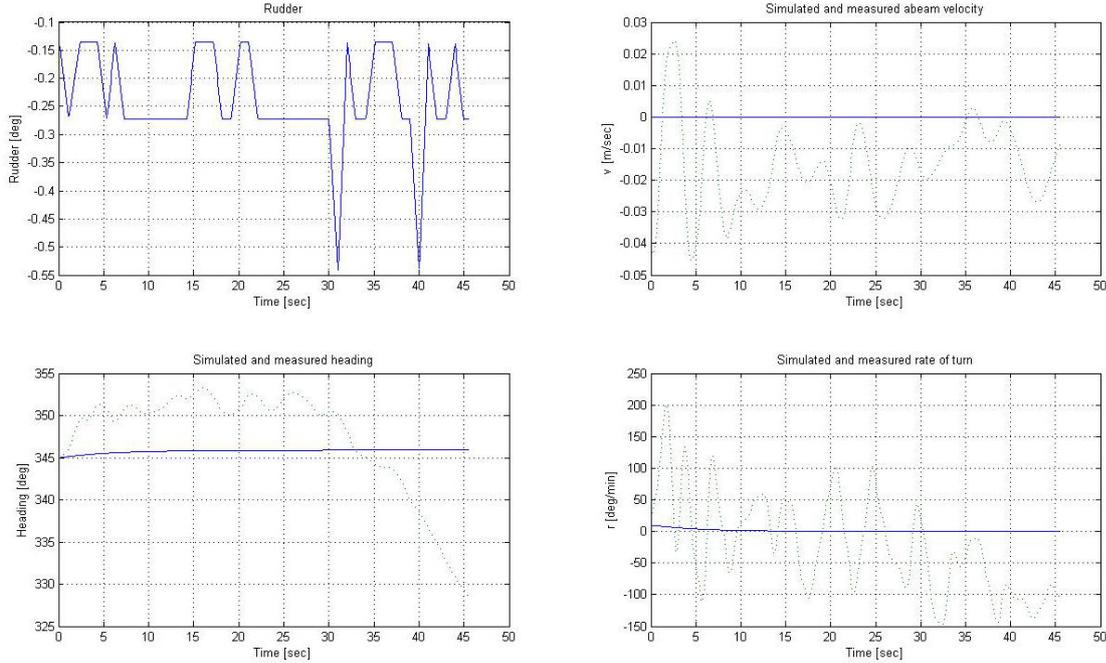


Figure 1: Results of natural stopping manoeuvre.

with the rudder, but these distortions are not included in the simulation model, while during the simulations the rudder signal of the trial is used as an input.

The acceleration manoeuvre is used to determine the thrust of the propeller. Usually a run is sailed in which a gradual acceleration is applied. For the REMUS this is not possible. The results of the acceleration trial are reasonable, see Figure 4. Again longitudinal velocity u against time is shown.

The position accuracy is very moderate. The reason is that it is extremely difficult to accelerate a vessel from zero speed without inducing a rate of turn. This has been difficult on a surface ship, let alone the REMUS. Secondary propeller effects for instance are not modelled. In addition, if the velocities of the REMUS equals zero the vessel comes up to the surface due to positive buoyancy.

Simulating a turning circle manoeuvre the accuracy of the positions will be especially important. We see in Figure 6 that the positions are well simulated. Due to added resistance from the fins the longitudinal velocity decreases during a turning circle manoeuvre. This is nicely simulated,

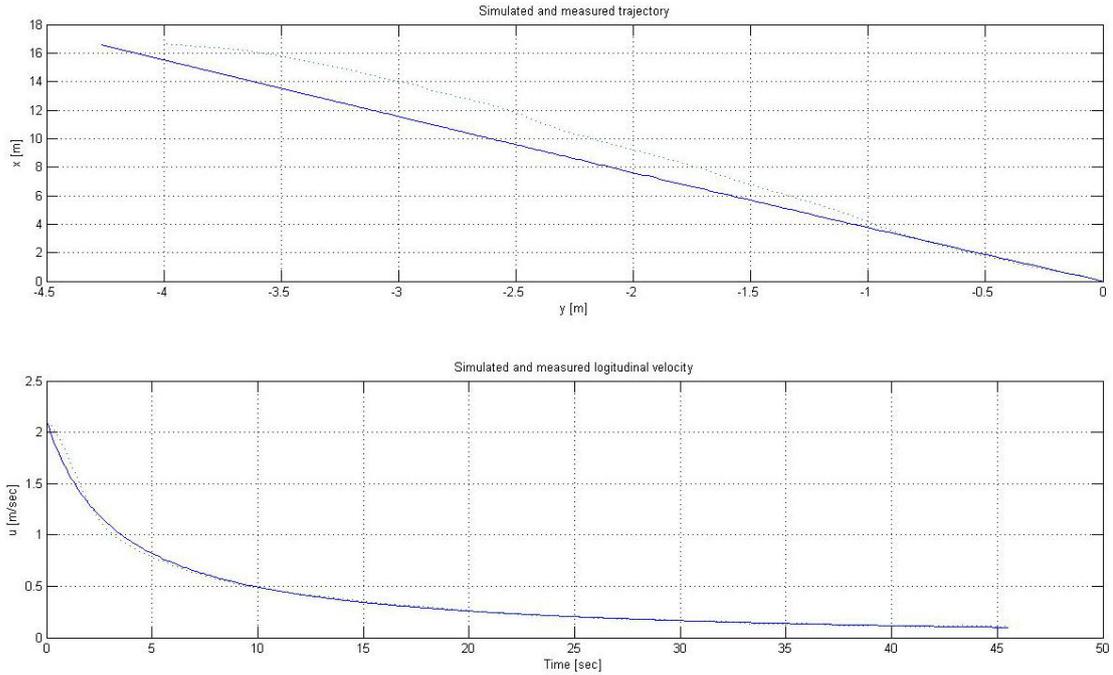


Figure 2: Results of natural stopping manoeuvre.

shown in the plot of the longitudinal velocity u against time.

The Y-equation

About the abeam velocity v we can be brief. In literature it can be found that the lateral velocity v can be written as a coefficient multiplied by the rate of turn of the ship. To describe this variable otherwise, a Nomoto expression can be used, but the Nomoto equation gives a number of additional coefficients to be determined.

The measured values of the abeam velocity v during the trials shows no correlation between a single parameter. For instance sailing a turning circle to starboard, an abeam velocity v to port is expected, none of the trials however shows this behaviour. This means that either the measurements of this variable during the trials is imprecise, or that the vessel did not get a clear transverse drift velocity v by sailing a turning circle. The latter seems unlikely. Because no correlation is found, c_{v1} is very small and as a consequence the values of the lateral velocity v lie around zero meters per second. See Figures 1, 3 and 5, v against time.

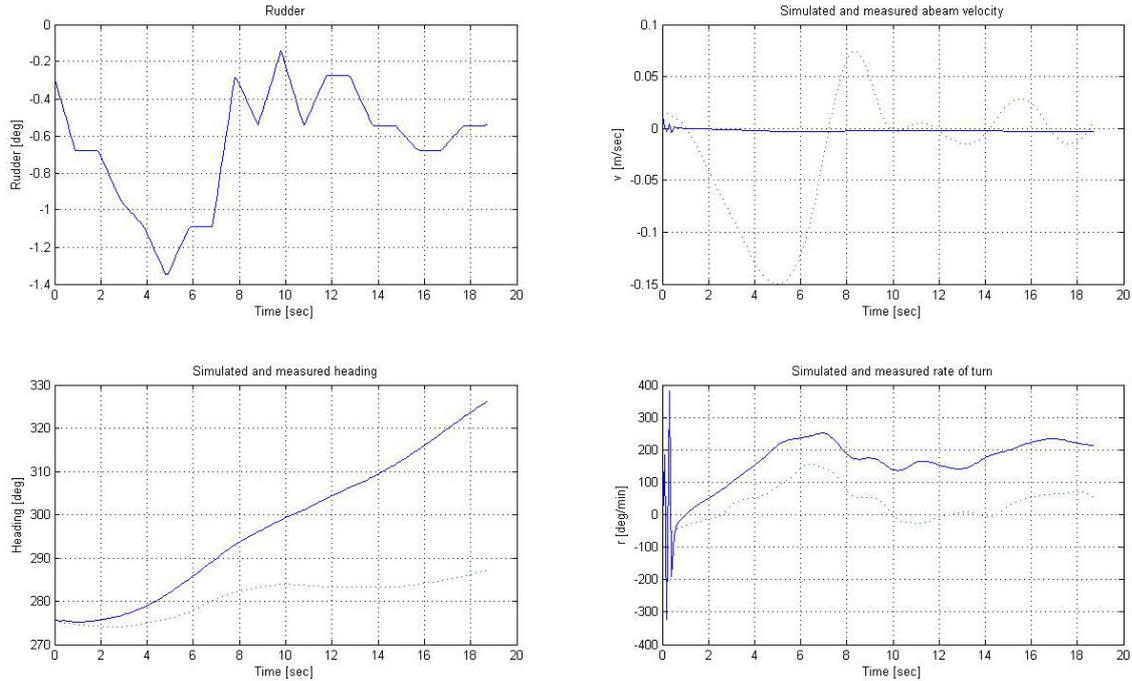


Figure 3: Results of acceleration manoeuvre.

The N-equation

Both during the natural stop manoeuvre and the acceleration manoeuvre little stirs are given. The tracks are almost straight. The rate of turn, heading and position in the turning circle trial is well simulated. The results of the simulation of the other turning circle trials are well comparable. An important criterion is the simulated behaviour of the rate of turn when the rudder is placed amidships, to see in Figure 6 the rate of turn r against time.

Some additional notes to the identification process

On a theoretical basis the value of coefficient c_{u1} is easy to calculate. The start value of 1.14 is the calculated value. In solving the system of differential equations a deviation of thirty percent is allowed. This percentage is arbitrary.

From the solution scheme in Table 1 it can be concluded that to calculate the final values of the coefficients of the model, only one last step, from eleven to twelve is needed. In practice this is somewhat optimistic, sometimes extra steps are needed to get the final set coefficients. This requires expertise and experience and often some additional research.

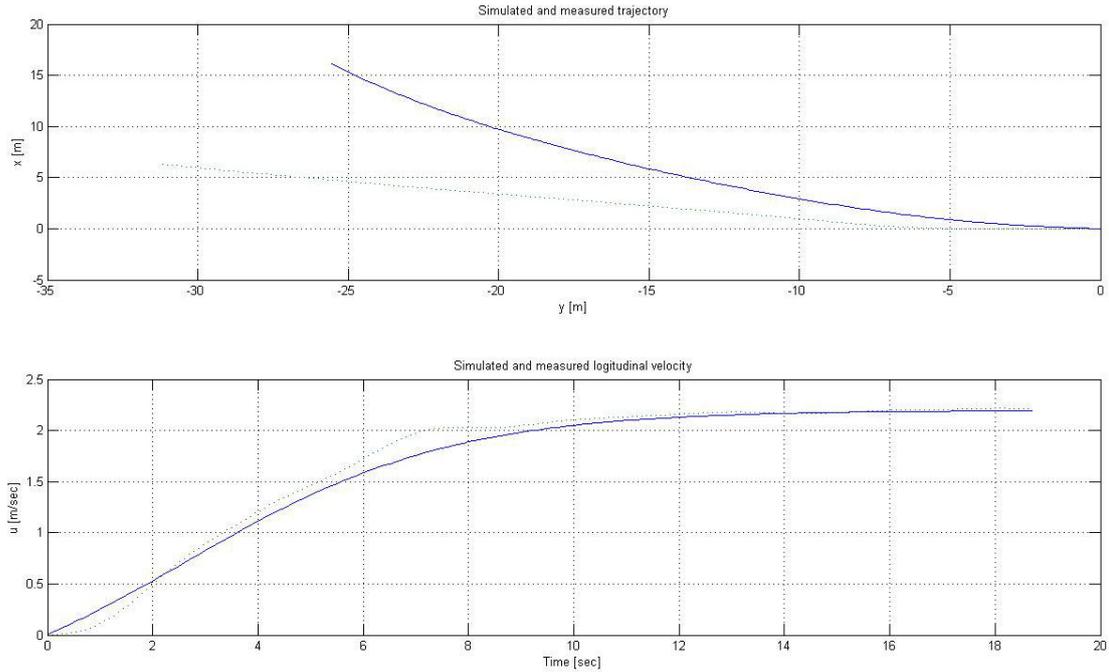


Figure 4: Results of acceleration manoeuvre.

To determine the coefficients involved in the turning of the ship it is not strictly necessary to take into account data from the natural stopping trials or acceleration trials. After all with that type of trial no rudder is given and so the rate of turn of the ship is very low. Yet there are two reasons to take into account the data of these types of trials. First of all, by conducting these trials often some rudder is given to hold the vessel on a straight heading. So the data file contains additional rudder information by which the values of the coefficients are ‘better’ calculated. This means the set of coefficient determined ‘better’ simulate the manoeuvring characteristics of the ship. The second reason is of more importance. When simulating the manoeuvring characteristics of the ship as stated you have to deal with a model and a set of coefficients. To simulate, the system of differential equations has to be solved. Depending on the initial condition of the variables and the values of the coefficients it may occur that the solution did not converge so the simulation delivered no solution. This occurs most frequently when in step 12 only the data of the turning circle trials and or zigzag trials are taken into account. Then, there is a change; neither the natural stop trial nor the acceleration trials can be well simulated.

In the case of the REMUS the error, which is minimized and which is

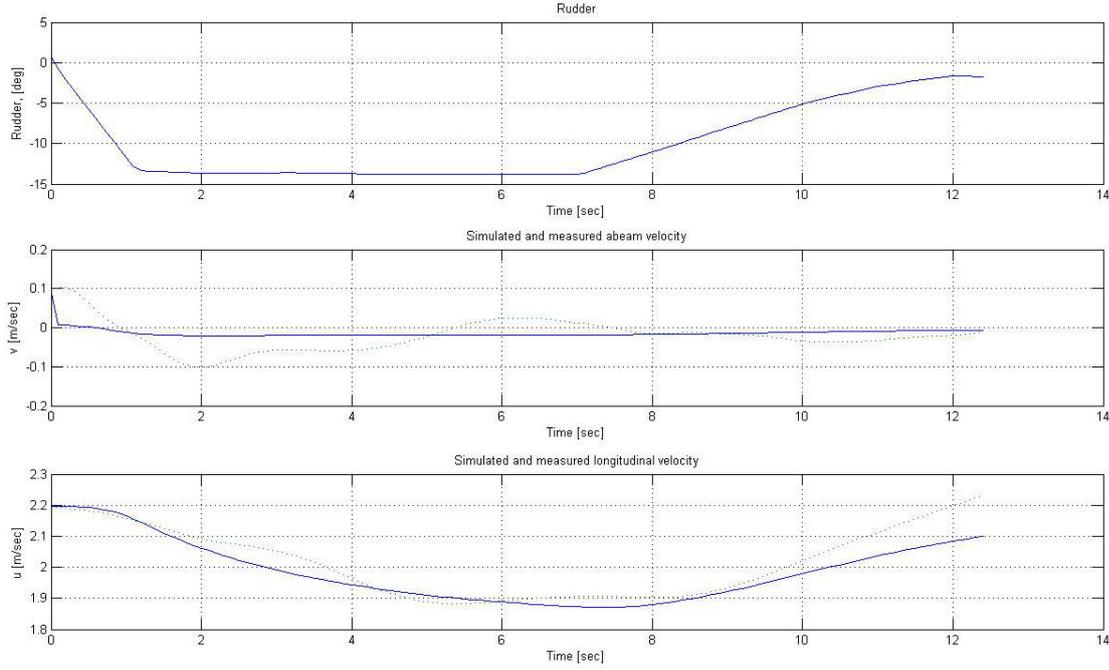


Figure 5: Results of turning circle manoeuvre.

important to find the final value of the coefficients, is split up into two errors. The first is based on the natural stopping trials and acceleration trials ($error_a$) and the second is based on the turning circle trials ($error_b$). Initially to optimize the coefficients involved by the turning of the ship only the error based on the turning circle trials ($error_b$) has to be minimized, so the error which will be minimized is defined as follows:

$$Error = \frac{error_b}{error_a} \cdot error_a + error_b$$

This seems absurd, but if $error_a$ becomes *not a number*, because during the integrating process the values of the variables do not converge, the total error also becomes *not a number* and the process to find the values of the coefficients will continue until the error is found with the lowest real value. Please note, by this choice of the error, only the error from the turning circle trials is used to optimize the determined coefficients. Then step 12 is performed again, but $error_a$ is divided by a factor. This factor is the last value of $error_b$ divided by $error_a$ from the previous step. Please note, by this choice $error_a$ is of the same order of magnitude as $error_b$, but

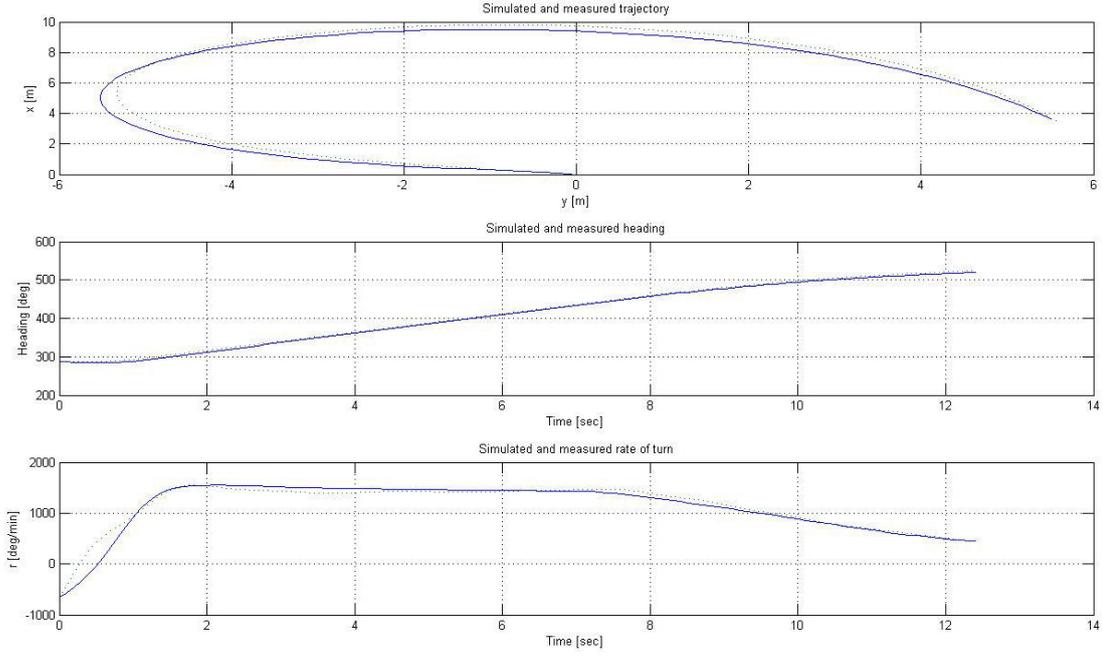


Figure 6: Results of turning circle manoeuvre.

the found values of the coefficients are based and optimized on the data of all the trials.

In the case of the REMUS, theoretically the propeller thrust can be modelled very simply in another way. The propeller of the REMUS has a fixed pitch so the velocity of the vessel is governed by the amount of revolutions per second n . The amount of revolutions can be measured during the trials. The x-equation becomes:

$$\dot{u} = c_{u1}rv + c_{u2}u + c_{u3}|u|u + c_{u4}un + c_{u5}n^2 + c_{u6}(\delta^2 + \delta_d^2)(uspeed)^{c_{u7}}$$

The propulsion section now has the form: $c_{u4}un + c_{u5}n^2$. By measuring the revolutions, c_{u5} and c_{u6} can be determined by using a LSM and SDE method in the same way as the other coefficients. It is expected that acceleration behaviour of the vessel will be better simulated.

The minimization routine is a standard call in the program MATLAB. In the past, to find a reasonable set of coefficients steps seven to twelve have to be conducted to find the final set. If in step twelve, the chosen initial values

of the coefficients equals zero and only c_{u1} gets an estimated or calculated value, no realistic set of coefficients was found. For the REMUS, as an additional research, this is tried again, see Table 4. The values of the final set so found (column five) shows little differences with the values found according the process from Table 3. This can be a matter of luck or the software version of MATLAB used. It seems that the minimization routine in this case is able to find the absolute minimum. This makes step seven until eleven redundant. This can lead to considerable time saving.

Table 4: Differences between the final values of the coefficients by different initial values, step 12.

Coefficient	Initial values steps 7-12	Final values Table 3	Initial values	Final values
c_{u1}	1.14	0.7980	1.14	0.7980
c_{u5}	-0.3619	-0.6639	0.0	-0.6563
c_{u6}	1.1541	0.8277	0.0	0.8386
c_{v1}	-0.0759	-0.0459	0.0	-0.0447
c_{r1}	-176.8837	-7.3560	0.0	-7.5152
c_{r2}	2.8177	0.7812	0.0	0.7917
c_{r3}	-2.8616	0.8896	0.0	0.9035
c_{r4}	-0.2983	-2.1965	0.0	-2.1538
c_{r5}	2.0092	0.5016	0.0	0.5116

As a last step, a large number of simulations should be performed to test whether in all (realistic) situations the model converges. This is not done as part of this brief research.

Interesting in this context is Figure 3. The figure of the rate of turn r and the lateral velocity v presents a small form of instability. The reason for that can be: the start values of the variables, the found values of the set coefficients or the used solution method of the differential equations. Also this issue is not further investigated.

Conclusions

The research question, whether for the REMUS the same type of simulation model can be used as for surface ships and whether a set of coefficients can be found based on full-scale trials, can be fairly positively answered. The fifteen trials as a basis for determining the coefficients can be reasonably well simulated with the set of coefficients found and the applied model. Please note, the modelling is only valid in the horizontal plane.

Some concern exists about the conducted full-scale trials. The REMUS navigates using a mission file. It should be investigated whether the vessel can be controlled by hand instead of the mission file, resulting in all relevant standard manoeuvring trials being performed.

The frequency of writing the data is too low. A rule of thumb states:

$$f = \frac{10u_{max}}{L_{ll}}$$

For the REMUS, this represents 16 [Hz]. During the full-scale trials, data is logged with a frequency of one hertz. By interpolation this is increased to ten hertz.

The discussed method to find a set of coefficients by solving the system of differential equations and to define an error as the difference between measured values and simulated values can also be applied if the values of the coefficients are determined otherwise. Then the method can be used to fine tune the set of coefficients of the manoeuvring model. The initial values of the coefficients are then the values based on an empirical formula or on towing tank trials. During the integration process the coefficients are allowed to differ from the initial value by an arbitrary margin.

References

- [1] S. Inoue S, M. Hirano and K. Kijima, “Hydrodynamic derivatives on ship manoeuvring,” *International shipbuilding progress*, May 1981.
- [2] M.A. Abkowitz, *Stability and motion control of ocean vehicles*. Massachusetts: M.I.T. press, 1969.
- [3] J.H. Wulder, “The implementation of a ship manoeuvring model in a integrated navigation system,” Ph.D. dissertation, Delft University of Technology, Delft, 1992.
- [4] D.J. Engelbracht, “Manoeuvreermodellen,” Lecture notes NLDA, Den Helder, 2008.

- [5] V.Bertram, *Practical ship hydrodynamics*. Oxford: Butterworth Heinemann, 2000.
- [6] S. Motora, “On the measurement of added mass and added moment of inertia for ship motions,” *Journal of the society of naval architects of Japan*, January 1960.
- [7] J.A.Steketee, “Lecture notes AERO I D-18,” Delft University of Technology, Delft, 1973.
- [8] J. Gerritsma Ed., “Scheepsbewegingen en sturen 2,” Delft University of Technology, Delft, Report 487-K., 1983.
- [9] J.P. Loog, “SDNE: lateral navigation accuracy,” M.Sc. thesis, University of Nottingham, Nottingham, 2004.
- [10] W.M. van der Hilst Karrewij, “Manoeuvring behaviour of the REMUS AUV, Identification and validation of a basic three degree of freedom simulation model,” B.Sc. thesis, NLDA, Den Helder, 2009.
- [11] T. Prestero, “Development of a six-degree of freedom simulation model for the REMUS autonomous underwater vehicle,” M.Sc. thesis, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, Cambridge (Ma), 2001.
- [12] A.S. Wessel, “Verticale (in-)stabiliteit van een REMUS AUV,” B.Sc. thesis, NLDA, Den Helder, 2009.

On Line Measurement of High Frequency Ship Motions of Channel Bound Ships

Dick Engelbracht & Wim van Buuren

Introduction

For the Dutch economy the ports of Amsterdam and Rotterdam are crucial. To continue the competitive positions of both ports versus foreign ports, the entrance channels to both ports have to be optimal. That said, it is however a problem. The largest ships calling at these ports have a very deep draught, so the ports are no longer accessible around the clock.

Still to receive such vessels, the competent authority worked with a so called tidal window. This means the ship steams at a certain time before the high tide (highest water level) and must have reached the port a certain elapsed time after the high tide. In other words there is an admission policy for the largest ships.

This has to do with the water depth of the fairway and the draught of the ship. This article discusses the draught of the ship. A ship sailing in waves has no constant draught but a varying draught due the ship motions, caused by the waves. These ship motions are: roll (rotation around the longitudinal axis), pitch (rotation around the lateral axis) and heave (translation in the direction of the vertical axis).

Criticism of the current admission policy

Rijkswaterstaat calculates, based on a deterministic- (past) or probabilistic calculation method, the tidal window for a specific ship. This calculation, taking into account IMO regulations, included [1]: the predicted water level, the draught, the squat, a safety margin for the ship motions due to

waves, the net under keel clearance, a margin for sedimentation between two soundings, sounding inaccuracies, extra dredged depth and the dredging tolerance. With these items you can calculate the net and gross keel clearance.

The master and to a certain extent the licensed maritime pilot, who advises the master about the navigation during the approach to the port area, will both be held responsible if the ship hits the bottom of the channel due the motions of the ship. Regarding the applied tidal window, they have to trust blindly on the calculations made by Rijkswaterstaat. That is, to say the least, not a desirable situation. The question arises: is it possible that the pilot during the transit across the channel can obtain an objective indication of the variation in the draught of the ship, including those resulting from the ship's motions by waves?

So this is a double edged sword. If it appears that the draught increase is larger than calculated, i.e. the tidal window is too wide, the passage should be aborted to avoid damage. On the other hand, if monitoring shows that the calculated tidal window is too short and that actually more time is remaining than calculated, this leads to a lengthening of the tidal window and hence an increase in the port accessibility.

Measuring the ship motions

The answer to the above question can to some extent be positive. Give the pilot a set of sensors, especially accelerometers, place these sensors on board and determine the draught changes by monitoring these motions. If this is practical, still there are some disadvantages. The pilot comes on board by helicopter or by the pilot boat, often after the necessary obstacles meaning that he or she must have hands free and all luggage should be light, small and well manageable. The priority is the navigation of the ship, so equipment is installed and has to be up and running as quick as possible. With the introduction in 2005 of the Portable Pilot Unit (PPU), some things have been speeded up.

The PPU is an electronic sea chart in the form of a laptop connected to a GPS/GLONASS receiver. It is a very technologically advanced navigation tool, entirely independent of the navigation equipment on board of the ship. The PPU presents the pilot all the navigation information, clearly and with

high accuracy, thus supporting him, along with his own observations, in his decision making process. Because use is made of a GPS/GLONASS receiver in a master-slave configuration, combined with real time position correction and a separate azimuth sensor, it is possible to calculate very accurately the position, the heading and one rotation angle of the ship. So this PPU has, at least partly, a sensor to determine the changes of the draught due to motions by waves.

Incidentally the Dutch Pilots in cooperation with the Netherlands Defence Academy has in the recent years developed a number of stand-alone programs which are installed on the laptop of the PPU, these are programs to calculate:

- Current forces on ships in confined channels and open sea [2];
- Anchor manoeuvring of vessels with very little under keel clearance;
- Forces and moments on the ship by wind;
- Squat [3];
- The anchor chain length, taking into account wind and current [4];
- Forces on mooring buoys;
- Wind forces on standard ship types [5].

Conducted research

To investigate whether the PPU actually can contribute to an objective determination of the draught changes, a number of studies are being conducted, two of them by the Netherlands Defence Academy KIM. These studies are:

- To investigate the accuracy of angular measurement of the PPU, the angular measurement is compared with the angle measurement of two different inertial navigation systems. This study was conducted on a frigate of the Royal Netherlands Navy in the IJ-channel (IJmuiden) [6].

- In the Euro-Maas-channel (Rotterdam) a bulk carrier was equipped with four PPU distributed over the ship, so the three unknowns mentioned above, heave, roll and pitch, can be solved mathematically [7] and [8].

The purpose of both studies is twofold, first of all to investigate how the measurements of the PPU are influenced by the location of the equipment on board and secondly, whether it is possible with a PPU to measure the ship's motions.

Outline of this article

This article does not explain the precise operation of the PPU. This is found in literature; see [6] and [9]. In section two, the underlying theory of motions of ship by waves is treated. Section three deals with the study of the bulk carrier. The article closes with section four, conclusions and recommendations.

Mathematical and physical explanation about ship motions by waves

Ship motions can be considered as output of a system [10-12]. The input is the wave field, the output, as stated, the actual motions of the ship and the system, the ship itself. The output depends amongst other things on the ship's natural frequencies and the natural frequencies in turn depend on the dimensions and mass distribution of the ship.

Description of the sea

Ship motions are a response to the resulting forces and moments which are exerted on the vessel. The resulting forces and moments in this case are caused by the sea. Sea can be seen as a superposition of many simple regular harmonic wave components each with its own period, phase angle, amplitude, length and direction of propagation, see Figure 1.

The properties of the single wave components should be known for more understanding of the wave field in which the ship is sailing. These properties, for instance the wave speed, include depending on the depth of the water [13]. Because we are dealing with harmonic signals, some advantage



Figure 1: The surface of the sea left may be described by a sum of an infinit number of waves with different amplitudes, frequencies en directions (right). Both figures: source Delft University of Technology.

can be obtained from performing the calculations in the frequency domain, instead of in the time domain. With a Fourier analysis the time signal is converted to the frequency domain, and represented as a so-called wave-spectrum, see Figure 2. The wave spectrum is to be transformed into a so-called encounter-spectrum, because the ship is sailing with a particular course and speed compared to the speed and direction of the waves.

A wave spectrum can be regarded as a distribution of the square of the wave heights in the seas over the frequencies. The square of the wave heights is a measure of the energy in the waves.

It is obvious that, if the input is described in form of a wave-spectrum, the output should be described, in the same form as a motion-spectrum.

Description of the ship motions

Three right-handed orthogonal coordinate systems are used to define the ship's motions:

- An earth-bound coordinate system $S(x_0, y_0, z_0)$. The (x_0, y_0) -plane lies in the still water surface, the positive x_0 -axis is in the direction of the

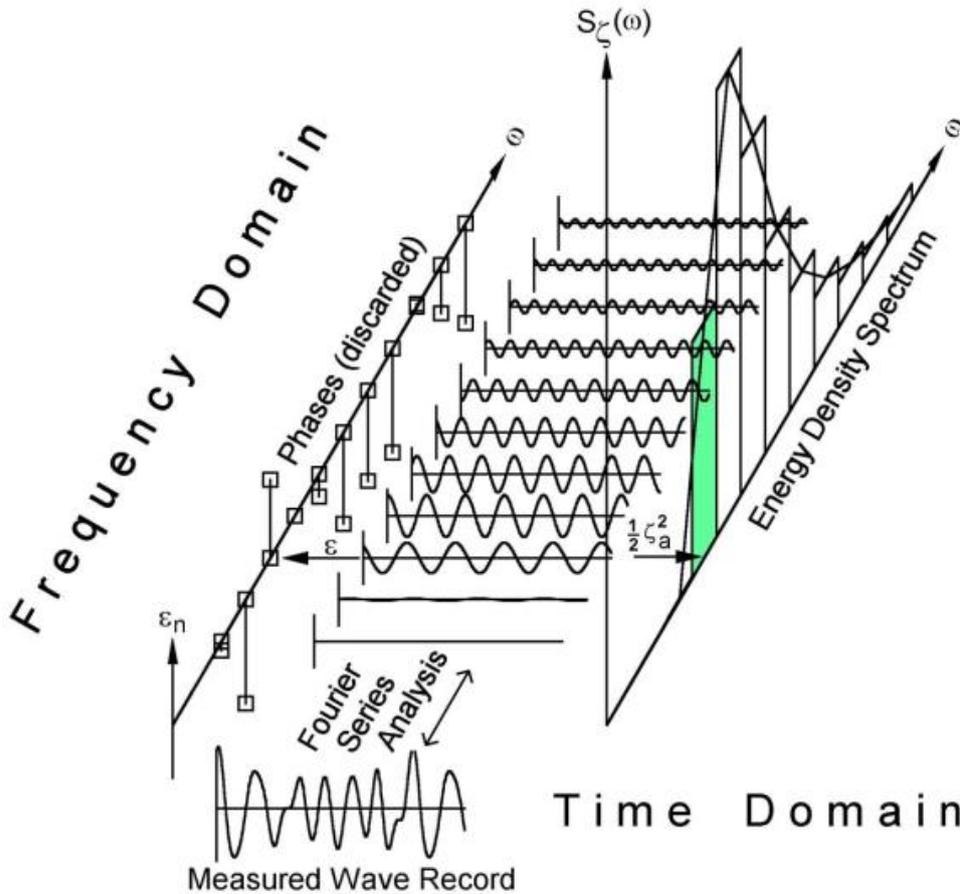


Figure 2: Conversion of a signal in the time domain to the frequency domain. Source: Delft University of Technology.

wave propagation. It can be rotated at a horizontal angle relative to the translating axis system $O(x,y,z)$. The positive z_0 -axis is directed upwards.

- A body-bound coordinate system $G(x_b, y_b, z_b)$. This system is connected to the ship with its origin at the ship's centre of gravity, G . The directions of the positive axes are: x_b in the longitudinal forward direction, y_b in the lateral starboard side direction and z_b upwards. If the ship is floating upright in still water, the (x_b, y_b) plane is parallel to the still water surface.
- A steadily translating coordinate system $O(x,y,z)$. This system is moving forward with a constant ship speed V . If the ship is stationary, the directions of the $O(x,y,z)$ axes are the same as those of the $G(x_b, y_b, z_b)$ axes. The (x,y) -plane lies in the still water surface with the origin O at, above or under the time-averaged position of the cen-

tre of gravity G. The ship is supposed to carry out oscillations around this steadily translating $O(x,y,z)$ coordinate system.

