IAIN 2015 CONGRESS PROCEEDINGS
Non-IEEE Full Papers

International Association of Institutes of Navigation
World Congress 2015

20–23 October 2015 | Prague, Czech Republic

The scientific content and standard of the papers is the sole responsibility of the authors.
# TABLE OF CONTENTS

**Session A: Satellite navigation systems, their augmentation, receiver technologies and special questions**

- The Navigation of Navigation ................................................................. 8
  *David Last*
- The future governance of GNSS ............................................................. 14
  *Serge Plattard*
- Considerations for Maritime Augmentation Service under Multi-constellation GNSS .................................................. 21
  *Sang Hyun Park, Ki Yeol Seo*
- EDAS (EGNOS Data Access Service) for Added Value Applications ................................................................. 26
  *Juan Vázquez, Elisabet Lacarra, Miguel-Angel Sánchez, Jorge Morán*
- Assumptions of Full Mission Ship Bridge Simulation Including EGNOS .......................................................... 32
  *P. Zalewski, L. Gucma, K. Urbanska, S. Schlueter, A. Born, M. Porretta*
- Application of faded harmonic subcarrier modulations in GNSS Modernization .................................................. 39
  *Wei Liu, Yuan Hu, Shengzhen Wang, Chaojian Shi, Baojia Xiao*
- Phase-Phase and Phase-Code Methods Modification for Precise Detecting and Predicting the GPS Cycle Slip Error ................................................................. 44
  *A. A. Elashiry, Mohamed A. Yousef, A. M. Abdel hamid*

**Session C: Signal processing in navigation systems** ...................................... 51

- Ship’s Heading Integrity using Multi-Compass ........................................... 52
  *Kazumichi Oi, Shigeyuki Okuda, Masatoshi Shimpo, Yasuo Arai, Yuji Okutomi*

**Session D: Marine navigation. Applications of GNSS and other technologies. Special problems** .............. 58

- A Study on Vessel Traffic Management System at Tokyo Bay using AIS ................................................................. 59
  *Tomoko Shirai, Kenji Inada, Nobuaki Kubo, Hitoi Tamaru*
- A Statistical Inference Approach to Extract Principal Fairways of Ship Passages through a Sea-strait .................. 65
  *Jinhai Chen, Feng Lu*
- Methodology of Creation Simulation Basin Based on the Channel Through the Vistula Spit .................................. 73
  *Piotr Zwolan, Krzysztof Czaplewski*
- Ship monitoring and location estimation based on spaceborne GNSS reflections .................................................. 79
  *Wei Liu, Shengzhen Wang, Jing Peng, Chaojian Shi, Baojia Xiao*

**Session E: Aviation and terrestrial navigation. Applications of GNSS and other technologies. Sensors and their data fuse. Special problems** ...................... 83

- Advanced-Receiver Autonomous Integrity Monitoring for Horizontal Civil Aviation Applications .................................... 84
  *Michel Mabilleau, Jonathan Vuillaume, Igor Nikiforov, Gerhard Berz, O.F. Bleeker*
- Enhancing the EGNOS NOTAM Proposals: increased automation and state-of-the-art prediction engine .......... 91
  *Jorge Morán, Juan Vázquez, Miguel-Angel Sánchez*
- EGNOS empowers tracking and tracing for logistics in Europe and around the Mediterranean .................................. 99
  *A. Di Fazio; D. Bettinelli, U. Celestino, C. Gracia*

**Session F: Spectrum management. Jamming and spoofing of navigation systems** ................. 106

- GNSS Spoofing Detection Scheme Based on the Combined Use of a Single Authentic Ranging Signal and RAIM ................................................................. 107
  *Hyoungmin So, Jaegyu Jang, Kihoon Lee, Kiwon Song, Junpyo Park*
Session H: Radar, electronic maps, special sensors

Characteristics of Echo Fluctuation and Echo Paint from the Small FRP Boat on the Marine Radar Screen

Tatsuto Yamada, Shinji Mizui, Akihiko Tsukisaka, Ruri Shoji

Availability of Navigational Fusion System for Small Domestic Vessels

Yoshifumi Ichikawa, Akihiko Hori, Shigeyuki Okuda, Yasuo Arai, Masatoshi Shimpo

3D ECDIS Implementation Base on Area Level of Detail Rendering Techniques

Shengzheng Wang, Chaojian Shi, Baojia Xiao, Jing Peng

Navigation with ECDIS: A Novel Approach to Navigational Calculations

Adam Weinrit

Session I: Special problems, other topics

Low-Cost GPS/INS/OBD-II Integration On an Embedded Linux Platform

Joonhoo Lim, Je Young Lee, Kwang Ho Choi, Hyung Keun Lee

Performance Analysis of Loosely Coupled RTKGNSS/IMU/Vehicle Speed Sensors in Urban Environment

Ren Kikuchi, Nobuaki Kubo, Shigeki Kawai, Ichiro Kato, Nobuyuki Taguchi

North Finding System with Real Time Attitude Determination for a Mortar

Jin Seung Lee, Jae Gun Choi, Suk Won Jang, Tae Gyoo Lee
Session A
Satellite navigation systems, their augmentation, receiver technologies and special questions
Abstract— Satellite navigation is an exceptionally successful technology. It benefits all modes of transportation plus industry, trade and telecommunications. Governments that historically administered these activities separately have struggled to make informed policy decisions for this single ubiquitous technology. This policy vacuum is becoming critical as the vulnerability of satellite navigation to system failures, solar weather, interference, intentional jamming and spoofing increasingly impacts users. Those governments that developed satellite systems independent of GPS are reluctant to recognise their shortcomings. They promote their individual constellations, while manufacturers provide users with receivers that benefit from all global navigation systems. This paper argues that, as the Era of Systems gives way to an Era of GNSS, there is marked lack of informed policy for the navigation of Navigation.

Keywords—Satellite navigation, GPS, GNSS, Galileo, Beidou, QZSS, IRNSS, WAAS, Enhanced Loran (eLoran), navigation policy

I. INTRODUCTION

Professionals in the navigation business are the stewards of an exceptionally successful technology. Satellite navigation has been one of the outstanding technical achievements of the late twentieth and early twenty-first centuries. Among science-based industries, it has been a star. It does not pollute the atmosphere, cause global warming, or involve fracking, natural selection, creationism, or the politics of the European Union: everybody loves GPS!

But satellite navigation has now raised challenges that navigation professionals - and especially their governments - appear unable to meet.

Life used to be simple: “proper” navigators were professionals. They wore uniforms and they had beards. Some were sailors. Others flew aircraft or navigated land vehicles. Matching government departments administered these separate activities and helped achieve international cooperation and standardisation. But the modes of transport remained apart: sailors had nothing to do with aviation technology or how others navigated on land.

All was well until along came satellite navigation and spoiled it! Soon, this single technology served navigators across all modes of transport. Then it escaped from the navigators and became a tool for many professions, and then simply a consumer product.

A report [1] by the United Kingdom’s Royal Academy of Engineering struggled to find a single area of transportation, commerce, industry or telecommunications in Britain that does not now employ satellite navigation.

My government, your government, and national governments around the world were completely unprepared to respond to this single technology on which depended (and from which profited) activities as diverse as missiles, farming and the stock market. Governments had separate ministries for each of these; and not only for the traditional modes of transport – land, air and sea – but also for industry, trade and communications. There were many such departments and agencies. No-one was responsible for setting national policy in navigation: there simply was no clear plan for the navigation of navigation!

Yet such leadership was essential, especially outside the United States where governments came to realise that this GPS, the technology on which so much in their economies depended, was controlled not only by a foreign power – the US - but ultimately by its military!

In response, those countries - or regions like Europe - that could afford to, set up their own satellite navigation systems. Thus, GPS, which much earlier had inspired GLONASS now begat Galileo and Beidou, QZSS and IRNSS, plus a host of augmentation systems: WAAS, EGNOS and other strange names.

Soon, these new “GNSS” became invested with immense national, or regional, pride. Their vast cost had to be justified by claims of technical superiority. In reality, we engineers know that their designers had no choice but to make them largely compatible with GPS, since GPS was decades ahead and had become the world standard. So these new systems had to squeeze alongside GPS in the narrow radio frequency bands allocated to navigation. Not surprisingly, indeed happily, all our GNSS turned out to look very like GPS: versions of the same technology - with just a hint of garlic here, or a whiff of curry there. This similarity is obvious to engineers and navigators, though rarely to politicians.

And now each of these new systems is following a similar trajectory to that of GPS. For its first decade, GPS was seen as the way to meet every significant navigation need; to replace all older aids across land, sea and air. That was the clear view of the US Government Accountability Office, strongly supported in Europe. And why not? The growth of GPS did indeed result in the demise of Omega, Decca Navigator, Datatrak and a host of national systems you have never heard of, that simply could not compete technically or commercially. But this triumph of satellite navigation had bred a certain hubris: that is, overbearing pride.
II. THE VULNERABILITY OF GNSS

Unexpected events began to shake confidence in GPS, and vulnerabilities appeared. Occasionally, an individual satellite would fail causing large position errors. As shown in Fig. 1, the final atomic clock in satellite SVN23 gave up the ghost with exciting consequences across Europe. Last year we saw a double failure of GLONASS – suddenly, with large errors in position fixes.

On another date the Sun emitted radio noise so intense that GPS receivers stopped working across the entire sunlit side of the Earth (Fig. 2). In a third event GPS navigation was lost, accidentally and without warning for two hours across the San Diego area. Many mobile cell-phone sites using GPS timing were affected.

Then intentional jamming appeared on the scene. A low-power jamming device tested at a British lighthouse (Fig. 3) disrupted GPS throughout the red zone shown in the top left picture, out across the North Sea to the horizon at 30km. The blue line in the top right picture is the track of a vessel: when crossing the red-bordered triangular jamming area it lost GPS entirely.

Outside that area, the jammer caused all the false positions shown. Other ships even appeared to track over land.

As illustrated in Fig. 4, the effect of jamming on a ship’s systems can be dramatic. A jammer of less than one milliwatt aboard the THV Galatea caused false positions on the chart displays; the autopilot would steer the ship quietly off course; the ship would report false AIS data to other ships nearby and to the shore VTS; it lost satellite communications; the distress system that raises alarms and guides in rescuers failed; even the ship’s clocks went wrong. And when the officers sensibly reverted to radar and gyrocompass, to their shock they found those affected, too. Ships nowadays, like so many of our critical systems ashore, have multiple GPS receivers embedded in multiple systems in ways no-one understands. When one receiver fails, they all fail.

We have seen a sovereign state launch prolonged high-powered GPS jamming attacks on a neighbouring state, causing maritime navigation to be blocked in just the way I have shown. Also affected were aviation systems, cell-phone services and critical military capabilities.
The small jammer shown in Fig. 5, sold world-wide, is hundreds of times more powerful than many earlier ones. But what is interesting about this device is how very carefully it has been designed to block all the new GPS frequencies, plus all the frequencies of Galileo, Beidou and QZSS. Further, it attacks all our augmentations, like WAAS.

Terrorists can buy or build a jammer like the one shown in that is powerful enough to affect large areas of a major city from a publically-accessible location (Fig. 6). Despite this, in many parts of the world there are now powerful myths: that the local version of GNSS is immune to GPS jamming.

As GPS developed, in the United States, growing concern among navigation professionals as early as 2001 culminated in the Department of Transportation’s Volpe Report. This clearly and officially recognised the multiple threats posed by the vulnerability of GPS [2]. It recommended independent backup systems. Since then, interference and jamming events have multiplied in all our countries.

A detector on the highway close to the threshold of a UK regional airport (Fig. 7), gets up to 200 jamming hits a month. Similar data has been recorded in France and the US. There are undoubtedly many jammers in use currently.

A paper in this conference will report that almost half of master mariners surveyed had experienced losses of position, navigation and timing in the past 12 months, which they knew or suspected were due to GPS outages. Almost none of these events had been reported via official channels. Over 90% of these senior mariners believed that GPS needed a back-up.

And now there is a new threat: “spoofing”: That is transmitting false GNSS signals to commandeer a receiver. Researchers from the University of Texas at Austin used a laptop and spoofer to lead the super-yacht shown Fig. 8 (left-hand picture) silently and gently off course. There was no alarm on the bridge navigation display to tell the crew that anything was wrong.

When criminals hijack a truck in the near future, they will use a low-cost spoofer to make its on-board tracker show it on course, when really they have diverted and robbed it. The Austin group have also demonstrated how to shift precise GPS timing using a spoofer.
Some finance specialists believe that this opens the door to fraud, by spoofing the automated systems of banks and stock exchanges with their million trades a second. It also may also allow the disruption of power grids. This year low-cost spoofers have appeared on the scene (Fig. 8 right-hand picture); any competent hacker can now build one and take over your GNSS receiver.

III. RESPONSES TO VULNERABILITY

So, what do we do about this vulnerability of satellite navigation, to jamming, interference, spoofing, solar weather or equipment failure? Well, first, we must recognise the problem and face up to the need for Resilient PNT. Almost without exception, engineers and practising navigators now do so. Almost without exception, politicians do not. Solving this problem is a key test of how effective our political systems are in dealing with navigation.

Of course, many recognise that we must harden our technology. We will use intelligent adaptive receiving antennas (Fig. 9) that favour satellite signals over interferers. The military already do that, and the very top of the civil market will in future. But these technologies are still only a remote possibility for the mass of vulnerable users already out there.

We can integrate satellite navigation with other technologies: dead-reckoning in land vehicles, stable clocks for precise telecomms timing. But the powerful solutions are navigation and timing technologies independent of GNSS yet complementary to it.

Aviation is rich in these, having maintained multiple independent technologies in the face of GPS. For example, London Heathrow runway 27 Left now has a GNSS Instrument Approach; but it is supported by an ILS, an MLS, DME, VOR, ADF, inertial navigation, radar and baro altimeters and magnetic compasses! That aviation GNSS has mandatory high standards, with RAIM and WAAS plus compulsory reversion to a legacy system as soon as GNSS is less than perfect. What a dramatic contrast with maritime and land practice!

Prompted by the Volpe Report, the US Federal Aviation Administration proposed and demonstrated Enhanced Loran (eLoran) [3]. By applying GPS digital techniques to the obsolete Loran-C low-frequency technology, they created a system that met the accuracy, integrity, availability and continuity standards of certain aircraft instrument approaches plus the demanding port entrance requirements of shipping. It could also deliver timing of GPS quality to support telecommunications. A high-level study group of industry leaders, led by Professor Bradford Parkinson, concluded that this was the only cost-effective substitute to GPS for US needs.

The Department of Homeland Security announced the adoption of eLoran as the US national backup to GPS [4] - and then completely failed to implement it.

Delivering a navigation system that benefits multiple areas of national life has turned out to be beyond the capabilities of governments. In Washington, no single department owned either civil GPS or this powerful backup. So when it came to cost-benefit analyses there was simply no one to aggregate the benefits across the whole of government, industry and commerce. Each department feared being landed with the costs. Some people called this dilemma the ‘Tragedy of the Commons’. Before it could be resolved, a budget cut closed down the obsolete precursor Loran-C system that had recently been modernised ready for the move to eLoran!

The United Kingdom and Ireland took this US concept and created a prototype system, re-using obsolete Loran-C infrastructure, stretching from the North of Norway to the South of France, and adding a new station. The system has achieved Initial Operational Capability with 10-metre accuracy and full compliance with IMO standards at the seven major UK ports shown in yellow in Fig. 10. Ship-borne equipment can switch automatically and seamlessly to eLoran when GPS is lost. Separately, a high-precision version of the technology has been developed for the maritime pilots at Rotterdam, Europe’s largest port.
What splendid news: apparently Europe has recognised GNSS vulnerability and adopted an insurance policy! Actually, no. These systems may never reach Full Operational Capability. Europe lacks any plan to respond to the vulnerability of GNSS – why, who needs that when Europe has Galileo and EGNOS? There has never been a Volpe Report on GPS vulnerability in Europe or indeed anywhere outside the US.

My belief is that in both the US and Europe the only route to success will be for the Loran infrastructure to be taken over and operated by industry. Its benefits will be sold to individual groups of users, inside government and outside. The market (and greed!) will provide the mechanism for realising the benefits, paying the costs and making a profit.

IV. THE NEED FOR A NAVIGATION POLICY

What this example of GNSS vulnerability and eLoran has demonstrated is the lack of any informed debate on this matter – let alone policy - in most countries. A third of a century after the launch of the first GPS satellite, there is still little recognition by governments anywhere in the world of how essential resilient Position, Navigation and Timing have become to the critical infrastructure of their nations. I cannot identify a single country that yet has a clear and realistic plan that encompasses applications from maritime navigation through telephone systems to banking transactions?

This is not a paper designed to sell eLoran. But arguing for eLoran has demonstrated to me a much wider truth: that our immensely successful navigation industry has simply out-stripped our systems of government.

Even now in the US, the country with the most sophisticated understanding of the civil benefits of satellite navigation, GPS funding decisions are still largely determined by the budget of a single part of the military: the Air Force.

I ask: would the US government have funded satellite navigation had there not been a Cold-War imperative? Would any other government have funded a GNSS had the US not developed GPS? I doubt that the case for GNSS would be strong enough to make them put their hands in their pockets?

II. INDIVIDUAL SYSTEMS – OR GNSS?

Our world is changing fast. We now have multiple satellite navigation systems. But, I suggest that we are approaching the end of an era: the “Era of Systems”.

Here is what I mean: this Conference will follow the tradition of starting with reports on the status of different the Systems: GPS, GLONASS, Beidou and so on, plus their augmentations. Judging from the authors, these will be excellent papers. But the view is of single systems, each vertically-integrated - with satellites, control systems, receivers, applications and users – overseen by a national or regional administration: there will be talk of Galileo markets and GPS markets, for example.

The relationships between these systems remains an area of friction: in Europe, might Galileo be mandated; in the US, is the reception of “foreign” GNSS illegal, immoral, un-American? The view is that of governments and diplomats: separate control, spheres of influence, geo-political competition for dominance.
On the left of Fig. 12 is our smart-phone as before. Then a low-power jammer is switched on in the vicinity. In a few seconds – as we see on the right - all the GPS, GLONASS and Beidou signals have disappeared!

We are now in the “Era of GNSS”; the constellations live together – and they die together! They have simply become components of a single GNSS.

V. CONCLUSIONS

I suggest that the time has come to stop focusing on systems, whether satellite constellations or terrestrial, and instead find ways to deliver to users world-wide the resilient PNT they need and deserve. Technically we can do that. But to implement it requires political will and wisdom. So, we need something else, we need it urgently and nationally and internationally and it is this: a way ahead, a clear path, a course, a direction and a flight-plan for: the navigation of Navigation!

References


The future governance of GNSS

Serge Plattard
European Space Policy Institute (ESPI)
Vienna, Austria
serge.plattard@espi.or.at

Abstract—By 2020 four global navigation satellite systems (GNSS) and two regional systems (RNSS) should be operational using over 100 MEO / GEO satellites beaming tens of L-band signals at definite frequencies and different shapes and modes for civil and military users. An estimated 7 billion installed base of GNSS devices should be active at the same date, 90% being mobile phones, supported by the pervasiveness of a continuously increasing number of applications. In this complex environment involving six different providers and billions of users, reliability and sustainability of the acquired data will be a central issue posing the question of an optimal governance schemes for GNSS. Several possibilities are looked at based on the work of the International Committee on GNSS, bilateral agreements, or international organizations dealing with data acquired from space-based systems. Drawn from this review, adoption of a non-legally binding Code of Conduct (CoC) is proposed a minima to pursue a set of guidelines / rules of the road guaranteeing a sustainable functioning of the different constellations, in particular vis-à-vis compatibility, interoperability, interference, integrity, and restricted services. Such a Code would protect each provider’s prerogatives together with transparent access to information, ensuring that any GNSS update or modernization remains in line with the overall de facto system of systems through multi-GNSS receiving capabilities. A loosely binding structure, inspired by the CoSars-Cospas organisation is also examined as a possible upgrade of the ICG informal status.

Keywords—Global Navigation Satellite Systems, UN, ICG, compatibility, interoperability, governance, Code of Conduct.

I. INTRODUCTION

Ubiquitous, the GNSS signals have become indispensable for economics, finance, commerce, transportation, defence and security, culture and leisure, knowledge activities… in short impacting our daily life and pointing out, once more, our increasing dependency on space systems. Figures speak for themselves: In 2014, 3.6 billion GNSS devices where in use around the world relying primarily on the American GPS and the Russian Glonass, a figure estimated to reach 7 billion in 2020 [1]. By the end of the decade, this base will be enlarged to two additional global systems that should enjoy full operational capability, the Chinese BDS (Beidou) and the European Union Galileo; and two regional systems, the Indian IRNSS and the Japanese QZSS. These latter constellations are expected to operate in addition to and in conjunction with the two historical founding systems that were declared fully operational for positioning, navigation and timing (PNT) in the mid 1990’s.

Such impressive growth of the GNSS receivers market (core revenue expected around 100 billion Euros in 2020 according to [1]) is also served by the enormous amount of mobile applications that rely on positioning information: Out of the 3.3 million apps in store at the end of December 2014, almost half of them use location information [1].

Looking at the providers’ side, and more specifically at the global PNT systems, the radio-frequency spectrum of available space-based radio signals transmitted in the L-band will be quite crowded around definite frequencies. In order to make sure that this will not be an issue on compatibility and interoperability of the various systems, there has been for the last ten years regular consultations and arrangements among the different partners on a range of technical questions. This process is still going on, aiming at having a robust operating environment by 2020.

After a short review of the radio signals landscape for the future GNSS/RNSS systems by this decade’s end, the paper will detail the existing mechanisms put in place to ensure a robust and secure reception of these signals indeed. It is then suggested that, based on the existing consultations mechanisms, to set up a governance scheme destined to maintain a sustainable functioning of the multi-GNSS de facto system of systems to be. Two possibilities are presented: i) adoption of an international code of conduct by all the space-based PNT providers; and ii) transforming the current International Committee on GNSS (ICG), an informal platform under the aegis of the United Nations, into a UN steered or autonomous entity that would ensure coordination among the different providers and foster wide dissemination of GNSS access and utilization worldwide.

II. THE NEED FOR A GNSS GOVERNANCE SCHEME

Initially the satellite navigation systems were relying on the GPS and on Glonass, two systems completely independent and for which the carrier frequencies of the transmitted radio-signals, on the one hand, and the mode of transmission on the other hand were set up to ensure maximum transmission performances and avoiding signal interferences. Moreover, the signal multiplexing methods to differentiate signals from various satellites belonging to the same constellation exploit signal characteristics, which differ from GPS and Glonass.
These proprietary differences also contribute to enhance the robustness of each of the initial systems. But with the arrival of two additional global systems by the end of the decade, Galileo and BDS, each of them providing signals at the same frequencies than GPS and Glonass, there might be an increased risk of interference between the different constellations, also known as a compatibility issue, and degradation of the quality of the signal transmission. In this new situation, the derogatory adage “two constellations are better than one, three are not really better than two, but four constellations are worst than three” has to prove wrong to guarantee a fully operational multi-GNSS serving a plethora of needs. Since some of these are determinant for the safety and security of a variety of land or space based critical infrastructures, it is indeed essential to ensure the compatibility and the interoperability of the aforementioned constellations.

Fig. 1 is an example of how packed the spectrum will be in the 1560-1610 MHz band during the coming years [2].

This example illustrates clearly the necessity to conduct coordinated actions when progressively populating the electromagnetic spectrum with new signals bunched around specific frequencies on a relatively narrow portion of the L-band. Thus, introducing the validity for establishing a governance scheme to be shared by the different global and regional providers. Signals generated for different usages by GPS, Galileo and BDS, and to some extent by Glonass, had to be properly distributed, spaced, and formatted to make sure that reception of any signal from one system was and will be free of interference from those transmitted by the other systems.

Parallel to these technical considerations implying the necessity for a sustainable governance scheme, socio-economic reasons pull also into the same direction. Indeed, the enormous market pressure mentioned above will create a forceful obligation on the providers since multi-GNSS receivers will become the norm. Moreover, it could even turn out that regulation, or consensual guidelines, will be imposed by the ever largest community of users, probably at least 80% of humankind, which will rely on a space based service that has become an integral and essential part of daily life.

III. GOVERNANCE, WHERE ARE WE TODAY?

Before reviewing the different tools that should help in setting up a governance scheme, we need a clear definition of what is governance.

A. A definition of governance

What is the meaning of governance when applied to GNSS and RNSS?

Generally speaking, governance refers to “all processes of governing, whether undertaken by governments, market or network, whether over a family, tribe, formal or informal organisation or territory and whether through laws, norms, power or language”[3]. Should the organisation be a formal one, governance is primarily about what the relevant “governing body” does. If the organisation is an informal one, such as a market, governance is firstly about rules and norms that guide the relevant activity. Whether the organisation is a geo-political entity, a corporate entity, or a socio-political entity, or an informal one, its governance is the way the rules and actions are produced, sustained and regulated.

For GNSS and RNSS, governments, precisely because of their instrumental role in setting up and exploiting the systems, together with the considerable involvement of the users for the development of the down-stream market, will be undoubtedly the drivers to establish a sustainable governance. To what extent and how, this is what needs to be explored indeed.

B. The specific nature of GNSS compared to other space-based information systems

The space-based production of information addresses basically four types of activities of a very different nature: Telecoms and broadcasting, Earth observation including remote sensing and meteorology, intelligence gathering, and positioning, navigation and timing. Space exploration is by itself an activity that does not fall strictly into the information production category. The first three types of

---

1 The substantive governance derives from the Greek verb κυβερνάω (kubernao) which means to steer
activities are based on collecting information generated by a passive or an active source, then received by a satellite which may or may not process it and which finally bounces it back to one or up to several million users. These systems are known to experience saturation, meaning that there is a limitation to the quantity of information they can transmit at a particular moment, comparable to ground networks that happen to be overloaded. These categories of systems can be considered in many respects as an extension of ground-based telecom / broadcasting networks for which government or market rules are well established for the 1st type, or as an extended domain of aerial photography for the 2nd type. Administrations, government agencies, and national or global companies are familiar with the environment in which they operate, where governance is not, to a larger extent, anymore a particular issue.

The case of space-based PNT systems is quite different. Indeed, here the primary information is generated in and transmitted from space, not collected from Earth or from anywhere else and then transmitted. Navigation satellites behave as beacons dispensing line-of-sight radio navigation signals emitted with a highly accurate time periodicity thanks to the use of on-board atomic clocks. Such an information system is unsaturated by essence, providing radio signals to an unlimited number of receivers, with no acquisition restriction for the civil signal, each receiver benefitting of the same amount of information compared to the others. Hence, this universal nature of the GNSS radio signals is certainly worth to consider them for a global public good labeling [4].

Because of the widely spread usage of GNSS primary signals, together with the status of the very limited number of providers to be, activities of these co-existing GNSSs have to go through a governance based on different grounds than those experienced in the information collecting systems mentioned earlier. There is a double rooted reason for that: i) of technical nature as aforementioned, intrinsic to the GNSSs, and differences between these systems and the telecom / broadcast space-based systems; and ii) of political nature, since three out four of the global providers are under the responsibility of their respective country ministry of defence which are essential for strategic forces and key to military operations and services.

At this stage no overall formal and agreed governance scheme has been established. But several instruments are used to exchange information, set up working groups to make progress on and harmonize some technical issues, including adoption of memoranda of understanding and agreements of technical nature. Such steps should pave the way for shaping a governance scheme satisfactory to the government providers as well as to the public and private users.