Ship motions are defined as the motions of the (x_b, y_b, z_b) axes relative to the (x, y, z) axes. However, the PPU calculate the position in the (x_0, y_0, z_0) system.

A free moving mass has six degrees of freedom, three rotations: roll ϕ , pitch θ and yaw ψ , and three translations: surge x , sway y and heave z . The height of any point n , $(x_b(n), y_b(n), z_b(n))$ on board, in the earth-bound coordinate system, mainly is influenced by the heaving, pitching and rolling motion of the ship. Ships are not totally rigid, so a seventh variable can be added, the vertical bending w_z of the ship. The motions, x , y , z , ϕ , θ and ψ are defined as the motions of the mass centre of gravity G. Regardless of the transformations of the body-bound coordinate system to the earth-bound coordinate system, and what direction is positive or negative, the height of a point for instance the position of the antenna of a GPS/GLONASS receiver is equal to:

$$z_0(x_b(n), y_b(n), z_b(n)) = z + z_b(n) + x_b(n) \cdot \theta + y_b(n) \cdot \phi + w_z(x_b(n)). \quad (1)$$

This equation is linearized and only valid for small rotation angles. In the equation above, $z_0(x_b(n), y_b(n), z_b(n))$ is measured by the master antenna of a GPS/GLONASS receiver at point n in the earth-bound coordinate system; $x_b(n)$, $y_b(n)$ and $z_b(n)$ are the coordinates of a receiver in the body-bound coordinate system. In the equation there are basically four unknown variables: z , ϕ , θ and the vertical bending w_z at the point n . By using four receivers, the four unknown variables can be solved. By placing the master and slave antennas, in transverse or longitudinal direction, ϕ or θ can also be determined so the number of unknown variables became less.

Research on board the bulk carrier Ferro Goa

During the run to the port of Rotterdam on board of MS Ferro Goa the next data of four PPU's were logged and written to a computer: time, position, angles etc. The passage to Rotterdam consists of two tracks, the ship starts on a track of 082.5° with a heading of 080° to compensate for current and wind. The second track of 112° is to be steered with a

heading of 105° , halfway the passage. During the total passage time the sea state has remained roughly constant, this means that the ship has met waves from two different directions relative to the ship. In other words, two different encounter-spectra can be made.

Details of the ship

In Table 1 one can see the loading condition of the bulk carrier Ferro Goa according to the onboard stability program.

Table 1: Loading condition of MS Ferro Goa.

L_{ll}	280.0	[m]	GG'	0.55	[m]	
L_{wl}	286.0	[m]	MG'	5.27	[m]	
B	45.0	[m]	Y_G	0.03	[m]	starboard
Δ	193938	[ton]	X_G	-7.43	[m]	for the half length
T_f	17.42	[m]	T_0	12.1	[sec]	
T_a	17.58	[m]	TPC	121.26	[ton/cm]	
T_m	17.50	[m]	ETM	2537.33	[tonm/cm]	
KM	18.73	[m]	Ben.mom.(sag.)	-1645770	[kNm]	42% of the max.
KG	12.91	[m]	Shear force	-37897	[kN]	44.3% of the max

The locations of the PPU

Four PPU are placed on board. In Table 2 the coordinates of the PPU positions in [m] relative to G are in the body-bound coordinate system.

Table 2: Coordinates of PPU positions in the body system.

	station	x_b	y_b	z_b
PPU1	one	114.37	-0.03	14.84
PPU2	two	-2.53	-0.03	14.84
PPU3	three	-115.33	22.47	27.44
PPU4	four	-115.33	-22.53	27.44

The PPU measures the heights in the earth-bound coordinate system, in addition, at station one and station four the pitch is measured and at station two and station three the roll. These rotations are not calibrated, nor to the earth-bound coordinate system nor to the body-bound coordinate system. The sampling frequency is equal to 5 [Hz].

The test conditions

The test conditions are summarized in Table 3.

The results of the measurements

By each PPU the height is measured and also at station two and station three the roll and at station four and station one the pitch. Some results can be seen for run one in Figure 3 and for run two in Figure 4.

Table 3: Conditions during the experiments.

Dataset		Run 1	Run 2
Length of the run	[hh:mm:ss]	05:46:35-05:52:34	06:00:59-06:10:59
Speed	[m/s]	5.00	4.58
Heading	[°]	080	105
Swell angle relative to the ship	[°]	055	030
Channel water depth	[m]	24	24
Swell direction	[°]	315	315
Water depth outside the channel	[m]	17	17

Taking a closer look at the measured signals we can see that the measured heights of all four stations and the measured roll signal of station two have a small bandwidth and that the signals are little affected by noise. The noise on the signal from the roll of station three and the signal of the pitch of station one is just a little larger. The noise on the signal of the pitch of station four is by far the largest.

The solution of the system of equations

As with four receivers, n_1 to n_4 , are measured at the same time at four different locations on board, the unknowns z , θ , ϕ and the bending can be solved out of a system of four equations, in matrix form the system of equation reads:

$$\begin{pmatrix} z_0(x_b(1), y_b(1), z_b(1)) - z_b(1) \\ z_0(x_b(1), y_b(1), z_b(1)) - z_b(2) \\ z_0(x_b(1), y_b(1), z_b(1)) - z_b(3) \\ z_0(x_b(1), y_b(1), z_b(1)) - z_b(4) \end{pmatrix} = \begin{pmatrix} 1 & x_b(1) & y_b(1) & 1 \\ 1 & x_b(2) & y_b(2) & \left(\frac{x_b(2)}{x_b(1)}\right)^4 \\ 1 & x_b(3) & y_b(3) & \left(\frac{x_b(3)}{x_b(1)}\right)^4 \\ 1 & x_b(4) & y_b(4) & \left(\frac{x_b(4)}{x_b(1)}\right)^4 \end{pmatrix} \begin{pmatrix} z \\ \theta \\ \phi \\ w_z(x_b(1)) \end{pmatrix} \quad (2)$$

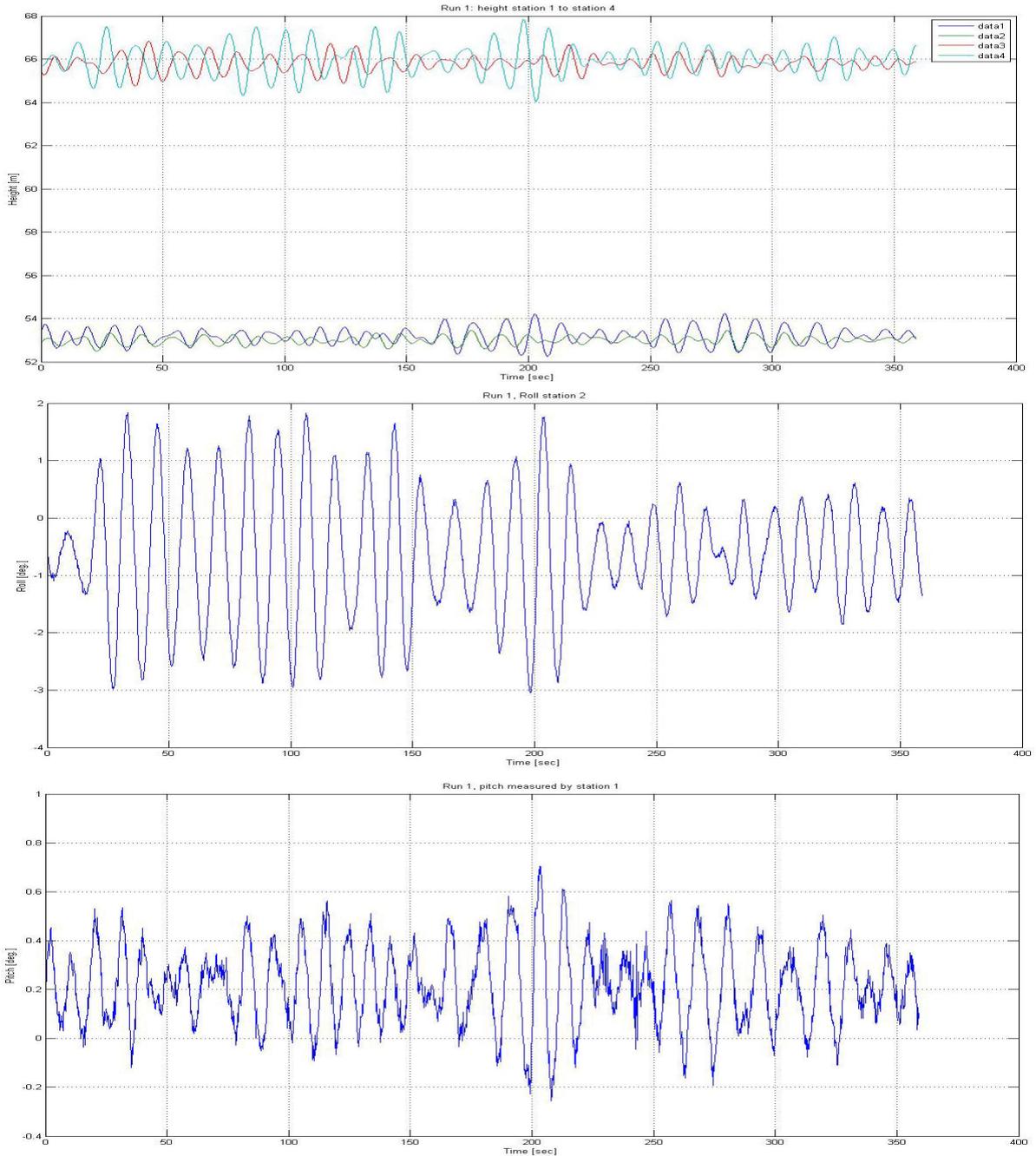


Figure 3: Run one, top: measured heights of the four stations; middle: measured roll of station two; bottom: measured pitch of station one.

The bending of the ship at various points can be determined by the bending at one point. This is based on common rules for the strength of materials:

$$w_x(x_b(n)) = \frac{x_b(n)^4}{x_b(1)} \cdot w_x(x_b(1))$$

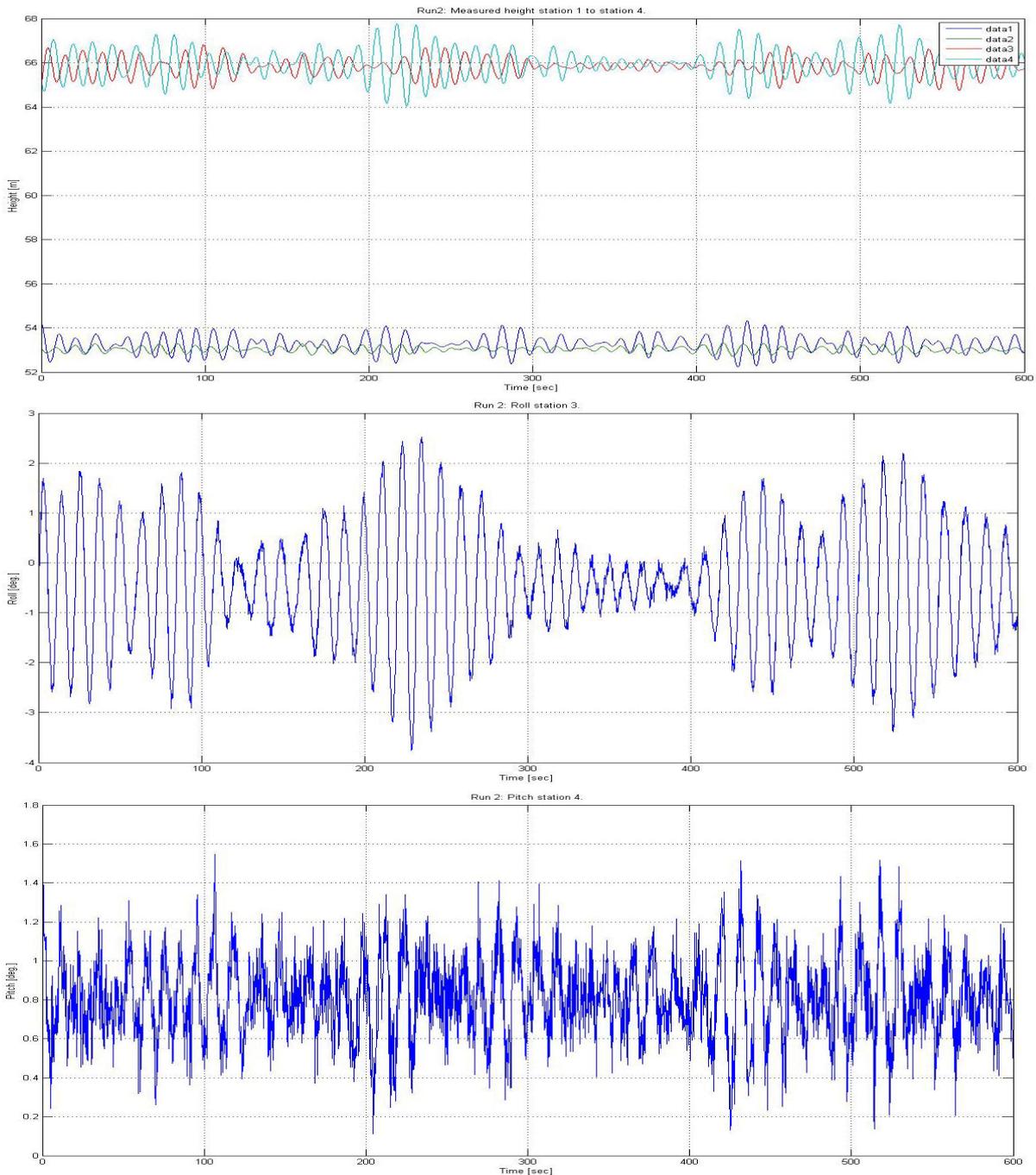


Figure 4: Run two, top: measured heights of the four stations; middle: measured roll of station three; bottom: measured pitch of station four.

On the left side of Formula 2 you can see the measured heights in the earth-bound coordinate system, corrected for the height of the antenna. On the right side of the system the first matrix includes the x and y coordinates of the receivers in the body-bound coordinate system and the expression for the bending of each location of the receivers. The second, most righthand

matrix contained the unknowns to be solved.

To get a first impression about the accuracy of the solution compared to the measured roll and pitch of the various stations, a number of figures can be compared.

First, the solution of the heave motion z is easy to compare with the measured heights of station two, because the height measurement of station two is barely affected by the roll and the pitch of the ship. The reason for that lies in the fact that station two is located very close to the centre of gravity G , see Figure 5.

Secondly, the measured roll and pitch must correspond to the pitch (θ) and roll (ϕ) motion that follows from the solution of the system, see Figure 6.

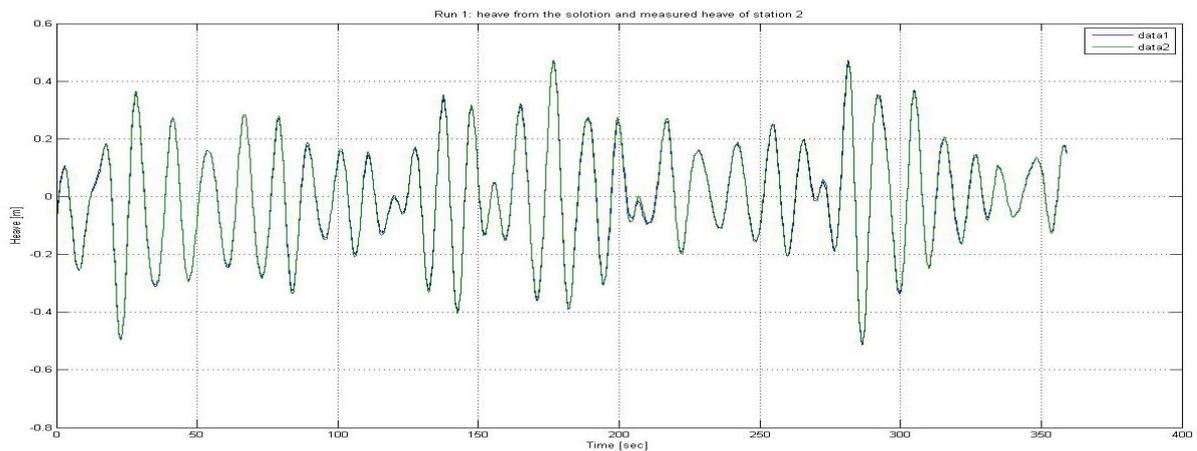


Figure 5: Run one, heave from the solution and the height measured by station two.

Accuracy of the pitch and roll measurement

As said before, every station not only measured the height, but also an angle i.e. pitch or roll. At station one, and station four the pitch is measured and at station two and station three the roll. Unfortunately these sensors are not calibrated, neither in the earth-bound coordinate system nor in the body-bound coordinate system.

The mean values which can be calculated out of the measurements seem unrealistic. This is based on two considerations. First of all, after the ship is moored, the draught fore and aft are read and the values of the trim

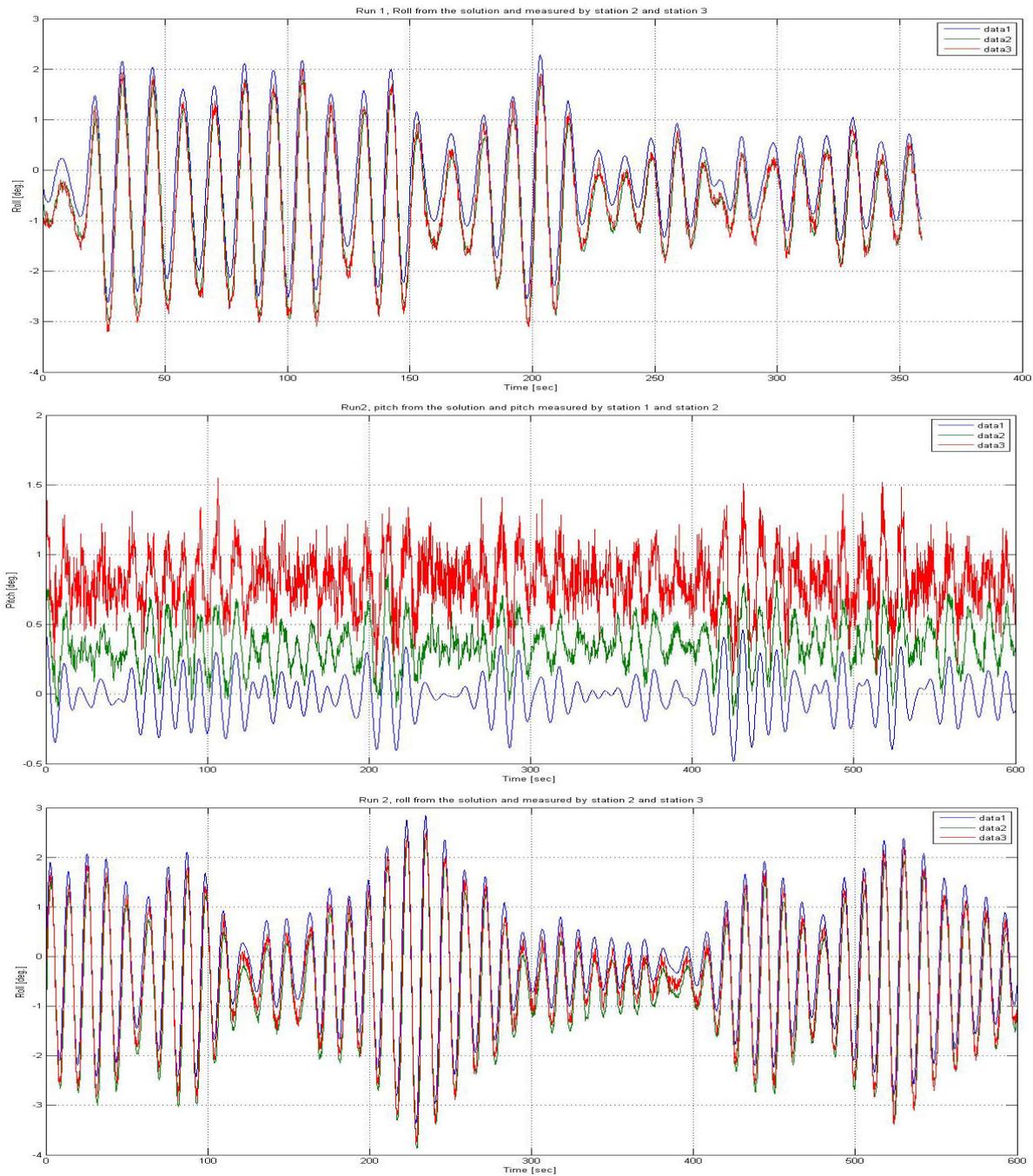


Figure 6: Top: Run one, roll from the solution and measured roll by station two and station three. Middle: Run two, pitch from the solution and the measured pitch by station one and station four. Bottom: Run two, roll from the solution and measured roll by station two and station three.

as a result of these draughts shows us a much lesser value than the mean pitch measured by station one and station four, secondly the leader of the experiments has a different feeling about the values, see Table 4.

Table 4: Mean values (in °) of rotations.

run	Mean roll equations	Mean roll station 2	Mean roll station 3	Mean pitch from equations	Mean pitch station 1	Mean pitch station 4
1	-0.2	-0.64	-0.62	-0.02	0.22	0.73
2	-0.15	-0.62	-0.45	-0.05	0.35	0.81

It is assumed that the list and trim angles that follow out of the solution of the system of equations (Formula 2) are more accurate than those which follow out of the measurements. The difference between the mean values is determined, a closer view is given to a possible phase shift, and last but not least, the ratio between the two amplitudes is determined.

To determine the accuracy of the angle measuring sensor, the time series of a motion, for instance the roll from station two, is given a little offset in time in relation to the reference signal as calculated by Formula 2. The mean for both signals is made zero, and the ratio α between the two signals is determined. An error is defined.

- $\zeta_{ref} = \alpha\zeta_{roll}$; the amplitude of the reference signal calculated with Formula 2 is a factor α multiplied by the amplitude of the measured signal.
- $error = \sum(\zeta_{ref} - \alpha\zeta_{roll})^2$; the total error equals the sum of the square of the values of the reference signal minus factor α multiplied by the value of the measured signal.

Table 5 gives the results for run 1.

Table 5: Comparison between the measured angles and the angles given by the system of equations, run 1.

dt	0.0	-0.2	-0.4	-0.6
Roll2, factor α	0.9706	0.9868	0.9916	0.9849
error	0.0299	0.0094	0.0032	0.0114
Roll3, factor α	0.9447	0.9483	0.9408	
error	0.0141	0.0091	0.0188	
Pitch1, factor α	0.9103	0.9271	0.9303	0.9210
error	0.0018	0.0014	0.0013	0.0016
Pitch4, factor α	0.5393	0.5534	0.5586	0.5572
error	0.0062	0.0058	0.0056	0.0056

Given a closer look at these results we can conclude that the angle

measuring sensors two tenths to four tenths of a second is slow compared to the reference signal. It also appears that the roll measured at station two, has a deviation to the reference signals of less than one percent and that the pitch measured at station four has the greatest deviation. The smallest value of the error is coupled to the maximum value of factor α (closest to one).

It is obvious that the overall error of the pitch is much smaller than that of the roll, this can be explained because the values of the pitch are about a factor five smaller than that of the values of the roll.

To visualize the accuracy of the angle measuring sensor both the signals are plotted. First the roll of the reference signal is plotted together with the measured signal of station two. The measured roll is corrected with the factor α , and the time shift. The same is done for the pitch of station four, both for run one, see Figure 7.

Individual contribution to the draught change

Now we can consider changes of draught by the individual contributions of the ship motions. Starting point is that the ship is upright with a zero trim. Use is made of linearized equations without taking into account the bending of the ship. To look at the total motion of a certain point on the ship, a point on the ship is chosen (115, 22.5, -12.91), i.e. the point lies 115 [m] before the centre of gravity G, 22.5 [m] to starboard and 12.91 [m] beneath G:

$$z_0(115, 22.5, -12.91) = z + -12.91 + 115 \cdot \theta + 22.5 \cdot \phi$$

The average heave again is reduced to zero. Negative results indicate a sinkage so increasing draught. Positive outcomes indicate a rise or decreasing draught.

In Table 6 maxima and minima are shown by the individual motions and the total motion. In Figure 8 this is visualized.

Due to the fact that the phases and periods of the individual motions are not equal to each other, the maximum of the total is not equal to the sum of the maximums of the individual contributions, as is the minimum. You can see in Figure 8 that the contribution to the total motion by the pitch and the roll are of the same order and that the heave is about half of

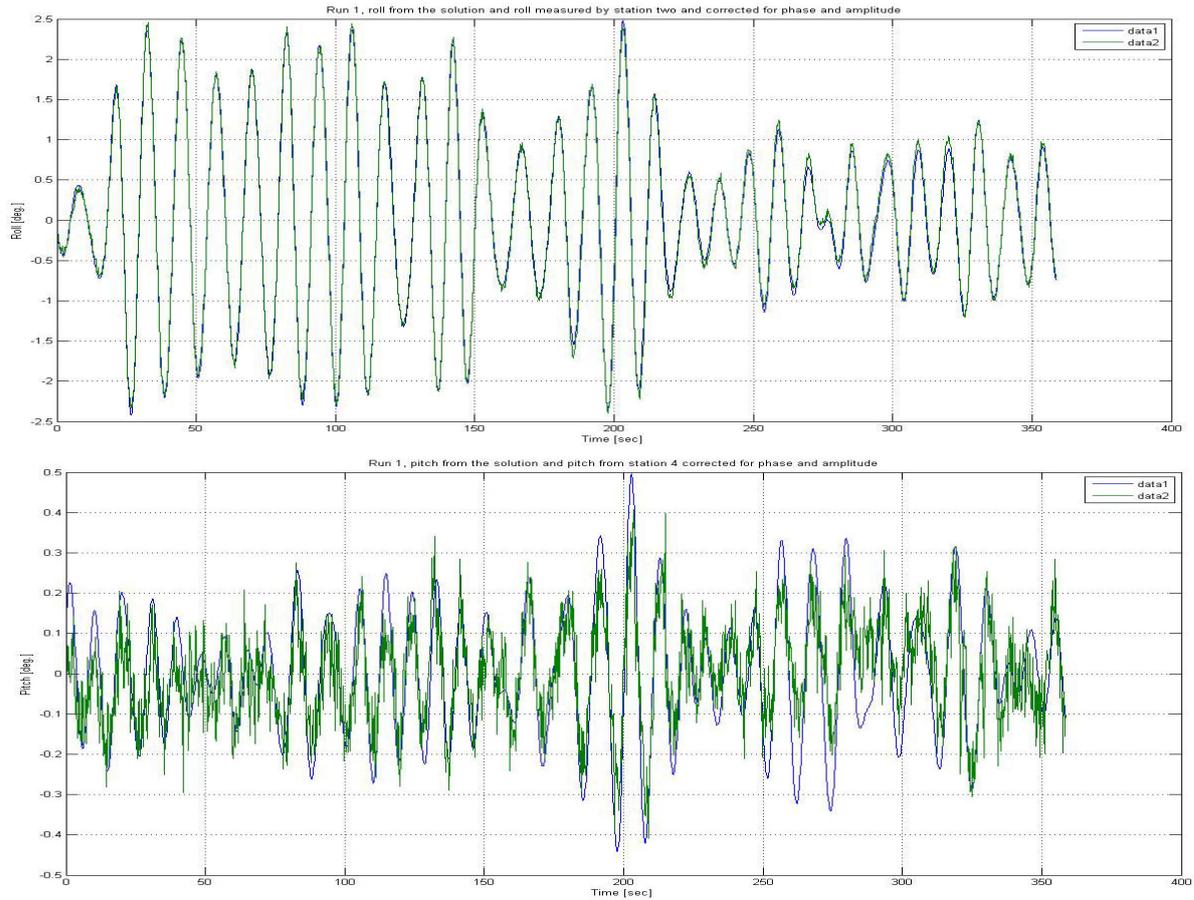


Figure 7: Top: roll from the system of equations compared to the roll measured by station two, the measured values are corrected for phase and amplitude (run one). Bottom: pitch from the system of equations compared to the pitch measured by station four, the measured values are corrected for phase and amplitude (run one).

Table 6: Individual contributions (in [m]) to the changes of draught by the motions of the ship.

		Heave	$115 \cdot \theta$	$22.5 \cdot \phi$	Total
Run 1	max. rise	0.47	1.00	0.97	1.93
Run 1	max. sinkage	-0.51	-0.89	-0.95	-1.74
Run 2	max. rise	0.29	0.94	1.18	1.78
Run 2	max. sinkage	-0.30	-0.96	-1.26	-1.86

it. The light blue coloured line of data 4 represent the signal of the total vertical motion at point (115, 22.5, -12.91). By comparing run one with run two, it is seen that the heave motion decreases and the roll motion increases. The pitching remains more or less the same as the total motion.

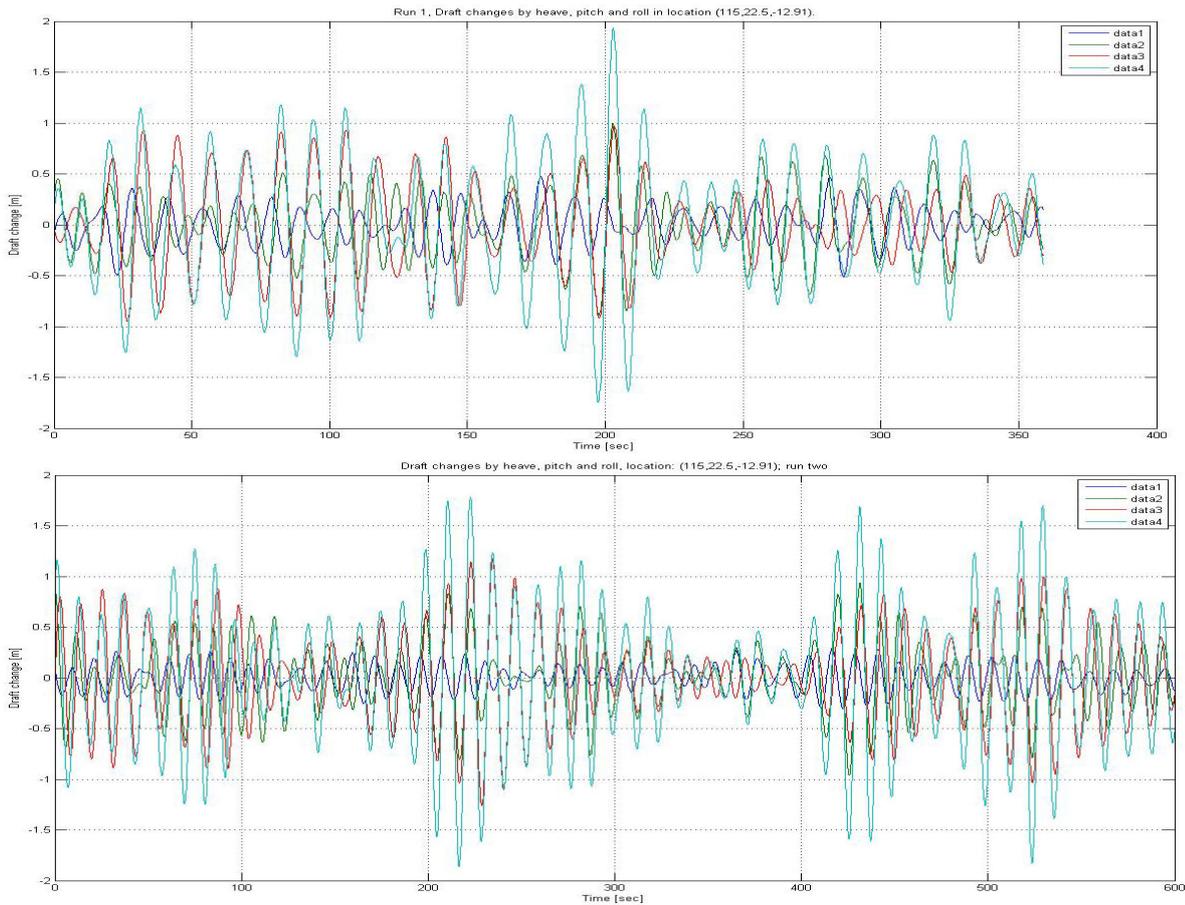


Figure 8: draught changes due ship motions during run one (top) and run two (bottom) respectively.

An analysis in the frequency domain

The question remains whether the findings above can be substantiated based on theoretical considerations. For the roll motion the natural roll frequency of the ship should be considered. For the heave and pitch motions the natural frequency is less important, but what should be considered is the relationship between the apparent wavelength and the length between perpendiculars of the ship.

Roll

If the encounter period between ship and sea is nearby the natural roll frequency of the ship ‘resonance’ can occur. In Figure 9 top, the wave-spectrum respectively the encounter-wave-spectrum of run one is shown. This encounter-spectrum is made based on the direction and speed of the vessel relative to the waves. A complicating factor is that the waterway

depth is neither deep nor shallow. Figure 9 bottom shows the spectra of the roll motions from the solution of the system of equations, run two.

Table 7: The wave data from the original spectrum; H in [m], T in [s].

$H_{significant}$	H_{mean}	$H_{maximum}$	T_2	$T_{visueel}$	T_{peak}	H_{peak}
2.05	1.28	2.61	5.3	5.8	9.1	0.52

Table 7 shows the statistical values of the original spectrum of the sea and the peak period. Table 8 shows the statistical values of the encounter spectra of run one and run two. Table 9 shows the statistical values of the roll spectra from the solution of the system of equations.

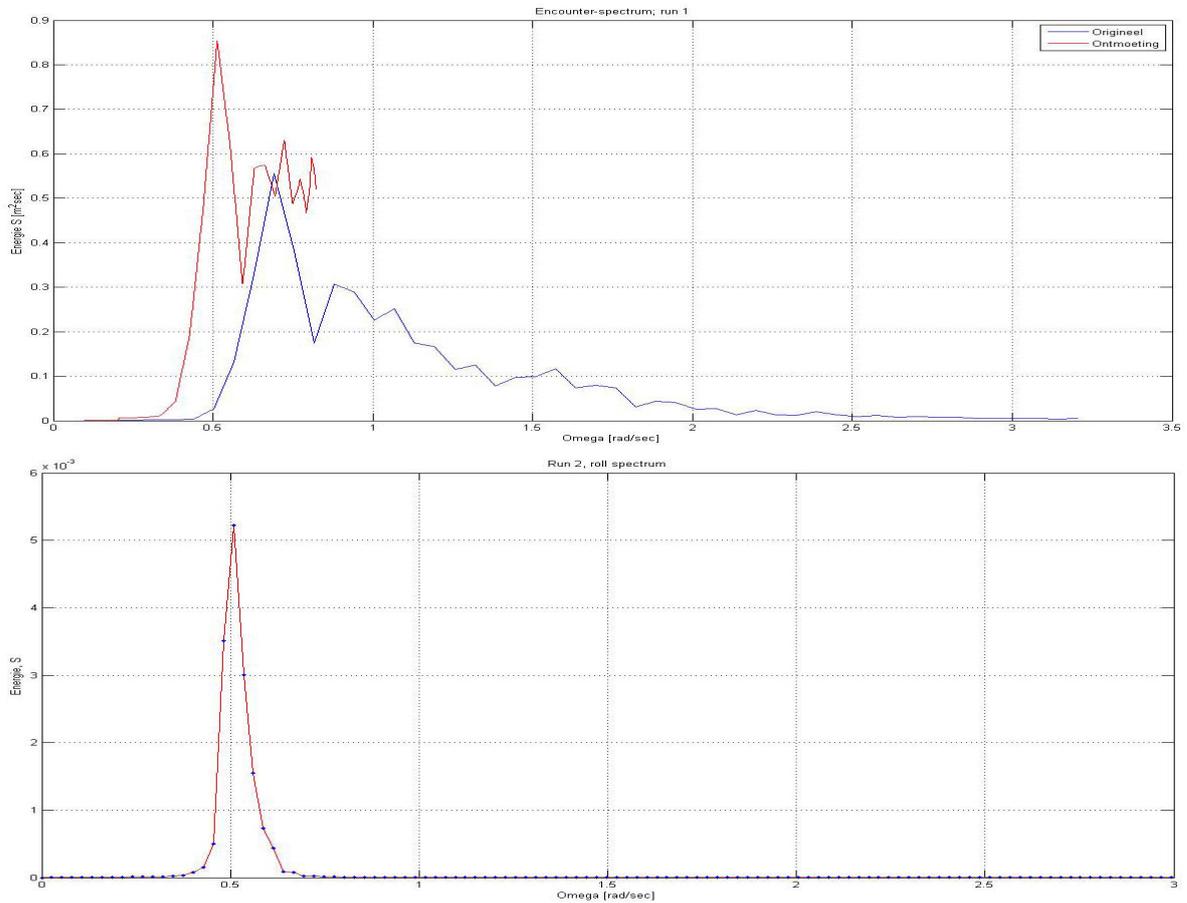


Figure 9: Top: encounter-spectrum, run one. Bottom: roll spectrum, run two.

Notably the spectrum of the roll is very narrow banded. A ship rolls more severely because the encounter frequencies of the waves excite the vessel close to the natural frequency of the roll motion. Here this is the case. The peak period of the encounter spectrum is equal to 12.2 [s] and

14.1 [s], while the peak periods of the motions are close by. Based on the figures the conclusion would be that during the second run the roll motion has to be slightly less intense, because the peak period of the encounter spectrum (14.1 [s]) is somewhat more on a distance relative of the natural period of the roll motion of 12.1 [s].

Table 8: Peak periods of the encounter spectra.

Encounter spectra	ω_{peak} [rad/s]	T_{peak} [s]
Run 1	0.513	12.2
Run 2	0.4454	14.1

Table 9: Statistical values from the roll-spectra.

Roll (solution)	H_{mean} [°]	$H_{maximum}$ [°]	T_2 [s]	T_{peak} [s]	ζ_{peak} [°]
Run 1	2.70	5.52	11.8	12.1	0.97
Run 2	2.93	5.98	12.1	12.4	0.96

This is not the case, the motion is slightly increased, see Table 9. An explanation can be that the form of the stern is extremely flat shaped, the frames run relatively more horizontal. With oblique waves incoming to behind the ship, the ship is more asymmetrically lifted by the hydrostatical pressure and gets a more severe rolling motion.

From a physical and mathematical point of view the motion-spectrum is the result of the encounter-spectrum of the sea multiplied by the square of the transfer-function of the ship theoretically the transfer-function can be determined, but it is highly dependent on the hull form and the distribution of the masses. Software is required to get the transfer-function e.g. the program ‘Seaway’ in use by DUT.

Pitch and Heave

The determining factor for the strength of the pitch and heave motions is the ratio between the length of the ship and the apparent wave length. The ‘resonance’ peak lies near $(L_{ll}/\lambda)^{0.5}$ equal to one. In Figure 9 the original wave-spectrum and the wave-encounter-spectrum is shown. The left part of the spectrum until the first peak can be regarded as a representation of the swell in the wave field. This part is very narrow banded, typical fore swell. It can be deduced that the mean wavelength that belongs to this swell is equal to 100.8 [m]. The apparent wavelength is equal to 175.7 [m]

for run one and 116.4 [m] for run two. The named quotient then is equal to 1.28 and 1.57 respectively, i.e. largely outside any resonance peak in the transfer-function. From Figure 8 you can see that the heave indeed is decreased, but the pitching remains more or less constant. Also this may be related to the shape of the ship.

For all three motions, roll, pitch and heave, from a physical and mathematical point of view, the strength of motions are equal to the square of the transfer-functions of the motions multiplied by the wave-encounter-spectrum. The transfer-function is highly dependent on the hull form and the mass distribution and can be determined on a theoretical basis (computer program).

Determining the motion of the ship with one receiver

The main research question remains, is it possible to determine the motions of the ship, roll, pitch and heave, with one receiver, not taking into account the bending of the ship. Looking at Formula 1 three unknowns must be solved. If θ or ϕ is measured two unknowns are left, z and ϕ or z and θ and with one receiver it is impossible to solve these two unknowns. Maybe a different approach is possible by taking into account the equations for the horizontal motions of the ship. This gives two extra equations but three extra unknowns (x , y and ψ'). One of these unknowns is measured namely the course ψ' , remains four unknowns (x , y , z and θ or x , y , z and ϕ). The following system of equations arises:

$$\begin{aligned}x(x_b(n), y_b(n), z_b(n))' &= x - y_b(n)\psi' - z_b(n)\theta \\y(x_b(n), y_b(n), z_b(n))' &= y + x_b(n)\psi' + z_b(n)\phi \\z_0(x_b(n), y_b(n), z_b(n)) &= z + z_b(n) + x_b(n)\theta + y_b(n)\phi\end{aligned}$$

Unfortunately these are three equations and four unknowns, so the system of equations is unsolvable. It should also take into account that some extra computation steps are required, i.e.: a correction for the current, a transformation of the earth-bound (x_0, y_0) coordinate system to the steadily translating coordinate system (x', y') and determination of the differences between the course and the mean course.

A next step could be that instead of one slave antenna, two slave antennas are used, one in the lateral direction of the ship to measure the

roll and one in the longitudinal direction of the ship to measure the pitch. To determine the last unknown, the heave, principally we can use only the equation for the heights or the system of equations includes the equations for the horizontal motions of the ship. Whatever is preferable will be examined.

Conclusions and recommendations

1. High-frequency ship motions are the motions of the mass centre of gravity G .
2. To determine the total sinkage of a point, the vertical displacement (heave), the roll and the pitch have to be determined.
3. The rotation sensors have to be calibrated relative to the body-bound coordinate system.
4. To estimate whether the motions of the vessel are reinforced or weakened, the wave-spectrum has to be transformed into an encounter-spectrum.
5. To determine whether the ship is responding, the quotient between the encounter period based on the dominant frequency of the waves and the natural period of roll has to be determined.
6. To determine whether the ship is responding, the quotient between the apparent wavelength of the dominant frequency of the waves and the length between perpendiculars of the ship has to be determined.
7. Most channel bound ships on their ways to the port of Rotterdam will find themselves in a sagging loading condition. This ensures that the draughts fore and aft will be less than in the midship. This is beneficial regarding the draught increased by motions, provided the starting point given the midship draught. Note that the draughts generated by stability programs usually do not take into account the bending of the ship.
8. With the current setup of the PPU, namely one receiver with one master and one slave antenna, it is not possible to determine all three components of the motions which affected the draught.

9. The position of the PPU on board seems to affect the angular measurement. Station two, near the mass centre of gravity, yielded the best results, station one and station four by contrast the worst.
10. Based on the results of the angle measuring determination of the roll seems to be more accurate than the pitch.
11. Examine the possibility to equip the PPU with two slave antennas, one in a lateral direction relative to the master and one in a longitudinal direction, making it possible to determine all three components of the motions.
12. Research which preferred approach to use: only the equation for the height, or a system of equations in which also the motions of the ship in the horizontal plane are taken into account.

References

- [1] R. Bouw, “Admittance policy tidal bound ships,” M.Sc. thesis, Delft University of Technology, Delft, November 2005.
- [2] W. van Buuren, “Laterale stroomkracht en langsscheepse scheep-sweerstand,” NLDA, Den Helder, Technical Report, 2009.
- [3] W. van Buuren, “Squat predictie,” NLDA, Den Helder, Technical Report, 2008.
- [4] W. van Buuren, “Ankeren,” NLDA, Den Helder, Technical Report, 2009.
- [5] W. van Buuren, “Windbelasting berekening voor gangbare scheepstypen,” NLDA, Den Helder, Technical Report, 2009.
- [6] A. van Dijk, “Dynamic Submersion Prediction in the IJ-geul,” M.Sc. thesis, University of Nottingham, Nottingham, 2006.
- [7] D.J. Engelbracht, “Kan het on line meten van de hoogfrequente scheepsbewegingen leiden tot een ruimer tijpoort voor geulgebonden schepen?,” NLDA, Den Helder, Technical Report, 2008.
- [8] V. Wolkenfelt, “GPS-Hoogtemetingen ten behoeve van scheepsbewegingen,” B.Sc. thesis, NLDA, Den Helder, 2009.

- [9] W. van Buuren, “Beschrijving van de NMS, Type ADX,” Loodswezen, Hoek van Holland, Technical Report, 2005.
- [10] C.A.J. Tromp, D.J. Engelbracht and T.O.H. Popma, “FYSICS OF FAILURE, Kans op bezwijken en het optreden van vermoeiingscheuren tijdens de levensduur van een schip,” Lecture notes, NLDA, Den Helder, 2005.
- [11] J.M.J. Journée and P. Naaijen, “HYDROMECHANICA 2, Introduction shiphydromechanics,” Lecture notes, Delft University of Technology, Delft, 2006.
- [12] J. Gerritsma, “SCHEEPSBEWEGINGEN,” Lecture notes, Delft University of Technology, Delft, 1989.
- [13] J.A. Battjes, “WINDGOLVEN,” Lecture notes, Delft University of Technology, Delft, 1982.

Cross Wind Loads on Ships and Complex Structures

Max van Hilten†, Dick Engelbracht & Paul Wolkenfelt

In memory of Max van Hilten, former registered Rotterdam maritime pilot and guest lecturer at NLDA, who we lost to cancer in March 2007

Introduction

A few years ago the Maritime Pilots Institute Netherlands (MPIN) was asked to carry out practical research aimed at developing a simple method to predict wind loads on huge constructions. For practical reasons the method should not be time consuming. Some reasons for the initiative were the following.

Quite frequently large vessels and huge complex constructions, varying from offshore constructions to container cranes, enter or leave Dutch ports. Wind has given rise to dangerous situations during manoeuvres with barges or ships carrying high and/or complex structures several times in the past, despite the fact that usually both master and pilot make at least rough calculations as to the wind loads. With the present size of the port of Den Helder and with the increasing size of the lateral area of naval ships such as the landing platform docks (LPD) and the joint logistic support ships (JSS), see Figure 1, these dangerous situations may also arise with ships of the Royal Netherlands Navy.

The second reason was the substantial difference in the results of calculations used by a major shipping company and the approach used by Netherlands pilots. Finally, during this study MPIN was asked to cooperate in a research program concerning the nautical consequences of the development of a new extension to the port of Rotterdam: ‘Maasvlakte-2’.

For this research MARIN (Maritime Research Institute Netherlands) and MSR (Maritime Simulation Institute Rotterdam) developed mathematical ship models of container vessels up to 385 m in length and of considerable height. A very important issue for these vessels was the calculation of the wind loads. This was a reason for MPIN to extend the study to wind loads on these large vessels as well.



Figure 1: Artist's impression of the Navy's new joint logistic support ship (JSS).

The most critical part of a passage often is the moment of the manoeuvre when the speed is nearly zero and the largest lateral area of the object concerned is exposed to the wind. At this moment enough propulsion power, either own ship's propulsion, tugboat's propulsion, or a combination of these, has to be available to, at least, stand the wind. Mostly the problem becomes less serious once the ship or barge increases speed. In this case a change of heading in order to correct for the set caused by the wind solves the problem as long as the fairway allows the required swept path. In this scenario low ship speeds should be taken into account: although increase of speed leads to hydrodynamical and rudder forces opposing the forces being exerted by wind, it also results in less effect from side thrusters and assisting tugs. These effects, amongst others, have to be taken into due consideration in daily practice. This article however focuses on the calculation of forces being exerted by the wind. The total power required to handle the ship safely will always be more than the calculated crosswind loads but may be based on these values.

For most floating objects however, as far as the authors know, the influence of air temperature and atmospheric pressure is not usually taken into account. Regarding normal ships the same goes for the use of a correct shape coefficient as well as a vertical wind profile. Variation of one of these parameters has its own specific influence (linear or quadratic). A combination of differences from standard values used may lead to results differing substantially from the wind loads being exerted on the object in reality. Especially the shape coefficient, the height of the object concerned and the air temperature play an important role. As stated above, several data are important in calculating the wind loads. The fact that not all of these parameters are used correctly in practice, if used at all, may result in unexpected wind loads, in some cases to a lesser and in some cases to a greater extent. Especially the latter, a greater load, might result in accidents. Some of the parameters are not easy to estimate, for instance the shape coefficient for the object concerned. For some of the others this is easier, the air temperature for instance. Even though a wrong assumption of the shape coefficient may contribute to wrong results for the most part, it is preferable (in the authors' opinion) to work as accurately as possible with regard to the other data. At least this avoids a wrong calculation result caused by the accumulation of a number of wrong assumptions.

Up until recently a lot of research has been carried out into wind loads on structures and ships. The scientists who carried out research in this area were, to name but a few, Davenport, Isherwood, Aage, Blendermann and Kareem. A lot of research was carried out regarding wind loads on offshore constructions as well, including wind tunnel tests. Furthermore, Engineering Sciences Data Unit (ESDU UK) supplies computer programs to calculate wind loads on structures. However, to use these programs one needs a sound theoretical background in engineering and also a lot of time to enter the necessary input.

Problem Definition and Basic Formulae

In trying to develop an approach to calculate the wind loads being exerted on ships or complex structures by cross winds the result should fulfil the following criteria:

1. Take into account, wherever possible, relevant parameters;

2. Easy to use on board for masters, navigation officers and pilots in daily practice;
3. Not be time consuming.

Fulfilling these preconditions requires analysis of the relevant components in determining wind loads and is only possible by the introduction of some tailor made software.

The forces and moments of force required in an equation of motion can be determined as a function of relative wind velocity, relative wind direction and 3-dimensional shape of the ship. However, the calculation models used are not always the same. For the components in the equation of motion representing the forces and moments of force caused by wind, X_{wind} , Y_{wind} and N_{wind} , we can use the basic form of the formula for forces being exerted on an object by a flow

$$X_{wind} = C_x \rho V^2 A_F / 2, \quad (1)$$

$$Y_{wind} = C_y \rho V^2 A_L / 2, \quad (2)$$

$$N_{wind} = C_N \rho V^2 A_L L_{oa} / 2, \quad (3)$$

with C the shape coefficients, ρ the air density, V the wind velocity, $A_{F,L}$ frontal and lateral projected areas and L_{oa} the length over all. The values of the shape coefficients depend on the three dimensional shape of the ship and the angle of attack of the relative wind. Since the calculation models used are not always the same, the values of shape coefficients for different calculation models may differ substantially. Therefore indiscriminate comparison of coefficient values from tests or research may result in wrong conclusions. A difference of calculation models comes to light at comparison of the theoretical approaches by Isherwood [1] and Blendermann [2,3]. The starting point of both Isherwood and Blendermann is the use of wind tunnel tests with a uniform flow. They take into account a boundary layer, but neglect fluctuations in speed and/or direction.

Most practical approaches are based on a simple form of the formula for dynamic pressure. The following examples of formulas often used in daily practice are of particular interest:

$$Y_{wind} = F_t = 0.075V^2 \cdot A_L / 1000 \text{ [ton force]},$$

$$Y_{wind} = F_t = 0.052V^2 \cdot A_L / 1000 \text{ [ton force]}.$$

For practical use in the offshore business most manuals advise calculating the loads for separate characteristic parts of the construction, followed by addition of the loads on these separate parts. They advise using height coefficients: factors for separate height intervals. This means they take into account the effect of a vertical wind profile for the most part. Shape coefficients for several 3-dimensional shapes of separate parts of the construction are given. Usually the air mass density is considered to be constant. A disadvantage of this method is that separate calculations have to be made for a number of height intervals for high parts of the construction. Most manuals advise adding to the load a certain percentage for small parts of the construction not included in the calculations (e.g. 20%).

A more Detailed Discussion on Parameters of the Basic Formula

The most elementary form of the formula for loads caused by wind was shown in (1) and (2). Here we will discuss parameters of the formula in more detail. For each of the parameters we will

- consider the theoretical background for determination of its value,
- refer to existing theoretical and practical approaches and,
- refer to the implementation in the computer program.

Wind velocity

Since the wind velocity varies with the height, a vertical wind profile has to be determined. Once this profile is laid down it is possible to calculate the wind velocities for any height above sea level. This is required for the determination of the variation (with the height) of the dynamic pressure. In this context it is important to be aware of the fact that wind velocities observed on board ships usually are those measured at the height of the sensor. The heights of wind sensors may differ considerably.

Meteorologists [4] usually lay down the vertical wind profile by means of a logarithmic function

$$\frac{V_{h_1}}{V_{h_2}} = \frac{\ln(h_1/z_0)}{\ln(h_2/z_0)}, \quad (4)$$

with V_h the wind velocity at height h . Further, the difference of wind profiles above a certain terrain (e.g. sea or land) is determined by the choice of the roughness length (z_0). The value of z_0 usually varies from 0.0002 [m] (open sea, absolutely flat surface) to more than 2 [m] (city centre). For open sea with waves, PIANC mentions a value of 0.004 in [5].

The vertical wind profile is sometimes also given by the power law

$$V_{h_1} = V_{h_2} \cdot (h_1/h_2)^\alpha. \quad (5)$$

For this approach the difference of wind profiles above a certain terrain is determined by the choice of the coefficient α . In [4] the power law is justified during neutral stability of the atmosphere, when α is determined by $\alpha \approx 1/\ln(\sqrt{h_1 \cdot h_2}/z_0)$.

From the above it is obvious, that determining α or z_0 correctly is important for the computations. Also, when wind velocities observed from wind sensors are used, the height of the sensor has to be taken into account because the observed wind velocities may differ substantially from those at the standard height (10 m), for which weather forecasts give the expected wind velocity.

The difference in wind loads on a rectangular shape, between model with and without vertical wind profiles should not be underestimated. Here, we consider the rectangular shape with a vertical wind profile defined by a power law. It turns out that a critical height exists, for which, depending on α , velocities of vertical wind profile flows and height independent wind velocity flows are the same. Using a vertical wind profile objects higher than this critical height will be subject to higher wind loads than calculated without a vertical wind profile. The opposite is true for objects lower than this critical height. The critical height h_C can be computed by

$$h_C = h_1 \cdot (2\alpha + 1)^{1/2\alpha}.$$

For the construction of offshore plants different methods are used in practice for wind profiles. Det Norske Veritas (DNV) uses a logarithmic approach [6,7] for the vertical wind profile as a function of height in combination with the averaging time interval, given by

$$\bar{V}(t, z) = \bar{V}(t_r, z_r) \cdot (1 + 0.137 \ln(z/z_r) - 0.047 \ln(t/t_r)). \quad (6)$$

Here z_r represents the reference height (10 m), t_r the reference averaging time interval (600 s) and t, z are the variable time and heights. We observe that the roughness length is not included in this formula, which implies that this formula only applies to situations at sea. In the next section we will have another look at this approach, when the phenomenon gust factor is introduced.

An approach by Hancox [8,9], used in practice in the offshore industry, is similar to the method mentioned above. There, the influence of the vertical wind profile is taken into account by the use of different height coefficients for separate height intervals. However, it seems that the effects of a vertical wind profile are not taken into account explicitly in formula used in daily practice for determination of wind loads on regular ships.

Our developed software takes into account the influence of height dependent wind velocities by including the defined vertical wind profile in the process of integration with respect to height, according to

$$Y_{wind} = C_y \rho / 2 \sum_{n=1}^{N_s} L_n \int_0^{H_n} V^2(z) dz, \quad (7)$$

where N_s sections of a ship with lengths L_n and heights H_n are used. In the case of different heights over the ship's length separate length sections can be used. The validity of using one shape coefficient for separate length sections will be discussed later.

Furthermore, we use in our software a 1/10 power law profile for regular ships for situations at sea. For 'not exposed' harbour situations, a logarithmic profile (4) with a roughness length of 0.2 m was chosen. The reason for choosing this approach is the use of a height dependent gust factor which is especially applicable to a logarithmic profile. This will be explained in the next chapter.

For high (offshore) structures at sea, formula (6) has been implemented in the software used. This method combines the use of a logarithmic vertical wind profile with the use of a already mentioned gust factor. Following this procedure the introduction of an error due to the use of height intervals is avoided.

Gust factor (GF)

An important issue regarding the wind velocity is the phenomenon gust factor (GF): a factor to be applied to average wind velocities for longer periods of time to find the maximum average wind velocity for shorter periods of time. Introduction of a gust factor extends the basic formula to

$$F = C_y \rho A (GF \cdot V)^2 / 2.$$

The wind velocity at a defined height is not constant in time. Variations in velocity as well as the duration thereof depend, amongst other things, on the height. Close to the water level so-called wake effects of surrounding obstacles or a rough sea will cause these variations. In this scenario the roughness length (z_0) is important. At higher levels variations due to higher wind velocities usually last longer and are mainly caused by instability of the boundary layer. With respect to wind loads on ships, given a defined height and geographical situation, the question arises which wind velocity, or rather the mean wind velocity of which duration, should be used in determining these loads.

For wind forecasts the World Meteorological Organization (WMO) advises the use of durations from 10 to 30 minutes. Usually the wind velocities referred to in weather forecasts are 10-minute mean values. In order to determine a mean velocity for shorter durations the 10-minute mean value is multiplied by the gust factor. When using actual readings of a wind velocity sensor it depends on the time interval at which the readings were observed. For longer observations the observer can take the peak velocities into account. Then the necessity for using a gust factor decreases.

In [10] Kareem defined the phenomenon gust factor by

$$\begin{aligned} GF(t, z) &= \frac{\bar{V}(t, z)}{\bar{V}(1\text{hr}, z)} \\ \bar{V}(t, z) &= \bar{V}(1\text{hr}, z) + g(t) \cdot \sigma(z), \end{aligned} \quad (8)$$

with $\sigma(z)$ the standard deviation at height z and $g(t)$ a peak factor. We observe that in (8) a reference time interval of 1 hour is used. In some cases other intervals are used, e.g. 10 minutes.

The standard deviation is directly proportional to the turbulence intensity. Given a wind velocity registration, the geographical situation and the

height, the values of both $g(t)$ and $\sigma(z)/\bar{V}(1\text{hr}, z)$ can be determined. The shorter the period of time, the higher the gust factor will be. A different interpretation of (8) with a reference time interval of 10 minutes yields

$$GF(10\text{min} \rightarrow 1\text{min}, z) = 1 + g(10\text{min} \rightarrow 1\text{min}, z) \cdot \frac{\sigma(z)}{\bar{V}(10\text{min}, z)}, \quad (9)$$

where $g(10\text{min} \rightarrow 1\text{min}, z)$ and $I(z) = \sigma(z)/\bar{V}(10\text{min}, z)$ represent the peak factor and the turbulence intensity respectively.

According to Wieringa and Rijkoord [4] this turbulence intensity may be estimated by

$$I(z) = 1/\ln(z/z_0). \quad (10)$$

Both, relations (9) and (10), clearly show the dependence of GF on the height z . From (10) the dependence of the roughness length z_0 can be seen. The value of the peak factor g depends on the height as well but that influence is of minor importance.

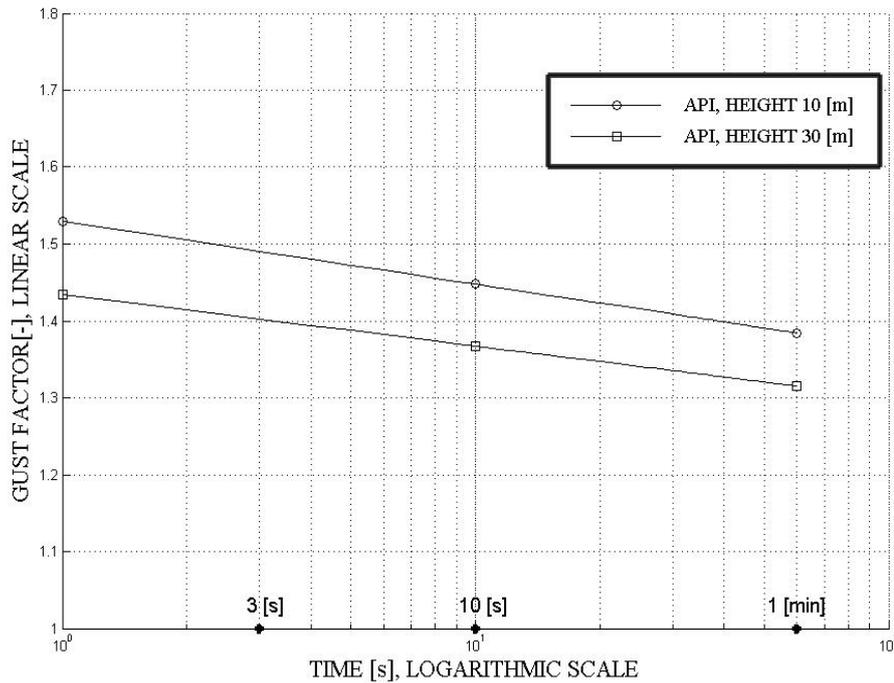


Figure 2: Comparing gust factor values of API for heights of 10 m and 30 m

As stated above, the value of a gust factor depends, amongst other things, on the height. This is shown clearly in Figure 2: the values used by the American Petroleum Institute (API) are shown for two different

heights. Values of gust factors as well as calculation methods mentioned in literature are not exactly the same, as most of these values are valid for the defined wind profile coefficients or logarithmic relationships used. Therefore gust factor values of e.g. API, DNV and PIANC should not be compared reciprocally!

Another important issue in the use of a gust factor is the response of a vessel to gusts. In particular the gust duration, which is relevant for the vessel and manoeuvre concerned, are of great importance. PIANC [5] states that intervals of more than 1 minute may be considered relevant for large vessels. This is an arbitrary value, since two vessels with exactly the same lateral area (for instance a car carrier and a loaded tanker), will not respond in the same way to the same gust because of their different displacement and added mass.

The gust wavelength is the product of the duration of the gust and the mean wind velocity in that gust [4]. In this way gusts have physical dimensions. This is important since it means that objects are often not completely exposed to the gust concerned.

There are several ways to take the influence of gustiness into account. The most uncomplicated approach is the multiplication of the wind velocity with a constant that only depends on the averaging time. This method was proposed by PIANC [5]. Table 1 gives an overview of the proposed GF values.

Table 1: Gust factors depending on averaging time durations [5].

Duration	GF
3 seconds mean	1.56
10 seconds mean	1.48
1 minute mean	1.28
10 minutes mean	1.12
30 minutes mean	1.05
1 hour mean	1.00

In many circumstances these values can be of great help. However, in reality the value of a gust factor depends on the height above sea level or a landmass as well as on the local roughness of the terrain. An example of an existing approach used in the offshore industry is the method of Det Norske Veritas (DNV) [6,7], based on formula (6).

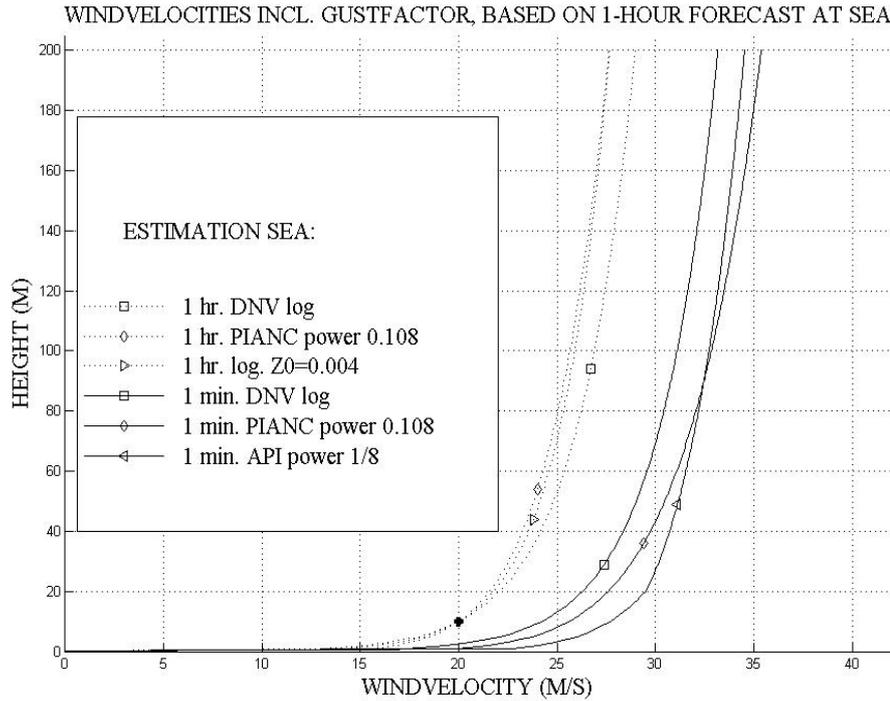


Figure 3: Comparison of some mean wind velocities at sea: 1-hour mean and 1-minute mean values.

Figure 3 shows a comparison of wind velocities according to different approaches discussed in this article: some 1-hour mean values ($GF = 1$) and some 1-minute mean values for situations at sea. In this figure ‘Log’ means defined by a logarithmic function and ‘power’ means defined by a power law.

Usually a gust factor is not applied in daily practice, when determining wind loads for normal ships. The same goes for the offshore industry. It is not known by the authors whether or not effects of gusts are included, in a general way, in safety margins used in daily practice.

Based on this study we have implemented two possible choices for GF :

- no gust factor at all ($GF = 1$), (probably used in most circumstances),
- a gust factor for a 1-minute mean wind velocity (based on [5]).

When the use of a gust factor is chosen the calculation methods are different:

- for ‘normal’ ships at sea or in exposed harbours, as proposed in [5] by PIANC,
- for complex structures at sea or in exposed harbours, the use of a height dependent gust factor in combination with the corresponding wind profile as proposed by DNV [7],
- for ‘normal’ ships and complex structures in ‘not exposed harbours’, the use of a height dependent gust factor as proposed by the Royal Netherlands Meteorological Institute (KNMI).

All of these calculations are based on weather forecasts providing a 10-minute mean wind velocity. Since the gust factor is height dependent, it is included in the process of numerical integration. As mentioned before the value of a gust factor depends, amongst others, on the roughness length (see formula (10)). Neither PIANC, nor DNV takes this into account. For this reason we searched for a method to determine a gust factor which can be used in ‘not exposed’ harbours where a different vertical wind profile is valid. From KNMI as well as from [11] we received valuable information about the value of the so-called peak factor (9) applicable in these situations. This information is especially applicable to vertical wind profiles defined by logarithmic functions. For this reason the software uses this logarithmic function.

Figure 4 shows a comparison of wind velocities according to different approaches discussed in this article. As in Figure 3 we show a comparison of wind velocities with and without the use of a gust factor, above land with a roughness length of 0.2 m. This has been depicted in Figure 4. In order to be able to compare Figures 3 and 4, the wind velocity of 20 m/s at sea is translated into an estimated velocity for the geographical position concerned, which is not straightforward.