C. The importance of the ITU and the ICG

In these harmonization processes, two fora play a significant and different role: The World Radio communication Conference (WRC), and the International Committee on GNSS (ICG). It is at the WRC, organized every 3-4 years by the International Telecommunications Union (ITU), that all radio frequencies are decided and attributed, including those carrying the different GNSS signals. ITU ensures the protection of the scarce and fragile electromagnetic spectrum resource. Although mobile phone operators are currently bidding for frequencies that could interfere with those used by GNSS providers, it is very unlikely they will get their way through because of the dual use nature of a number of frequency slots in service on three national global constellations.

ICG for its part is totally devoted to a range of GNSS issues and is driven by a Providers’ forum. Because of its specificity and its conductivity generated within the participants to this committee for the last 10 years, a detailed description of its objectives and working methods deserve to be presented in a more detailed fashion.

The ICG was formally established in December 2005 as a result of a series of actions following the Vienna Declaration “Space Millennium for Human Development” heralded at the UN level during the UNISPACE III conference in 1999. The declaration called, among other matters, for a universal access to and compatibility of space-based navigation and positioning. Through information exchanges among the different GNSS / RNSS providers, government officials, experts, and scientists, the ICG acts as an informal exchange information platform to eventually propose rules of the road to ensure the compatibility of the different PNT systems and their interoperability. This committee also makes sure that through appropriate measures, the deployment of new constellations and their modernization will not jeopardize permanent availability and integrity of the space-generated signals. Clearly, there is a concerted effort among the GNSS providers to lay down a set of practical standards allowing the different systems to co-exist, but at this stage nothing formal or of a soft binding nature exists.

Four ICG Working Groups (WG) cover the following areas:

- **WG-A, Compatibility and Interoperability**, deals with the following issues: compatibility, open service and performance, spectrum protection and interference detection and mitigation (IDM), interoperability, and international GNSS monitoring and assessment (IGMA). For compatibility, the objective is to agree on a common set of parameters for signals characteristics, and to prepare a convention for specified maximum power received, preventing a too powerful signal from one system to interfere with others. To harmonize open service performance, efforts are still underway among the different providers to eventually agree on templates specifying signal and system information policies of provision, and the minimum levels of performance offered for the open services. Formal data exchanges should be identified and interferences recorded automatically; an approach to practical modalities needs to gather ITU-R experts, spectrum regulators,
representatives of major industrial and transportation sectors using GNSS, and IGS and GNSS reference station developers. Concerning interoperability, answers to a range of important questions are still being worked out to guarantee full interoperable systems. Namely: means to adopt a common time scale steered by the Universal Coordinated Time (UTC), common / diverse carrier frequencies, common standards for geodesy and time references, monitoring and disseminating time offsets between each time system, multi-constellation / multi-frequency tracking of the IGS network, value of a common third open signal... The basic problem for IGM is to determine the service parameters to be monitored, make sure that the different GNSS are synchronized as closely as possible to UTC, and that the reference frame attached to each of these systems is in conformity with the International Terrestrial Reference Frame (ITRF).

- **WG-B, Enhancement of Performance of GNSS Services**, looks after improvement implementation for enhancing the performance of GNSS. For instance: alignment of different models used to estimate the spread in delay times of signals reaching a receiver after their propagation through the ionosphere and the troposphere, in order to have this spread in time being almost identical for each GNSS; mitigation of radio frequency interference with GNSS radio signals, an important source of degrading accuracy of user positioning together with multi-path propagation caused mainly by reflecting surfaces near the receiver; signal integrity covering satellite autonomous integrity, ground integrity and (Assisted) Receiving Autonomous Integrity Monitoring- (A) RAIM. For the latter, the objective is to establish a cooperative global multi-constellation monitoring network to support ARAIM, including the role of augmentation systems, which needs to be clarified. Other points considered by this working group are related to applications (GNSS services for indoor navigation, use of navigation satellites in disaster management, supporting GNSS applications in the Space Service Volume (SSV) domain).

- **WG-C, Information Dissemination and Capacity Building**, aims at exposing potential user communities to GNSS applications via, among other means, supporting organisation of workshops in developing countries.

- **WG-D Reference Frames, Timing and Applications**, focuses its activities on the alignment of the different geodetic reference frames and time scales that are currently in use, an essential goal to strengthen a robust interoperability. Indeed, a unique terrestrial reference frame to determine the position of a GNSS receiver or of a navigation satellite, together with one time scale steered by the UTC would be ideal since eventually there is a logic that space and time measurements refer to one reference system each. This WG has recommended to the ICG a progressive approach to that end, the details of which can be found in [5]. Thanks to a number of workshops, regular meetings of WG-D, close cooperation with WG-A, involvement in an International GNSS Service Multi-GNSS Global experiment, and the active participation of the BIPM (Bureau International des Poids et Mesures), a number of preparatory steps have been taken. One of them being to adopt at the next WRC-15 a continuous reference time-scale, whether by modification of UTC or some other method. Another step aims at having acceptance by the providers to monitor GNSS offsets between GNSS times, since this is the basis of interoperability and combined operations of the different GNSSs.

In short, ICG working groups activities cover issues that are essential to achieve a sustainable, reliable and secure use of GNSS applications, in particular those involving open services.

It should be noted too that preparatory actions for the future governance of GNSS have been made also through a number of bilateral agreements comparable to, for instance, the US-EU agreement of 2004 [6], or joint statements / announcements between the United States and Australia, China, EU, India, Japan, Russia, UK. In addition, the United States, for example, work on international issues through multilateral bodies including ICAO, IMO, ITU, COPUOS, APEC, NATO and WTO.

**IV. PATHS FOR A FUTURE GNSS GOVERNANCE**

**A. Preliminary considerations**

Before laying the ground for some governance propositions, two considerations should be highlighted:

1) Compared to other space-based information systems, GNSS is the only one generating original information directly from space destined to more than 99 % of the habitable Earth regions. The constellations behave as a network of individual lighthouse beacons, the position of which in space are constantly known with high accuracy. Like standard land-based lighthouse services, these beacons provide the necessary means to acquire a position fix and allow for a few meters-class navigation accuracy. Up to now, these beacons are of public nature, provided by governmental entities or agencies, differing significantly in nature -technically and legally- from space-based telecom and broadcast carrier beams.

These remarks, together with the limited number of global and regional providers (six in a foreseeable future, four of them being operated under the responsibility of a ministry of defence), call at first glance for a tailored governance scheme.

2) As already mentioned, each GNSS / RNSS was set up to respond to national (European in the case of Galileo) needs and has laid out its own set of managing and exploitation rules. In addition, the legacy of GPS and Glonass over 40 years makes difficult today to consider the possibility to dilute some prerogatives of each of the global systems, even if they were affecting open services only. Indeed, civilian and defence exploitation services rely on systems and subsystems, in space as well as on the ground, where information gathering,
processing and dissemination originating or destined to either military or civilian users relies on close technical intimacies, if not common. Moreover, each GPS, Glonass and BDS satellite features both defence and civilian functionalities, using the same atomic clock; compelling facts which make difficult any sharing or partial sharing of the constellations management. Indeed, since three out of four of the GNSSs are considered as a major instrument for guaranteeing the sovereignty and the independence of the United States, Russia and China, respectively, it is difficult to imagine a kind of minimum sharing of the constellations’ management among these three nuclear powers. It should be reminded also that for Galileo, security and sovereignty of Europe were the two basic motivations to start this new endeavour. At best, and this is the pursued objective for a sound governance, the GNSSs have to co-exist independently from each other, under a sustainable regime, delivering primary open signals shared by billions of users.

B. From an International Code of Conduct to a UN steered agency or an autonomous entity?

These preliminary observations relevant to the very nature of GNSS push for a light governance scheme based on an International Code of Conduct (ICoC) to be adopted. They pertain to a more general ensemble of guidelines that call for sustainable, reliable and secure space activities that are currently under discussion in several international fora, including the United Nations Committee on the Peaceful Uses of Outer Space (COPUOS). For GNSS, one of the ground rules would be for each provider to keep its prerogatives upon its own individual system, perfectly aware that his system will have to coexist independently from each other, under a sustainable regime, delivering primary open signals shared by billions of users.

1) At the lower end of the spectrum, a set of guidelines and rules of the road consisting of an ICoC to be adopted by the COPUOS and/or by the UN General Assembly. This could be presented at the initiative of one or several Members operating a GNSS. The signatories would be committed to follow a set of guidelines covering the following areas:
   • **Compatibility.** With the collaboration of ITU-R, maintain the necessary protection of the radio frequencies spectrum used by the GNSS; report systematically interference issues in full transparency and exchange information on interference mitigation procedures; agree on a common set of performance parameters for open services as well as GNSS monitoring capabilities. The overall objective should be to ensure complete compatibility over time of the systems, including when going through updates and modernization, and also when facing occurrence of possible new threats of electromagnetic nature;
   • **Interoperability.** Achieving sustainable interoperability among the different GNSS would require from the operating States and EU a number of actions to be implemented as rapidly as possible. i) Decide about a realistic time line to eventually have the same satellite time-scale for every constellation steered by the UTC; ii) decide, along the same principles, how and when an identical terrestrial reference frame will be used by all the providers. Although there are historical reasons and national resistances to give up on some currently adopted time-scales and terrestrial reference frames that can be harmonized through a chain of corrections, good will, common sense and operational technical aspects should prevail to end up with simplified and robust time-scale and terrestrial reference frame; iii) fix a date for making mandatory to broadcast systematically in navigation messages the time-scale offsets of each system and lay the steps to achieve this objective; iv) coordinate schedules among the different providers for putting into service new civil channels; v) harmonize practices for making public release of new interface control documents (ICD) needed *inter alia* by manufacturers for producing up-to-date chipsets.

   • **Vulnerability.** This is a rising preoccupation that concerns all the GNSS providers. Jamming and spoofing are developing rapidly; jamming prevents a position fix in a particular area, and spoofing results in an erroneous position assignment. The ICoC would require each provider to report such incidents under common agreed formats, as a first step, thereafter appropriate measures to be taken globally could be studied and lead to some recommended implementation. Burden sharing between stakeholders is much advised to save time, money and come to the best possible solutions. Considering the possibility for having to accommodate for enhanced and long lasting threats, backups are currently studied and tested at a national level, such as eLORAN in the United States and in Europe. ICoC should encourage large exchange of views among providers of such backup systems.

   • **New GNSSs.** Since plans for preparing the next GNSS generation, for instance post GPS III in the United States and Galileo 2.0 in Europe are underway, at least two sets of measures should be promoted by the ICoC: i) Ensure that a new constellation design displays functionalities that are consistent with existing ones so that the end user continues to take full advantage of all the available navigation constellations, ii) encourage shared architecture design of new constellations in order to maximize for the user the best geometrical viewing of transmitting satellites in different orbital planes, thus providing him with increased range accuracy. These optimizations can be made at no cost since the designers are planning anyway these new constellations architectures.

   • **Users requests.** The phenomenal growth of the GNSS embedded devices market coupled with the plethora of apps offering new services is giving a considerable leverage to the end users on the type of management that the providers should guarantee. Yet, the ICG currently considers a Providers’ forum only, refraining, as of today, to set up a Users’ forum because users’ concerns and requests are already taken care through actions recommended by the different Working Groups. Whereas
for its part, the US National space-based PNT Executive Committee takes into account the requests of institutional users represented by government departments only sitting in this Committee. Given the diverse nature of the community of users, ICoC should identify and categorize the needs of the different users first, followed by the set up of a feedback loop, thereby enabling the providers to be regularly informed by the demand side. One of the difficulties of this process will be to select the relevant and agreed representatives of the various groups of users. Time moving on, it seems unrealistic indeed to keep this community, made essentially of civil users, outside of the full GNSS picture. The ICoC will have to propose early on a suitable mechanism to bring in the user community on a permanent basis.

2) **Considering a more formal frame for the governance.**

In a nutshell, an International Code of Conduct is a set of rules commonly agreed by the different stakeholders, based on good will and on best efforts made in the common interest of the parties. The nature of such a Code is not legally binding. Some providers may consider that an ICoC, if necessary, should be sufficient to ensure a sustainable, secure and safe deployment, exploitation and modernization of the GNSSs. Others may prefer a more formal frame engaging the responsibility of the providers on the continuous availability and quality of services, arguing that the public nature of the deliverables, comparable to a global public good, addressed to and accessible by almost the entirety of humankind requires a structure with a minimum of binding rules. In order to be examined by the providers, who for the time being consider the ICG as a well-suited platform for the issues they are dealing with, the propositions need to be realistic, demonstrating a tangible added value compared to the current situation. In addition, the proposed format will have to be adapted to the limited number of providers.

Along this line, an organisation similar to Cospas-Sarsat, the well known programme providing accurate timely and reliable distress alert and location data to help search and rescue (SAR) authorities assist persons in distress, based on satellite dedicated payloads and worldwide disseminated beacons [7], may be interesting to consider. Indeed, this organisation consists of a Council of the four founding representatives, meeting at least once a year, supported by a light Secretariat assisting the Council. Cospas-Sarsat has 26 associated States, 10 using countries and works with 3 international organisations (IMO, ICAO, and ITU). The Council is in charge of administrating the programme, the management of the SAR system, the relations with the participants, and with the international organisations. Definitely, a light, reactive and efficient structure centered on a core mission.

Under a similar scheme, one could imagine to have the current ICG morphing into a comparable set up, having a Council in which the four GNSS and the two RNSS providers would sit, headed by a President elected for a limited term belonging to one of the providers’ nations. The ICG working groups would become dedicated programmes, with Programme Leaders, submitting their recommendations within a well-defined schedule to be approved by the Council. A Secretariat would support the Council in its different tasks. Such a transformed ICG should count also a number of participants representing communities of users and international organisations directly involved in the assignment and the usage of GNSS frequencies and signals. A Joint Providers-Users Committee would submit proposals and recommendations to the Council. For the years ahead, adoption of an action plan by the Council could be directly inspired by the aforementioned ICoC areas. This rather loose model of governance leaves to each provider the full management responsibility of its constellation and augmentation system.