In [4] a formula is presented to estimate the velocity at geographical positions, given the velocity at another position. First the expected velocity at reference height z at sea is translated into a velocity at the so-called blending height (for the blending height 60 m is chosen). When the distance

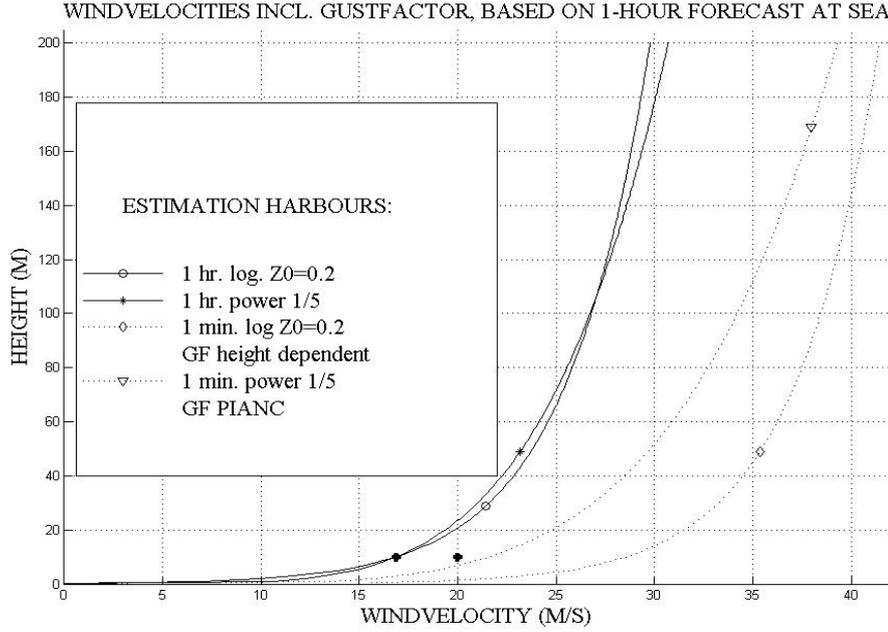


Figure 4: Comparison of some mean wind velocities ashore: 1-hour mean and 1-minute mean values.

between the two positions is not too long (about 5000 m), the wind velocity at blending height is supposed to be the same for both positions. Then the wind velocity at blending height above the ‘not exposed’ harbour is translated into a velocity at the local reference height according to the local vertical wind profile, see Equation (11):

$$V_{\text{land}} = V_{\text{sea}} \cdot \left(\frac{\ln(60/z_{0,\text{sea}}) \cdot \ln(z_{\text{land}}/z_{0,\text{land}})}{\ln(60/z_{0,\text{land}}) \cdot \ln(z_{\text{sea}}/z_{0,\text{sea}})} \right). \quad (11)$$

For a height of 10 m, with roughness lengths of 0.004 m (sea) and 0.2 m (land), this results into $V_{\text{land}} \approx 0.843V_{\text{sea}}$. Formula (11) was used when compiling Figure 4, i.e., to estimate the wind velocities at 10 m height above land. For comparing Figures 3 and 4 two stars are plotted at 10 m height: one at velocity 20 m/s and one at about 16.9 m/s. The first one represents the wind velocity at sea at 10 [m], the second one the estimated wind velocity in a nearby ‘not exposed’ harbour. From a practical point of view the remarkable increase of wind velocity caused by the use of a gust factor for a vertical wind profile above land, especially at lower heights, is a very important conclusion.

One may argue, that the gust factors used are arbitrary values, but at least they decrease the risk of calculating wind loads based on weather forecasts too low. The chosen values result in a considerable increase of load since the latter is determined by the square of the velocity. In most cases however, in practice the use of a gust factor will be omitted since wind velocities will be observed on local sensors and for a longer period of time.

Shape coefficient (C)

The value of the shape coefficient depends for the most part on the three dimensional shape of the object concerned and on the angle of attack of the relative wind approach. Since there is a great variety in three dimensional shapes of ships, the only realistic way to determine shape coefficients is by means of experiments like full-scale tests or wind tunnel tests using scale models. In this way the effect of mutual interference of nearby individual parts of the complete object is included in the coefficient. A disadvantage of full-scale tests is that laboratory-like circumstances are hard to create: disruptions of stable wind velocity situations are nearly always present. Gathering experimental data is time consuming and expensive, therefore these data are only available for a restricted number of ship types.

We will define the angle of attack for bow wind as 0° . In the context of this article we will use the value of the shape coefficient for relative crosswind for the most part: in that case the relative angle is $\pm 90^\circ$. The value of the shape coefficient also depends on the computational model used. Therefore, the values of a shape coefficient for these methods may differ substantially.

In [5], PIANC proposes a rough estimate of the shape coefficient for ‘normal’ ships of 1.1 for cross wind for general use in combination with formula (2). Although not mentioned explicitly, we assume in this case the wind velocity at a height of 10m, where velocities at other heights can be derived by formula (5). For head wind an estimated value of 0.8 is mentioned.

The Oil Companies International Marine Forum (OCIMF) [12,13] has published coefficients for VLCCs and large liquefied gas carriers. The data for VLCCs is based on wind tunnel tests conducted at the University of Michigan in 1975. The data for LNG carriers are based upon tests carried

out in the industry. Both publications suggest that the values mentioned therein might be rather conservative. The reason mentioned is that the coefficients must be applicable to a general range of vessel forms. Also their calculation method is based on formula (2) using the wind velocity at a height of 10m and a modification of the velocity by formula (5) using a height exponent $\alpha = 1/7$. The coefficients we obtained from there differ substantially from a study by Blendermann [2]. The difference in wind loads shows a ratio of about 1.4. It is difficult to indicate the reason for this but it should be kept in mind that, although LNG carriers with spherical tanks were used for both tests, the shapes of the vessels show significant differences in dimensionless ratios like e.g. L_{oa}/B (Length over all divided by Breadth).

In the offshore industry, it seems to be common practice to calculate the loads for separate characteristic parts of the complete construction followed by addition of the loads on these separate parts. Both, DNV [7] and Hancox [8,9], use different shape coefficients for these separate parts. Some values mentioned in [7] are given in Table 2.

Table 2: Shape and shape coefficient as in [7].

Shape	Shape coefficient (C)
Spherical	0.4
Cylindrical	0.5
Large flat surface (hull, deckhouse)	1.0
Clustered deckhouses	1.1
Wires	1.2
Drilling derrick	1.25
Isolated shapes (crane, beam)	1.5

Using a vertical wind profile which is defined by height (z) combined with a certain averaging time interval (t) and different shape coefficients for individual parts of the construction leads to:

$$F_{\text{wind}} = \sum_{n=1}^{N_s} C_{s,n} \rho L_n / 2 \int_{h_{1,n}}^{h_{2,n}} V^2(t, z) dz. \quad (12)$$

For ‘normal’ ships the derived software offers the user different levels of accuracy. Besides, it offers the use of one or more length sections: sections of different heights and reaching from sea level. Furthermore, the following methods of application of a shape coefficient are offered:

1. A method requiring a minimum of time for calculating wind loads makes use of different preset values of the components used in the calculations; in this case a value of 1.0 is used for the shape coefficient.
2. More advanced methods offer:
 - a) A value of the shape coefficient of 1.1 (\approx PIANC),
 - b) A value of the shape coefficient based on a mean value of comparable types of ships [2], converted to values to be used in formula (7).
 - c) Values of the shape coefficients (for different angles of wind approach) for specific ships mentioned in [2].

In method 2c the software uses the calculation method proposed by Blendermann. In this case the projected lateral area is represented by one rectangular reference area. A distribution of the cross wind load to forces on bow and stern is available for several angles of attack of the relative wind. The longitudinal wind load is shown as well. When the type of vessel concerned is not available in [2], the vessel may be divided into separate length sections followed by application of method 2a or 2b offering a distribution of forces on bow and stern as well.

We observe that the distribution of forces at bow and stern is impossible in case of use of only one rectangular section. When using an estimated coefficient for more than one length section, the same coefficient is used for all sections. This is not fully correct since separate length sections will usually not be subject to exactly the same shape coefficient. The sum of the forces will still be correct. The results of the calculation of the distribution of forces on bow and stern however, may be affected. A comparison of calculation results of methods 2b and 2c for some of the vessels available in [2] show promising results: differences in forces at bow and stern of at most 4%.

For complex (offshore-) structures the computer program uses the values of shape factors mentioned by DNV [7]. The outlines of the calculation method proposed by DNV are applied: calculation of loads for separate characteristic sections followed by addition of these separate loads. The selection of sections may be based on different heights, different shape factors and/or different contours of the projected lateral area. Since the effect of mutual interference of nearby individual parts of the complete object is neglected, this method gives a rough indication only.

Comparisons of results of full-scale measurements, wind tunnel tests and the method mentioned above show that wind loads found by this method would usually be too large [14]. For use in practice the availability of a safety margin is an advantage since it reduces the risks. A disadvantage is the extra charges for tugs. On the other hand these structures do not call at ports frequently.

Air mass density (ρ)

The density of the air flowing around the vessel is a function of the height and depends on the air temperature at sea level (T_0), the rate of change of temperature with respect to the height (a), the atmospheric pressure at sea level (P_0) and the humidity.

The humidity factor will not be addressed in this article since this item is beyond the original scope of this research. Using the standard atmosphere as a starting point, the effect of all of the other factors mentioned above is included in one single formula [15,16] given by

$$\rho(z) = \frac{P_0}{RT_0} (1 + az/T_0)^{-(1+g_0/aR)},$$

with R the gas constant (in dry air) and g_0 the gravity constant at mean sea level.

The rate of change of atmospheric pressure with the height plays a role too and depends on the temperature. It plays a minor role but greater than the role of the rate of change of temperature (with the height) itself.

Using a realistic temperature range and atmospheric pressure range as starting point, the possible differences of temperature have a far greater influence on the air mass density than the possible differences of atmospheric pressure. The influence of these factors is often neglected in daily practice. A constant value often used is 1.225 kg/m^3 . Most classification societies neglect this influence, which is rather surprising, since the temperature factor especially may contribute to a relatively high percentage of wind loads.

All of these contributions are included in the computer program where $\rho(z)$ is included in the integrand. The necessary input data are the temperature at sea level, the atmospheric pressure at sea level and the height interval.

Projected lateral area (A_L)

In many cases, especially when constructions are considered as a combination of separate sections, each with its own contour and shape coefficient, it is useful to express the width of the lateral area concerned as a function of the height, i.e., $dA = L(z) dz$. In our software we only use this method for complex constructions, as for ‘normal’ ships a concatenation of rectangles can be used. A variety of contours of the projected lateral area is shown in the menu driven program, offering the user choices. In accordance with methods used in books on practical oil field seamanship 20% is added to the projected lateral areas concerned for small parts, which are not included in these areas. Since the addition of 20% is represented by a constant factor of 1.2, it does not influence the process of integration and is applied to the total of wind loads.

Resulting Formula and the Software

As explained in the previous subsections, ρ , GF , V and often also L , are functions of the height. The formula that combines all the previous considerations, taking (12) as a starting point, may now be written as

$$F = C_y/2 \int_{h_l}^{h_u} \rho(z)(GF(z) \cdot V(z))^2 L(z) dz.$$

In the derived computer program this force is computed by means of numerical integration.

In order to offer masters, navigation officers and pilots the possibility of estimating the wind loads of cross winds for a great variety of ships / complex objects in daily practice, MPIN and NLDA developed a menu supported computer program. One of the criteria was that the method should not be time consuming. The software offers the choice between calculation of wind loads on ships and the calculation of wind loads on complex constructions.

The effects of air density, a vertical wind profile and the possibility to use a gust factor (mostly height dependent) are included in both methods. Depending on the required accuracy different paths can be followed. Report [17] gives more background on the structure of the program and the computing methods used. The program is also used within the Royal Netherlands Navy [18].

Conclusions

The subject of determining wind loads on ships appears to be very complex. The most common used methods to determine wind loads in daily practice are often based on simplified models. By doing this too many risks may be taken: risks for damage, or worse, for human life and the environment. In most cases the formula used by the Dutch pilots shows reasonable results. For low or extra high objects however substantial differences may appear.

We observe, that there are only two ways to determine correct wind coefficients for objects like ships: full scale trials and wind tunnel tests. Furthermore, despite the existence of scientific approaches it remains very difficult to determine the actual forces being exerted by wind on a ship to a high level of accuracy. Taking account of the influence of air temperature, atmospheric pressure as well as the rate of change in both, on the air density, a higher level of accuracy is reached. Also, taking into account a realistic vertical wind profile instead of using no wind profile at all, leads to a higher level of accuracy.

In daily practice the influence of gusts may cause serious problems: a good reason to take into account a value for the gust factor. Especially when wind loads are calculated based on wind forecasts. Gusts depend to a great extent on local circumstances and appear to be an extremely complex subject. Existing gust models require an extensive study for each particular geographical situation and wind direction. Within the derived software choices had to be made as to the application of these models. Since 'easy to use in daily practice' is one of the preconditions in this study, simplification of gust calculations could not be avoided. From a practical point of view the remarkable increase of wind velocity caused by the use of a height dependent gust factor for logarithmic profile with higher roughness lengths, especially at lower heights, is a very important conclusion.

Attention should be paid to the fact that values of wind loads depend to a certain extent on the actual vertical wind profile. This can be taken into account either in a way as proposed by Blendermann [2,3] or leading to the same result, by adjustment of the transverse force coefficient. As a result of this, indiscriminate comparison of coefficient values from tests or research may result into wrong conclusions.

Recommendations

In the authors' opinion further study should address the following aspects:

- Better insight into the effect of wind on large high-sided vessels, in particular the large container vessels expected to come. These vessels with their containers high-stacked on deck have such a height (and width) that the wind coefficients used now may not reflect reality. With respect to this aspect wind tunnel tests might be very useful.
- The same applies to other ships and floating structures, of which construction differs significantly from what has been seen as normal up until now. An example is the type of cruise vessels with cabins located in such a way that a rough 'indented' surface of shipsides is created with regard to wind.
- Better insight in appropriate wind spectra, in particular with respect to wind gusts, and their influence on ships and floating structures. A question to answer is what effect the gust wave length has on the forces being exerted by the wind on a ship, taking into account type, size, displacement and hydro dynamical characteristics of the vessel concerned.
- Extension of computer programs to calculate wind loads, based on the results of the previous recommendation, by offering the use of different gust durations.
- The possibility of implementation in the computer program of the probabilistic and spectral modelling approach as published by Blendermann [19].
- It might be useful to study whether the technique of computational fluid dynamics can be used to determine wind coefficients for ships and offshore structures.

References

- [1] R.M. Isherwood, “Wind resistance of merchant ships,” in *Proceedings RINA*, 1973.
- [2] W. Blendermann, “Wind Loading of Ships - Collected Data from Wind Tunnel Tests in Uniform Flow,” Hamburg University, Hamburg, Report Nr. 574, 1996.
- [3] W. Blendermann, “Estimation of wind loads on ships in wind with a strong gradient,” in *Proceedings OMAE 95*, Volume I-A (Offshore Technology), 1995.
- [4] J. Wieringa and P.J. Rijkoort, *Windklimaat van Nederland*. Royal Netherlands Meteorological Institute, Staatsuitgeverij, 1983.
- [5] PIANC, “Report of working group 1 (International Commission For The Reception Of Large Ships),” annex to Bulletin No. 32 Vol. I, PIANC, 1979.
- [6] Det Norske Veritas, *Classification notes*, No. 30.5, 2000.
- [7] Det Norske Veritas, *Offshore Standard DNV-OS-C301*, 2001.
- [8] M. Hancox, *Towing. Oil Field Seamanship Vol. 4*. Oil field publications Ltd., 1994.
- [9] M. Hancox, *Jackup Moving. Oil Field Seamanship Vol. 2*. Oil field publications Ltd., 1994.
- [10] J.D. Riera and A.G. Davenport (eds.), *Wind Effects on Buildings and Structures*. Rotterdam: Balkema, 1998.
- [11] J.W. Verkaik, “Evaluation of two gustiness models for exposure correction calculations,” *Journal of Applied Meteorology*, vol. 39 no. 9, 2000.
- [12] Oil Companies International Marine Forum, *Prediction of Wind and Current Loads on VLCC's*, 2nd. edition. Witherby & Co Ltd., 1994.
- [13] Oil Companies International Marine Forum, *Prediction of Wind Loads on Large Liquefied Gas Carriers*, OCIMF and SIGTTO, 1985.

- [14] H. Boonstra, “Wind tunnel tests on a model of a semi submersible platform and comparison of the results with full scale data,” in *Proceedings Offshore Technology Conference*, Houston, 1982.
- [15] F.M. White, *Fluid mechanics*, McGraw-Hill, 1999.
- [16] De Jong, *Electromechanical instruments*, Dutch National Aviation Association.
- [17] M.J. Hilten, “Rapportage orienterend onderzoek windbelasting op schepen en hoge transporten,” Loodswezen, Hoek van Holland / Royal Netherlands Naval College, Den Helder, Technical Report, 2001.
- [18] D.J. Engelbracht, “Windkracht uitgeoefend op het LCF,” Report MHKC, Royal Netherlands Naval College, Den Helder, 2003.
- [19] W. Blendermann, “Probabilistic and spectral modelling of the wind loads on ships,” Hamburg University, Report Nr. 615, 2001.

Applying NEC to UAS Operations Using an Evolutionary Approach

Jochum Tadema, Armand Goossens & Erik Theunissen

Introduction

Unmanned Aerial Vehicles (UAVs) are controlled from UAV Control Stations (UCSs). Part of the information that is contained in the data transmitted over the control link is the result of communication with Air Traffic Control (ATC) and with Command & Control (C2). The Unmanned Systems roadmap [1] foresees an integration of UAVs, UCSs and C2 in a larger network, enabling seamless access to the desired platform and payload on a time-shared basis. The resulting Network Enabled Capabilities (NEC) should enable the command to achieve coherent effects through the effective use of all observation and weapon capabilities. Clearly such an increase in capabilities will not happen overnight. A stepwise approach is expected in which NEC will evolve as the connectivity between systems is increased.

Similar to the embracement of NEC in the military community, future Air Traffic Management (ATM) concepts rely on System Wide Information Management (SWIM) [2,3]. Connectivity between aircraft and ATC is one of the cornerstones of future ATM, and a specific challenge concerns the seamless integration of unmanned aircraft into controlled airspace. Here too, an evolutionary approach is being used which relies on existing networks and a stepwise increase in capabilities that are enabled through upgrades of the functionality of the aircraft and ground systems.

An important commonality between the use of NEC in the military environment and SWIM based ATM is that it concerns the coordinated navigation of many entities and a need for local synchronization. The

goal of the research described in this paper is to initiate an evolutionary approach for the integration of a UCS with ATC and C2 systems by taking advantage of existing technologies which are not yet fully exploited, taking into account the similarities with the developments in civil aviation.

Where to start and how?

A brief summary of the envisioned future NEC architecture could be ‘*a range of functions that can be used to manage and share data in a SWIM environment in order to achieve coherent effects*’. Such a capability is being pursued in the development of the Global Information Grid (GIG¹). Although the capabilities of the desired future system are often outlined in terms of high-level requirements, the roadmap describing how to get there contains holes between the current situation and the envisioned one. At present, the required SWIM environment is not available, but the potential connectivity that can be achieved using existing networks provides opportunities to already realize NEC for a range of functions. Clearly, a difference in capabilities will exist between systems that were retrofitted with some networking capability and systems that were designed around a networking concept. The research discussed in this paper focuses on the opportunities that arise when connectivity from a UCS with ATC and command and control C2 is realized. These opportunities are discussed in relation to the functions needed to integrate the data into the existing systems.

Because many, if not most of today’s UCS, ATC and C2 systems were not designed with connectivity to the other ones in mind, the development, evaluation and refinement of the required functions that will benefit from the future connectivity is far from trivial. The challenges are similar to the ones identified by Hazlett [4] who states: ‘*In the near term we must live with the separate systems we have today. But we can take steps, using modeling and simulation, to test and tune future integration*’. To achieve the desired simulation environment needed to test and tune future integrated concepts, the recommendations include:

¹The GIG is defined as a “globally interconnected, end-to-end set of information capabilities for collecting, processing, storing, disseminating, and managing information on demand to warfighters, policy makers, and support personnel”.

1. Start to use modeling linkages to tie together the disparate elements that make up our non-system of systems, to begin to develop the non-existent interchanges that take advantage of potential synergies.
2. Use models and simulations to develop ‘wrappers’ to encapsulate unruly and uncooperative system elements so that they can interact with other elements in the most opportune manner.
3. Use simulations as ‘fillers’ or ‘placeholders’ for not-yet-developed system elements, to take the fullest possible advantage of asynchronous system developments, allowing system elements to come ‘on-line’ when they are ready, rather than waiting for the entire system(s) maturation.

The results of such an approach contribute to the definition of a roadmap for the functions that will benefit from an increase in connectivity (e.g. in terms of bandwidth, security, availability, integrity) and refine the requirements for the final SWIM environment. In this way, an evolutionary, spiral-based approach to NEC can be achieved.

UAS operations, NEC and C2

An important goal of future UAS operations is the seamless integration into controlled airspace. The use of a network to share trajectories between the aircraft and the ground as foreseen in both the Next Generation Air Transportation System (NextGen²) and the Single European Sky ATM Research (SESAR), is the main enabler to achieve this goal. Similar to the advantages that result when connectivity is realized between ATC and the UCS, the integration of UCSs into a C2 network can significantly reduce the amount of voice communication needed to coordinate and synchronize events. This translates into the possibility to act faster.

This paper starts with an overview of related developments in commercial aviation. After this, different levels of connectivity are discussed and the NEC levels are related to functions and connectivity. Next, both the concept and the simulation environment are discussed in more detail. For the experiments that addressed connectivity with ATC and C2, it will be

²NextGen is the Federal Aviation Administration’s (FAA) plan to modernize the National Airspace System (NAS) through 2025

illustrated what NEC levels have been achieved for the functions that have been implemented. Based on the results, an analysis of existing datalink standards, the feasibility and the potential of the evolutionary approach will be discussed.

Network Developments in commercial aviation

The network-based approach to future ATM started in the early nineties. Based on the results from the research performed in the Program for Harmonized Air Traffic Management in Europe (PHARE), the need for ‘*a generic protocol for information sharing between system components and their offered services, so that interaction of services can be standardized and operate on a global basis*’ is stated [5]. It is concluded that the Total Information Sharing Protocol (TISP), a generic software protocol for client-server software architectures offers a solution. Nowadays, System Wide Information Management (SWIM) is being heralded as the enabler for future Air Traffic Management [6]. SWIM has been described as: ‘*an international concept resulting from FAA and European recognition of the need for network centric operations to meet future traffic demands*’, see [2]. In [3] it is stated that ‘*the core of SWIM is a framework enabling authorized applications and services to reliably and securely share information*’.

Use of the network

For the envisioned future concept of operations, the benefits obtained through the sharing of information result from the increase in accuracy of information (trajectory data) which provides increased predictability, thus allowing optimization over a longer time horizon. In [7] this is explained as follows: ‘*By letting the aircraft Flight Management System (FMS) communicate with the ground the Air Traffic Controller could receive information about what the aircraft intends to do, i.e. what flight path it will take including the time at the different positions. By providing exact four dimensional data to the ground, the pilot and the Air Traffic Controller have the same accurate information about the aircraft flight path*’. In terms of data, the required interaction is of low bandwidth. This allows for an early implementation with gradually increasing capabilities on existing infrastructure through evolving software.

In [7] the following results are reported: ‘*On March the 19th 2006, the first Green Approach was performed by flight Scandinavian Airlines (SAS)*

SK007 from Lulea to Stockholm Arlanda. The flight took 58 minutes and when the flight reached the runway at Arlanda it was only 2 seconds after the time that had been reported from the aircraft and FMS 42 minutes earlier'. The same concept is being pursued for the operation of UAVs in controlled airspace. Mueller and Jardin [8] discuss 4-D operational concepts for UAV/ATC integration. In [9] it is indicated that *'The developments in the area of 4D operations do not only pertain to manned aviation. Uninhabited Aerial Vehicles (UAVs) are planning machines par excellence, producing predictability by definition. A SWIM/CDM based ATM concept using shared 4D data would therefore principally enable interoperability between traditional and UAV traffic*'. In 2009, GE Aviation demonstrated a UAV flight controlled with a modified commercial FMS with 4-D trajectory capability [10].

Initial implementations

To test initial implementations, connectivity with the ground system is required. In the commercial aviation domain this has been partly addressed through the use of existing networks that, although not meeting the requirements envisioned in the future SWIM environment, already provide a significant leap in capabilities. In [7] it is reported that *'For the initial flight trials the ACARS was used to communicate the 4DT to the ATCC*'. and *'When the new VMMR is certified and approved all the messages will be sent over VDL Mode 4 instead*'. In [11] a similar connectivity challenge was addressed for the simulation based evaluation of a network enabled concept for enhanced airport surface navigation. As a result we have, that rather than waiting for the SWIM environment to 'happen', existing datalinks and networks are being used to demonstrate the potential of SWIM.

From Connectivity to NEC

Table 1 provides an overview of the five levels of NEC as presented in [12]. From level 2 to 5, connectivity is the basic requirement, but the actual level is determined by the ability of the networked participants to synchronize their local processes with the higher-level processes. An analogy in the control theoretical domain is the synchronization of multiple closed-loop processes in a larger control loop, which in turn can be part of another loop. An analogy with Star-Trek would be the Borg that with their collective mind may qualify for NEC Level 5.

Table 1: NEC Levels [12].

Level	State	Action
1	Isolated	Exchange of information through conventional means
2	De-confliction	Limited coordination, No common picture of the situation
3	Coordination	Coherent and efficient communication, Information sharing, Common picture of the situation
4	Integration	Integrated and coherent cooperation Efficient, interactive planning and execution
5	Coherent effects	Effective use of all observation and weapon capabilities

When starting to tie together systems that were not designed with the envisioned networking concept in mind, the resulting connectivity that can be achieved may be more limited than desired. Yet, it often allows the NEC level to be increased from 1 to at least 2. Hence, to take advantage of existing technologies which are not yet fully exploited, it is important to understand the possibilities and limitations for a particular configuration. Figure 1 illustrates the different types of connectivity for a UCS that have been explored.

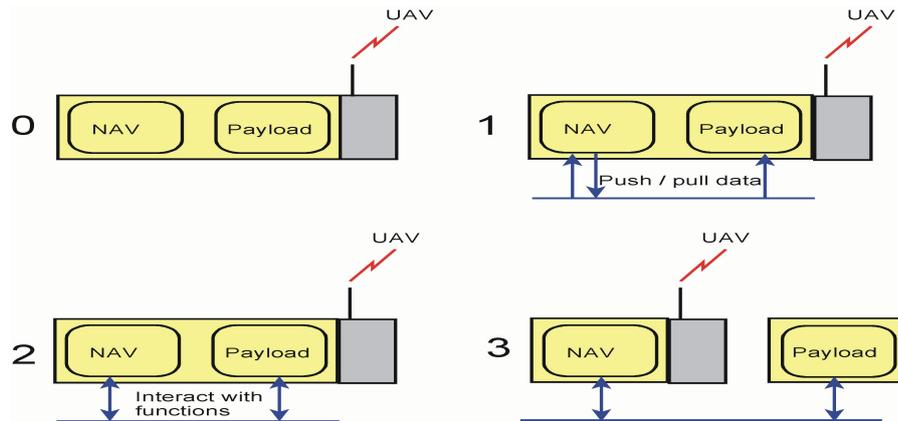


Figure 1: Different types of connectivity between a UCS and a network.

In Figure 1, (0) represents a UCS configuration without any connectivity. Both the navigator and the payload operator communicate with ATC and C2 by means of voice. (1) represents a configuration where there is a possibility to use additional data, but no functionality to interact with other participants on the network. An example is the ability to connect to a network on which data about other traffic, weather and similar data is available. Also, a data-out possibility may exist, in which for example the position of the UAV is put onto the network, providing ATC with additional surveillance data. The important characteristic of (1) is that no

coordination with other systems is required to obtain or provide data, so no real dialogue capability which is needed to coordinate and synchronize events is available and hence no interaction is possible. Configuration 1 typically can be found with systems that started in configuration (0) and obtained connectivity during an upgrade, based on the availability of existing datastreams. (2) represents a configuration that contains functionality to interact with ATC and C2 systems on the network, and thus can both request and provide data. The connectivity shown in (2) is the basis for NEC levels 3 to 5. The specific NEC level is determined by the overall integration of the systems, and the ability to coordinate and synchronize execution of events. In (0), (1) and (2), the navigator and the payload operator are co-located and communication is direct.

Typically, the control of the functions (payload and navigation) is performed from a single location. This has always been the case with the manned counterparts, and if a single manned platform could not perform all required functions, multiple platforms would be used. Although in general an important advantage of an unmanned platform is that the environment from where it is controlled is less constrained than the typical cockpit, in certain situations the need arises to minimize the footprint of the control system. One possibility to reduce the system footprint is to try to minimize the overall set of functions that require an operator in the loop and also minimize hardware. Alternatively, one can consider separating the functions in such a way that only those that are required at the location with footprint constraints are available, while the other ones are managed from a separate location. Configuration (3) enables such a reduction in local system footprint through a re-allocation of functions that are presently co-located. It represents a configuration in which the navigator and the payload operator are geographically separated and the network is also used to facilitate communication between them.

Development of NEC functions

Connectivity does not automatically provide NEC, but there is no NEC without connectivity. To relate potential improvements in terms of NEC to the available connectivity, the research is structured to identify and explore opportunities for configurations 1, 2 and 3. To explore the potential of such an evolutionary integration, a simulation environment consisting of both real and simulated systems, connected over a network has been created.

The UCS baseline system (which represents configuration 0 in Figure 1) is research UCS that has been developed and refined in the context of several research projects at the Netherlands Defence Academy.

UCS baseline system

The research UCS shows a graphical depiction of the position of the UAV, the planned route and the environment³ in both two and three-dimensional reference frames [13]. A graphical specification and modification of the route using drag-and-drop functions to insert or move waypoints is available. Figure 2 provides a schematic overview of the functionality that is implemented to manage the route.

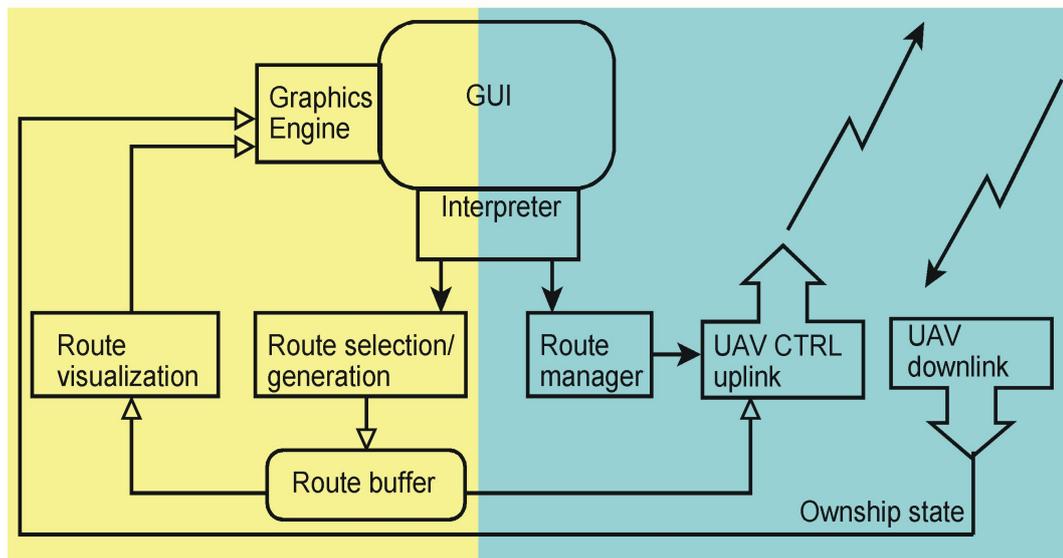


Figure 2: Overview of the functionality used to manage the route of the UAV.

The functions in the left block represent the functionality to support real-time interaction with the route (through direct manipulation). The functions in the right block comprise the functionality that ensures that the selected route (contained in the route buffer in the right block) is uploaded to the navigation system of the UAV. The position of the UAV that is received from the UAV downlink is used for route conformance monitoring.

To allow direct manipulation of the route through the graphical user-interface (GUI), the process in the left block has a high update-rate during

³Data that is used to describe the environment comprises Digital Terrain Elevation Data, data about the location and type of threats and specific data about the target environment.

the route definition and route modification process, and the definition of the draft route in the buffer will change with every change in location of a waypoint and/or constraints. The actual navigation, guidance and control loops of the UAV are not connected to this loop. Once an acceptable route has been defined, the draft route is changed into the active route and subsequently uplinked to the UAV.

This route is used by the Guidance, Navigation and Control (GNC) system onboard the UAV to close the position, directional and orientation loops. When a mission starts, the GNC system is loaded with a route. During a mission, the operator will need to make changes to this route by ATC and by C2. In today's operations (configuration 0 in Figure 1) the information about required changes is obtained through voice communication. The required changes are communicated as a set of vectors (speed, direction, altitude) or a target location. Even if at ATC or C2 a more strategic definition of the required changes is available (e.g. a path defined by waypoints), this will be communicated as a set of vectors.

Opportunities for different levels of connectivity

As indicated in the previous section, a main characteristic of configuration 1 is the lack of sufficient functionality to interact with other systems in the network. Because the ability to interact with other systems requires functionality on both sides, a situation can also exist in which from a UCS perspective configuration 1 applies in relation to ATC and configuration 2 in relation to C2 or vice versa.

In general, configuration 1 can be used to provide the system with a larger amount of real-time data than can be input by the user. This provides the opportunity to use the system functions to integrate this real-time data with other, related data, serving as an enabler for increased Situation Awareness (SA), reduced workload and support for decision making.

The interaction that is enabled in configuration 2 provides the possibility to request specific data and/or actions. This can be used to more effectively use the available bandwidth of the network, and to coordinate and synchronize events. The separation between navigator and payload-operator shown in configuration 3 provides opportunities to distribute the system footprint, which may be of benefit in environments where there are limitations in this area. To create a roadmap from today to a future with

GIG-like connectivity, our research aims to systematically explore the potential of all three types of connectivity. Table 2 provides an overview of the functions that have been designed and implemented for three different applications. The functions are referenced by the capitals A to G and will be addressed in the following subsections.

Table 2: Functions implemented on top of network to support an application.

	Connectivity						
Connected with	ATC			C2			payload nav
Config (Fig.1)	1	2		1	2		3
Application	From	To	Interaction	From	To	Interaction	Interaction
Airspace integration	A A1	B	E E1				
Faster OODA loop				C	B	F F1	
Control from geographically separated locations	A						G

Connectivity with ATC

Network connectivity with ATC forms the basis for the integration of UAVs into tomorrow’s controlled airspace. In terms of the potential to realize the desired connectivity with ground systems, UAVs are actually ahead of commercial aviation. However, this advantage is not yet being exploited for ATC purposes because current UCSs lack direct connectivity with ATC. Many of today’s UCSs were not designed with network connectivity in mind. Still, possibilities to communicate certain information often exist, but the likelihood that several systems all use the same protocol is rather low. In the future this is expected to change because of the emergence and acceptance of standards to provide interoperability such as STANAG 4586 [14] and the developments in the area of SWIM. But also for today’s systems, so-called wrappers can be implemented that provide connectivity with which level 2 and level 3 NEC is achievable. Concerning the connectivity with ATC, the following assumptions have been made:

- In the minimum network configuration, the UCS has access to a network on which data about the location of other traffic is available. This forms the basis for blocks [A] and [A1] in Figure 3.