It is debatable if the credibility and legitimacy of such a structure would require the steering and/or the support of the United Nations, or could do without it.

Guided by the necessity of proposing a light, simple and reactive scheme for the future governance of GNSS, the only one that could be eventually acceptable by the ICG Providers’, and that still remains to be seen, inspiration from larger organization was examined also. Hence, the World Meteorological Organization (WMO), the International Maritime Organisation (IMO), or even the International Mobile Satellite Organisation (IMSO) display an inappropriate format, unlikely to enhance the ICG to an agreed formal status. Indeed, the number of stakeholders with provider roles is too large, the decision-making and consultations process too cumbersome to fulfill the aforementioned requirements. Such models would probably send a wrong signal to the ICG Provider’s to come up with a consensual governance scheme.

3) **Some concluding remarks**

Although the ICG has proven its merits for the last ten years in building up convergences on a number of issues in a constantly evolving environment fed by the introduction of new constellations and the modernization of the legacy systems, it probably needs to be relooked at from the view point of the 2020 horizon, introducing mechanisms that indeed maximize the concretization of some essential deadlines. Precisely for these reasons, a firm up governance scheme would definitely help in reaching these goals. Two ways to move forward may be investigated: Either the COPUOS, after a debate among its members, recommends to the ICG to offer a set of governance modalities within a fixed deadline; or it could be up to the ICG to propose a viable scheme to be endorsed by the COPUOS, also under a deadline constraint. It is also anticipated that the new governance structure should address a number of forthcoming technical as well as socio-economic issues that will impact on the usage of GNSS, including: Assisted-GNSS for Indoor-outdoor navigation, miniaturized devices for inertial navigation that could challenge some of the GNSS services, long-term evolution of the space-based PNT systems, exploring new applications in close consultation with the users’ communities...
References


Considerations for Maritime Augmentation Service under Multi-constellation GNSS

Sang Hyun PARK  
GNSS Research Center  
Korea Research Institute of Ships and Ocean Engineering  
Daejeon, 305-343, Republic of Korea  

Ki Yeol SEO  
GNSS Research Center  
Korea Research Institute of Ships and Ocean Engineering  
Daejeon, 305-343, Republic of Korea

Abstract—DGPS is the representative augmentation method for maritime navigation. The maritime DGPS, which broadcasts GPS pseudorange corrections via medium frequency beacon signal, is a ground-based augmentation system as well as one of the most-used augmentation systems in maritime applications. Recently as the various satellite navigation systems are becoming available such as GLONASS, BDS, Galileo, and QZSS, there is growing need to provide maritime DGNSS service and establish maritime DGNSS related standards. For this reason, this paper deals with considerations for maritime augmentation service under multi-constellation GNSS. Firstly this paper introduces the structure and the operation process of the existing maritime DGPS. And then, the differences are shown between maritime DGPS and DGNSS. Finally it is presented that these differences can affect various parts of maritime augmentation service. It is anticipated that these analysis results will be utilized as a base study for maritime augmentation service under multi-constellation GNSS.

Keywords—multi-constellation; maritime augmentation; DGNSS; correction ages; integrity monitoring

I. INTRODUCTION

Differential Global Positioning System (DGPS), which is the representative augmentation method for maritime navigation, has been widely used to enhance GPS positioning performance due to its minimal burden on transmitting the correction data and the ease of correction processing to improve position accuracy. The maritime DGPS, which broadcasts GPS pseudorange corrections via medium frequency beacon signal, is a ground-based augmentation system as well as one of the most-used augmentation systems in maritime PNT (Positioning, Navigation, and Timing) applications. Such trend has increased the importance of the role of DGPS as a national integrity assurance system. Recently as the various satellite navigation systems are becoming available such as the GLObal Navigation Satellite System (GLONASS), the BeiDou Navigation Satellite System (BDS), and the Quasi-Zenith Satellite System (QZSS), there is growing need to provide maritime differential Global Navigation Satellite System (DGNSS) service and establish maritime DGNSS related standards.

This paper deals with considerations for maritime augmentation service under multi-constellation GNSS. Firstly, this paper introduces the structure and the operation process of the existing maritime DGPS based on software RSIM (Reference Stations and Integrity Monitors). And then, the differences are shown between maritime DGPS and DGNSS. Finally it is presented that these differences can affect various parts of maritime augmentation service, such as latency time, and considerations for multi-constellation GNSS on maritime GNSS will be discussed.

II. EXISTING MARITIME DGPS

Maritime DGPS consists of reference station (RS), integrity monitor (IM), and control station (CS) as shown in Figure 1 [7, 13]. The RS uses a reference receiver and an MSK modulator to generate the Radio Technical Commission for Maritime Service (RTCM) messages for broadcasting. The IM receives the RS broadcast messages and verifies that the transmitted corrections are within tolerance. The IM routinely provides positive system feedback to the RS to indicate normal, monitored operation. The IM generates alarms during out of tolerance conditions. The most important alarms are the corrected position error alarm and the pseudorange correction (PRC) alarms. These require a simple radial error check of the DGPS navigational solution against a known position and an analysis of the accuracy of individual ranges from the satellites to the known fixed position of the IM. The CS receives the alarm reports, isolates the points of failure, and initiates the corrective actions as appropriate. It manages the service by setting or resetting the equipment parameters, including data recording intervals, modes of operation, alarm thresholds and intervals, radiobeacon almanac information, and the RTCM message broadcast schedule.

Fig. 1. DGPS RSIM architecture.
For re-capitalization of maritime DGPS radiobeacon, we have considered the flexibility to support future service requirements and future GNSS improvements such as use of GLONASS, BDS, and Galileo, and lifetime of at least 10 years. However the hardware dedicated off-the-shelf RSIM has lack of meeting the above consideration for re-capitalization because it is necessary that the hardware dedicated off-the-shelf RSIM should replace the hardware with similar dedicated RSIM to meet future requirements and to change in the core GNSS within lifetime. For this reason, several studies have been made on effective re-capitalization method and plan [1, 2, 11].

The representative result among the previous studies was software RSIM presented by Ferguson et al. [2] as shown in Figure 2. Software RSIM is to implement the RS, IM, and CS in software, using commercial off-the-shelf hardware GNSS receivers. This has the advantage of flexibility as the hardware would be easy to replace and reconfigure. But software RSIM also has the weakness on flexibility for future GNSS signals compared with software radio integrated software RSIM by which software radio technology is used instead of hardware GNSS receiver and beacon receiver [11]. Software radio integrated software RSIM, on the other hand, has the weakness on maturity of its technology compared with software RSIM.

![Fig. 2. The architecture of software RSIM proposed by Ferguson et al. [2](image)](image)

The main functionality of integrity monitoring for the DGPS RSIM is to generate and verify the pseudorange corrections. The maritime DGPS checks maximum PRC and rate of change of the pseudo range correction (RRC), RSIM feedback message, user differential range error (UDRE), minimum number of satellites being tracked, correction age, pseudorange/pseudorange rare (PR/RR) residuals, absolute position error, DOP, message error ratio (MER), and signal strength (SS) to monitor the integrity. Based on the RTCM RSIM document [13], the performance requirements related to the integrity monitor are summarized as shown in Table 1.

<table>
<thead>
<tr>
<th>Alarm Parameters</th>
<th>Threshold</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum number of satellites</td>
<td>4SVs</td>
<td>RS/IM</td>
</tr>
<tr>
<td>Maximum PRC</td>
<td>100m</td>
<td>RS</td>
</tr>
<tr>
<td>Maximum RRC</td>
<td>4m/s</td>
<td>RS</td>
</tr>
<tr>
<td>IM feedback message</td>
<td>12s</td>
<td>RS</td>
</tr>
<tr>
<td>Correction age</td>
<td>30s</td>
<td>IM</td>
</tr>
<tr>
<td>PR residual</td>
<td>5m</td>
<td>IM</td>
</tr>
<tr>
<td>RR residual</td>
<td>0.5m/s</td>
<td>IM</td>
</tr>
<tr>
<td>Absolute position error</td>
<td>10m</td>
<td>IM</td>
</tr>
</tbody>
</table>

The core functionality of the integrity monitor for maritime DGPS is to transmit feedback messages to the reference station. The transmitted feedback information includes the correction age, the PR/RR residuals, and the absolute position error. The RTCM correction data received via radio beacon receiver are interpreted by the RTCM decoder to generate corrected pseudorange to perform integrity monitoring on correction data. Then, the PR/RR residuals and corrected positions are calculated and compared to the known position. This is followed by the creation of IM information, which includes position flags as well as PR residual flags, transmitted to the RS system via RSIM #20 message. The feedback message with the alarm information using the position flag is transmitted to the reference station, and sets the ‘do not use’ value in the RTCM message upon anomaly occurrence.

### III. CONSIDERATIONS FOR MULTI-CONSTELLATION GNSS

As the interest in reliability of PNT information becomes deeper than that of PNT accuracy, the number of maritime navigation equipment considering multi-constellation GNSS including GPS is on the rise. The international organizations, such as International Maritime Organization (IMO) and International Association of marine aids to navigation and Lighthouse Authorities (IALA), are also discussing multi-system shipborne navigation receivers containing maritime augmentation service under multi-constellation GNSS, which are a kind of maritime navigation equipment based on multiple radio navigation systems, in order to achieve the stable resilient PNT acquisition [5]. Now, GPS, GLONASS and BDS are available in Korea, and it is also possible to receive test signals of QZSS and Galileo before full operational capability. Within five years, the number of valid GNSS will be increased in Korea. It raises a question of course about the possibility of maritime multi-constellation service on the existing DGPS software RSIM. Above all things, the standard of maritime
DGNSS messages and service operation should be determined, which is already under way. This paper focuses the technical feasibility on broadcast of DGNSS augmentation data apart from the standards related with maritime DGNSS.

Maritime beacon messages are largely into RTCM SC104 type-1 message and type-9 message. Type-9 message serves the same purpose as type-1 message, in that it contains the primary differential corrections. However, unlike type-1’s, type-9 messages do not require a complete satellite set. The content and the format of type-9 message are identical to that of type 1 message, except that the number of satellites \(N_S\), and the number of 30 bit words \(N\) will be much smaller. Maritime beacon service does not provide RTCM SC104 type-1 message but broadcast type-9 message though type-1 message contains the corrections of all visible satellites. The reason is to maintain a maximum correction age under slow data links in the presence of noise, such as that encountered in radiobeacon operation. Figure 4 and Table 2 describe the gap of broadcast scheme between type-1 and type-9 messages and the difference of maximum correction age.

(a) First word of each RTCM SC104 message.

| 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 |
| Preamble | Message Type (Frame ID) | Station ID | Parity |
| 0 1 1 0 0 1 1 0 | MSB | LSB | MSB | LSB |

(b) Second word of each RTCM SC104 message.

| 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 |
| Modified Z-count Sequence No. | No. of Data Words Station Health | Parity |
| MSB | LSB | MSB | LSB |

(c) Type-1 message for differential GPS corrections.

Fig. 3. The message structure of Type-1 message [12].
(a) The broadcast schemes of type-1 messages (N₆ = 4).

(b) The broadcast schemes of type-9 messages (N₆ = 4).

Fig. 4. The broadcast schemes of RTCM SC104 correction messages.

TABLE II. THE DIFFERENCE OF MAXIMUM CORRECTION AGE BETWEEN TYPE-1 AND TYPE-9 MESSAGES UNDER 100bps.

<table>
<thead>
<tr>
<th>No of Satellites</th>
<th>Difference of maximum correction age (Type-1 minus Type-9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0 sec. (5.4 – 5.4)</td>
</tr>
<tr>
<td>5</td>
<td>0.6 sec. (6.6 – 6.0)</td>
</tr>
<tr>
<td>6</td>
<td>0.9 sec. (7.2 – 6.3)</td>
</tr>
<tr>
<td>7</td>
<td>0.9 sec. (8.4 – 7.5)</td>
</tr>
<tr>
<td>8</td>
<td>1.5 sec. (9.6 – 8.1)</td>
</tr>
<tr>
<td>9</td>
<td>1.8 sec. (10.2 – 8.4)</td>
</tr>
<tr>
<td>10</td>
<td>1.8 sec. (11.4 – 9.6)</td>
</tr>
<tr>
<td>11</td>
<td>2.4 sec. (12.6 – 10.2)</td>
</tr>
<tr>
<td>12</td>
<td>2.7 sec. (13.2 – 10.5)</td>
</tr>
</tbody>
</table>

The correction messages of GNSS except GPS will be inserted after sending type-9 correction messages of all visible satellites if the current broadcasting scheme is used in multi-constellation GNSS. So the correction age cannot help rising depending on an increase in the number of GNSS (N₆) which is needed to broadcast corrections. This phenomenon stimulates our curiosity about whether it could affect the integrity monitoring described as Table 1. Figure 5, which shows the maximum correction age according to the variation of the number of satellites (N₆₀) and the number of GNSS (N₆) under data rates of 100bps, gives the answer of the curiosity.

In the case of GPS standalone augmentation service for 8 satellites, the average maximum correction age is 6 seconds. However In the case of 3-GNSS augmentation service, the average maximum correction age has tripled. Eventually, growing the correction age is involved in corrected range errors and corrected position errors. We can say in other words that maritime augmentation service through multi-constellation GNSS enhances the reliability of PNT, but weakens the accuracy of PNT from the DGNSS user side. An increase of the correction age additionally could have a strong influence on the performance of time-to-alarm (TTA). From the TTA point of view, Maritime DGNSS service is normally possible under only up to 6 satellites for 2-GNSS.

Fig. 5. The maximum correction age according to the variation of the number of satellites and the number of GNSS under data rates of 100bps.