- This network allows trusted entities to provide information about traffic, as illustrated by [B] in Figure 3.
- With the appropriate functionality on the UCS and ATC side, the network can be used to exchange route data, forming the basis for [E] and [E1] in Figure 3.

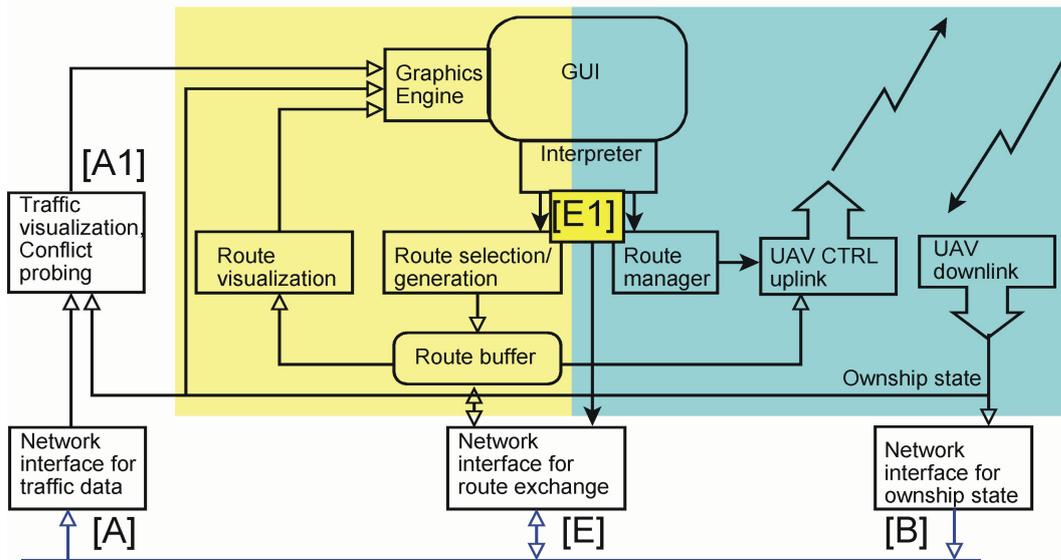


Figure 3: Connectivity to the system shown in Figure 3.

Module [A] reads all traffic data and filters out the traffic that is not relevant (based on distance and altitude). Level 2 traffic awareness is supported through the depiction of the location of the traffic in the same reference frame as own ship and the current route. This goes beyond the information that is available to pilots on today’s Traffic alert and Collision Avoidance System (TCAS) displays, and is similar in presentation to a Cockpit Display of Traffic Information (CDTI) as proposed in RTCA DO-317 [15]. Also, conflict probe functions such as discussed in [16] are implemented in module [A1]. The conflict probe provides data to the operator about the impact of a change in current direction in the separation with other traffic. Through the integrated presentation of the results level 3 traffic awareness is supported⁴. Hence, even limited connectivity (configuration 1) can already yield a significant increase in SA, which in turn can contribute to a more efficient interaction with ATC.

⁴Level 3 Situation Awareness is indicative of the ability to anticipate how the situation will develop.

The connectivity in module [B] is intended to contribute to the surveillance capability of ATC. A particular strength of this configuration is the exchange of the planned and the measured own ship position in case of a lost down link. In such a situation, the estimated location of the UAV (obtained through the primary radar used by ATC) can be presented to the operator, and the planned location of the UAV (computed in the UCS from the 4-D route and the current time) can be provided to ATC. In this way, it is still possible to perform conformance monitoring. In the section ‘Simulation Studies’ the functionality in the UCS that supports this capability will be discussed.

Similar to the developments in commercial aviation, a big leap in capabilities is expected once a dialogue capability for the exchange of route data becomes available. On the research UCS, the required level of connectivity to explore this concept is realized by functions in the modules [E] (the network interface) and [E1]. Module [E1] represents the additional functionality that has been implemented in the GUI to realize a digital dialogue capability with ATC.

Connectivity with C2

Like other platforms, a UAV is an asset that contributes to a successful closure of the Observe-Orient-Decide-Act (OODA) loop. An important factor that determines the time within which the OODA loop can be closed is the time spent in the Orient and Decide phases. At present, the communication between the UCS and C2 is performed by means of voice, and the limited information exchange bandwidth that can be achieved also imposes constraints on the bandwidth of the OODA loop.

Regarding the OODA loop of armed UAVs, Gibbs [17] indicates that *‘The desire to compress the kill chain time line led to discussions about the targeting cycle CONOPS, especially regarding time sensitive targets (TSTs). Leadership decided that, to improve success against pop-up and especially mobile targets, the targeting cycle of find, fix, track, target, engage, and assess (F2T2EA) must be reduced from hours to minutes’*. In [17] it is also indicated that for the armed Predator operations this has been achieved by computer enhancements and changes in processes.

Besides the possibility to significantly reduce the time within which the OODA loop can be closed, connectivity between the UCS and C2 provides

the possibility to achieve more accurate coordination and synchronization. Similar to the connectivity with ATC, this is achieved through the use of an accurate description of the desired location of the UAV as a function of time, defined in the 4-D route. The first phase of our research has focused on exploiting these advantages for Intelligence, Surveillance, Target Acquisition and Reconnaissance (ISTAR), Battle Damage Assessment (BDA) and Time-Sensitive Re-tasking.

The connectivity that has been realized between the UCS and C2 is similar to that between the UCS and ATC depicted in Figure 3. Besides the modules [E] and [E1] shown in Figure 3, two additional modules, [F] and [F1], are implemented. Module [F] contains functionality to provide the UCS with data concerning updated target locations, target areas and threat locations. The use of 4-D routes now serves as the enabler for better synchronization of assets on the C2 side. Module [F1] contains the functionality needed to support the dialog with C2. The route buffer and the route visualization function have been extended with the capability to store and visualize targets and target areas. Although the implemented functionality mainly yields a replication of existing procedures by digital means, the resulting coherent and efficient communication increases the accuracy, efficiency and flexibility. The fact that all data is digitally available instead as pieces of information in the memory of the operators (because it was communicated by voice) makes it possible to integrate the data into the frame of reference used at the UCS and C2, supporting a common operational picture.

Reduction in local system footprint

To combine maximum flexibility in terms of payload control with a minimum total manning and equipment footprint on a naval vessel, the concept that is explored using configuration 3 in Figure 1 pursues geographically separated control of the UAV and its payload. In this concept of operation, the navigation process is not managed from the vessel, but at a central, off-board location. Similar to current UAV operations that are controlled from a ship (e.g. the RQ-2 Pioneer), the management of the payload is performed from the naval vessel.

In the foreseen concept, the main tasks of the navigator remain the same. They comprise coordination with Air Traffic Control and translat-

ing the requests from the payload operator into route segments that can be inserted into the current flightplan within all existing constraints and flight safety issues (e.g. emergencies). Because the payload operator and the navigator are not co-located, the direct, intuitive way of communication that is used to change the flightplan based on new, payload-related requirements is no longer possible. This reduces shared Situation Awareness (SA) and operational efficiency. Furthermore, it is no longer guaranteed that both crew members have the same information available, which also has a negative effect on shared SA and operational efficiency. The goal of the research was the development of a dialogue capability to provide the payload operator and the navigator with a means of interaction through their networked control stations that supports a level of coordination equivalent to that of co-located operators.

The connectivity between the navigation and the payload management station allows a replication of all relevant data on both systems. The navigator can provide the payload operator with a specification of the target area within which the payload operator has the freedom to plan the route needed for (optimal) employment of the various sensors. Part of this route definition comprises the insertion of specific flight patterns that can be scaled in terms of leg-length, spacing between legs and height. The concept of direct manipulation of the route, including the ability to insert the payload-based navigation patterns which are scaled based on specific payload properties and information requirements, enables the payload operator to specify the desired path in the target area. This path is provided to the navigator who integrates it into the overall flightplan of the UAV.

Simulation studies and concept demonstrations

At the beginning of this paper it was pointed out how this has been done in the civil aviation domain. In the military domain, similar approaches using linkages and wrappers to integrate multiple ‘stove-pipe’ systems into a single, distributed simulation environment for UAV missions have successfully been applied. Twesme and Corzine [18] describe the development and initial use of a distributed simulation infrastructure to develop and exploit the capabilities and interoperability of UAVs and UCAVs. In their evaluation the command and control (C2) node was essentially excluded. For our simulation studies, existing interfaces of a MASE C2 system have been used for information exchange with a research UCS.

UCS-C2 simulation infrastructure

To evaluate the ideas and concepts discussed in the previous sections, the identified functions have been designed and implemented. The simulation studies that have been performed addressed the coordination between ATC, C2 and the UCS regarding airbase operations, traffic deconfliction, ISTAR and BDA missions, time-sensitive retasking and lost comms procedures.

Figure 4 provides a schematic overview of the components of the simulation environment. It comprises several separate simulation systems, which through a combination of wrappers and linkages are connected to each other. As can be seen, several types of message wrappers are used to enable interaction between the different subsystems.

Given the limited possibilities to modify the MASE simulation functionality, the UAV simulator and the traffic simulator needed to provide the data in the format expected by the MASE simulator, the Distributed Interactive Simulation (DIS) protocol. The UAV position and altitude including Mode 3A were provided by the UCS to the MASE simulation program and converted into the Radars Southern Region and Portugal (RSRP) protocol. This is the protocol that the MASE uses to process radar data from different sensors. In this way, the UCS was connected to the MASE in a similar way as the normal simulation assets. The second one mimics a real link that provides the UCS with an air picture (simulated or live). This air picture data is based on primary and secondary sensor data from the MASE sensor(s) which is also used for display at the MASE (C2 and ATC) consoles. Furthermore, a virtual link for 4-D route and airspace info (UAV routes, C2 and ATC commands and airspace boundaries) to support strategic deconfliction is used. This functionality was available at the UCS but, due to MASE interface limitations, had to be pre-loaded at the MASE. In this way the UAV routes and areas could be displayed for the ATC and C2 controller at the MASE.

With this setup it is possible to use the (simulated and live) traffic provided by the MASE system, in this way integrating a simulated UAV into a mission with otherwise live, real players. As an alternative it is also possible to use the UAV research environment to simulate other traffic and provide it to the MASE. This allows for more controlled simulated missions.

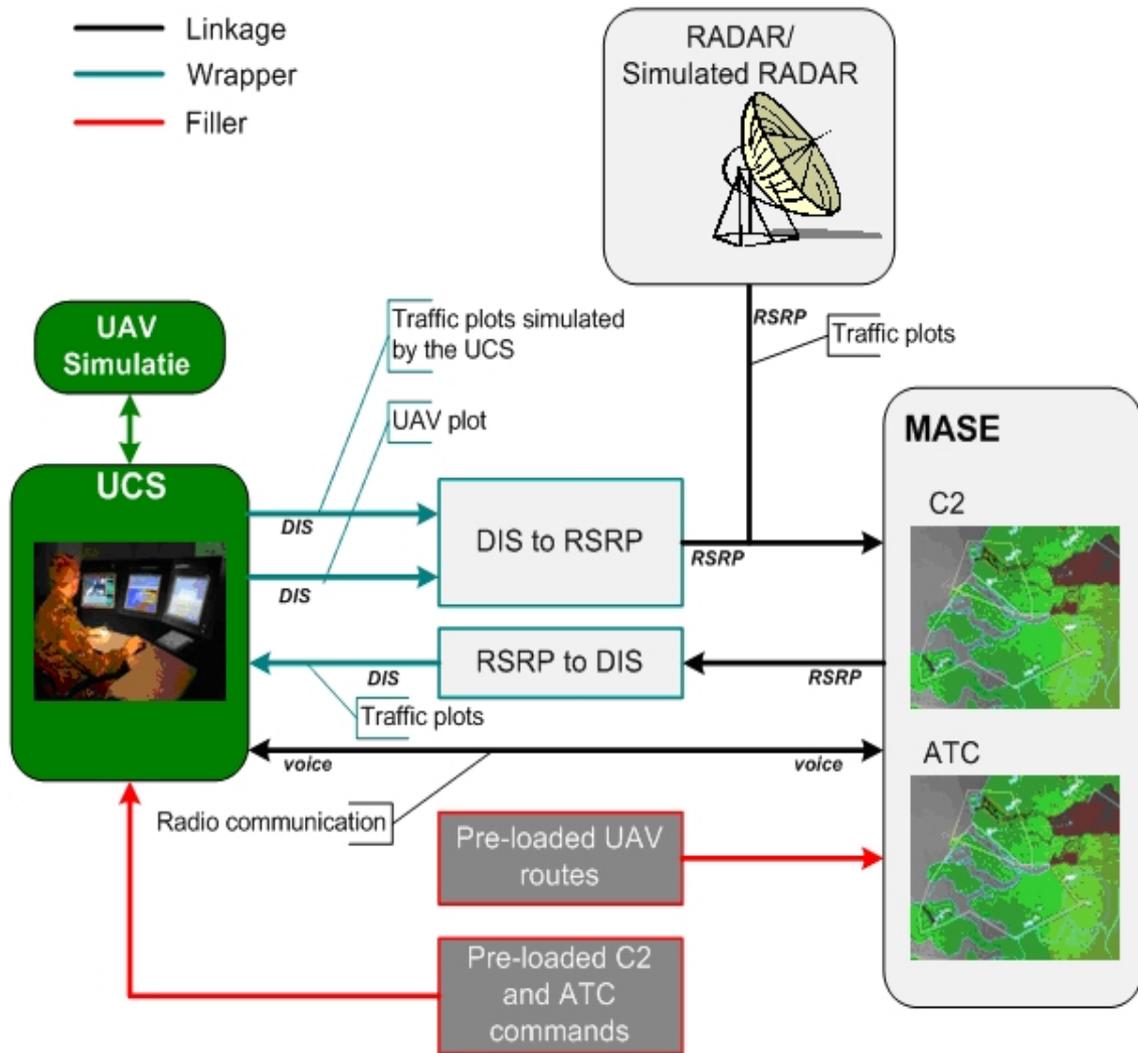


Figure 4: Overview of the different simulation components and how they are connected.

Topics

To explore the different types of capabilities discussed in the previous section, several simulation-based missions have been performed. During a simulated mission above Zeeland and Rotterdam, airspace integration, IS-TAR, BDA and time-sensitive retasking have been addressed using scripted scenarios to trigger a range of events. The five-level NEC scale was used to rate the level of each function.

Airspace Integration

To explore concepts for traffic deconfliction as a function of available connectivity and functionality, four different set-ups were used. In the first set-up the conventional means of voice communication was used to decon-

flict traffic, and connectivity of type 1 (modules [A] and [B] in Figure 3) for the exchange of traffic data. On the UCS side, the traffic is integrated into the plan view display and as a conformal sensor overlay. This supports the UAV operator with the visual acquisition once the traffic is within the field of view of the sensor. Although the communication between ATC and the UCS is still performed by voice, the common picture of the situation supports the coordination and it was concluded that the functionality provided through modules [A] and [B] allows NEC level 3 to be reached for normal deconfliction.

The second and third set-up are triggered by the ACAS system in the UAV and use the available information generated by the ACAS and ATC system for SA and coordination. Because the ACAS information is not directly available for ATC, for collision avoidance only NEC level 2 is achieved.

In the fourth set-up, the connectivity represented in Block [E] of Figure 3 is also used. This enables conflict resolution to be achieved through integrated and coherent coordination between ATC and the UCS. If the conflict prediction function indicates that at a future location along the planned path a loss of separation will occur, the operator can prevent this by modifying the flightplan (through a direct manipulation of the location of waypoints or the reference speed along one or more legs of the flightplan). The modified flightplan is passed to ATC for approval. If approval is received, the flightplan is uplinked to the UAV. Also, ATC has the possibility to add some additional constraints to the flightplan, before it is passed to the UCS. This process of interactive planning, made efficient through the possibility of digitally exchanging 4-D flightplans of available information yields NEC level 4 for this setup.

To explore lost comms procedures, a scenario was designed in which the UAV operator would lose all status information about the UAV and its location. By receiving the UAV position information from ATC or via TDL, the UAV operator could assess whether the UAV adhered to the pre-planned route. This is also the case for the ATC controller who will have the lost link routes in his system and can monitor the UAV behavior by means of the radar or transponder returns. Both operators still have a common picture of the situation, yielding NEC level 3 for the lost comms situation.

ISTAR, BDA and Time-Sensitive Retasking

In the ISTAR part of the scenario, the UAV is flying towards the target area for which the UAV operator will insert an observation pattern (by means of the touchscreen). Upon entering the loiter area, the UAV will provide last minute target information to a flight of two F-16s. These aircraft simulate a targeting run after which the UAV operator will descend in the loiter area to perform a Battle Damage Assessment (BDA) run.

To support a common operational picture between the UCS and C2, during the ISTAR task and BDA phases, an exchange of the orbits of the UAV and the location of the target is performed between the C2 controller and the UAV controller using the functionality in module [F]. Furthermore, module [A] is used to provide the data needed to display the location of the cooperating assets (e.g. the F-16 flight). The associated functions in modules [A] and [F] contribute to achieving NEC level 3. To evaluate the functionality for Time Sensitive Retasking, the scenario includes a phase in which a re-tasking is received from the higher command echelon to support a calamity in the Rotterdam harbour area. To determine whether the re-tasking can be accepted, the UAV operator checks the request against the constraints (e.g. remaining endurance and reaction time). Only feasible scenarios were used. Both the data specifying the manoeuvring area and the area of interest are received through block [F] and translated into geographical objects that are subsequently presented on the display used to plan and modify the flightplan.

Similar to the process used to deconflict with traffic through the modification of the 4-D flightplan, the UAV operator re-routes the flightplan to the new area by moving and adding waypoints using the touchscreen. This new route is linked back to the ATC system for approval. Between the entry and exit waypoints of the target area, an optimized sensor pattern is inserted by the UAV operator over the area of interest to minimize crew workload during the support of the calamity. Once the UAV enters the pattern the operator can focus on the operation of the sensor and providing on-scene commanders on the ground with the requested information.

The coordination of the re-tasking by exchanging of the areas and the re-planned route during the Time Sensitive part of the mission confirmed the expected efficiency of this cooperative interactive planning. The associated

functions for modules [F] and [F1] contribute to achieving NEC level 4.

Control from geographically separated locations

Clearly, shared control from geographically separated locations is only possible if the connectivity allows for a level of interaction between the navigator and the payload controller that is similar to the interaction when co-located. So, whereas the connectivity in the previous two applications is compared to a baseline that only comprises voice communication through a radio channel, in this case the baseline is the situation where the two operators are sitting next to each other when coordinating how payload-driven navigation requirements can be realized through modifications to the flight plan.

To test the concept, two existing research operator stations have been extended with functionality to support the payload operator and enable the communication between the two stations. One of the research operator stations (shown in Figure 5) is located at the Netherlands Defence Academy (NLDA) in Den Helder.



Figure 5: UCS research station at NLDA in Den Helder.

As can be seen in Figure 5, the UCS research station has two positions. The left one is for the pilot flying, and the right one for the payload operator. For the current research, only the payload operator position was used. The second UCS research station is located at Delft University of Technology. This research station has a single position and was configured

for the pilot flying.

The mission management software used in the research operator stations already had the capability to connect to other systems. Only minor enhancements were needed to establish the basic dialogue capability needed to exchange routes and areas. Although the communication between the two stations did not use the STANAG 4586 protocol, the message content used was quite similar. An analysis of the current version of STANAG 4586 indicates that with minor additions to the message set, the basic functionality to support geographically separated, payload driven navigation can be achieved.

Initial tests were performed in April 2006. Those tests were mainly focused on the technical implementation of the concept, rather than evaluation. During those tests, an interesting observation was that when only the datalink for cooperative control is present, and no voice communication, the feedback on the status of requests becomes very important. When operators are co-located, their dialogue includes timing information (e.g. on when to expect a particular result). Especially when a certain task takes some time, e.g. the re-planning of a route, timing information becomes important for the operator awaiting the results. When no voice communication is available, this information must be provided through the network connection. Such process-related timing information is not included in the current datalink message set, but it is regarded as an important addition.

The tests demonstrated the feasibility of the concept, but also showed the need for further evaluations to be able to better identify the additional information needed to support the dialogue.

Demonstration of connectivity with ATC and C2

Between October 2005 and March 2006, the simulation environment was used in a total of four demonstration sessions to various groups of subject matter experts. These included representatives from the Chief of Royal Netherlands Air Force Command, Military Air Traffic Control Centre, Defence Materiel Organization (Projects Branch), C2 Knowledge Centre (Army, Navy, Air Force), Defence Research & Development and National Knowledge Centers and Laboratories.

These demonstration sessions covered airspace integration, ISTAR, BDA



Figure 6: Setup in the auditorium of the AOCS-NM used for the demonstrations of the simulated missions using a UCS networked with C2 and ATC.

and time-sensitive re-tasking. Figure 6 shows the setup in the auditorium of the Air Operations Control Station Nieuw Milligen (AOCS-NM). The research UCS was located in the auditorium and the audience could observe the C2 and ATC controllers on a large video screen. Also, some of the UCS displays were replicated on larger screens. During these demonstrations, the subject matter experts commented on the high degree of operational realism that was achieved.

Demonstration of geographically separated control

The concept as outlined above was successfully demonstrated at the UAV thematic day of the NLDA in May 2006, with the payload operator located at the NLDA in Den Helder and the pilot flying in Delft. The demonstration used a scenario comprising a retasking of the UAV to a new mission area and was situated in the environment of Kabul, Afghanistan to create a plausible UAV reconnaissance environment.

Discussion

The concepts discussed in this paper are not new, but the implementation of the integrated simulation environment and the subsequent use to explore these concepts is still quite rare.

In terms of open systems and standards, the future is not yet here.

Standards for the protocols needed to realize the envisioned NEC concept of operations do not yet exist. Also, the current generation C2 systems still have proprietary interfaces. The increased adherence to standards for the exchange of route data will reduce the amount of wrappers needed to integrate different, non-standard systems into a common network. Once this is achieved and interaction between systems becomes possible, the challenge becomes how to evolve to NEC level 4 for all functions. Achieving NEC level 5 goes beyond the integration of a UCS with C2 and ATC and requires a consideration of the overall system of which all these elements are part.

Impact of standards and open systems

Today's UCSs are dedicated to one type of UAV and typically use proprietary protocols for communication with that particular type of UAV. Similar to manned platforms, different mission types will yield a range of different unmanned platforms, optimized for a particular set of missions. Hence, on the platform side it will be hard to benefit from an economy of scale. The physical separation between the platform and the UCS provides the opportunity to achieve an economy of scale on the UCS side. This requires a change from the proprietary communication protocols to standardized protocols. To allow the UCS to be independent from the type of UAV, STANAG 4586 has been introduced. When UAVs comply with this standard, a handover of navigation and payload control between different (types of) UCSs is possible.

The use of STANAG 4586 for the control link protocol is more and more becoming a requirement in the acquisition process. Current developments in the area of UCSs also point towards a future where an open system architecture is required. Within certain certification constraints, this will allow third party software to be added to the Original Equipment Manufacturer (OEM) UCS, a development which is already taking place in the world of commercial avionics. The impact of this development is that it significantly increases the possibilities to enhance overall system capability during the lifecycle.

Beyond NEC level 3

At present, the two driving factors for standardization are to achieve system interoperability and to enable information sharing needed for increased

battlespace awareness and a Common Operational Picture (COP). However, information sharing in itself will only allow NEC level 3 to be reached. Level 4, integrated and coherent cooperation, requires the development of concepts defining how multiple users interact with the data. Clever use of ICT-enabled opportunities will result in a force-multiplier effect once NEC level 4 is achieved. Clearly, these capabilities will not just ‘happen’. Focused research is needed to identify possibilities and explore them.

Summary and Conclusions

Connectivity does not automatically provide NEC, but there is no NEC without connectivity. The integration of UCSs into a C2 network can significantly reduce the amount of voice communication needed to coordinate and synchronize events. The integration of UCSs into an ATM network provides the basis for UAV operations in future controlled airspace, as foreseen in NEXTGEN and SESAR. By taking advantage of existing technologies which are not yet fully exploited and taking into account the similarities with related developments in civil aviation, an evolutionary approach for the integration of a UCS with ATC and C2 systems is being pursued. When starting to tie together systems that were not designed with the envisioned networking concept in mind, the resulting connectivity that can be achieved may be more limited than desired. Yet, it often allows the NEC level to be increased from 1 to at least 2.

In order to identify feasible opportunities, it is important to understand the possibilities and limitations for a particular configuration. To explore NEC for UAV missions, existing connectivity together with wrappers, linkages and fillers has been used to create a simulation environment in which C2, ATC and a UCS are integrated. In this way, the opportunities can be explored for a range of possible configurations in terms of connectivity and functions.

The results illustrate that the opportunities which can be realized using existing connectivity already provide significant operational benefits. Hence, it is concluded that waiting with the development and implementation of functions that increase the NEC level until the ‘promised’ connectivity becomes available, will unnecessarily delay the moment at which significant operational gains can be realized.

References

- [1] U.S. Department of Defense, *Unmanned Systems Roadmap 2007-2032*, 2007.
- [2] J.S.Meserole and J.W. Moore, “What is System Wide Information Management (SWIM)?,” *Proc. 25th Digital Avionics System Conference*, Portland, 2006.
- [3] M.S. Taylor, “System-Wide Information Management for Aeronautical Communications,” *Proc. 23rd Digital Avionics System Conference*, Salt Lake City, 2004.
- [4] J.A. Hazlett, “Modeling and Simulation and C4I: The Quest for Dominant Battlespace Awareness,” *Proc. AIAA Flight Simulation Technologies Conference*, San Diego, pp. 31-36, 1996.
- [5] R. Ehrmanntraut, “Enabling Air-Ground Integration: Definition of a Total Information Sharing Protocol,” *Proc. 22nd Digital Avionics System Conference*, Indianapolis, 2003.
- [6] D. Harkness, M.S. Taylor, G.S. Jackson and R.W. Stephens, “An Architecture for System-Wide Information Management,” *Proc. 25th Digital Avionics System Conference*, Portland, 2006.
- [7] N. Friberg, “Using 4DT FMS Data for Green Approach, A-CDA, at Stockholm Arlanda Airport,” *Proc. 26th Digital Avionics Systems Conference*, pp. 1B3.1-1B3.9, 2007.
- [8] Mueller, E.R. and M.R. Jardin, “4-D Operational Concepts for UAV / ATC Integration,” *Proc. 2nd AIAA Unmanned Unlimited Systems, Technologies, and Operations Conference*, San Diego, 2003.
- [9] E. Theunissen, R.M. Rademaker, O.F. Bleeker and K. Wichman, “Aircraft Trajectory Based Network Centric Applications, ” *Proc. 26th Digital Avionics Systems Conference*, 2007.
- [10] http://www.aviationtoday.com/regions/usa/First-Trajectory-Based-Flight-of-UAS_65328.html
- [11] E.Theunissen, G.J.M. Koeners, F.D. Roefs, R.M. Rademaker and O.F. Bleeker, “Evaluation of an EFB Airport Map with Integrated

- Routing in a Distributed Simulation Environment,” *Proc. AIAA Modelling and Simulation Technologies Conference*, 2005.
- [12] Defensiestaf, Dutch Ministry of Defence, *Netwerkend Optreden*, (in Dutch).
- [13] J. Tadema, J. Koeners and E. Theunissen, “Synthetic Vision to Augment Sensor Based Vision for Remotely Piloted Vehicles,” *Proc. SPIE Conference*, Orlando, 2006.
- [14] NATO: STANAG 4586 Edition 2 , *STANDARD INTERFACES OF UAV CONTROL SYSTEM (UCS) FOR NATO UAV INTEROPERABILITY*, 2005.
- [15] RTCA, *Minimum Operational Performance Standards (MOPS) for Aircraft Surveillance Applications System (ASAS)*, DO-317, 2009.
- [16] J. Tadema and E. Theunissen, “ A Concept for UAV Operator Involvement in Airborne Conflict Detection and Resolution,” *Proc. 27th Digital Avionics Systems Conference*, IEEE/AIAA, 2008.
- [17] D.G. Gibbs, “The Predator in Operation IRAQI FREEDOM – A Pilot’s Perspective,” *Proc. AIAA Infotech @ Aerospace*, Arlington, 2005.
- [18] J. Twesmen and A. Corzine, “Naval Air Systems Command (Navair) Unmanned Aerial Vehicle (UAV) / Unmanned Combat Aerial Vehicle (UCAV) Distributed Simulation Infrastructure,” *Proc. 2nd AIAA Unmanned Unlimited Systems, Technologies, and Operations Conference*, 2003.

Self Separation Support for Unmanned Aircraft Systems

Jochum Tadema & Erik Theunissen

Introduction

Recent guidance from the U.S. Federal Aviation Administration (FAA) [1] suggests that self separation needs to be a component of an Unmanned Aircraft System (UAS) Sense and Avoid (SAA) solution in order for UAS to behave similarly to manned aircraft. Conceptually, UAS self separation is the protective (or conflict avoidance) method that precludes a threat aircraft from ever triggering a time-critical collision avoidance maneuver. Under self separation, aircraft should remain ‘well clear’ of other aircraft or airborne hazards by following the priority rules of 14 CFR 91.113 [2], which determine which aircraft that has ‘right-of-way’.

Self separation is also a key component of future Air Traffic Management (ATM) systems envisioned for both North America and the European Union that shifts some separation responsibilities from ground-based air traffic controllers to pilots. While self separation for manned aviation is being driven primarily by the anticipated increase in traffic density, self separation is more of a critical enabling technology for UAS, given how they operate. That is, unlike most manned aircraft which conduct point-to-point flights, unmanned aircraft more often conduct aerial work (e.g. aerial mapping or surveillance missions) which may be ad hoc in nature. Without the ability to self separate from other aircraft, UAS will be challenged to effectively conduct such operations.

Due to communications latency and reliability concerns, the time-critical collision avoidance component of SAA may require a fully autonomous ca-

pability, at least to account for the possibility of lost link. However, the greater time horizon associated with self separation allows for pilot-in-the-loop (PITL) operation and, in fact, the nature of self separation demands more pilot involvement. Furthermore, it is likely that ATM authorities will be more accepting of an initial PITL SAA system that can serve as a stepping stone to a fully autonomous capability.

The ability to effectively conduct PITL self separation will be critically dependent on decision aides and advanced displays that allow pilots to make accurate and timely maneuvering decisions. This paper presents a concept for a UAS self separation system, including a traffic conflict probe and concepts for how the results of the conflict probe might be presented to the UAS pilot.

Requirements and Criteria

For navigation, Required Navigation Performance (RNP) has been introduced to allow airspace requirements (use of airspace) to be satisfied independent of a specific method [3]. Similar to RNP, rather than mandating a single conflict/collision avoidance maneuvering algorithm¹ that could serve all UAVs, a set of performance-based requirements could serve as the basis for future separation assurance and collision avoidance systems. Criteria that should be considered when defining such requirements comprise:

- Method used for resolving conflicts (prescribed², optimized³, or combination⁴ thereof);
- Ability to handle multiple intruders;
- Ability to integrate constraints imposed by other hazards;
- Allowable types of maneuvers (speed changes, lateral, and/or vertical maneuvers);
- Degree of coordination among own ship and intruder(s), ranging from completely independent to coordinated and cooperative;

¹Like what was done for the Traffic Alert and Collision Avoidance System (TCAS)

²Prescribed maneuvers are determined a priori based on a set of procedures (e.g., ‘right of way rules’)

³Optimized maneuvers involve a rule-based decision among several avoidance options that minimizes a given cost function

⁴Combined approaches determine optimal maneuvers that follow prescribed rules, if possible.

- Level of autonomy (ranging from manual pilot-in-the-loop to fully autonomous);
- Ability of the pilot to assess the situation;
- Technology maturity level.

In previous work, the use of conflict probing has been proposed as a concept to support the pilot in maintaining separation with other traffic. A conflict probe display contains information that both informs the pilot about a predicted loss of separation and provides guidance on how to maintain or restore separation. The current paper uses these eight criteria to better identify the potential and limitations of conflict probing.

Conflict Probe and Pilot Displays

Earlier research has addressed the potential of conflict probing to support UAV pilots with the Detect, Sense and Avoid task [4] and the integration of vehicle maneuvering constraints [5,6]. PITL simulations, in which the potential of conflict probing to support level 3 traffic awareness was evaluated, have demonstrated advantages in terms of safer and more efficient maneuvering decisions [7].

Method for resolving conflicts

Conflict probing consists of predicting the future separation between own ship and hazards for a set of own ship velocity vectors -representing possible combinations of Track, Flight Path Angle (FPA) and Speed- up to a predefined prediction horizon or look-ahead time. Separation requirements result in a cylindrical volume. Using predefined separation criteria, e.g. thresholds used to define a loss of separation or a collision hazard, the probing data indicates which own ship velocity vectors will lead to a future conflict and what the corresponding time to loss of separation is.

To illustrate the concept of probing, Figure 1 shows a top view of an example conflict geometry resulting from the presence of converging traffic. Vehicle A (center) represents own ship, vehicle B is the intruder aircraft; the dashed lines represent the current Tracks. Initial Bearing of the intruder is 290, initial Range is 5 NM; the intruder's Track is 050, flying level at a Speed of 250 kts. Both airplanes are at the same Flight Level.

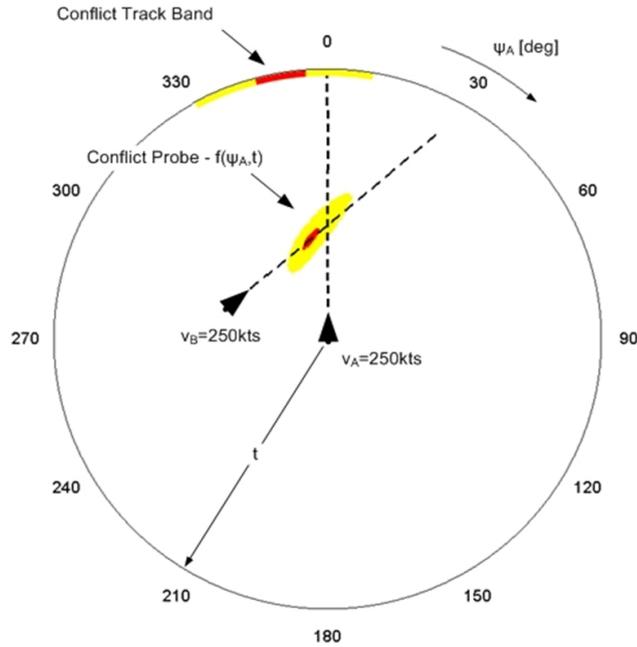


Figure 1: Top view of an example conflict geometry and Conflict Probing in the Track (Ψ_A) domain. The conflict probe indicates the own ship tracks that will result in a violation of the separation (yellow) or collision hazard criteria (red).