IV. CONCLUSIONS

This paper described the structure and the operation process of the traditional maritime DGPS in order to look into effects of multi-constellation GNSS on maritime augmentation service and the differences between maritime DGPS and DGNSS. It was presented that these differences could affect various parts of maritime augmentation service, such as correction age. We have finally reached the conclusions that maritime augmentation service through multi-constellation GNSS enhances the reliability of PNT, but could weaken the accuracy of PNT and the performance of time-to-alarm as well. It is anticipated that these analysis results will be utilized as a base study for maritime augmentation service under multi-constellation GNSS.

ACKNOWLEDGMENT

This work was supported by the Korea Research Institute of Ships and Ocean Engineering (KRISO) (No.PES1960).
REFERENCES


EDAS (EGNOS Data Access Service) for Added Value Applications

J. Vázquez, E. Lacarra, M.A. Sánchez, J. Morán
ESSP SAS (European Satellite Services Provider) Madrid, Spain
Juan.Vazquez@essp-sas.eu, Elisabet.Lacarra@essp-sas.eu, Miguel-Angel.Sanchez@essp-sas.eu, Jorge.Moran@essp-sas.eu

1. INTRODUCTION TO EDAS

EGNOS, the European Satellite Based Augmentation System (SBAS), provides corrections and integrity information to GPS signals over a broad area centered over Europe and is fully interoperable with other existing SBAS systems (e.g. WAAS, the American SBAS).

ESSP (European Satellite Services Provider) SAS is the EGNOS system operator and EGNOS Service provider for the following three services:

- **EGNOS Open Service (OS)**, freely available to any user.
- **EGNOS Safety of Life (SoL)** Service, that provides the most stringent level of signal-in-space performance for safety critical applications.
- **EGNOS Data Access Service (EDAS)**, which is the EGNOS terrestrial data service offering free of charge ground-based access to GNSS data to authorised users.

![Figure 1: EDAS High-Level Architecture](image)

As it can be observed in Figure 1, EDAS is the access point to the data collected and generated by the EGNOS ground infrastructure through the EGNOS stations network. In other words, EDAS gathers all the raw data coming from the GPS, GLONASS and EGNOS GEO satellites collected by all the receivers located at the EGNOS stations. There are currently 39 ground stations (Ranging and Integrity Monitoring Station - RIMS) and 4 uplink stations (Navigation Land Earth Stations - NLES), mainly distributed over Europe and North Africa. EDAS disseminates this GNSS data in real time to EDAS users and/or Service providers that distribute the data locally or to specific set of applications. In consequence, EDAS allows users to "plug in" to the EGNOS system by providing access to GPS/GLONASS satellite navigation and observation data, along with the EGNOS messages received by EGNOS ground stations, supporting a variety of other services, applications and research programs.

The European Commission, as owner of EDAS system, officially declared an initial set of EDAS services available to EU users in July 2012. Currently, the services provided by EDAS are as follows (please refer to the EDAS Service Definition Document (EDAS SDD) [1] for a detailed description:

- **Main Data Streams**: GNSS data is provided through Internet in real-time in ASN.1 format [2] (Service Level 0 (SL0) and Radio Technical Commission for Maritime Services (RTCM) 3.1 [3] format (Service Level (SL2).
- **Data Filtering**: Filtering capabilities to allow GNSS data to be received from only certain subsets of RIMS stations when connecting to EDAS Service Level 0 (SL0) and/or 2 (SL2). Currently 6 different groups of RIMS stations are defined according to geographical criteria.
- **SISNeT Service**: EGNOS messages provided in real time using the SISNeT protocol [4] defined by ESA.
- **FTP Service**: Historical GNSS data available through an File Transfer Protocol (FTP) site including:
  - EDAS SL0, SL2 raw data.
  - GPS/GLONASS navigation and observations (RINEX format [6])
  - EGNOS messages (EMS [7] + RINEX-B formats)
  - Ionosphere information in IONEX [8] format.
- **Ntrip service**: GNSS measurements in real time using Ntrip protocol [5], delivered in RTCM 3.1 [3], 2.3 [10] and 2.1 [9] formats. Within the data delivered by the Ntrip service, differential GNSS corrections as long as additional messages for RTK (Real-time kinematic) implementation are provided.
The following table summarises the types of data that can be retrieved via the different EDAS services.

<table>
<thead>
<tr>
<th>Mode</th>
<th>EDAS Service</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Time</td>
<td>SL0&amp;2</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>DF&amp;2</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>SISNet</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Ntrip</td>
<td>X</td>
</tr>
<tr>
<td>Archive</td>
<td>FTP</td>
<td>X</td>
</tr>
</tbody>
</table>

EGNOS data coming from the EDAS Services can be used for the development of applications based on GNSS data streams, or for the provision of added-value services based on EDAS. As examples, EDAS services are currently used for tracking dangerous goods, high precision positioning, management of airport fleets, engineering activities in the EGNOS programme, monitoring of GNSS performances, atmospheric investigation and R&D activities.

EDAS is accessible upon registration through the connection to a dedicated Internet domain (egnos-udas.eu). EDAS is free of charge and can only be used for non-safety critical purposes. In order to request an EDAS account, users should follow the steps detailed below:

1. Register in the EGNOS User Support Website:
   http://egnos-user-support.essp-sas.eu
2. Fill and submit the EDAS registration form (only accessible upon registration in the web):

After the verification of the provided data, the EGNOS Helpdesk will provide the user with the credentials and configuration details necessary to connect to the requested EDAS Service. Additionally, the website credentials will allow the user to download user oriented documentation and SW, such as the EDAS Client SW User Manual and the user information packages for each EDAS Service.

EDAS users are welcome to contact the EGNOS Helpdesk (egnos-helpdesk@essp-sas.eu) for EDAS registration and for any request or question related to EDAS.

II. EDAS SERVICES PERFORMANCE

The EDAS SDD [1] defines the committed performance levels for EDAS (levels that should always be met in a nominal situation) in terms of availability and latency:

- **Availability**: percentage of time in which EDAS is providing its services according to specifications (e.g. format as per applicable standard). The availability of EDAS services is measured at the EDAS system output (excluding user access network performance).
- **Latency**: time elapsed since the transmission of the last bit of the navigation message from the space segment (EGNOS and GPS/GLONASS satellites) until the data leaves the EDAS system (formatted according to the corresponding service level specification). EDAS latency is a one-way parameter defined for real-time services.

Based on the above definitions, the tables below provide EDAS services’ minimum availability and maximum latency:

<table>
<thead>
<tr>
<th>Performance</th>
<th>SISNeT GEO 1</th>
<th>SISNeT GEO 2</th>
<th>FTP</th>
<th>Ntrip</th>
<th>Data Filtering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability</td>
<td>98.5%</td>
<td>99%</td>
<td>98%</td>
<td>98%</td>
<td>98%</td>
</tr>
<tr>
<td>Latency</td>
<td>1.3 sec</td>
<td>1.450 sec</td>
<td>1.150 sec</td>
<td>N/A</td>
<td>1.75 sec</td>
</tr>
</tbody>
</table>

The availability and latency achieved from September 2014 to May 2015 are shown in the figures below. For the case of latency, when a service is composed of multiple data streams (e.g. EDAS SISNeT service provides data for both operational EGNOS GEO satellites), the average delay is shown.

As seen above, EDAS availability has been above 98% over the last year and the latency has been consistently below the target delays, always below 1 second. Hence, it
can be easily seen that the commitment values have been met for all the services throughout the whole period.

III. EDAS FOR ADDED-VALUE APPLICATIONS

A. EDAS SISNeT for Real-time EGNOS performance monitoring

A straightforward application of the EDAS SISNeT service is its use for real-time EGNOS performance monitoring. EDAS SISNeT provides access to the EGNOS messages over the Internet and is robust to local effects (interference, user visibility) and failures of specific receivers: as long as at least one EGNOS stations is tracking the target GEO satellite, its messages will be output through EDAS.

ESSP, as the EGNOS services provider, is also using EDAS SISNeT service as one of the data sources for the provision of the EGNOS real-time performance information, which is available at the EGNOS User Support Website (http://egnos-user-support.essp-sas.eu/). As an example, the availability map (Figure 4) is updated in real time using EDAS SISNeT Service.

Figure 4: 24 hours availability map on the EGNOS User Support Website

B. EDAS for EGNOS-based Virtual Reference Stations

IALA (International Association of Lighthouse Authorities) DGNSS is the internationally accepted method of providing differential GNSS corrections and integrity information to maritime users. There are more than 300 IALA DGNSS stations in over 45 countries throughout the world including the majority of European Member States.

However, it is noted that some of the DGNSS networks existing in Europe are becoming obsolete. In this line, IALA has recently recommended [11] that members operating DGNSS services should consider modernization, to ensure that levels of service can be maintained and future requirements can be met.

In this scenario, EDAS can provide a cost-effective solution for being used as a primary or backup source for the generation of differential GNSS corrections. For instance, one possible alternative for the modernization of the beacon DGNSS service would be to broadcast the EGNOS corrections in RTCM SC-104 [9] through IALA Medicum Frequency beacons.

The architecture of this solution could consist on the following elements:

- **EDAS SISNeT (or SL2) interface**: Responsible for retrieving in real-time the EGNOS corrections in SBAS [12] format as well as the GPS ephemeris from the EDAS SISNeT service.
- **Main Processing Facility**: For each of the IALA beacons comprising the DGNSS network, the central facility would compute the GPS satellites in view (based on the GPS ephemeris obtained from EDAS). For each of the satellites in view by each beacon, the pseudorange correction would be computed using the EGNOS corrections (satellite orbit and clock, ionosphere).
- **IALA Beacons**: The EGNOS corrections are broadcast in RTCM SC-104 to the maritime users through a network of IALA beacons.

![Figure 5: EGNOS-based Virtual Reference Stations Architecture](image-url)

It is worth to remark that, for those countries with an existing DGNSS infrastructure, EGNOS can be used as a resilient solution for the traditional DGNSS service. In case of failure in one of the RSIM stations (Reference Stations and Integrity Monitors), the EGNOS corrections can be used as a backup source of data for the generation of the RTCM SC-104 messages.

Apart from being used as a primary or backup source of data for a DGNSS beacon service, EGNOS and in particular EDAS can also provide a cost-effective solution for the generation of the AIS Message Type 17. The Automatic Identification System (AIS) is a ship borne radio communications system introduced by IMO for exchanging static, dynamic and voyage related data between equipped vessels, AtoN and onshore base stations. Particularly, the message type 17, allows AIS Base Stations to broadcast differential GNSS (DGNSS) corrections to the vessels within the range of coverage. The contents and format of the AIS message 17 is in accordance with the RTCM SC-104 standard, but
excluding the parity and preamble information included in the RTCM messages.

To this regard, it is noted that some commercial solutions (e.g. Kongsberg AIS BSC 420) already have the capability of transmitting EGNOS corrections (converted to RTCM-104) through the AIS Message 17. Moreover, it is also important to highlight that the vast majority of GNSS receivers are able to receive and process RTCM SC-104 messages. Hence, the dissemination of EGNOS corrections in RTCM SC-104 format, could benefit those receivers not able to receive the EGNOS SIS, but capable of processing the RTCM SC-104 messages.

Within the framework of the IRIS Europe 3 project (www.iris-europe.net), the Hungarian project partner RSOE (National Association of Radio Distress-Signalling and Infocommunications) implemented a pilot application to assess the performance of various types of DGNSS corrections to be used in inland waterway navigation. One of the goals of this pilot test was to identify correction sources that could satisfy the requirements of the Hungarian PannonRIS system. Alberding GmbH, a leading developer and distributor of professional GNSS system solutions has been contracted by RSOE to deliver the required services for the pilot application.

Different types of DGNSS corrections have been used in a position performance monitoring test during the first months of 2014. One of the correction sources used was the EDAS SL2 service. The Alberding DataConv software has been used to convert EDAS SL2 based RTCA data (EGNOS corrections) to RTCM 2.3 format differential corrections and output that for a PannonRIS base station location in Budapest, Hungary. Correction data availability and positioning accuracy have been analysed using two Hungarian GNSS monitor stations located by the river Danube and the Alberding Monitor software provided by Alberding GmbH. The achievable positioning performance was assessed at two monitor stations located at 10 km and 100 km distance from the AIS station, respectively. The test showed sub-meter level horizontal position accuracy over the short baseline (mean error of 0.52 m, 2σ of 0.70 m) and slightly higher errors over the longer baseline (mean error of 0.71 m, 2σ of 1.01 m).

The Budapest pilot test confirmed that both the availability and accuracy of the EDAS-based corrections are suitable for typical inland waterway navigation positioning requirements in Central Europe. Considering the pilot test results it could be concluded that EGNOS/EDAS may be used either as a backup solution or even as a primary correction source of an inland waterway DGNSS service.

C. EDAS Ntrip for high accuracy applications

The EDAS Ntrip service provides, through an Ntrip caster, raw measurements in different RTCM formats from all the EGNOS stations. In particular, this service provides the DGPS corrections and RTK messages that enable users to obtain a DGPS or RTK position fix using EGNOS stations as reference stations. The application of these techniques provides accuracies in the sub-meter to centimetre level. Of course, the aforementioned techniques have a baseline limitation (maximum distance between user and reference stations) which means that, taking into account the distribution and density of EGNOS stations, an European wide coverage cannot be provided. However, thanks to the use of well known standards and protocols (RTCM, Ntrip), EDAS could be a solution for high accuracy applications in specific geographical areas or act as a complement to other existing networks of reference stations.

In consequence, EDAS could be a solution for surveying and agriculture. For example, mining usually takes places in clear-sky (free from obstructions) and requires centimetre level accuracies, which is in line with the accuracies obtained applying DGPS and RTK techniques based on EDAS Ntrip Service.

Warsaw was the selected location in order to show the obtained performances using high precision techniques with EDAS Ntrip Service.
RTCM 3.1) and the correction data stream was taken from the BOR1 station in Borowiec, Poland (EUREF permanent network, http://www.epncb.oma.be/) which provides RTCM 2.3 DGPS corrections. The distance between both locations is around 300 km meaning that this is a medium/long baseline solution for DGPS. The map below further illustrates the scenario.