The depicted conflict Track band and conflict probe result from probing for a range of variations of the own ship Track angle (Ψ_A), for the current FPA and Speed. The separation criteria (yellow) used in this example are 1 NM lateral and 1 kft vertical. The collision hazard criteria (red) used in this example are 0.25 NM lateral and 500 ft vertical.

Note that although the separation requirements result in a cylindrical volume around the intruder, and the top view would yield a circle, the probing area shown in Figure 1 does not represent this circle. The contours of the yellow area indicate the instants at which the separation circle will be penetrated, and the red area at which the collision hazard circle is penetrated, when own ship changes its track in that particular direction.

In [8] an analytical approach is presented to determine six solution vectors (two track changes, two FPA changes and two speed changes) that prevent an intersection with the cylindrical volume. To illustrate the similarity between the probing approach and the approach discussed in [8], consider the situation depicted in Figure 1. In this example, the required separation distance, represented by the circles, is 2 NM. The conflict space, with a separation below 2 NM, is visualized by mapping the conflict probes

to a spatial reference frame, using the grey scale on the right.

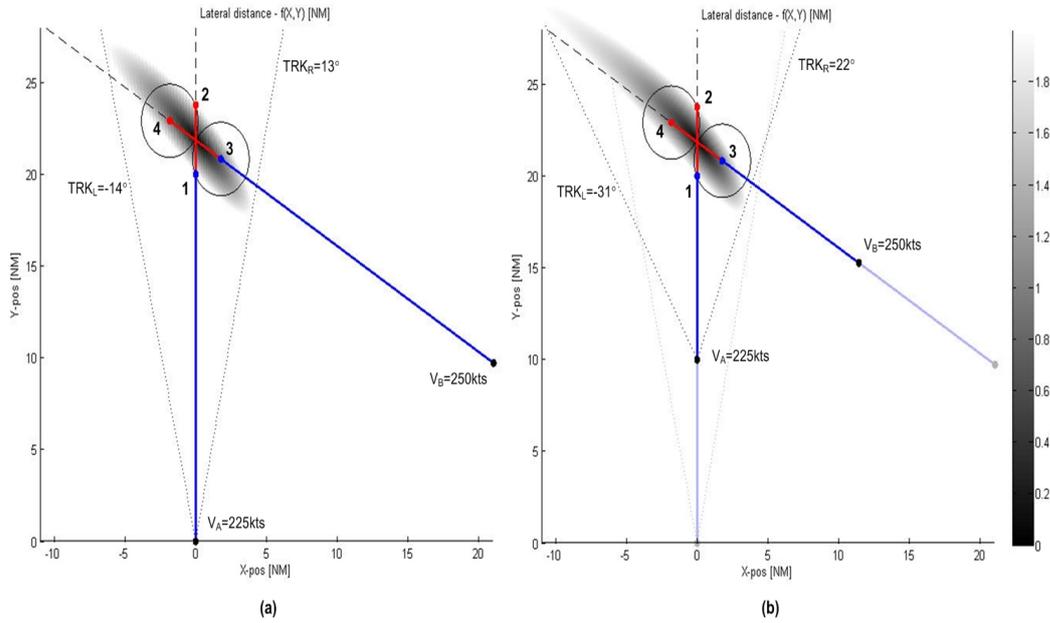


Figure 2: Probe-based visualization of an encounter and the resulting lateral options.

In Figure 2, the intruder comes from the right and will reach its Closest Point of Approach (CPA) exactly on the current track. If both aircraft continue with the current track and speed, actual loss of separation will take place at t_{LOS} with own ship in location 1 and the intruder in location 3 (own ship location on circle around intruder). If both vehicles maintain track, speed and FPA (and would not collide), the loss of separation will continue until the intruder reaches the location identified by point 4. Own ship will again be on the circle, this time at point 2. To prevent a loss of separation, own ship needs to maneuver in such a way that it stays outside of this (moving) circle. Assuming speed remains constant, own ship can avoid the circle by changing its track to the left or the right before t_{LOS} . The minimum change in track that is needed to avoid a loss of separation increases with a reduction in distance towards the location where the loss of separation is predicted. In Figure 2b, own ship has travelled 10 more miles compared to the situation depicted in Figure 2a. Since the probing area visualizes the locations where a loss of separation is predicted to occur for each particular track, it follows that the minimum required track change to avoid a loss of separation yields a track that just misses the area. In Figure 2 these options are visualized as TRK_L and TRK_R . In [8] it is stated that the *'NextCAS II technique provides a novel approach to solving*

the collision detection and resolution proble. It differs from other methods we have examined in that it does not require a fixed look-ahead time (i.e., RA alert threshold) to declare a conflict situation'. Analysis of the proposed algorithm reveals that this is achieved by computing the time t at which the aircraft has to maneuver from a pre-defined upper limit on vehicle maneuvering constraints which in turn yields maximum values for $TRK_L(t)$ and $TRK_R(t)$. In Figure 2 this is the time at which own ship reaches the location on the current path where the track line that represents the avoidance maneuver has an angle with the current track that is equal to the computed threshold.

Limitations and solutions

The result of a state-based prediction is only valid for the time that the state of the involved aircraft does not change. This limits the useful look-ahead time. However, if intent information is available, it can be used to filter out false positives (predicted loss of separation that will not occur because the state of the intruder will have changed before the point is reached where a loss is predicted to occur). Furthermore, intent information can be used to identify those locations where a loss of separation will occur once the intruder has maneuvered. An example implementation of an Airborne Separation Assurance System (ASAS) Cockpit Display of Traffic Information (CDTI) that uses such a hybrid concept to control a conflict heading band was developed by NASA Langley and is discussed in [9].

To eliminate false positives when intent information is available, the time at which the intruder vector is specified to change (at the transition point) is used as the look-ahead limit for the state-based prediction that uses the current state. To prevent false negatives, the state from the transition point is used by the probing function beyond this time, and up to the total look-ahead time.

Ability to handle multiple intruders

When there is more than one intruder, the probing algorithm will generate additional areas where a loss of separation is predicted. The amount of the areas that will be displayed are a function of the look-ahead time (i.e. the prediction horizon); the size of the areas is a function of the separation criteria and data uncertainty. For the same traffic density, the amount of areas will increase with an increasing number of aircraft, but the location

of these areas will also be further away from own ship. Because the effect of uncertainty in track and speed increases with an increase in look-ahead time, the size of the area will also increase.



Figure 3: Plan-view conflict probe display, depicting an example scenario with four other aircraft and an impending Loss of Separation.

Figure 3 shows an example of a plan-view conflict probe display in the situation where a loss of separation is predicted on the current path and four other aircraft are in the vicinity. Probe-area A indicates where a loss of separation with aircraft 1 (TRF3-1), flying in from behind, is predicted to occur. The conflict resolution system has computed that a change in track to the left is a feasible solution (indicated by the dashed magenta line, S). Figure 4 shows the same situation, depicted on a HUD with integrated Track-FPA probe. In a latter subsection, addressing the ability of the pilot to assess the situation, we discuss the depicted elements in both displays in more detail.

When computing a manoeuvre to prevent a loss of separation that is predicted to occur at a time T_{AL} ahead, it is important that the solution does not yield a new conflict within a pre-determined time. The approach discussed in [8] does not include a requirement that prevents the situation where a computed resolution almost immediately yields a conflict with another aircraft. It is indicated that in future work the ‘*resolution will be expanded from pair-wise to multi-aircraft situations*’. In [10] it is illustrated how by using an alert limit of T_{AL} seconds, and a search algorithm in the

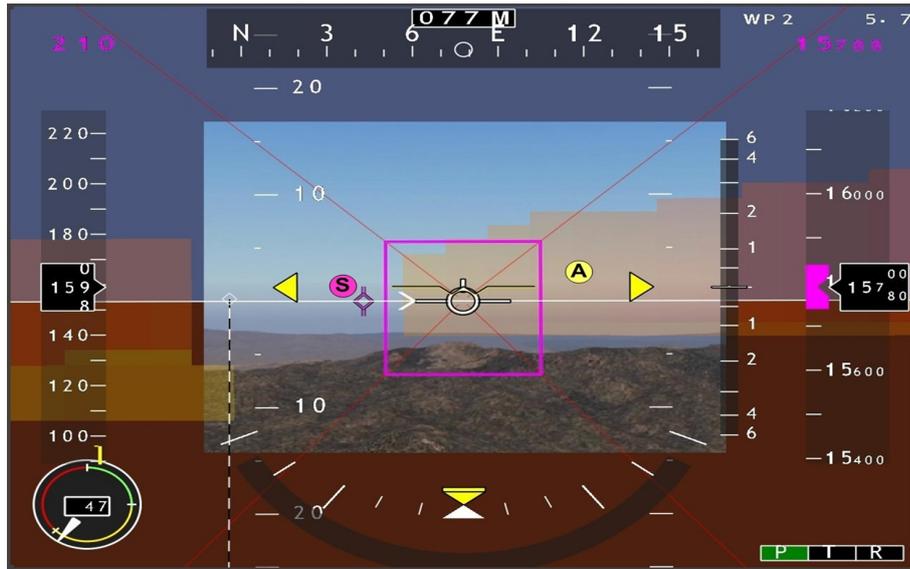


Figure 4: HUD with integrated Track-FPA probe, depicting the situation as displayed in Figure 3, from an egocentric perspective viewpoint.

probing matrix that looks for vectors that are conflict free for at least T_{LA} seconds, a solution is generated that is conflict free for at least $T_{LA} - T_{AL}$ seconds.

Integration of constraints and other hazards

In [11] the design and evaluation of an integrated avionics alerting system is described. Regarding the lack of integration, it is pointed out that *current alerting systems lack a common framework to share intent and integrate and prioritize information*. Furthermore, the lack of integrated strategic information for predictive SA and planning ahead is identified. In [12] it is discussed how conflict probing can provide a common framework for the computation of coordinated conflict avoidance maneuvers that include integration of multiple types of hazards and constraints such as vehicle performance and right-of-way rules. To illustrate this, Figure 5 depicts the elements and information flow of a conflict avoidance concept that uses conflict probing as a framework for integrating data.

The *integrated conflict structure* contains the conflict space, defined by the output of the conflict probes, and additional constraints to the maneuver space. Each cell in the structure represents a specific combination of Track, Flight Path Angle (FPA) and Speed. The value of the cell indicates whether that velocity vector leads to a conflict or whether it has to be

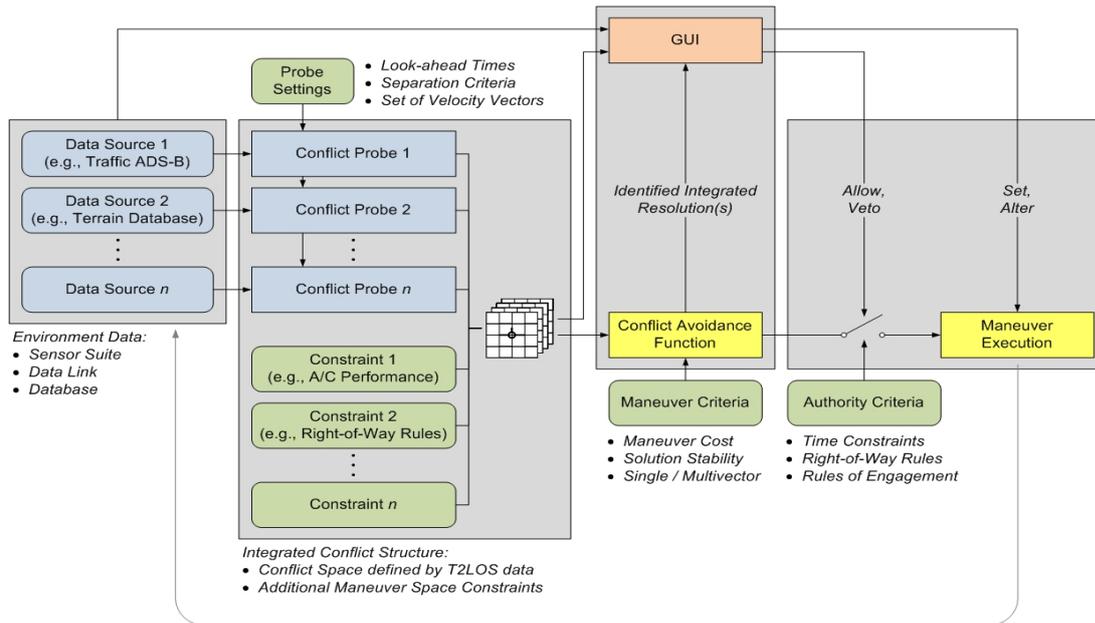


Figure 5: Elements and information flow of an integrated conflict avoidance concept based on probing.

excluded for any other reason. For example aircraft performance characteristics pose a fundamental constraint on the solution space. Procedural constraints such as the right-of-way rules [2] can be regarded as a set of situation dependent ‘no-go’ directions, and can be implemented as such, as one of the structure’s layers.

In the approach presented in [8], performance limitations are specified as constraints for each of the six vectors that are computed. To maximize the available time until a loss of separation can no longer be prevented, conflict resolution should be able to use the full performance capabilities of the UAV, rather than command standard resolution maneuvers designed to accommodate the worst performing class of UAVs. The available 3D space for conflict resolution can be maximized by combining vertical and lateral maneuvers, and utilize the ability to convert the available speed margin relative to V_{min} or V_{max} (excess kinetic energy) into altitude (potential energy). For humans it is almost impossible to maximize the maneuvering performance in this way without violating maneuvering constraints such as angle of attack, stall speed, load factor and bank angle. In [6] we illustrated how information on the maximum safe maneuvering authority is integrated into the conflict prevention/resolution function of a probe using integrated control authority allocation and envelope protection functionality.

Allowable types of manoeuvres

In Figure 1, an example of conflict geometry was used to illustrate the concept of conflict probing from a single dimensional (Track) perspective. For that conflict geometry and separation criteria, the corresponding time to loss of separation (T2LOS) for a range of variations of all three velocity vector dimensions is depicted as a colour coded object in Figure 6a. This 3D data structure is one of the layers of the integrated conflict structure discussed in the previous subsection. The depicted ‘volume’ represents the conflict space in the velocity vector domain, providing a ‘translation’ of the relative motion problem, to a set of velocity vectors that should be avoided. All combinations of own ship Track (Ψ_A), FPA (γ_A) and Speed (v_A) that lie outside of this volume represent possible conflict prevention/resolution maneuvers. Hence, information about the trade-offs and *interdependencies of changes of the velocity vector components* can be readily obtained from the structure, supporting multi-dimensional conflict resolution. Figures 6b-d show ‘slices’ of the 3D data structure for the original speed (v_0), FPA (γ_0) and Track (Ψ_0). For example, Figure 6b represents the conflict space in the Track-FPA domain. This data can be conformally integrated in a HUD or SV PFD (as was illustrated in Figure 4), providing a readily actionable representation of the maneuver space that is available with respect to traffic.

As probing is performed in real-time, the conflict space (i.e. its ‘position’, shape and T2LOS content) is continuously updated while the situation develops. Should intruder aircraft maneuver, this will be reflected by a corresponding change of the conflict space. By storing the probing data as presented here, as T2LOS records for the variations of the own ship velocity vector, the temporal characteristics of potential conflicts are retained, allowing prioritization and timing of avoidance maneuvers.

Degree of coordination

Current collision avoidance systems use the TCAS transponder to coordinate the resolution advisories that are provided. At present, no guidance exists for the coordination of maneuvers between two separation assurance systems. One possibility to enable cooperation between two systems is to provide the intended maneuver as intent information to the other system.

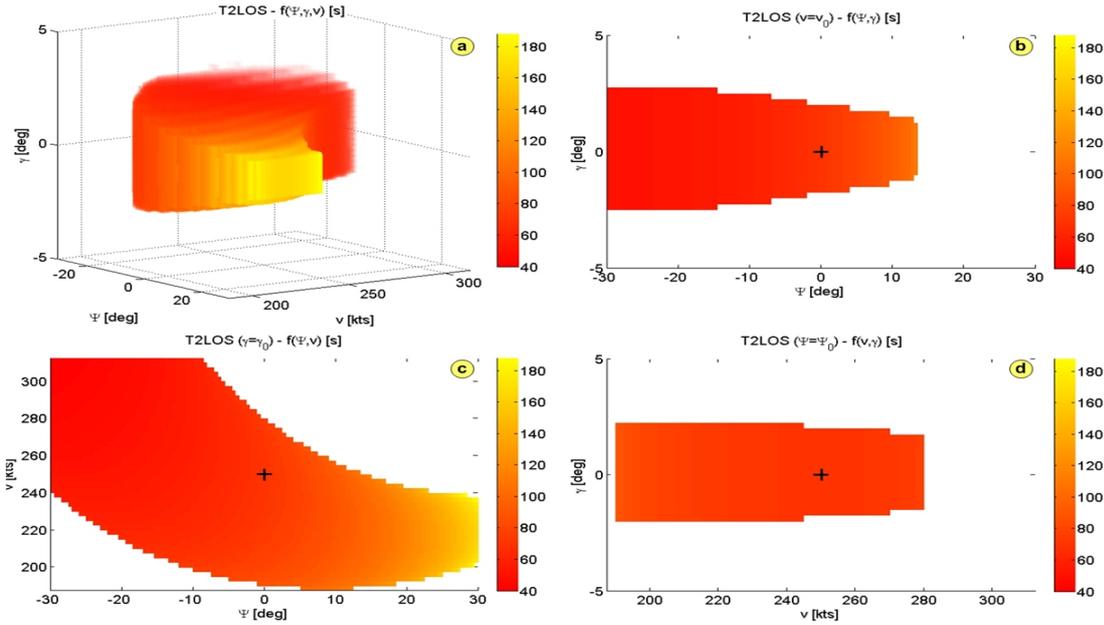


Figure 6: Example conflict space presented in the three domains constituting the flight path vector.

Assumed level of autonomy

In earlier work it has been argued that, although collision avoidance systems for UAS need to have sufficient authority to autonomously execute maneuvers in loss-of-link and/or time-critical situations, all other situations provide for the possibility of pilot involvement in the maneuver selection and execution. This enables a more evolutionary approach to UAS integration, allowing a gradual increase in separation system authority⁵ as these systems mature and the complexities associated with automatic maneuver selection and execution are better understood.

With the probing concept and an appropriate Graphical User Interface (GUI), at the lowest level of system autonomy the pilot can derive the maneuvering decision from the location of the conflict areas and other constraints. At a higher system autonomy level, the identified avoidance maneuvers can either be used as suggestions, serving to support the pilot's decisions, or they can be executed automatically based on predefined authority criteria (triggered by e.g., time-critical or lost link conditions). In [4], the different options for the Level of Autonomy (LOA) are discussed.

⁵Autonomy indicates what the system can do without pilot involvement, authority indicates what the system is allowed to do without pilot consent.

corresponding HUD format.

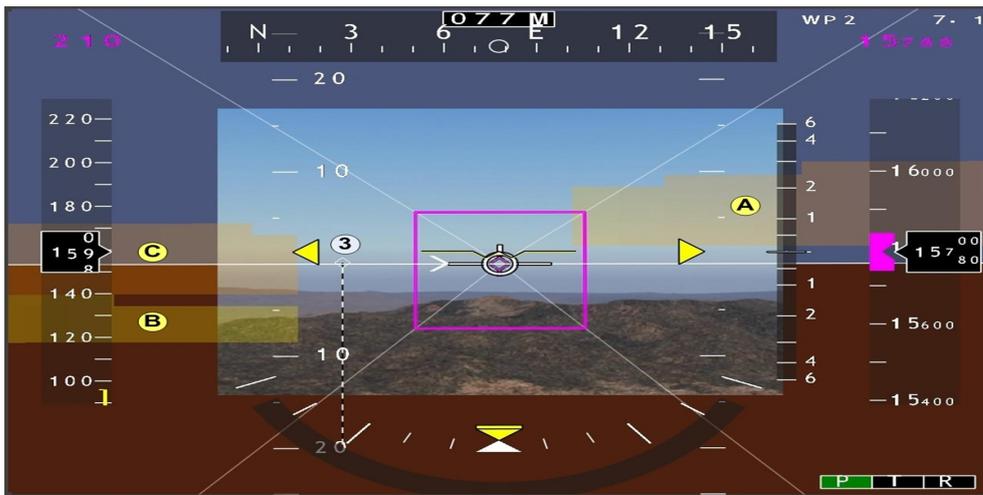


Figure 8: HUD with integrated Track-FPA probe (2-D probe).

The probing data shows that for the current path, no loss of separation is predicted with the other traffic. The proximity of probe area C to the current path provides the pilot with awareness that if he has to depart from the current path, maneuvering to the left can quickly lead to a loss of separation. On the CDTI, probing area D represents the location where a loss of separation with aircraft 4 will occur if the track were changed in that direction. Area C indicates this for aircraft 3. For the current flight path angle, no loss of separation is predicted with aircraft 1 and 2.

The HUD view in Figure 8 shows three probing areas. Area C corresponds to area C in Figure 7, but now presented in the Track-FPA domain. Area D from Figure 7 lies outside the field-of-view of the HUD (represented by the dashed white lines in Figure 7). The CDTI did not show areas A and B because these areas lie respectively above and below the current FPA. Area A represents the directions in which a loss of separation with aircraft 1 from Figure 7 will occur and area B provides this information for aircraft 2.

From the HUD view, the pilot can see that climbing to the right will lead to a loss of separation with the overtaking traffic 1. This cannot be derived from the CDTI, since the altitude where the loss of separation is predicted is more than 1000 ft above the current flight path. This emphasizes the point that the depiction of the location where a loss of separation is expected to occur (level 3 awareness) is more relevant than the depiction of the current location of traffic itself (level 2 awareness).

Technology maturity level

Probing in itself is a concept, not a technology. The data needed for the computations depends on the technology used to obtain information about other traffic and is the same as the data that is needed by comparable concepts such as proposed in [8]. The range of options in terms of display, level of integration and level of autonomy make it a scalable concept, and hence the possibility for implementation is influenced by the choices made regarding these options. The current research addresses a range of options, varying from current state-of-the-art digital avionics used in commercial aircraft to the anticipated open-systems approach used in future UAS Control Stations. The scalability allows for an evolutionary approach, where the level of integration with other hazard data can be increased in a step by step fashion.

Conflict probing is currently being considered as a means to integrate the alerts from multiple hazard detection systems in the context of research performed for the NASA IIFD program [13]. Figure 9 shows the integration of a two-dimensional (Track-FPA) probe into an existing Synthetic Vision System (SVS) Primary Flight Display (PFD) used in this research. The functionality of the existing SVS PFD is represented by the blue block and the probe functionality by the green one. The inputs to the traffic probe comprise own ship and traffic data and the thresholds for the minimal required separation.

Several definitions for these thresholds exist. In [14], the threshold that is used to declare a loss of separation is the Assured Normal Separation Distance (ANSD). Regarding this threshold, it is stated that *‘the ANSD may be altered by the flight crew to mimic the current airspace separation standards. The parameters determine the separation that the CD uses to define a conflict (i.e., violations of separation) and generate alerts’*. The Conflict Detection Zone (CDZ) alerts indicate an expected loss of separation and the Collision Avoidance Zone (CAZ) alerts indicate that a collision situation is imminent. Hence, a second threshold can be defined based on the CAZ criteria. In the UAV Sense and Avoid (SAA) concept being developed [1], a so-called Self Separation Threshold (SST) and a Collision Avoidance Threshold (CAT) are used. The SST is defined as *‘the boundary at which the self separation function declares that action is needed to preclude a threat aircraft from penetrating the collision avoidance threshold,*

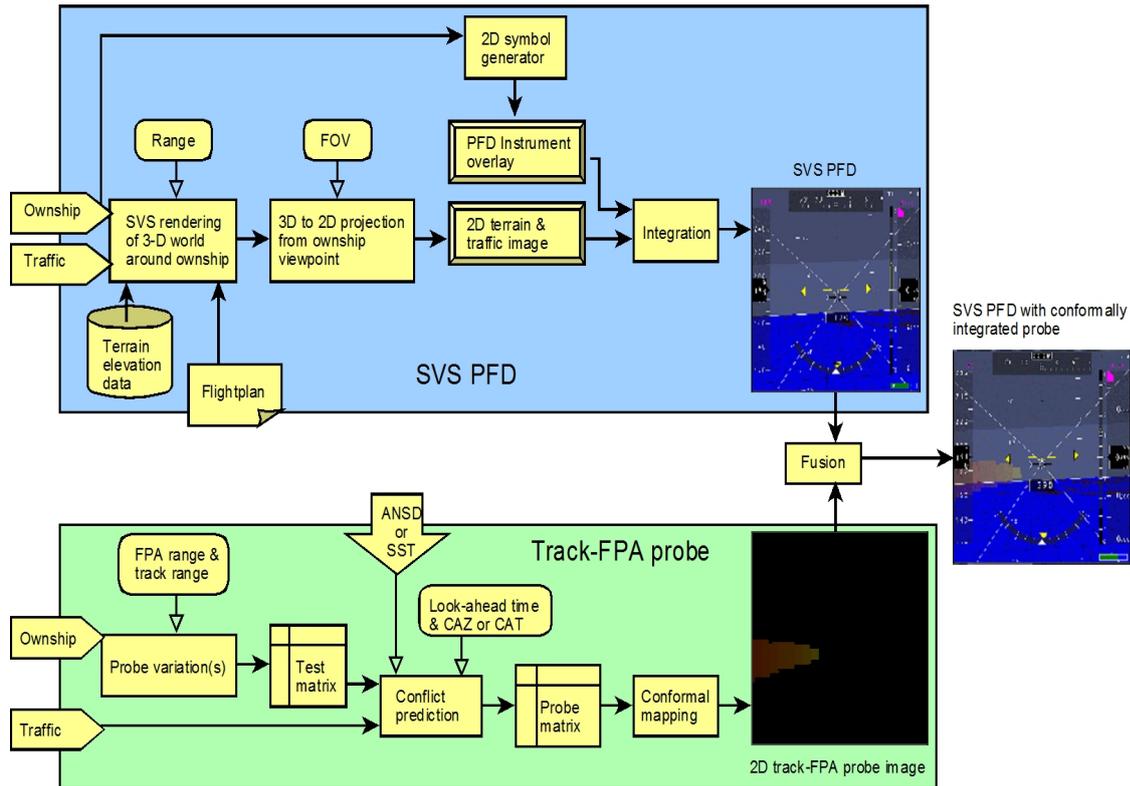


Figure 9: Implementation and integration of directional traffic probe data into an SVS PFD.

thereby maintaining “well clear”’. The CAT is defined as ‘the boundary at which the collision avoidance function declares that action is necessary to avoid a collision and prevent the threat aircraft from penetrating the collision volume’. The ‘collision volume’ is defined as ‘the volume of airspace around the UAS that, if penetrated results in a MAC or NMAC’ [1].

The Test matrix shown in Figure 9 is filled with vectors comprising all combinations of Track and FPA specified in the block FPA range & Track range. These vectors are tested against the traffic in the conflict prediction functions, and the result (time to loss of separation for the particular Track-FPA) is used to update the probe matrix. Using a pre-defined colour coding scheme, the values of time to loss of separation are translated into a colour. The resulting object is conformally integrated into the SVS PFD.

Figure 10 shows how the one-dimensional Track probe is integrated into a CDTI. The inputs to the traffic conflict probe are the same as those in Figure 10. The test matrix now only comprises a range of Tracks for the current FPA.

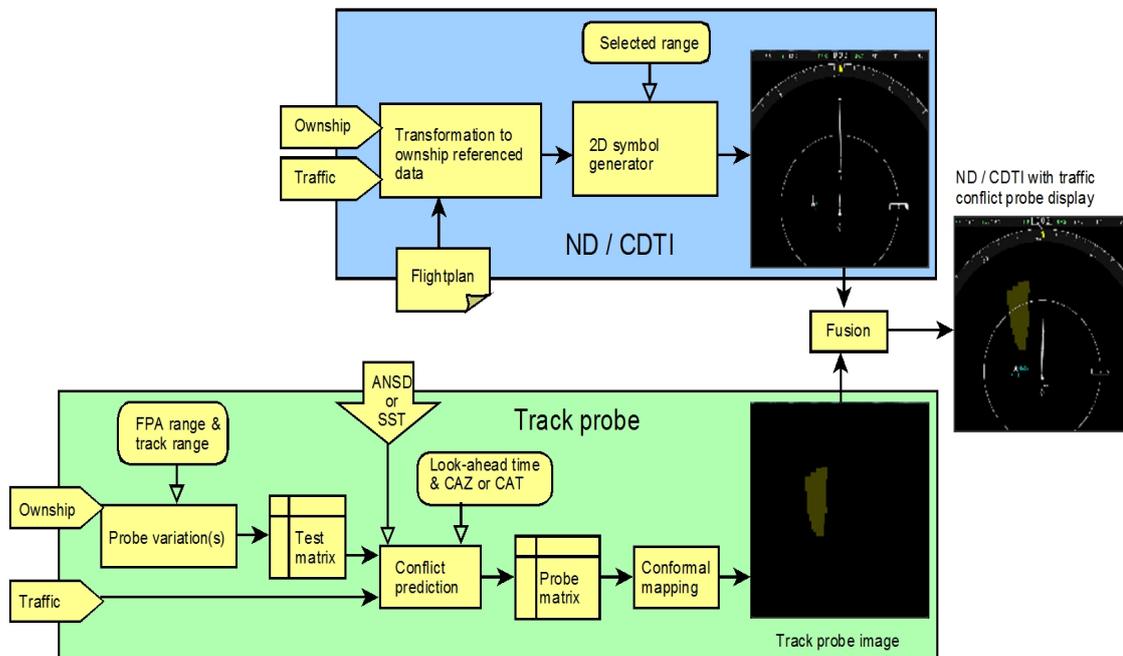


Figure 10: Implementation and integration of Track traffic probe data into an ND or CDTI.

Implementation Options

The SVS PFD and ND implementation presented in the previous section require the presence of an SVS as a baseline system and assume the availability of accurate speed and directional data of other traffic. As illustrated in Figure 5, the probing concept is scalable in terms of probe dimensions, types of hazards, prediction algorithms and look-ahead times. The scalability enables a range of possible implementations, specifically matched to the available data, interfaces and displays. The implementation of the conflict probing concept does not require an SVS as a baseline system. In order to determine the possibilities to integrate a conflict probe function and display into an existing system, it must be determined:

1. Whether and, if so, how the required data can be obtained;
2. How the results from the conflict probe can be integrated into the existing PFD/HUD and ND.

Obtaining the data

Most algorithms for future conflict detection and separation assurance require speed and directional information of the intruders. Typically the availability of ADS-B information is assumed [8]. The FAA's proposed deadline for the mandatory equipping of an ADS-B out capability is 2020, but there is a bill being debated in US congress to mandate ADS-B OUT as early as 2015 [15]. For near term implementation of a conflict probe display as a situation awareness aid this raises the question to what extent basic TCAS information, possibly augmented with information from passive EO or IR sensors can be used.

When obtained through an ARINC 735 bus, the TCAS range data has a resolution of 1/16 nmi [16]. Also, the relative bearing is rather inaccurate, only 9 degrees RMS is required [17]. Hybrid systems that combine output from an EO sensor with TCAS provide a better bearing accuracy. In [18] it is reported: *'The real value of sensor fusion of EO and TCAS tracks can be seen in the improved relative bearing estimate of the EKF, where the fusion of the EO-based line-of sight information eliminates the nearly 10 degree bearing error that was reported by TCAS'*. With the availability of range and bearing measurements from a hybrid system, it becomes possible to estimate the velocity and track of the intruder. The limited range resolution and the bearing inaccuracy will yield a 95% containment interval for the velocity and track estimates that is considerably larger than when this data is obtained through ADS-B. Furthermore, the 95% containment interval will be influenced by conflict geometry. Yet, with adequate filtering, it may be possible to use this concept for the computation of the conflict probe areas. To evaluate the impact of the limited range resolution and bearing inaccuracy on the location of the conflict probe areas, the setup shown in Figure 11 has been created.

Using this setup in the simulation environment, the 95% containment interval of estimated intruder speed and track is determined for a range of conflict geometries. The impact of an error in speed and/or track will increase with an increase in prediction time. To be able to evaluate the effect, the plan-view conflict probe display has been enhanced with the capability to compute the contour of the area for given 95% containment intervals of speed and track. Figures 12 and 13 show a research CDTI with a configuration panel to change the values of the 95% containment

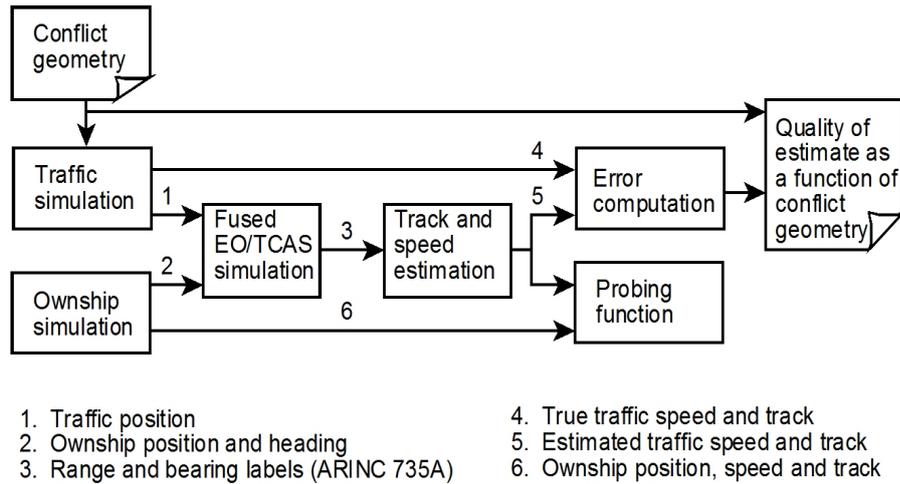


Figure 11: Simulation setup used to determine impact of conflict geometry on accuracy of intruder Speed and Track estimates.

interval for the intruder tracks and speeds. An example scenario involving two other aircraft (TCAS symbols and labels) is depicted. In Figure 12, the 95% containment interval for the intruder track errors is 1 degree and for speed 1 kt. In Figure 13, these values have been increased to 5 degrees and 5 kts.

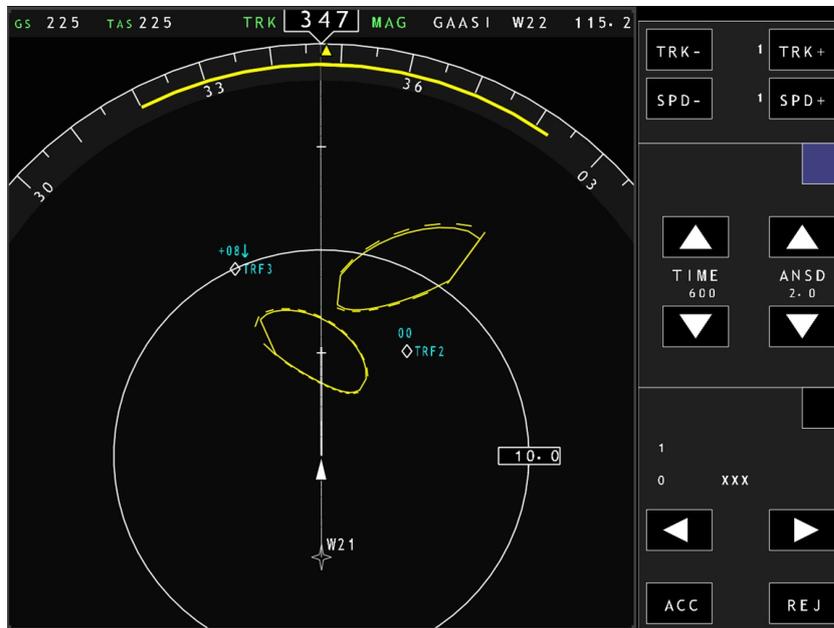


Figure 12: Impact of sensor inaccuracy (dashed) on the predicted conflict area (solid); intruder data uncertainty interval: 1 deg, 1 kts.

The solid contours in Figures 12 and 13 depict the area where loss of

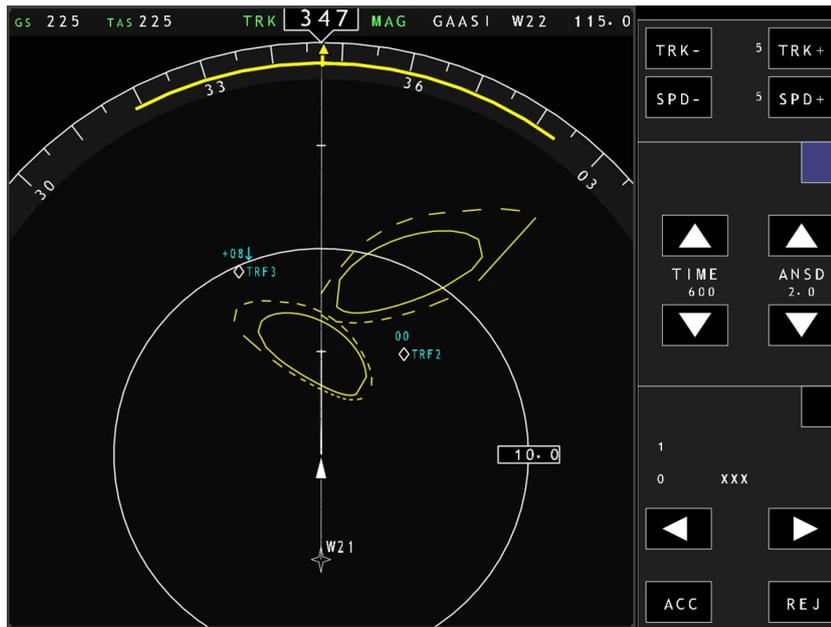


Figure 13: Impact of sensor inaccuracy (dashed) on the predicted conflict area (solid); intruder data uncertainty interval: 5 deg, 5 kts.

separation is predicted based on the true state of the intruder. The dashed line indicates the contour that contains all possible areas that result from different combinations of track and speed that lie within the predefined 95% containment interval. As can be seen from Figure 13, the impact of the uncertainty is asymmetrical. During the next phase of the research it will be determined whether it is possible to use a look-up table containing the data which relates the magnitude of the 95% containment interval to the particular conflict geometry to compute the conflict area.

Integrating the presentation

The demonstrator that has been created using the SVS setup shown in Figure 9 performs the fusion in memory by translating the probing data into a texture which subsequently is conformally mapped into the SVS format. This was possible because the software of the SVS prototype could be enhanced with the required functionality. Alternatively, an SVS or HUD that has the capability to integrate live sensor video can use such a capability to integrate a probing image.

Similar to how the output from TAWS can be used on the weather radar input of an ND for depiction of terrain, the output that contains the probe data as a function of Track angle and Range can be translated to

ARINC 708A [19], as depicted in Figure 14. In this way, the probe function depicted in Figure 10 can be integrated as a separate device, using only existing interfaces for input and output, enabling a one-dimensional traffic probe to be displayed on a conventional ND.

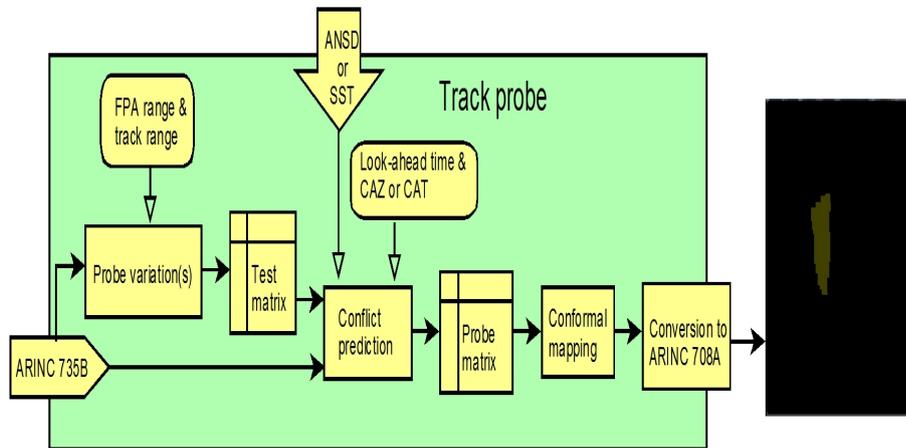


Figure 14: Depiction of conflict probe on ND using WXR input.

Summary and Conclusions

In this paper, conflict probing has been proposed as an element of a UAS self separation concept. Conflict probing consists of predicting the future separation between own ship and hazards for a set of own ship velocity vectors -representing possible combinations of Track, Flight Path Angle (FPA) and Speed- up to a predefined prediction horizon or look-ahead time.

The potential solution space generated by conflict probing comprises all possible vectors that will prevent a loss of separation with a pre-defined time within the particular probing dimension. The dimensionality of the solution space is determined by the dimensionality of the probe. The conflict avoidance vector is obtained through a search in the solution space using a set of pre-defined criteria which can include a cost function, rules of the road and thresholds in time or space. The solutions that can be obtained from the proposed conflict probing concept include those that would be generated by the NEXTCAS system described in [8]. The ability to generate conflict avoidance vectors which remain conflict free for a specified amount of time is obtained through appropriate selection of the alerting time and the look-ahead time.

Possible implementations range from the depiction of status information (the conflict areas) to the computation of a single conflict prevention command. At the lowest level of system autonomy, the pilot can derive the maneuvering decision from the location of the conflict areas and other constraints. At a higher system autonomy level, the identified avoidance maneuvers can either be used as suggestions, serving to support the pilot's decisions, or they can be executed automatically based on predefined authority criteria. This allows a gradual increase in separation system authority as these systems mature and the complexities associated with automatic maneuver selection and execution are better understood.

The probing concept is scalable in terms of probe dimensions, types of hazards, prediction algorithms and look-ahead times. This scalability enables a range of possible implementations, specifically matched to the available data, interfaces and displays.

References

- [1] S. George, presentation at FAA UAS Sense and Avoid Workshops, April 2009.
- [2] FAA, "Right-of-way rules: Except water operations," 14CFR91.113, 2001.
- [3] J.M. Davis and R.J. Kelly, "RNP Tunnel Concept for Precision Approach with GNSS Application," *Proc. 49th Annual Meeting ION*, Cambridge, Massachusetts, pp. 135-154, 1993.
- [4] J. Tadema and E. Theunissen, "A Concept for UAV Operator Involvement in Airborne Conflict Detection and Resolution," *Proc. 27th Digital Avionics Systems Conference*, IEEE/AIAA, 2008.
- [5] J. Tadema and E. Theunissen, "Exploring the potential of energy and terrain awareness information in a Synthetic Vision display for UAV control," *Proc. SPIE Enhanced and Synthetic Vision 2009*, Vol. 7328, pp. 9.1-9.13, 2009.
- [6] A.A. Lambregts, J. Tadema, R.M. Rademaker and E. Theunissen, "Defining Maximum Safe Maneuvering Authority in 3D Space Required for Autonomous Integrated Conflict Resolution," *Proc. 28th Digital Avionics Systems Conference*, IEEE/AIAA, 2009.

- [7] J. Tadema and E. Theunissen, “Design and Evaluation of a GUI for Operator Involvement in Airborne Conflict Detection and Resolution,” *Proc. 27th Digital Avionics Systems Conference*, IEEE/AIAA, 2008.
- [8] R. Chamlou, “Future Airborne Collision Avoidance - Design Principles, Analysis Plan and Algorithm Development,” *Proc. 28th Digital Avionics Systems Conference*, pp. 6E2.1 - 6E2.17, 2009.
- [9] R. Barhydt, T.M. Eischeid, M.T. Palmer and D.J. Wing, “Use of a Prototype Airborne Separation Assurance System for Resolving Near-Term Conflicts During Autonomous Aircraft Operations,” AIAA 2003-5401, 2003.
- [10] J. Tadema and E. Theunissen, “Integrated Conflict Prediction and Multi-Dimensional Prevention/Resolution,” *Proc. AIAA Infotech @ Aerospace Conference and AIAA Unmanned...Unlimited Conference*, Seattle, pp. 1-12, 2009.
- [11] M.C. Dorneich, P.M. Ververs and M.D. Good, “Design and evaluation of an integrated avionics alerting system,” *Proc. 20th Digital Avionics Systems Conference*, IEEE/AIAA, 2001.
- [12] J. Tadema and E. Theunissen, “An integrated conflict avoidance concept for aviation,” *Proc. 28th Digital Avionics Systems Conference*, Orlando, pp. 5C6.1-5C6.11, 2009.
- [13] M. Uijt de Haag, R. Bezawada, S. Vana, K. Venable, T. Schnell, M. Cover, D. Haverkamp, T. Etherington, M.P. Snow, E. Theunissen and R.M. Rademaker, “Design, Development, Verification and Validation of an Integrated Alerting and Notification Function for an Intelligent Integrated Flight Deck - Phase 1,” NASA CR, *to be published*.
- [14] RTCA, “Minimum Operational Performance Standards (MOPS) for Aircraft Surveillance Applications System (ASAS),” DO-317, 2009.
- [15] S. Bill, 1451 FAA MODERNIZATION AND SAFETY IMPROVEMENT ACT.
- [16] ARINC, “Mark 2 Traffic Alert and Collision Avoidance System (TCAS),” *ARINC Characteristic 735A-1*, Aeronautical Radio Inc, Annapolis, 2003.

- [17] RTCA, “Minimum Operational Performance Standards for Traffic Alert and Collision Avoidance System II (TCAS II) Airborne Equipment,” DO-185A, 1997.
- [18] V. Raska, O. Shakernia, S. Graham, W. Chen, A. White and J. Zvanya, “Sense And Avoid (SAA) Flight Test and Lessons Learned,” *AIAA Infotech*, paper 3003, 2007.
- [19] ARINC, “Airborne Weather Radar with Forward Looking Windshear Detection Capability,” *ARINC Characteristic 708A-3*, Aeronautical Radio Inc, Annapolis, 1999.

Crew Resource Management Training and Research in a Military Organization

Hilde van Ginkel

Introduction

Military operations, whether in the field, in the air or at sea, are characterized by their complexity, the high risks that are involved and the specialized teams who have to reach their goals under these circumstances. The improvement of high-tech equipment and the increasing complexity and uncertainty of the environment necessitates flexibility of the military involved. Cooperation with international teams in a specific environment (e.g. Afghanistan) complicates the operations even more. The big challenge is to train both individuals and teams to operate in such an environment.

Crew Resource Management (CRM) training is one of many different trainings to develop skills necessary to face the variety of situations described above. In this specific training, students become aware of the importance of team-work and the skills involved in working as a team under high-risk conditions. CRM skills are supportive skills, they are not a goal in themselves. Improvement of these skills helps to improve the effectiveness, efficiency and safety of a team operating in a high-risk environment.

This article consists of two parts. In the first part we will show the importance of CRM in military operations and the necessity to lay a scientific foundation for CRM training and coaching. In the second part of this paper the research done within the Netherlands Defence Academy (NLDA) will be explained. One of the specific models in this research will be highlighted and some preliminary results will be shown. How these results can help improve CRM training and therefore help to prepare military service men and women for their missions will be explained in the discussion.

CRM in a military environment

2005, Afghanistan

The Netherlands Air Force Chinook-D104 helicopter was on transit from Mazar-e-Sharif to Kandahar Air Field with personnel and supplies on board. During the flight, the pilot in command decided to take a shortcut through the mountains to save time and fuel. The shortcut led them through a mountainous area into a cull valley with a high mountain ridge at the end. Due to lack of preparation and miscommunication during the flight a faulty indication of the height of the ridge was presumed. Although the pilot often asked and received feedback, the helicopter (technically perfect) had to make a sharp avoidance manoeuvre (it could not climb to the height of the ridge) and crashed into the mountains. The crash caused the loss of the helicopter and injured one of the people transported on board (see Figure 1).



Figure 1: Crashed Chinook helicopter in Afghanistan.

2006, Curaçao

During a demonstration frigate NLHMS Van Speijk with visitors on board collided with support ship NLHMS Pelikaan on open water near Curaçao (see Figure 2). During the collision a Lynx helicopter with 3 persons on a

rope was forced to fly away from the ship to avoid danger to the visitors, thereby endangering the personnel on the rope. Luckily, there were no casualties and the ships were both able to sail back to the harbour. Investigation showed the cause of this accident was due to human errors within the team sailing the Van Speijk and due to miscommunication during the preparation of the demonstration. The overall conclusion from the board of investigation was that the accident happened due to a lack of CRM.



Figure 2: Left: HNLMS Pelikaan; middle: HNLMS van Speijk; right: Fastrope under a Westland Lynx Helicopter.

Both examples show the danger of human error in military situations. All equipment was working perfectly, the crew was experienced, the weather conditions were reasonable. Still, when looking at both preparation and execution, these were accidents waiting to happen.

The collision between HNLMS Van Speijk and HNLMS Pelikaan prompted the Commander in Chief of the Royal Netherlands Navy to demand improvement of CRM within the naval fleet. This was the start signal for the NLDA to consider its current Bridge Resource Management (BRM) training and propose a plan for both an update of BRM and development of a new (CRM) program from a different perspective.

Both BRM training of the Naval Academy and the Aircrew Coordination Training (ACT; predecessor of CRM) of the Netherlands Naval Fleet Air Arm were developed in 1993 with at its core the improvement of teamwork in the real world. The trainings given were an answer to the questions from practice based on experience, often with very limited scientific background. Both BRM and ACT offer practical training, focused on the specialized community (i.e. BRM was training for bridge officers, ACT

was training for the crew of Westland Maritime Lynx Helicopters and P3C Maritime Patrol Aircraft of the Netherlands Naval Fleet Air Arm). SAS Flight Academy Stockholm developed BRM for seafarers in international collaboration with several pilot services, merchant shipping schools and naval academies. The instructor courses were given at the SAS Flight Academy (which facilitated both Cockpit and Bridge Resource Management); the NLDA was certified to train their own students in BRM. ACT was developed by the US Naval Fleet Air Arm, the courses given in Pensacola, USA. Although both trainings had the same goal, enactment was different. BRM was a course of one week, divided in workshops and ending with a simulator sortie. ACT was a week course as well, but the theory and practice were (and are) afterwards coached throughout the entire education period. It was incorporated in the training assessment sheets as well, forcing both instructors and students to acknowledge the importance of ACT behavior and train the appropriate skills. Therefore, after their education, the entire flight crew 'breathed' ACT.

Between 1993 and 2010, the ACT developed into CRM training. Until 2004, the Royal Netherlands Naval Fleet Air Arm followed this development and updated their assessment (the update from error avoidance to error management was made). BRM was also updated in 2004, introducing the module 'threat and error management'.

Even with these trainings, accidents due to human error occurred (these trainings were in use when the accidents described above happened). Therefore, a new approach to CRM was made. In order to structure the present training and coaching within the Naval Fleet Air Arm and start coaching in a similar way at the NLDA, education at the NLDA was reformed (based on the same approach as within the Netherlands Naval Fleet Air Arm, applied to the sailing community) and scientific research was started to combine practical expertise with scientific knowledge of topics related to CRM.

Research

Originally, this research was defined by the practical question of improvement of CRM training within the Royal Netherlands Naval Fleet. At the NLDA, this question was translated in a two-way program: the BRM course had to be remodeled and updated, and the course had to have

a scientific base, founded in scientific research.

The method for training BRM was by using workshops and case studies. These didactic tools proved to be effective and were readopted in the new course. The content was based on a different training: the most up-to-date version of the CRM course as offered by the US Naval Fleet Air Arm in Pensacola. This training is on a high abstract level and thus applicable for a wide range of students. Therefore the name of the course changed from BRM to CRM. “CRM can be defined as a management system which makes optimum use of all available resources - equipment, procedures and people - to promote safety and enhance efficiency of [flightdeck] operations” [1]. The basic content of the course is based on the ‘skills’ described in the team competence model of Cannon-Bowers, Tannenbaum, Salas and Volpe [2] and includes the following issues:

- Decision Making
- Assertiveness
- Mission Management
- Communication
- Leadership
- Adaptability/ Flexibility
- Situation Awareness
- Fatigue
- Error management
- Stress

The topics Fatigue and Stress are not skills. However, they are causes for human limitations, and therefore related to CRM.

Research is done into different aspects of CRM (i.e. [3-6]). All the different aspects of CRM are not yet measured together in a field team working in a complex high-risk environment.

The goal of this research is to determine the influence of individual and team characteristics on CRM behavior and team performance in a team working in a high-risk environment. The influence of the situation and interventions on CRM behavior and team performance will also be determined.

Practically, this information is very useful as support of CRM training and coaching. The scientific relevance lies in joining individual and team characteristics of teams working in a high risk environment in a military (non-academic) setting, combining models from different types of research in one new model (see Figure 3). This research builds a bridge between scientific knowledge and practical application in the military field.

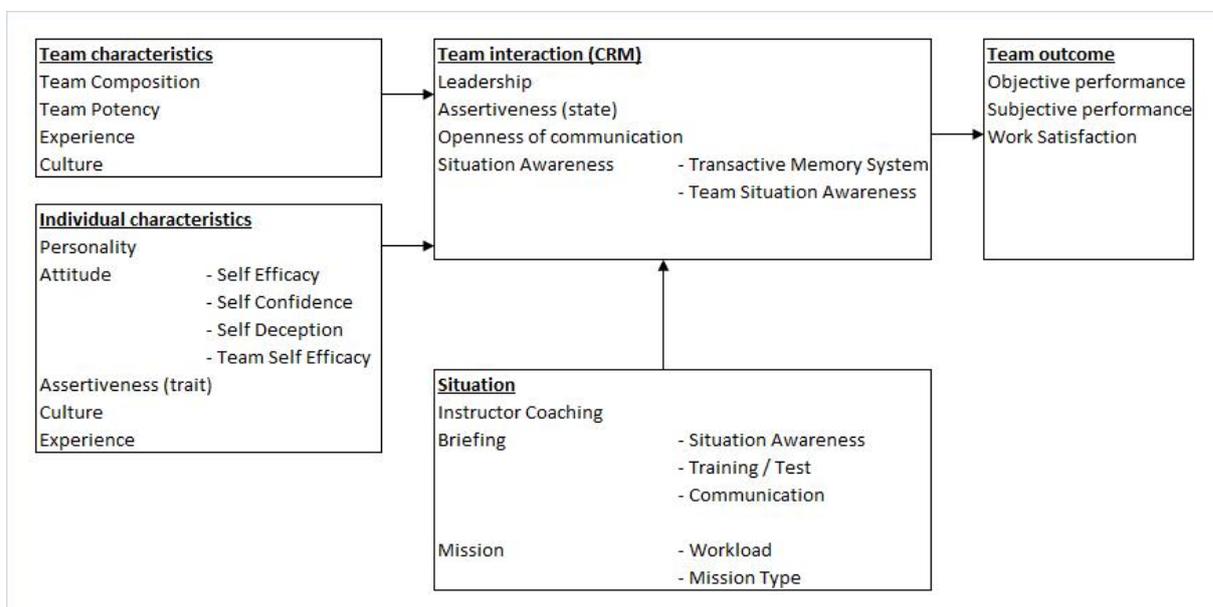


Figure 3: CRM Construct model (CC-model).

The CC-model of Figure 3 shows all the topics related to CRM that are discussed in this research. As can be seen there are 3 major questions in this research:

1. What is the influence of team characteristics on team interaction and team outcome?
2. What is the influence of individual characteristics on team interaction and team outcome?

3. What is the influence of the situation on team interaction and team outcome?

Each of the research questions consists of several models containing the aspects described in the CC-model. One of the aspects specifically important in military settings is the use of specialized teams (heterogeneous team composition). In a team with specialized individuals team knowledge is important for team performance. One construct especially relevant for understanding team knowledge processes is Transactive Memory System (TMS). Therefore, this is a very important construct of CRM.

The Influence of Team composition on Transactive Memory System and Performance

The increasing complex environment, in which service teams operate during missions, necessitates specialization of individual team members in order to execute the mission objectives. This specialization and differentiation of team members (team composition - team specialization) adds a new challenge into the decision making process of the team: besides knowledge of the mission field and situation, knowledge about both the specialization of the team members (TMS) and what kind of team composition to use during different mission types is vital. A TMS is the cooperative division of labor for learning, remembering, and communicating relevant team knowledge [7]. It is a cognitively interdependent system for encoding, storing and retrieving information that combines the knowledge possessed by individual members with a shared awareness of who knows what [8]. Recent research indicates the effect of TMS on team effectiveness, decision making and situation awareness [9,10]. Empirical evidence of dangers of ‘illusory’ TM on team performance in ambiguous situations shows the importance of TMS in military teams working in unknown, complex conditions [11].

Figure 4 shows a part of the CC-model, indicating the influence of several Team characteristics on one of the CRM skills (TMS) and Performance.

“TMSs [can] be discerned from the differentiated structure of member’s knowledge (specialization), member’s beliefs about the reliability of other member’s knowledge (credibility), and effective, orchestrated knowledge processing (coordination)” [7]. Therefore, to examine the impact of team composition on TMS and of TMS on performance, we need to examine the

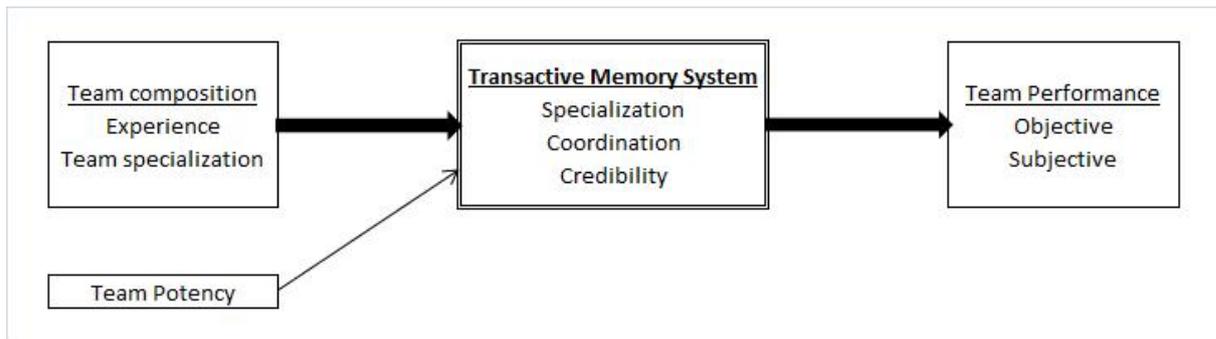


Figure 4: Influence of team composition and team potency on TMS and team performance.

impact of team composition on each of these constructs, and the impact of each of these constructs on performance.

Relationship between Team composition and TMS

A team is “two or more people with different tasks who work together adaptively to achieve specified and shared goals” [12]. The environment in which a team has to achieve its goal is of importance for the team composition. Military operations often take place in a high-risk environment. A high-risk environment is defined as an operating area with a high level of uncertainty. As the situation becomes increasingly complex, a team has to have a wide range of knowledge and skills to reach the mission objectives. Therefore, team members will have their own specialization in order to meet this goal. This specialization reduces the redundancy within the team, while increasing the flexibility of the team to face uncertain and unpredictable situations.

As specialized team members each have unique knowledge necessary for goal achievement, coordination of the resources in the team is required in order to create team knowledge. Team composition-Specialization will therefore be positively related to TMS coordination. With highly specialized team members it is more difficult to know exactly what another team member knows. This means that TMS specialization will decline with increased differentiation of team composition-Specialization (TCS) [13].

When a team is more specialized, members have to rely on each other’s specific knowledge. Because individual members cannot judge other member’s specific knowledge, they have to rely on their credibility. Whether someone is credible depends on several factors. Expertise is important for

credibility [13], but it can also be a self-fulfilling prophecy: when someone believes he is credible, he will act that way (i.e. share information), inviting team members to think he is credible [14]. In teams with more differentiated specialization, individual members are forced to act credible, as there is no redundancy. Therefore, team composition will be positively related to TMS credibility.

Hypothesis 1a: Team composition-specialization will be negatively related to TMS specialization.

Hypothesis 1b: Team composition-specialization will be positively related to TMS coordination.

Hypothesis 1c: Team composition-specialization will be positively related to TMS credibility.

When team members have worked with each other more often, they know each other's specific knowledge better than when they work together as a team for the first time. Members of an experienced team also know the reliability of each other's specific knowledge and therefore have better credibility of each other. When experience is limited, the complexity of TM can create confusion; especially when expertise is in dispute and important information falls through the cracks [13]. Experts (team members with specific knowledge) relate to the structure that is at the core of the situation instead of the problem itself. Therefore, they will tend to direct and coordinate when it is appropriate [15].

Hypothesis 2a: Experience will be positively related to TMS specialization.

Hypothesis 2b: Experience will be positively related to TMS coordination.

Hypothesis 2c: Experience will be positively related to TMS credibility.

Team potency is the mutual confidence in the effectiveness of the team [16]. Transactive Memory is an antecedent of team potency [17]. Practice also indicates the reverse: confidence in team effectiveness can be the effect of team composition and team experience, which will result in positive TMS credibility and positive TMS specialization. As confidence in the effectiveness can result in incorrect assumptions about the coordination,

the relation between team potency and TMS coordination will be two fold [13].

Hypothesis 3a: Team Potency will be positively related to TMS specialization.

Hypothesis 3b: Team Potency will be related to TMS coordination.

Hypothesis 3c: Team Potency will be positively related to TMS credibility.

Relationship between TMS and Team Performance

When team members with distinct roles have an overlapping knowledge amongst themselves, this causes redundancy in information. In a specialized team, team members are more efficient in cognitive processing on their specific knowledge, as only the individual assigned to a particular expertise attends to the relevant information. This frees up other team members to concentrate on their specific tasks and improves information processing in the entire team, resulting in better team performance [14].

Coordination is critical for team performance, and effective TMS will only come from effective coordination of team members. TMS coordination helps in increasing the storage capacity of the group and makes retrieval more efficient [14]. This will improve team performance.

It is not exactly understood how credibility improves performance. It has been found that individuals perceived as experts engage in more information seeking than perceived non-experts. They actively share their expertise as well as engage in seeking out unique information held by minority members [14]. Therefore, a positive relation between TMS credibility and team performance is expected.

Hypothesis 4a: TMS specialization will be positively related to team performance.

Hypothesis 4b: TMS coordination will be positively related to team performance.

Hypothesis 4c: TMS credibility will be positively related to team performance.

Methods

The hypotheses are tested during real time training and practice session. These sessions take place on the Maritime Westland Lynx Full Mission Flight Trainer (FMFT) at Maritime Air Base De Kooy in Den Helder. This high-end simulator contains a Maritime Westland Lynx Helicopter Cockpit and can be set in motion. The instructor manages the scenario and the aircraft from the ‘instructor station’ (see Figure 5).



Figure 5: Left: Maritime Westland Lynx Full Mission Flight Trainer (FMFT); middle: FMFT instructor station; right: FMFT cockpit view.

The participants were 61 pilots (operational (N=13), students (N=4) and instructors (N=15)), TACTical COOrdinators (operational (N=9), students (N=1) and instructors (N=6)), SENSor OPERators (operational (N=3), students (N=1) and instructors (N=6)), technicians (operational (N=2)) and Test Flight Engineers (operational (N=1)), average age $m = 32$, 60 male, one female, from the Maritime Lynx community of five countries (Netherlands, Germany, Denmark, Portugal, Norway). The team size varied, with an average member total of 2.2, flying a total of 29 flights in the simulator. The teams differed in composition and experience. All the flights contained a briefing directly before the flight and a debriefing directly after the flight. The questionnaire was administered after the flight before the debriefing. No explicit reference was made about TMS or Team Potency. When participants completed the questionnaire, they referred to the crew (Oxford Compact Dictionary defines crew as “a group of people who work on and operate a ship, boat, aircraft, or train”) they flew with during the simulator flight. The questionnaire was designed to be completed anonymously. Participants were ensured of confidentiality.

Measures

Transactive Memory was measured using the scale developed by Lewis [7] for measuring this process in field studies. This scale contains 15 items designed to assess the three constructs of TMS (specialization, coordination, credibility). Each item was scored on a five-point Likert-type scale ranging from one (strongly disagree) to five (strongly agree).

Team Potency was measured using an adaptation of the scale developed to assess Team Potency of Civil Aviation Crews [18]. The scale contains three items designed to assess team confidence and team spirit. The responses were given on a five-point Likert-type scale ranging from one (strongly disagree) to five (strongly agree).

Experience for individuals is defined by the amount of flight hours in a Westland Lynx Helicopter cockpit. Flight hours in the FMFT count as flight hours in a Lynx helicopter. As all individual flight hours are in crew composition, experience in this research is defined as the mean of total individual Flight Hours of the crew members.

Objective Performance was measured using the instructor evaluation. The instructor assessed team performance on three issues: navigation (Pnav), procedures (Pproc) and teamwork (Pteam). The responses were given on a four-point scale (O = poor, S- = below standard, S = standard, S+ = above standard).

Subjective Performance was measured using an adaptation of the scale developed to assess perceived team performance of Civil Aviation Crews. The scale contains three items designed to assess perceived team performance. The responses were given on a five-point Likert-type scale ranging from one (strongly disagree) to five (strongly agree).

Preliminary Results

The data shown are for a small sample as the assessment and data collection are currently in progress: all data are preliminary results, which can only give indications for answers on the hypotheses. Therefore, significant results and trends should be used cautiously. As there were 6 sorties with 3 persons on board, results for the third person were too limited to include in this analysis.

For all data analysis, the Statistical Package for Social Sciences (SPSS 16.0) was used. Before analysis, all scales were checked on internal consistency (see Table 1). The table shows an average alpha for Specialization. As there was a different outlier for the Right Seat (RS) and the Left Seat (LS), no items were removed from the lists.

Table 1: Reliability analysis for all scales.

			Cronbachs alpha
TMS	Specialization	RS	0.55
TMS	Specialization	LS	0.62
TMS	Coordination	RS	0.86
TMS	Coordination	LS	0.77
TMS	Credibility	RS	0.60
TMS	Credibility	LS	0.78
Team Potency		RS	0.80
Team Potency		LS	0.80

Combined RS/LS TMS scales of specialization, coordination and credibility were created: TMS specialization: $\alpha = .61$; TMS credibility: $\alpha = .68$; TMS coordination: $\alpha = .87$. Also, a total RS/LS Team Potency scale was made: $\alpha = .82$. A total RS/LS scale of experience was not created: $\alpha = -.34$.

The three objective performance measures each assess a different part of Team performance ($\alpha = .31$). Therefore, the 3 objective performance measures are analyzed separately. A total scale of subjective performance (RS: $\alpha = .77$; LS: $\alpha = .50$) was created: $\alpha = .80$.

Hypothesis 1

There is a significant relationship between TCS and TMS Specialization, $r = .32$, $p(\text{one tailed}) < .05$.

There is no significant relationship between TCS and TMS Coordination, $r = .10$, $p(\text{one tailed}) > .05$.

There is a significant relationship between TCS and TMS Credibility, $r = .34$, $p(\text{one tailed}) < .05$.

Hypothesis 2

There is no significant relationship between experience and TMS Specialization, RS: $r = .27$, $p(\text{one tailed}) < .10$; LS: $r = -.22$, $p(\text{one tailed}) > .05$.

There is no significant relationship between experience and TMS Coordination, RS: $r = .08$, $p(\text{one tailed}) < .35$; LS: $r = -.08$ $p(\text{one tailed}) > .05$.

There is no significant relationship between TCS and TMS Credibility, RS: $r = .10$, $p(\text{one tailed}) < .30$; LS: $r = -.08$ $p(\text{one tailed}) > .05$.

Hypothesis 3

There is no significant relationship between Team Potency and TMS Specialization, $r = .23$ $p(\text{one tailed}) > .05$.

There is a significant relationship between Team Potency and TMS Coordination, $r = .79$, $p(\text{two tailed}) < .01$.

There is a significant relationship between Team Potency and TMS Credibility $r = .65$, $p(\text{one tailed}) < .01$.