For the RTK solution in Warsaw, both the user location and reference station data come from EDAS. In this case, the EGNOS RIMS A and B sites were used, accessing the RTCM 3.1 mountpoints for the respective locations through EDAS Ntrip service.

Below, a summary table provides the following results during a 24 hours period from August 15th to August 16th 2014:

- **Availability**: percentage of time in which a positioning solution was available.
- **Accuracy**: in terms of error in the North, East, height and horizontal directions. For these errors, the mean and 2-sigma (= percentile 95) value of the accuracy throughout the observation time are provided.

<table>
<thead>
<tr>
<th></th>
<th>WARSAW (DGPS)</th>
<th>WARSAW (RTK)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>2σ</td>
</tr>
<tr>
<td>North [m]</td>
<td>-0.41</td>
<td>0.38</td>
</tr>
<tr>
<td>East [m]</td>
<td>0.14</td>
<td>0.42</td>
</tr>
<tr>
<td>Height [m]</td>
<td>-0.15</td>
<td>1.22</td>
</tr>
<tr>
<td>Horizontal [m]</td>
<td>0.49</td>
<td>0.34</td>
</tr>
<tr>
<td>Availability [%]</td>
<td>99.19</td>
<td>98.83</td>
</tr>
</tbody>
</table>

TABLE 3 shows that, as expected, the 2σ accuracies in the horizontal directions are around 50 cm and below 15 cm for the DGPS and RTK solutions respectively. In terms of availability, a high level of robustness is observed with a positioning solution being achieved more 98.5% of the time in both cases, which is due to stability and quality of the input data streams used for the analysis.

In order to further illustrate the results, the figures below show the geographical distribution of position error in the horizontal plane. DGPS results are shown in blue colour whereas RTK ones are given in red. The three concentric red circles show the 1σ, 2σ and 3σ limits of the horizontal accuracy.

For the case of agriculture, it is a common practice to use the pass-to-pass accuracy value as the key parameter to illustrate the performance level of a given positioning service. The pass-to-pass accuracy corresponds with a short-term dynamic performance, which is determined from off-track errors along straight segment passes occurring within a 15 minute time frame. According to key actors in this domain, the Pass to Pass accuracy is used for assessing the precision of guidance equipment.

ISO 12188-1 [13] was used as the reference for the calculations in this paper. Part 1 of ISO 12188 provides a procedure for evaluating and reporting Pass to Pass accuracy of navigation data, which is determined using positioning devices that are based on GNSS. It specifies common performance parameters that can be used to quantify performance of different positioning devices. The Pass to Pass accuracy results presented in this paper are based on ISO 12188 [13] Section 4.2.5, which describes this performance parameter and the way to compute it using static measurements.

The table below shows the maximum value of the pass to pass accuracy (between North and East directions) for the week from June 24th to June 30th 2014. It shows that, using differential GPS, the pass-to-pass accuracy is always below 11 cm whereas when the RTK solution is taken, it is kept below 2 cm for all the days.

<table>
<thead>
<tr>
<th></th>
<th>WARSAW (DGPS)</th>
<th>WARSAW (RTK)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass to pass (m)</td>
<td></td>
</tr>
<tr>
<td>24/06</td>
<td>0,10050</td>
<td>0,01475</td>
</tr>
<tr>
<td>25/06</td>
<td>0,09989</td>
<td>0,01717</td>
</tr>
<tr>
<td>26/06</td>
<td>0,10803</td>
<td>0,01752</td>
</tr>
<tr>
<td>27/06</td>
<td>0,10589</td>
<td>0,01576</td>
</tr>
<tr>
<td>28/06</td>
<td>0,10693</td>
<td>0,01547</td>
</tr>
<tr>
<td>29/06</td>
<td>0,106010</td>
<td>0,01684</td>
</tr>
<tr>
<td>30/06</td>
<td>0,10517</td>
<td>0,01592</td>
</tr>
</tbody>
</table>

The achieved results using EDAS data are in line with the expected performance level to be provided by each positioning technique. Also, the fact that an external
correction source (independent from EDAS/EGNOS) has been used for the DGPS solution demonstrates the alignment of EDAS RTCM data streams with the standard. In addition, the very high availability performance gives a feeling on the stability of EDAS services.

In conclusion, EDAS, although not able to support a high accuracy service European-wide, is indeed capable of contributing to the existing networks of GNSS data streams and support high accuracy positioning techniques within a certain distance from the EGNOS stations.

IV. CONCLUSIONS

EDAS is the access point to the data collected and generated by the EGNOS ground infrastructure through the EGNOS stations network in real time and off line in form of a FTP archive. EDAS has been freely available for the GNSS community in the European Union since July 2012, with a commitment of minimum availability of 98.5% for the main data services and of 98% for the rest of the EDAS Services [1].

EDAS has already proven to be a versatile service, supporting professional users in different commercial applications. In this paper, the added value that EDAS can bring has been analysed in three different scenarios:

- **EDAS for EGNOS real-time performance monitoring**: ESSP is using EDAS as source for EGNOS messages retrieval for performance analysis/monitoring in real-time, due to its robustness to local issues, high stability and low delay.

- **EDAS based Virtual Reference Stations for maritime navigation**: the potential use of EDAS data to compute differential GPS corrections for maritime navigation has been discussed. A R&D project currently ongoing in Hungary has been referred, which preliminary results indicate that the achievable performance level is sufficient for inland waterway navigation and that EGNOS/EDAS may be used either as a backup solution or even as a primary correction source of an inland waterway DGNSS service.

- **EDAS for high accuracy applications**: EDAS could be used for high precision applications, such as agriculture and surveying domains. Performance assessment in Warsaw has been presented, showing centimetre-level accuracy when EDAS data is used to feed DGPS and RTK positioning techniques, hence proving the quality and stability of EDAS products. Even if the EGNOS network is not designed to provide a European wide DGNSS/RTK service (not sufficient number/density of stations), EDAS could still support users in the high accuracy business for specific locations or as a complement to other existing networks.

Hence, EDAS potential to support users in different application domains has been shown. The above examples are provided with the aim of encouraging GNSS community to analyse the EDAS services portfolio and identify the specific services and/or products which could better suit their needs. For that purpose, European citizens are welcome to register to the EDAS Services (http://egnos-user-support.essp-sas.eu) or contact to the EGNOS Helpdesk (egnos-helpdesk@essp-sas.eu) for further information.

ACKNOWLEDGMENTS

Firstly, the authors would like to acknowledge the efforts done by European Commission (EC) and the European GNSS Agency (GSA) for continuously supporting the EDAS program and for launching the evolution project that included the enhancements to the EDAS support services and the addition of a new set of added value services.

The authors would like to acknowledge Tamás Horváth (Alberding GmbH) and RSOE for contributing to this paper with the very interesting information on the IRIS Europe 3 project.

Finally, the authors would like to thank their ESSP colleague Pedro Gómez for his invaluable support to this paper.

REFERENCES

[8] IONEX: The IONosphere Map EXchange Format Version 1
Assumptions of Full Mission Ship Bridge Simulation Including EGNOS

P. Zalewski, L. Gucma
Marine Traffic Engineering Centre, Maritime University of Szczecin (MUS) Szczecin, Poland

K. Urbanska, S. Schlueter
EGNOS Project Office European Space Agency (ESA) Toulouse, France

A. Born
Institute of Communications and Navigation German Aerospace Center (DLR) Neustrelitz, Germany

M. Porretta
European Space and Technology Research Centre (ESTEC) European Space Agency (ESA) Noordwijk, The Netherlands

Abstract—This paper presents the assumptions of the Full Mission Bridge Simulation (FMBS) studies which aim to define the role of European Geostationary Navigation Overlay System (EGNOS) from a marine user perspective. The work focused on the possible benefits of Satellite Based Augmentation Systems (SBAS) integrity data presentation within the ship’s bridge environment. At the present stage of research, the MVPA (Marine Vessel Protection Area) model is under development for ships navigating in limited water areas and utilizing precise navigation by fusion of EGNOS integrity data with course/heading error estimates. The model takes into consideration several aspects, influencing protection area dimensions. These include signal coverage aspects, ship’s size, ship’s heading accuracy, position of the Global Navigation Satellite Systems (GNSS) antenna, and GNSS/EGNOS integrity data. The output from this model can be displayed graphically on the desired route within the Electronic Chart Display and Information System (ECDIS) environment as ship’s uncertainty area (MVPA in ECDIS). This ECDIS layer constitutes an EGNOS integrity display within bridge environment showing graphical representation of the ship and integrity information (protection levels in conjunction with alert limits). The plan of experiments to be conducted in the controlled environment of navigational FMBS is discussed with regard to possible benefits of proposed graphical representation of EGNOS integrity data.

Keywords—GNSS Augmentation Systems, EGNOS, Full Mission Bridge Simulator, FMBS, Marine Vessel Protection Area, MVPA, EMPONA

I. INTRODUCTION

The research, the assumptions of which are presented in the paper, is carried out in the “EMPONA” (Implementation of EGNOS in the Maritime Domain as Effective Augmentation System for Positioning in Inland and Pilot Navigation) project financed by ESA. The expected outcomes are:

1) Description of opportunities and approaches for utilization of EGNOS within the maritime Positioning, Navigation and Timing (PNT) system.

2) Performance results based on real experiments of EGNOS in comparison to IALA Beacon Differential GNSS (DGNSS) in transition service areas and in test beds.

3) Model of GNSS/EGNOS ship’s uncertainty area by fusion of EGNOS integrity data with ship’s heading estimated error.

4) Specially designed layer of ECDIS showing ship’s uncertainty area based on EGNOS integrity data.

5) Simulation results of influence of navigational conditions and marine operations on Marine Vessel Protection Area (of proposed acronym MVPA).

6) Method of Alert Limit (AL) estimation and prediction in various navigational conditions or marine operations. In particular, AL recommendations will be derived based on the actual performance of the proposed system with respect to fixed values of safety limits like the “safety isobath” or “cross track error (XTE)” in the Electronic Chart Display and Information System (ECDIS).

The 5th and 6th outcomes will be achieved by FMBS studies under the assumptions that they reflect with high fidelity the technical Human Machine Interface (HMI), external environment and dynamics of marine vessel. These assumptions comprise integration of data from real experiments with simulated environment and construction of credible mathematical model of the ship’s MVPA. The proposed model takes into account several aspects, influencing resultant MVPA dimensions, like coverage factors, ship’s size, ship’s heading accuracy, position of the GNSS antenna, and GNSS/EGNOS integrity data. The output from this model can be displayed graphically around ship’s contour within the ECDIS, together with other alphanumeric integrity data.

The mathematical model and MVPA construction algorithm is based on an integrity concept which is derived from EGNOS provided data [1]. The model is under development for systems using real data of GPS augmented by EGNOS and for systems using simulated data in FMBS.
II. POINT POSITIONING MARINE INTEGRITY CONCEPT OF GPS AUGMENTED BY EGNOS

The EGNOS augmented GNSS point positioning integrity concept in the maritime domain has been adapted from the model defined within the Appendices A & J of the Minimum Operational Performance Standards (MOPS) for airborne equipment [4]. The concept is based on the broadcast of differential GPS corrections in message types MT1-5, 7, 9, 17-18, 24-26 and corresponding integrity data in MT2-6, 10, 24, 26-28 transmitted by EGNOS geostationary satellites (PRN120, PRN124, and PRN126). The input quantities for the integrity algorithm on the user side are:

1) Geometry between GPS satellites and user derived position from observations of the GPS satellites (the geometry matrix G of size nx4):

\[
G(n \times 4) = \begin{bmatrix}
- \cos e_i \sin A_{i1} & - \cos e_i \cos A_{i1} & - \sin e_i & 1 \\
- \cos e_i \sin A_{i2} & - \cos e_i \cos A_{i2} & - \sin e_i & 1 \\
\vdots & \vdots & \vdots & \vdots \\
- \cos e_i \sin A_{in} & - \cos e_i \cos A_{in} & - \sin e_i & 1 
\end{bmatrix}
\]  

(1)

where \(e_i\) and \(A_{i}\) are the elevation and azimuth angles between the receiver antenna and the \(i^{th}\) satellite (\(i=1,2,...,n\)), and \(n\) is the number of visible satellites, respectively.

2) User differential range error \(\sigma_{UDRE}[m]\), transmitted by the EGNOS satellite (MT2-6, 24) via UDREI, indicator (conversion according to the table A-6 defined within section A.4.4.4 of [4]).

3) Grid ionospheric vertical error \(\sigma_{GIVE}[m]\), transmitted by the EGNOS satellite (MT26) via GIVEI, indicator (conversion according to the table A-17 defined within section A.4.4.5 of [4]).

4) Residual tropospheric error parameter \(\sigma_{d}\) [m], calculated according to the model defined within section A.4.2.5 of [4]:

\[
\sigma_{d} = 0.12 \frac{1.001}{\sqrt{0.002001 + \sin^2 e_i}} 
\]  

(2)

5) Error of marine (shipborne) receiver \(\sigma_{mr}[m]\), depending on receiver properties, derived by analogy to the model defined within section J.2.4 of [4]:

\[
\sigma_{mr} = \sqrt{\sigma_{mr}^2 + \sigma_{dultpath}^2 + \sigma_{divg}^2} 
\]  

(3)

where:

\(\sigma_{dultpath}\) is estimated multipath error [m] depending on \(i^{th}\) satellite elevation \(E_i\), the reference model of this error can be found in [4] but its verification in marine environment will be necessary according to the methodology presented in [5];

\(\sigma_{divg}\) is estimated error [m] induced by the steady-state effects (divergence) of shipborne receiver smoothing filter assumed to be identical to the one presented in [4];

\(\sigma_{noise}\) is estimated error [m] associated with GNSS receiver for \(i^{th}\) satellite, including receiver noise, thermal noise, interference, inter-channel biases, extrapolation, time since smoothing filter initialization, and processing errors; assumed to be identical to the one presented in [4];

On the basis of the input quantities the weight matrix \(W\) is built under assumption of uncorrelated, EGNOS corrected, measurements characterized by the inverse variances (total estimated errors) of the distances to the observed satellites:

\[
\sigma_i^{2} = \sigma_{mr}^{2} + \sigma_{GIVE}^{2} + \sigma_{dultpath}^{2} + \sigma_{divg}^{2} 
\]  

(4)

\[
W = \begin{bmatrix}
\frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\
0 & \frac{1}{\sigma_2^2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{1}{\sigma_n^2}
\end{bmatrix}
\]  

(5)

Where, in (4):

\(\sigma_{mr}^{2}\) is the model variance for the residual error associated to \(\sigma_{dult\text{imr}}\) as defined in section A.4.5.1 of [4] [m²];

\(\sigma_{GIVE}^{2}\) is the model variance for the slant range ionospheric error associated to \(\sigma_{mr\text{GIVE}}\) as defined in sections A.4.4.10 and A.4.5.2 of [4] [m²].

Finally the point positioning covariance matrix is found:

\[
\begin{bmatrix}
s_E^{2} & s_{EN} & s_{EU} & s_{ET} \\
s_{EN} & s_N^{2} & s_{SU} & s_{ST} \\
s_{EU} & s_{SU} & s_U^{2} & s_{UT} \\
s_{ET} & s_{ST} & s_{UT} & s_T^{2}
\end{bmatrix} = \left(G^T W G\right)^{-1} 
\]  

(6)

where:

\(s_E^{2}\) is the variance of the antenna receiver Easting measurement in the local reference frame centered on the GPS antenna (East, North, Up, ENU) [m²];

\(s_N^{2}\) is the variance of antenna receiver Northing measurement in the local reference frame (ENU) [m²];

\(s_U^{2}\) is the variance of antenna receiver vertical measurement [m²];

\(s_T^{2}\) is the variance of receiver time correction measurement multiplied by speed of light [m²]; and, finally, the mixed terms (e.g. \(s_{EN}\) etc.) are the co-variances of respective measurements [m²].

The “elliptical” assessment of the point positioning user integrity can be given as a protection ellipse (\(PE_{conf}\)), which is specified by 4 parameters, i.e.: (1) semi-major axis of the estimated position error ellipse, \(d_h[m]\); (2) semi-minor axis of the error ellipse, \(d_o[m]\); (3) orientation of the error ellipse, \(\Phi\); and (4) coverage factor, \(k\), based on the confidence intervals. The rationale for the \(k\) factors comes from the assumption of uncertainty normal distribution in both the North and the East directions of position parameters. For example, to create a 95% confidence ellipse from the 1σ error ellipse, a factor of \(k=2.45\) must be used; in contrast, to get a 99.8% confidence level, a factor of \(k=3.5\) must be used, and so on. As an initial assessment, a \(k\) factor of 4.9 (or larger) will be used. This
corresponds to an integrity risk of $10^{-5}$, which is the value recommended by IMO for Future Global Navigation Satellite systems, [3]. The integrity risk (or probability of Hazardous Misleading Information, HMI) is the probability that the user will experience a true position outside the protection ellipse, $PE_{mr}$. In summary:

$$PE_{mr} = \left[ \frac{kd_x}{kd_y} \right] \Phi$$

(7)

where:

$$d_x = \sqrt{\frac{s_E^2 + s_N^2}{2} + \left(\frac{s_E^2 - s_N^2}{2}\right)^2 + s_{EN}^2}$$

(8)

$$d_y = \sqrt{\frac{s_E^2 + s_N^2}{2} - \left(\frac{s_E^2 - s_N^2}{2}\right)^2 + s_{EN}^2}$$

(9)

$$\Phi = \frac{\pi}{2} - \frac{1}{2} \cdot \text{atan}^2 \left(2s_{EN}, s_E - s_N^2\right)$$

(10)

$\Phi$ is a clockwise angle of rotation from North either of the semi-major ellipse’s axis (if $s_E > s_N$) or of the semi-minor axis (if $s_E < s_N$); and \text{atan}^2 is the arctangent function with two arguments $(y, x)$.

III. CONCEPT OF MARINE VESSEL PROTECTION AREA BASED ON EGNOS INTEGRITY DATA

The concept of Marine Vessel Protection Area (MVPA) determination with EGNOS augmented GPS as positioning source is based on the assumption that the vessel is a 2-dimensional (2D) spatial object and as such is presented in electronic chart system (ECDIS) integrated into Full Mission Bridge Simulator (FMBS). The pilot has to monitor the position of the vessel contour relative to declared safety isobaths (navigation dangers). As a result, the vessel position cannot be treated as a point. The mathematical model describing how the vessel is presented on the bridge navigation display can be expressed by two observation equations:

$$x_{j,N} = x_j - x_{GPS} + d_j \cos(\psi + \alpha_j)$$

(11)

$$y_{j,E} = y_j - y_{GPS} + d_j \sin(\psi + \alpha_j)$$

(12)

where:

$$d_j = \sqrt{x_j^2 + y_j^2}$$

(13)

$$\alpha_j = \frac{\pi}{2} - \text{atan}^2 (x_j, y_j)$$

(14)

$x_j, y_j$ are the calculated coordinates of consecutive $j$ points of ship’s contour in the body-fixed reference frame (this is fixed to the marine vessel at the common reference point of aft perpendicular with positive $x$ axis up, $y$ axis right, following the convention used in marine craft hydrodynamics and simulations [2] – Fig. 1);

$x_{GPS}, y_{GPS}$ are the coordinates (offsets from 0 at aft perpendicular, see Fig.1) of EGNOS augmented GPS receiver antenna in the body-fixed reference frame;

$x_{j,N}, y_{j,E}$ are the calculated coordinates of consecutive $j$ points of ship’s contour in the local reference frame (ENU);

$x_N, y_E$ are the recorded positions of EGNOS augmented GPS receiver antenna in the local reference frame (ENU);

$\psi$ is the heading of marine vessel counted clockwise from North in the local reference frame (ENU);

$d_j$ is the $j^{th}$ distance between GPS antenna and $j^{th}$ point of ship’s contour; and

$\alpha_j$ is the $j^{th}$ angle between GPS antenna and $j^{th}$ point of ship’s contour counted clockwise from $x$-axis in body-fixed reference frame.

Prior to presentation of the vessel contour in navigation chart system the following requirements must be fulfilled:

1) Vessel’s contour points $(x_j, y_j)$ need to be set and uploaded to the navigation system. Depending on the size and shape (curvature) of the vessel hull the number of points describing vessel’s contour may vary. For most applications 14 to 16 points $(j=1,2,...,14)$ have been assumed as sufficient. The origin of the vessel’s body-fixed coordinate frame is referred as CCRP (consistent common reference point) and usually in ECDIS-es it is fixed either to the craft’s aft perpendicular or to the navigator’s conning position. The reason for CCRP fixing to the conning position (front-centre of the bridge) is to minimize the parallax error resulting from different placement of various electronic and optical position fixing devices in the ship.

2) GPS antenna offset should be measured with the highest possible accuracy in the prevailing circumstances,
preferably with use of land survey techniques (e.g. a Total Station Theodolite, TST).

The errors of parameters in equations (11) and (12) will propagate into the final MVPA according to Gauss’s Error Propagation Law. Systematic errors of \( x_{\text{GPS}}, y_{\text{GPS}}, d_{j} \) and \( \alpha_{j} \) can be minimized to a negligible magnitude by precise dimensional control. Therefore, only propagation of other parameters’ errors (\( x_{j}, y_{j} \)) is taken into account in the MVPA determination according to the formula:

\[
C_{j,PA} = J_{j}C_{j}^{T}
\]

(15)

where:
\( C_{j,PA} \) is the covariance matrix of derived quantities:
\[
C_{j,PA} = \begin{bmatrix} s_{x,j}^{2} & s_{x,j,EN} \\ s_{x,j,EN} & s_{x,j,N}^{2} \end{bmatrix}
\]

(16)

where:
\( s_{x,j}^{2} \) is the Easting variance of consecutive \( j \) points of ship’s contour in the local reference frame (ENU) [m²];
\( s_{x,j,N}^{2} \) is the Northing variance of consecutive \( j \) points of ship’s contour in the local reference frame (ENU) [m²];
\( s_{x,j,EN} \) is the covariance of \( j \) points respective coordinates [m²];
\( J_{j} \) is the Jacobian matrix (matrix of all first-order partial derivatives) of equations (11) and (12) excluding \( x_{\text{GPS}}, y_{\text{GPS}} \) due to their negligible errors:
\[
J_{j} = \begin{bmatrix} 1 & 0 & \sin(\psi + \alpha_{j}) & d_{j} \cos(\psi + \alpha_{j}) \\ 0 & 1 & \cos(\psi + \alpha_{j}) & -d_{j} \sin(\psi + \alpha_{j}) \end{bmatrix}
\]

(17)

\( C \) is the covariance matrix of observations:
\[
C = \begin{bmatrix} s_{x}^{2} & s_{x} s_{y} & 0 & 0 \\ s_{x}s_{y} & s_{y}^{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{\nu}^{2} \end{bmatrix}
\]

(18)

where \( s_{\nu}^{2} \) is the marine vessel heading variance, usually fixed for specific attitude / heading equipment (the typical values are in range 0.5°-1°); and
\( J_{j}^{T} \) is the transposed Jacobian matrix (17).

Estimated error of each \( j \)th ship’s contour point involves the errors of two jointly distributed variables of \( s_{j},y_{j} \) coordinates. Thus, the positional error follows a bivariate normal distribution. Taking above into account, to fully describe the estimated error of each \( j \)th point, it is necessary to determine the orientation \( \Phi_{j} \) and lengths of the semi-major \( d_{j,a} \) and semi-minor axes \( d_{j,b} \) of the \( j \)th error ellipses according to the formulas anallogical to (8)–(10):

\[
d_{j,a} = \frac{s_{x,j}^{2}+s_{y,j}^{2}}{2} + \frac{\left( \frac{s_{x,j}^{2}+s_{y,j}^{2}}{2} \right)^{2}}{s_{x,j,EN}^{2}}
\]

(19)

\[
d_{j,b} = \frac{s_{x,j}^{2}+s_{y,j}^{2}}{2} - \frac{\left( \frac{s_{x,j}^{2}+s_{y,j}^{2}}{2} \right)^{2}}{s_{x,j,EN}^{2}}
\]

(20)

where \( \Phi_{j} \) is a clockwise angle of rotation from ship’s body-fixed \( x \)-axis either of semi-major ellipse’s axis (if \( s_{x,j} > s_{y,j} \)) or semi-minor axis (if \( s_{x,j} < s_{y,j} \)).

Each of the determined \( j \) ellipses can be further enlarged to established confidence level (0.99999) by multiplying \( d_{j,a} \) and \( d_{j,b} \) by coverage factor \( k \) (5), in analogy to (7).

Knowing the parameters (19) – (21) of uncertainty ellipses centred in \( j \) points of ship’s contour the next step is to find the extreme outer points of these ellipses in order to construct the MVPA. In order to do this, the maximum vertical values for a generalised (rotated) ellipse in the Cartesian reference frame fixed to the \( j \)th segment of ship’s contour (\( x_{j} - axis between consecutive \( j \) and \( j+1 \) points), i.e., the upper bounding line or tangent for such ellipse has to be calculated (Fig. 2).

\[
\Phi_{j} = \frac{\pi}{2} - \frac{1}{2} \left( \tan 2 \left[ \frac{d_{j,a} - d_{j,b}}{d_{j,a} + d_{j,b}} \right] + \psi \right)
\]

(21)

Fig. 2. Construction of two tangent points (red colour) to the \( j \)th error ellipse.

The algorithm is as follows:

1) The angle \( \beta_{j} \) of the line leading through \( j \) and \( j+1 \) points in case of \( j \) achieved) counted clockwise from \( x \)-axis in body-fixed reference frame is determined according to the formula:

\[
\beta_{j} = \frac{\pi}{2} - \frac{1}{2} \left( \tan 2 \left[ s_{x,j} - s_{y,j} \right] \right)
\]

(22)

2) Tangent points of ellipses with lines of slope \( \beta_{j} \) are determined according to the formulas:

\[
\Phi_{j,c} = \frac{\pi}{2} - \Phi_{j} + \beta_{j}
\]

(22)

\[
R_{j} = \begin{bmatrix} \cos \beta_{j} & -\sin \beta_{j} \\ \sin \beta_{j} & \cos \beta_{j} \end{bmatrix}
\]

(23)

\[
A_{j,3} = d_{j,b}^{2} \sin^{2} \Phi_{j,c} + d_{j,a}^{2} \cos^{2} \Phi_{j,c}
\]

(24)
\[ A_{j,k} = d_{j,k}^2 \cos^2 \Phi_{j,k} + d_{j,k}^2 \sin^2 \Phi_{j,k} \]  
(25)

\[ I_{j,3} = \frac{-A_{j,3} d_{j,0}^2 d_{j,1}^2}{\cos^2 \Phi_{j,3} \sin^2 \Phi_{j,3} (d_{j,0}^2 - d_{j,1}^2)^2 - A_{j,3} A_{j,2}} \]  
(26)

\[ T_{j,3} = \begin{bmatrix} I_{j,3} & -I_{j,3} & 0 \\ -I_{j,3} & I_{j,3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \]  
(27)

\[ T_{j,3} = T_{j,2} R \]  
(28)

\[ \begin{bmatrix} y_{j,p} \\ x_{j,p} \\ y_{j,m} \\ x_{j,m} \end{bmatrix} = \begin{bmatrix} T_{j,3,1} + y_j & T_{j,3,2} + x_j \\ T_{j,3,1} + y_j & T_{j,3,2} + x_j \end{bmatrix} \]  
(29)

where:
- \( \Phi_{j,c} \) is the counter-clockwise angle of the \( j \)th ellipse rotation to the \( x \)-axis in standard Cartesian \( 0xy \) reference frame.
- \( x_{j,p}, y_{j,p}, x_{j,m}, y_{j,m} \) are the coordinates of consecutive \( j \) tangent points in body-fixed reference frame, the extreme outer points are either \( x_{j,p}, y_{j,p} \) if \( \beta_j > 0 \) or \( x_{j,m}, y_{j,m} \) if \( \beta_j \leq 0 \).