Hypothesis 4

There is no significant relationship between TMS Specialization and Objective Performance: Pnav, $r = .17$, $p(\text{one tailed}) > .05$; Pproc, $r = .11$, $p(\text{one tailed}) > .05$; Pteam, $r = -.14$, $p(\text{one tailed}) > .05$.

There is no significant relationship between TMS Coordination and Objective Performance: Pnav, $r = .10$, $p(\text{one tailed}) > .05$; Pproc, $r = .09$, $p(\text{one tailed}) > .05$; Pteam, $r = -.07$, $p(\text{one tailed}) > .05$.

There is no significant relationship between TMS Credibility and Objective Performance: Pnav, $r = -.05$, $p(\text{one tailed}) > .05$; Pproc, $r = .18$, $p(\text{one tailed}) > .05$; Pteam, $r = -.09$, $p(\text{one tailed}) > .05$.

There is no significant relationship between TMS Specialization and Subjective Performance, $r = -.09$, $p(\text{one tailed}) > .05$.

There is a significant relationship between TMS Coordination and Subjective Performance, $r = .42$, $p(\text{one tailed}) < .05$.

There is a significant relationship between TMS Credibility and Subjective Performance, $r = .30$, $p(\text{one tailed}) < .05$.

Conclusions

Since the data collected represent a small amount of simulator flights, the results can give implications for answers on the hypotheses, but no final conclusions can be made.

The hypothesis that Team Composition Specialization is negatively related to TMS Specialization is not confirmed. Instead, the data indicate a positive relationship between TCS and TMS Specialization. Although the result is significant (p (one tailed) $< .05$), the correlation is .32, indicating there are more factors of influence on TMS Specialization than TCS. The result can be explained by looking at the items of TMS Specialization. Of these 5 items, 4 items are about specialized knowledge, and 1 item is about the knowledge content of the other crew member. Therefore, the results indicate that the crewmembers were aware of their colleagues' expertise in a different area than their own and confirmed the necessity of that expertise in the flight. Whether the crew knew the contents of each other's expertise is neither confirmed nor rejected by these results.

The expected positive relationship between TCS and TMS coordination is not confirmed. This indicates that team coordination does not vary according to the composition of the team. This can be due to the setting in which the measurement took place. There are strict procedures about communication and task handling in the cockpit crew of a Lynx Helicopter. Therefore, coordination is fixed in procedures. This hypothesis might be confirmed when looking at a specific flight type in which the tasks between the pilot and pilot-non-flying, tacco or sensop differ widely. There is not enough data available to examine this explanation.

The hypothesis of a positive relationship between TCS and TMS Credibility is confirmed with these results. However, the correlation $r = .34$ is indicating there are more factors of influence on TMS Credibility. This cannot be due to experience of the team, as the hypothesis of a positive relationship between experience and TMS is not confirmed.

The results also show no trends that indicate any correlation between experience and TMS specialization, TMS coordination or TMS credibility. This can be due to the way experience is measured in this research. In the simulator, team members vary with each flight: no experienced flight crews train in the simulator after flying operationally. Therefore, experience is

measured as the mean of total individual flight hours of all crew members, reasoning that all flight hours are in crew connection which indicates experience in teamwork. It might be that measurement in real flights will show different results on this hypothesis. Therefore, this hypothesis cannot be rejected.

The results show no indication of a relationship between TMS and objective performance. However, the results indicate a significant relationship between TMS Coordination and subjective performance ($r = .42$, p (one tailed) $< .05$) and between TMS Credibility and subjective performance ($r = .30$, p (one tailed) $< .05$). The measures of objective and subjective team performance assessed different aspects of performance, which explains the difference in results: the objective performance measure assesses navigation, procedural and teamwork performance, while the subjective performance measure assesses whether everything went according to plan. The results show no relationship between TMS and objective performance indicating no relationship between TMS and the flight criterion. However, as TMS Coordination and TMS Credibility are significantly related to subjective performance, there is indication of a relationship between TMS coordination and perceived performance and TMS credibility and perceived performance. This indicates that when a team perceives good coordination and credible colleagues, they perceive good performance. As the results show no clear relationship between TMS and performance, further research is needed to explain this issue.

Although these preliminary data show some promising results, there is an issue that should be considered: team size is very limited in this research. As the cockpit of a Westland Lynx Helicopter is small, the operational team usually consists of 2 persons, with a third person (sensop) in the cabin when necessary. This simplifies team interaction as there are relatively few resources to manage. In testing complicated models like the CC-model this is an advantage. However, it should be taken into account when applying the results to larger teams.

Discussion

The results of this research, together with field experience can provide a solid base for the design of CRM training and coaching. The model highlighted in this article as part of the CC-model explains and clarifies

TMS. It indicates that team composition and team potency are related to TMS and that TMS is related to (subjective) performance. This means that these topics should be part of CRM training for specialized teams that have to operate in a complex, high-risk environment.

Within CRM training, the results can support the Situation Awareness and Mission Analysis sections. Situation Awareness can be defined loosely as knowing what is going on around you (CRM training reader: the degree of accuracy by which one's perception of the current environment mirrors reality). One's perception starts with information gathering and selection. In a team with differentiated knowledge, this includes knowing what information can be provided by other team members. This is where TMS is of importance.

Mission Analysis is defined as the ability to make short-term, long-term and contingency plans and to coordinate, allocate and monitor crew resources (CRM training NLDA). This occurs during three phases: briefing (pre-mission planning), Mission monitoring and updating and debriefing (post-mission review). In specialized teams, the planning of tasks within the crew prior to the mission, the allocation, adaptation and accomplishment of the mission and the review afterwards are all influenced by the correct allocation of specializations within the team at the appropriate moments. Therefore, within CRM training, TMS can be incorporated in the Mission Analysis section.

The results can also be used in practice. For example, when the preliminary results are confirmed in the final analysis, lessons can be learned from the way communication in a flight crew is handled. As TCS is not related to TMS coordination in flight crews, the flight crew coordination set-up is robust to specialization within the team. This means that their way of using procedures, short, relevant checklists, two-way communication and verification of information when the reliability of information or clear communication is in doubt, ensures TMS coordination. Extended research of implementation of this type of communication and procedures in other differentiated teams (i.e. commando teams in Naval ships or medical teams in hospital OR) is needed prior to adjustment of CRM training. In this extended research, team size has to be considered as well.

The results from this research support and improve practical CRM train-

ing for specialized teams operating in a complex environment. As the data are obtained in a military, complex environment with military operators, the results have face validity. This improves the acceptance of the results and the implications that the results might have on CRM training, coaching and practice in the field.

Acknowledgements

My special thanks to the flight crews who participated in this research and the instructors who allowed me use their expertise. I would also like to thank the international partners (German Navy, Danish Navy, Portuguese Navy and Norwegian Air force) for allowing me to assess their flight crews. I owe the civil staff of CAE my gratitude for their technical support. My thanks to the head of the FMFT Lt. Rogier van Kralingen for giving me access to the FMFT and his personnel and students.

References

- [1] United Kingdom. Civil Aviation Authority. *CAP 737, Crew Resource Management (CRM) Training Guidance For Flight Crew, CRM Instructors (CRMIS) and CRM Instructor-Examiners (CRMIES)*. 2006. <http://www.caa.co.uk/docs/33/cap737.pdf>.
- [2] E. Salas, C.A. Bowers and E. Edens, *Improving teamwork in organizations applications of resource management training*. Mahwah, N.J.: Lawrence Erlbaum, 2001.
- [3] J.A. Cannon-Bowers and E. Salas, "Making Decisions Under Stress: Implications for individual and team training," *American Psychological Association*, 1998.
- [4] Helmreich and Merritt, "Safety an error management: the role of crew resource Management," in *Aviation Resource Management*, B.J. Hayward and A.R. Lowe, Eds. Aldershot: Ashgate Publishing Ltd., 2000, pp. 107-119.
- [5] P. Essens, A. Vogelaar, J. Mylle, C. Blendell, C. Paris, S. Halpin and J. Baranski, "Military Command Team Effectiveness: Model and Instrument for Assessment and Improvement," NATO RTO Technical Report TR-HFM-087, 2005.

- [6] R. Flin, P. O'Connor and M. Crichton, *Safety at the sharp end: a guide to non-technical skills*. Aldershot: Ashgate Publishing Ltd., 2008.
- [7] K. Lewis, "Measuring Transactive Memory Systems in the Field: Scale Development and Validation," *Journal of Applied Psychology*, vol. 88 no. 4, pp. 587-604, 2003.
- [8] S. Mohammed and B.C. Dumville, "Team mental models in a team knowledge framework: expanding theory and measurement across disciplinary boundaries," *Journal of Organizational Behavior*, vol. 22 no. 2, pp. 89-106, 2001.
- [9] J.S. Prichard and M.J. Ashleigh, "The effect of team-skills training on transactive memory and performance," *Small Group Research*, vol. 38, pp. 696-726, 2009.
- [10] K.A. Smith-Jentsch, K. Kraiger, J.A. Cannon-Bowers and E. Salas, "Do familiar teammates request and accept more backup? Transactive Memory in air traffic control," *Human Factors*, vol. 51, pp. 181-192, 2009.
- [11] F. Tschan, N.K. Semmer, A. Gurtner, L. Bizzari, M. Spychiger, M. Breuer and S.U. Marsch, "Explicit Reasoning, Confirmation Bias, and Illusory Transactive Memory: A Simulation Study of Group Medical Decision Making," *Small Group Research*, vol. 40, pp. 271-300, 2009.
- [12] M.T. Brannick and C. Prince, "An Overview of Team Performance Measurement," in *Team Performance assessment and measurement: Theory, Methods, and Applications*, M.T. Brannick, E. Salas, and C. Prince, Eds. Mahwah, N.J.: Lawrence Erlbaum, 1997, pp. 3-16.
- [13] D.M. Wegner, "Transactive Memory: a contemporary analysis of the group mind," in *Theories of Group Behavior*, I.B. Mullen and G.R. Goethals, Eds. New York: Springer-Verlag, 1987, pp. 185-208.
- [14] A.K. Nandkeolyar, "How do teams learn? Shared mental models and transactive memory systems as determinants of team effectiveness," working paper pp. 1-82, 2010.
- [15] R. Glaser and M.T.H. Chi, "Overview," in *The nature of expertise*, M.T.H. Chi, R. Glaser and M.J. Farr, Eds. Hillsdale, NJ: Lawrence Erlbaum, 1988, pp. xv-xxix.

- [16] A.E. Akgün, H. Keskin, J. Byrne and S.Z. Imamoglu, “Antecedents and consequences of team potency in software development projects,” *Information & Management*, vol. 44, no. 7, pp. 646-656, 2007.
- [17] E.A. Mannix, T. Griffith and M.A. Neale, “The phenomenology of conflict in distributed work teams,” in *Distributed Work*, P. Hinds and S. Kiesler, Eds. Cambridge, MA: Massachusetts Institute of Technology, 2002, pp. 213-234.
- [18] M.A. Campion, E.M. Papper and G.J. Medsker, “Relations between work team characteristics and effectiveness: a replication and extension,” in *Personnel psychology*, vol. 49, 1996, pp.429-451.

Training Tools for Research

Rico Bloemberg

Introduction

In the Navigation Technology Department, a special interaction exists between scientific education and research on the one hand and skill based education on the other. The skill based education consists besides theoretical courses of a lot of practical courses. The NLDA can provide their students with some special training facilities. These include a Warfare Electronic Chart Display and Information System (WECDIS) trainer, a radar simulator, a training sailing vessel, a training vessel (see Figure 1) and a Full Mission Bridge Simulator (FMBS).



Figure 1: Navy training vessel Van Kinsbergen.

Some of these facilities are used for research. The training vessel Van Kinsbergen is also used for research in manoeuvring behaviour. The FMBS

is used for research as well. Most of the bridge simulator research is focused on the so called “non-technical skills” which are used in Crew Resource Management. An example is the dissertation by Letty Aarts [1]. This dissertation focussed on the factors that influence the overview in a bridge environment. Recently, Lt Beemsterboer executed an experiment on the bridge simulator to measure Situation Awareness (SA) effects as part of his bachelor thesis [2]. Lt De Jonge is currently working on a bachelor paper that is based on the findings of Lt Beemsterboer. The bridge simulator and the training vessel Van Kinsbergen are being used for this research.

The advantages of using these training facilities for research are numerous. Students can use training facilities they are familiar with for their scientific research. Using the facilities during normal training hours will normally have little impact on the running training program. Results acquired by the research can be incorporated in the current training programs at very short notice.

To give an impression of the research done with our training facilities, a summary of the bachelor paper of Lt Beemsterboer follows.

The thesis of Lt Beemsterboer focusses on SA on the bridge of a ship. The aim of this paper is to indicate the influence of the increasing number of navigational means for the SA for the Officer of the Watch by means of an experiment.

What is SA?

The definition of Situational Awareness used in this paper is the following:

“Situational Awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the future” [3]

SA is knowing what is happening around you [4, p2]. This is of the utmost importance for the Officer of the Watch as this is the basis of his decision making. As a consequence, sound decision making results in safe navigation.

We can identify three levels of SA:

- Level one: Perception

- Level two: Comprehension
- Level three: Projection in the future

Level one SA is monitoring important data and elements from the environment. This is fundamental for requiring good information which is vital for a good SA. When monitoring is insufficient, there is a big chance that a wrong perception of the situation occurs.

The way important data and elements are acquired is displayed in Figure 2. Besides input from other team members and own observation, systems can deliver input as well. When used properly, all these aspects will result in a good SA level one.

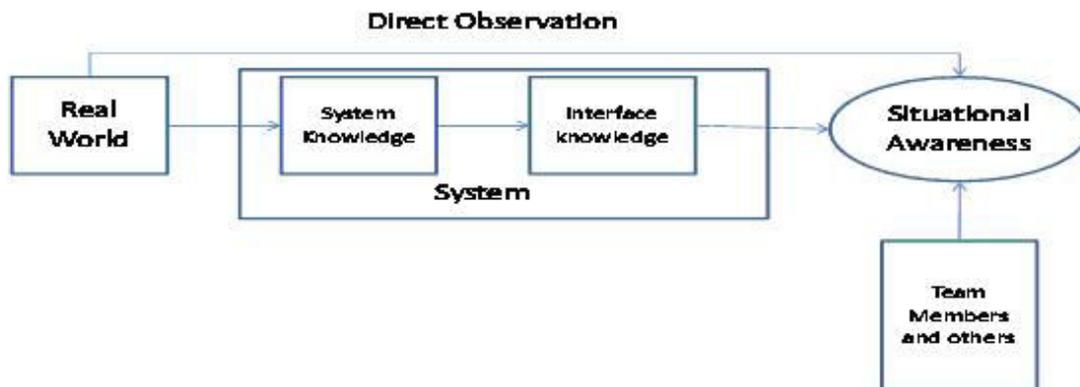


Figure 2: Conceptual model of Situational Awareness [4, p7].

Level two SA goes further than just perceiving vital information. It incorporates combining, interpreting, storing and maintaining this vital information. The Officer of the Watch receives his information from looking outside, the chart, radar and other sensors. All this information needs to be integrated to form a good assessment of the situation. The SA of the Officer of the Watch will improve when observation of another ship is a combination of a compass bearing and radar. When the Officer of the Watch observes more ships, he will be able to make priorities.

A good level two SA can only be achieved with a good level one SA. When the perception of a situation is wrong, the comprehension of this situation will be wrong as well. The decision to reduce speed for another ship

to avoid a collision when missing the ship that is at your stern overtaking you, may result in an even more dangerous situation. The importance of level one SA is stated in the following:

“A person with level two SA has been able to derive operationally relevant meaning and significance from the level one SA data perceived”.

A level three SA is projecting the current SA in the future. An Officer of the Watch has achieved the highest level of SA when he is able to predict how a situation will develop in the future. The Officer of the Watch is now able to anticipate future situations. The coming together of all three levels of SA is shown in Figure 3.

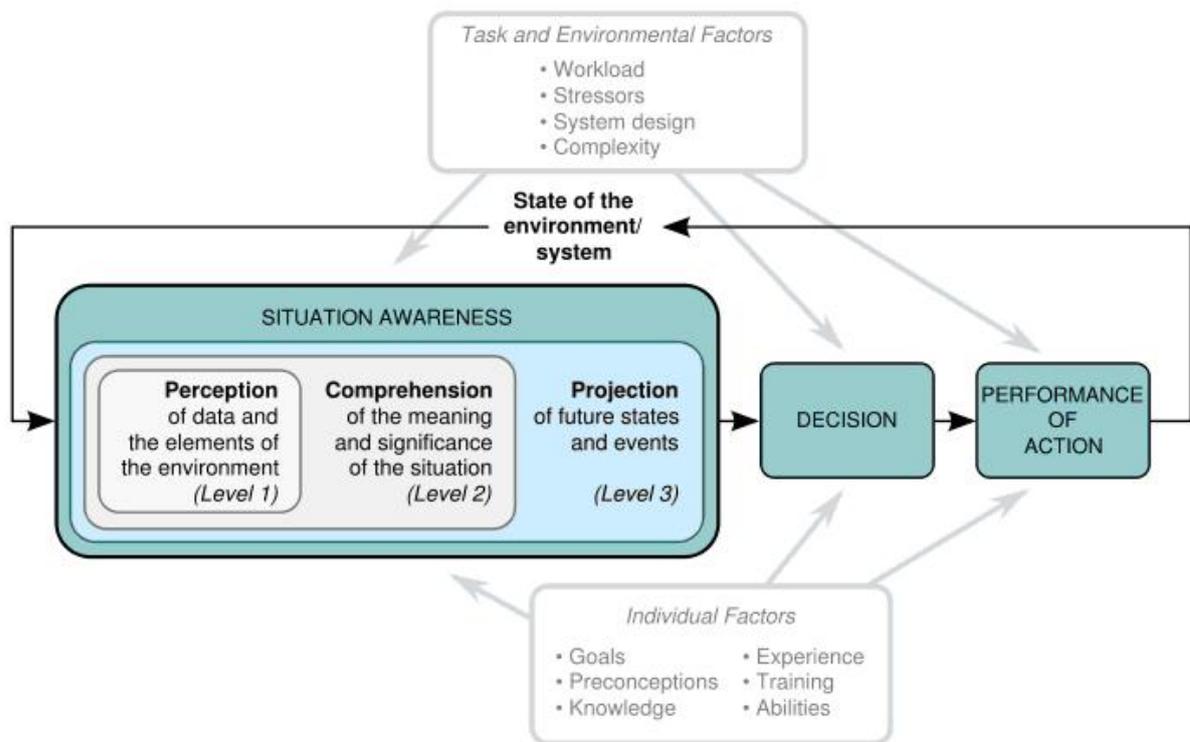


Figure 3: The three levels of SA. These three levels form the basis for a sound decision making. The building of SA itself is subjected to individual task factors and task and environmental factors. [4, p6]

How do we measure SA?

To measure SA in the experiment, the Situation Awareness Global Assessment Technique (SAGAT) method was used [3]. The SAGAT method was primarily developed for measuring the SA of pilots.

The SAGAT method is best explained by the following seven steps.

1. The pilot flies a certain scenario.
2. The simulation is paused at a certain moment.
3. The pilot must answer a questionnaire in order to determine his knowledge of the situation. The questionnaire is derived from the desired SA qualifications for a pilot. The answers are processed in a computer.
4. It is impossible to make a questionnaire that includes all the SA requirements. Hence it is a random selection of all the SA requirements.
5. The answers are compared with the real situation, which is easily measured in a simulator.
6. The answers will generate a SAGAT score which is divided in three zones; immediate, intermediate and long-range.
7. This process will be repeated several times to meet the statistical requirements.

Because the SAGAT method was specifically designed for pilots, some adjustments had to be made for the experiment on the bridge simulator. Another reason to adjust the experiment is that not all the required means for a full SAGAT method are available on the FMBS.

The Experiment

On the bridge of a ship, several systems are used to acquire and maintain a good SA. On board the new Royal Netherlands Navy ships, these systems are normally integrated. This study does not focus on the integrated bridge system. The purpose of this study is to analyse the effect of the increasing number of navigational aids on the SA of the Officer of the Watch.

It appears that despite the known knowledge of SA, there has not been any research into the effect of increasing aids to navigation and the SA of the Officer of the Watch on board a ship. This research was therefore executed on the bridge simulator in Den Helder (see Figure 4).

The experiment has used students who were nearly finished with their study. They had a limited amount of experience as watch keeping officers,



Figure 4: Full Mission Bridge Simulator of the NLDA at Den Helder.

mainly obtained on the training vessel van Kinsbergen and the Full Mission Bridge Simulator. Measuring the SA of these people was the main focus of the experiment. Besides the persons acting as officer as the watch, two important questions have to be answered:

1. What is the role of the Officer of the Watch on board of a navy ship?

The Officer of the Watch has to maintain the safe navigation of the vessel and is responsible for the daily routine on board. The Officer of the Watch can even be put in charge of the operational deployment of the ship. The following definition of navigation will be used:

“Navigation is the art of moving a vessel from a given position to a desired position following predetermined criteria.” [5]

The Officer of the Watch is responsible for the navigation. It is paramount for the Officer of the Watch to have a clear picture of navigational, meteorological and operational matters. The Officer of the

Watch has to take all measures necessary to ensure the safety of the ship.

2. What does SA mean for the Officer of the Watch?

Every watch keeping officer on the bridge of ship needs to have a good knowledge of the Rules of the Road. These rules are internationally agreed by and used by all seamen on the high seas. Rule number 5 of the Rules of the Road states:

“Every vessel shall at all times maintain a proper look-out by sight and hearing as well as by all available means appropriate in the prevailing circumstances and conditions so as to make a full appraisal of the situation and of the risk of collision.”

This rule applies to all circumstances. All three levels of SA can be traced in this rule. Maintaining a proper look-out at all times is level one SA. To make a full appraisal of the situation and of the risk of collision is level two and three of SA. By making a full appraisal of the situation, level two results in determining whether there is a risk of collision. Knowing the moment of the risk of collision is level three. Other rules can be used in the level three SA by supporting the decision making process of the Officer of the Watch. Rule 8, describing the actions to avoid collision, states:

“Any action taken to avoid collision shall be taken in accordance with the Rules of this Part and shall, if the circumstances of the case admit, be positive, made in ample time and with due regard to the observance of good seamanship.”[6]

Making the right decision in ample time means that a good level three SA is required.

Method of research

The nucleus of this research is the experiment on the bridge simulator to measure the effect of increasing the number of available navigational aids on the bridge. Due to the different nature of the bridge simulator, the SAGAT method had to be adjusted. The (increasing) number of systems available will be the variables in the experiment. The setup of this experiment contained the following system combinations:

1. Bearing compass and paper chart.

2. Bearing compass, paper chart and GPS.
3. Bearing compass, paper chart, GPS and radar.
4. Bearing compass, paper chart, GPS, radar and Warfare Electronic Chart Display and Information System (WECDIS).

There were 12 persons available for the experiment, so all system combinations could be tested three times. Every simulator run took 60 minutes and during every run two measurements were made. An overview of the scenario is presented in Figure 5.

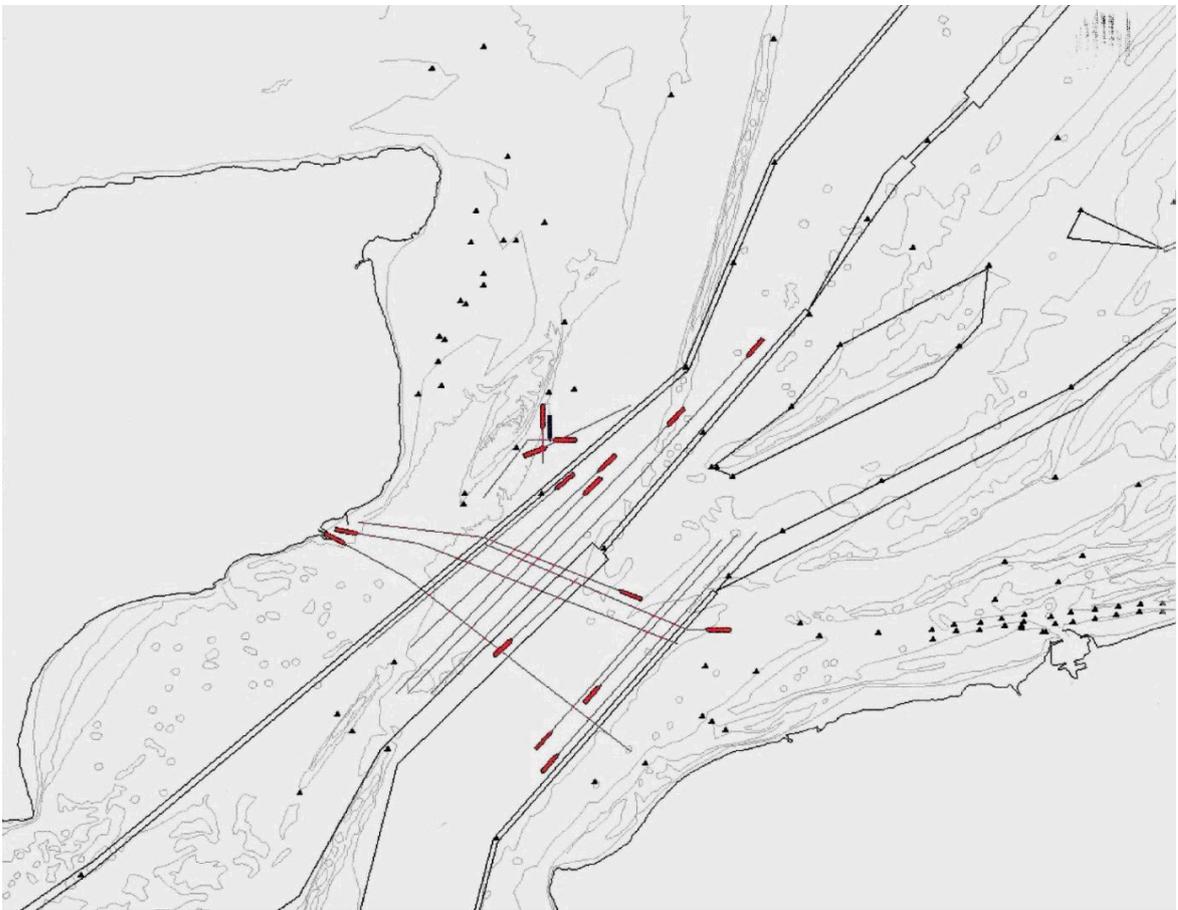


Figure 5: Overview of the scenario. Own ship is blue and the other shipping is presented by the red ships with planned tracks.

The scenario contained the following points:

- Close to land to provide landmarks for position making;
- Two tracks had to be sailed;

- The initial speed was 12 knots;
- Visibility was 20 Nautical Mile (NM);
- Other shipping consisted of sixteen contacts;
- Great variety in shipping;
- Different movements of other shipping;
- Presence of navigational dangers like wrecks and shallow waters;
- At the start of the run, no dangerous shipping in the area in order to provide time to build up a good SA.

Due to the different settings and constraints delivered by the bridge simulator, the SAGAT method had to be adjusted. In the experiment, the pilot is of course replaced by an Officer of the Watch. Point three of the SAGAT method had to be adjusted as well with the following points:

- When the simulation was stopped at time t_1 , the Officer of the Watch was required to fill in two plot charts. In plot chart one, the current situation (t_1) is plotted. In plot chart two, the anticipated situation of $t_1 + 30$ minutes is plotted.
- The Officer of the Watch is required to fill in a plot chart at $t_1 + 30$ min. Only the current situation is plotted.
- Besides the plot charts, a questionnaire is answered with questions concentrating on positions of own ship, buoys and other shipping.

By using the plot charts, point four of the SAGAT method is not applicable.

Quality control

The following three aspects are used for quality control.

Reliability Can we use all the measurements we made for our results? We have tried to prevent corrupt data in our results by using the SAGAT method.

Validity Are the measured data, the data that we need? By using the plotting charts and the questionnaire, we tried to get the most accurate picture of the three levels of SA.

Generalization To what degree do the results of the experiment represent other groups? We used midshipmen in their third year as officers of the watch. Although their background and education is the same as other officers of the watch, their experience level is of course less.

Restraints of the research

The following restraints were encountered during the research:

1. The experiment was executed with twelve officers of the watch. This is a relatively small number to draw conclusions. Due to practical reasons, it was impossible to increase the number of officers of the watch.
2. Despite the fact that all the officers of the watch are from a homogenous group, difference in performance due to the individual may influence the results. Although the four experiment groups were made at random, there may be a chance that one group exists of 'better' performers. The only way of eliminating this problem is by using more officers of the watch.
3. It is very hard to estimate distances on the bridge simulator due to the use of projectors. This does not help the process of assessing the positions of objects.
4. In the last experiment configuration, the WECDIS is used as an extra navigational aid. None of the officers of the watch are experienced WECDIS users. Better results will probably be attained with experienced users.

Data analysis

Data of the different runs has been collected from every system combination and was compiled from the questionnaires and plotting charts. An area of tolerance was used in order to allocate values to the collected data. We can compare the system combinations with the use of tolerance areas. Not all

the questions from the questionnaire have the same tolerance but all are the same for every system combination. We have to refrain from taking major conclusions due to the small group of persons used in the experiment.

Based on the tolerance areas, every combination received a score. The score for combination two was higher than combination one and combination three scored higher than combinations one and two. It was expected that system combination four would give the best results due to the use of the WECDIS. In this combination the best technical means were available for acquiring and maintaining a good SA. The score was however lower than system combination three. There are two possibilities that might be the cause of the lesser results scored in combination four:

1. None of the officers of the watch in the experiment have followed a WECDIS course. The knowledge of the system was limited. The WECDIS was probably not used to its maximum effectiveness.
2. Another theory might be the lack of workload. People with a reduced workload will lose focus and get bored [7]. In this way reduced workload can result in a loss of SA. In combination four, the WECDIS will reduce the workload of the officer of the watch significantly and might have contributed to a lack of SA.

Conclusions and recommendations

At the end of the experiment, it seems there is an effect in increasing the number of available navigation systems and an increasing Situation Awareness.

We have to be careful of making any conclusions due to the small number of measurements. Doing more measurement with larger groups will add more weight to the results and will reduce the effects of failures and individual effects. Another recommendation is to do an experiment with increased workload for the Officer of the Watch, especially with regard to the use of the WECDIS system. Results in this area may lead to new insights and procedures for using the WECDIS.

Currently, Lt de Jonge has used the experiment of Lt Beemsterboer as the basis for his research in comparing students on the bridge simulator

and on the training vessel van Kinsbergen. We are looking forward to the results of this study as they may help us to enhance our training program or even adjust current bridge procedures. All these experiments are very closely related to Crew Resource Management. Situation Awareness is one of the seven skills we teach our students in our Crew Resource Management course. Scientific research in this area will enhance our understanding and support our Crew Resource Management courses.

There is a very fine line between scientific research and the actual use of the results of this research in our courses, navigational developments, Crew Resource Management and standing procedures on board our ships. Due to the increasing automation of the bridge with integrated systems, this research will only grow in importance. The introduction of a new Full Mission Bridge Simulator in the near future with possibilities of camera monitoring, sound recording and extensive data collection and debrief facilities will greatly enhance our research possibilities, especially in the area of Crew Resource Management.

References

- [1] L.T. Aarts, "Overview maintenance in man-machine environments, applications in ship navigation," Ph.D. dissertation, University of Amsterdam, Amsterdam, 2004.
- [2] W. Beemsterboer, "Situational Awareness op de brug," B.Sc. thesis, NLDA, Den Helder, 2009.
- [3] M.R. Endsley, "Situation awareness global assessment technique (SAGAT)," in *Proceedings of the National Aerospace and Electronics Conference (NAECON)*, pp. 789 -795, 1988.
- [4] M.R. Endsley and D.J. Garland, Eds., *Situation Awareness Analysis and Measurement*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
- [5] Y. Draaisma, J.J. De meester, J.H. Mulders and J.A. Spaans, *Leerboek Navigatie*. Bussum, 1979.
- [6] K. van Dokkum, *Verkeersregels op zee*. Meppel, 2005.
- [7] C.D. Wickens and J.G. Hollands, *Engineering Psychology and Human Performance*. New Jersey, 2000.

Contributing Authors



Contributing authors from NLDA's Navigation Dept.

- Rico Bloemberg** Lecturer Navigation Technology
Crew Resource Management
ja.bloemberg@nlda.nl
- Wim van Buuren** Guest lecturer
Ship Manoeuvring
Dutch Pilot
w.v.buuren@nlda.nl
- Dick Engelbracht** Assistant professor Hydromechanics
Naval Architecture, Ship Manoeuvring
dj.engelbracht@nlda.nl
- Hilde van Ginkel** PhD Student NLDA
Lieutenant Royal Netherlands Navy
Crew Resource Management
hta.v.ginkel@mindef.nl

- Armand Goossens** Program Manager HDIO IV
Lieutenant Colonel Royal Netherlands Air Force
Information Management, NEC, Command and Control
aahe.goossens@mindef.nl
- Cees de Groot** Assistant professor Navigation Technology
Inertial Navigation, Integration of Navigation Systems, ECDIS
ca.d.groot@nlda.nl
- Benoit Muth** PhD student NLDA / Delft University of Technology
GNSS, Signal Processing, Statistics, Detection, Estimation
B.J.Muth@tudelft.nl
- Patrick Oonincx** Full professor Navigation Technology
Signal Processing, GNSS, Applied Mathematics
pj.oonincx@nlda.nl
- Anton Scheele** Assistant professor Navigation Technology
Geodesy, Navigational Data Processing, eLoran, Military Systems
ca.scheele@nlda.nl
- Jochum Tadema** PhD student NLDA
Lieutenant Royal Netherlands Navy
GNC, SVS Technology, Human Interactive Aspects of Automation
j.tadema@nlda.nl
- Erik Theunissen** Full professor NLDA / Delft University of Technology
Avionics
e.theunissen@tudelft.nl
- Christian Tiberius** Associate professor Delft University of Technology
GNSS, Satellite Navigation
Participant NLDA-TUD GNSS DSP project
c.c.j.m.tiberius@tudelft.nl
- Daniela Vaman** PhD student NLDA / Delft University of Technology
Telecommunication, Signal Processing, (Terrain) Navigation
D.Vaman@tudelft.nl
- Paul Wolkenfelt** Assistant professor Mathematics
Applied Mathematics, Numerical Analysis, Mathematical Modeling
phm.wolkenfelt@nlda.nl