The MVPA is constructed by linear connection of the resultant tangent points. This way the bounding spline representing the furthest points of ellipses in respect to the ship’s hull is found. In order to achieve acceptable level of ellipses’ areas coverage by the MVPA (or to minimize the linear spline approximation error) the number of tangent points can be increased adding extra tangent lines of slope angles in the range between \( \beta_j \) and \( \beta_j + 1 \).

In the Fig. 3 the resultant MVPA is presented in red color, built from 3 tangent points for each uncertainty ellipse in blue color, for input parameters: \( \psi = 45^\circ, s_E = 1 \text{m}, s_N = 2 \text{m}, s_{EN} = 0 \text{m}^2, s_\psi = 2^\circ, k = 1 \). It can be noticed that, in a body-fixed reference frame, the ellipses are rotated according to heading. Specifically, the ellipses’ semi-major axes are approximately oriented towards North, i.e. the direction where the position error has the largest variance \( (s_N) \). The semi-major axes are bigger in the aft part due to heading error propagation, as the GPS antenna is assumed to be in the vessel’s fore part.

Assuming non zero covariance of \( s_{NE} \), \( s_\psi \) the MVPA will change gradually as presented in the Fig. 4 and 5. The resultant MVPA in the Fig. 4 is built from 3 tangent points for each uncertainty ellipse, for input parameters: \( \psi = 45^\circ, s_E = 1 \text{m}, s_N = 2 \text{m}, s_{EN} = 1 \text{m}^2, s_\psi = 2^\circ, k = 1 \). The ellipses have changed their dimensions and rotation in comparison to Fig. 3 due to East-North (EN) positive covariance.

In the Fig. 5 the MVPA is built for input parameters of: \( \psi = 45^\circ, s_E = 1 \text{m}, s_N = 2 \text{m}, s_{EN} = 2 \text{m}^2, s_\psi = 2^\circ, k = 1 \). This means perfect positive correlation of position estimation errors in the East and the North directions.

Fig. 3. Example of MVPA around ship heading to 45° in body-fixed reference metric frame and antenna position in fore part with zero EN coordinates’ covariance.

Fig. 4. Example of MVPA around ship heading to 45° in body-fixed reference metric frame and antenna position in fore part with positive non-zero EN coordinates covariance.
IV. FMBS EXPERIMENTS

The objective of the Full Mission Bridge Simulation (FMBS) including EGNOS functionality is to show the influence of EGNOS system on marine navigation performance and safety during inland and pilot manoeuvring. The results of FMBS should lead to identification of possible benefits of SBAS integrity data presentation within the ship’s bridge environment.

The FMBS includes high-fidelity hydrodynamic and visual models of vessels and navigation areas (Fig. 8). It can be used to perform research activities in a realistic environment [6]. In particular, high risk scenarios can be simulated, where pros and cons of additional data presented to navigator as MVPA can be evaluated. The main assumption of the proposed research is the use of simulated GPS/EGNOS data and its presentation in ECDIS type navigation display. In the worst case scenarios, this input data can be limited to the covariance matrix of observations, $C$ in (18). The time and the position related distributions of parameters in $C$, and parameters of instantaneous position and heading errors will be modelled based on data gathered during real experiments on board of the Maritime University of Szczecin (MUS) research-training vessel, “Nawigator XXI.” The trials will be carried out for various environmental conditions and satellite segment configurations. As a reference, the positions’ measurements available from suitable Real Time Kinematic/Precise Point Positioning (RTK/PPP) receivers will be used.

In summary: the pseudo-random number generators based on the data gathered during real experiments in marine environment will provide:

1) instantaneous values of position and heading variances ($d_E^2$, $d_N^2$, $d_\psi^2$) applied to error/bias free position and heading derived from FMBS system before input to ECDIS; and

2) estimates of position variances and their covariance (as EGNOS simulated data) and gyro-specific estimate of heading variance ($s_E^2$, $s_N^2$, $s_{\text{EN}}$, $s_{\psi}$) input to the ECDIS as MVPA parameters.

The operator will have the possibility of influencing these instantaneous variances and EGNOS simulated estimates by adding extra systematic errors (e.g. to simulate random walk of EGNOS augmented GPS position due to satellites’ geometry changes and multipath). Also, a constant increase of position errors is planned to be applied during each scenario. Specifically, the increase will be added until the ship is either unable to navigate or grounds as a result of critically high position errors. Furthermore, intentional standard positioning errors (bias) applied to port and starboard will enable worst case scenarios testing. This way, the assumed value of the $k$ factor can be analysed as a function of the maximum, acceptable, probability that the MVPA does not include the whole ship’s silhouette associated with the actual position of the vessel. This maximum, acceptable probability will be identified for several types of waterways and ships by analysis of the MVPA dimensions that trigger alert limit in case of safety isobaths or other safety line crossings.

The basic rule is that the simulation trials will be conducted twice: with and without presentation of information coming from EGNOS in ECDIS display (Fig. 6 & 7).

![Fig. 5. Example of MVPA around ship heading to 45° in body-fixed reference metric frame and antenna position in fore part with 100% positive EN coordinates correlation.](image)

![Fig. 6. Figurative ECDIS display with EGNOS integrity information (MVPA, ship’s waterplane and red / green alert mark).](image)

![Fig. 7. Figurative ECDIS display without EGNOS integrity information.](image)

The dashed line, grey filled, ship’s silhouette in Fig. 6 & 7 is the biased position (by $d_E$, $d_N$, $d_\psi$) of the vessel as derived from GPS augmented by EGNOS. The solid line, colour filled, areas around grey filled ship’s silhouette in Fig. 6 are MVPA areas of different $s_E$, $s_N$, $s_{\text{EN}}$, $s_{\psi}$, $k$ factors. The black ship’s silhouette is the real (unbiased) marine vessel position. It will not be presented to navigator in FMBS studies, but in the figures it gives indication of situation awareness benefit coming from the EGNOS integrity information. The red colour of the round alert mark in Fig. 6 means crossing of the safety line or another navigation object by MVPA. The green
V. CONCLUSIONS

The mathematical model of Marine Vessel Protection Area (MVPA) based on EGNOS provided integrity data is an innovative solution of marine navigation safety evaluation. The output from this model can be displayed graphically on the desired route within the Electronic Chart Display and Information System (ECDIS). Such an ECDIS layer constitutes an EGNOS integrity display within bridge environment showing graphical representation of the ship and its MVPA together with alert triggering lines.

The proposed simulation approach will allow for identification of this solution benefits from the user perspective in various marine operations and nautical tasks.

The further work will lead to the proposal of meaningful requirements for a hybrid PNT system integrated in ECDIS (position & heading integration, etc.) and consolidation of requirements of such PNT system with special account to its integrity in specific marine operations. It should also justify a future integration of SBAS in e-Navigation.

REFERENCES

Application of faded harmonic subcarrier modulations in GNSS Modernization

1*Wei Liu, 2Yuan Hu, 1Shengzheng Wang, 1ChaoJian Shi, 1BaoJia Xiao

1Merchant Marine College
Shanghai Maritime University
Shanghai, China
*Email: liuwei@navisagere.com

2College of Engineering Science and Technology
Shanghai Ocean University
Shanghai, China

Abstract—New signal modulation design for next-generation of GNSS must offer improved performance and the opportunity for spectrum compatible with existing and planned signals. This paper presents a class of particularly attractive modulations called Faded Harmonics subcarrier (FHS) modulations, and demonstrates their essential characteristics and performance for GNSS. Firstly, the principle of FHS modulation is first described. Then, the general mathematical model and the power spectral density of FHS modulations are derived. The FHS modulations present important characteristics for GNSS and provide superior performance with regards to specific secondary lobe suppression when compared to other BOC modulations. Finally, examples are presented to illustrate the construction and performance of Faded Harmonics subcarrier modulations for GNSS including spectra and correlation characteristics, code tracking, multipath and spectral separation performance. FHS modulations can provide potential opportunities for GNSS modernization and construction.

Keywords—GNSS, Signal design, GNSS Modernization

I. INTRODUCTION

As quickly growing of global navigation satellite system community, there is increasing interest in new signals for military and civilian uses. New signals must provide more robustness, higher performance, and greater capacity than already-existing signals and compliant with the radio frequency compatibility rule with existing and planned signals.

Faded Harmonics subcarrier (FHS) is an interplex modulation that uses non-binary signals to better suppress the specific secondary lobe. The FHS modulations were shown to offer opportunities for higher spectral suppression by the first work of A.R Pratt [1] in 2003. To reduce the power spectrum around the GPS M-code and ensure higher compatibility with the rest of GPS signals, a solution received the name of 8-PSK BOC(2,2) or BOCd(2,2) was proposed [1]. However, the BOCd(2,2) signal divides the code chip waveform in equal length segments. This could make the implementation more difficult and limit the degree of freedom in the signal design. In addition, a four-level FHS modulation was proposed by Dr. José Ángel Ávila Rodríguez [2]. Indeed, we can design the navigation signals with any level and arbitrary length of code chip waveform. As we can imagine, this will open new possibilities to extend the concept to a new modulation family, where each of segments can be adopted any arbitrary length and any level.

This paper presents the extension of the Faded Harmonics subcarrier modulation in GNSS signal design. The principle of Faded Harmonics subcarrier modulations is firstly described. Then, we create a reference framework by dividing the code chip in k subchips and introduce the fading parameters indicated any arbitrary length. The general mathematical model of Faded Harmonics modulations is proposed. Examples are presented to illustrate the construction and performance of Faded Harmonics subcarrier modulations for GNSS including spectra and correlation characteristics, code tracking, multipath and spectral separation performance. The draw backs of the FHS modulations will be discussed. The Faded Harmonics Subcarrier modulations can provide capacity for compatible with other signals while offering advantages for spectral power suppression, and the opportunity for flexible receiver design.

II. FADED HARMONICS SUBCARRIER MODULATION

A. Signal model

According to the Fourier series theory [3], any BOC subcarrier modulation is indeed a periodic signal and can be expressed as a Fourier series with only odd coefficients [2]:

\[ s(t) = \sum_{k=-\infty}^{\infty} \left[ a_k \cos(2\pi kf_s t) + b_k \sin(2\pi kf_s t) \right] \]  \hspace{1cm} (1)

where \( b_k \) is the odd coefficients of the sine terms, and all the cosine terms are zero.

Figure 1 shows one example of Fourier series. Fig.1(a) is a sine function with frequency \( f_s \) in Time-domain. This sine function and the other three sine functions together form a new Time-domain function. As shown in Fig.1(c) and Fig.1(d), this new function only has \( 1^{\text{st}}, 3^{\text{rd}}, 5^{\text{th}}, \) and \( 7^{\text{th}} \) harmonics.
As we can imagine, if we want to suppress the $k^{th}$ harmonic, we can just subtract the BOC signal above a subcarrier frequency and amplitude equal to that of the harmonic which we want to suppress. However, the procedure will be more complex when many harmonics need to be suppressed. Indeed, similar to the construction of Fig.1(c), we can create a five-level waveform and introduce the fading parameters for suppression any harmonics in the signal design.

According to [4], a direct sequence spread spectrum (DSSS) signal, $s(t)$ can be represented as follows

$$s(t) = \sum_{k=-\infty}^{\infty} a_k q(t-kT_c)$$  \hspace{1cm} (2)

where $\{a_k\}$ represents the spreading sequence, $q(t)$ is the spreading symbol, and $T_c$ is the spreading code period.

For the Binary Coded Symbol (BCS) modulations, the spreading symbol is divided into $K$ segments, each of equal length $T_c/K$. Then the spreading symbol is given by [5]

$$q(t) = \sum_{k=0}^{K-1}s_k p_{T_c/K}(t-kT_c/K)$$  \hspace{1cm} (3)

where $s_k$ is the sequence with value +1 or -1, $p_{T_c/K}(t)$ represents the shape of each five-level waveform.

Based on the five-level waveform, we can create a reference framework by dividing the code chip in $k$ subchips and introduce the fading parameters indicated any arbitrary length. Fig. 2 shows the five-level waveform to realize the fading effect, where Fig. 2 (a) and Fig. 2 (b) can suppress the 3$^{rd}$ and 5$^{th}$ harmonics, respectively. $\rho_1$, $\rho_2$, $\rho_3$ are the fading parameters. $\cos(\varphi)$ is the amplitude parameter.

According to the five-level waveform, a modified version of the BCS modulations, namely the Faded Harmonics Subcarrier modulation (FHS), denoted by $FHS(m,l,\Theta,\varphi)$, is defined, where $\Theta = \{\rho_1, \rho_2, \cdots, \rho_N\}$.

### B. Spectra and correlation characteristics

Time domain representation of the Faded Harmonics Subcarrier modulation is shown in Fig. 2. According to the Fourier Transform theory, the frequency spectrum of the FHS can be written as

$$Q^{FHS}(j\omega) = P_{\varphi}(j\omega)e^{j\omega f_2k} \sum_{k=1}^{K} s_k e^{-\frac{j\omega f_2k}{K}}$$  \hspace{1cm} (4)

where $P_{\varphi}(j\omega)$ is the frequency spectrum of the five-level waveform.

According to the expression (4), the general expression of the power spectral density (PSD) for the $FHS(m,l,\Theta,\varphi)$ can be expressed as

$$G^{FHS}(f) = \frac{1}{\rho^2} \left[ P_{\varphi}(f) \right] \sum_{k=1}^{K} s_k e^{-\frac{j2\pi kf}{K}}$$  \hspace{1cm} (5)

where

$$P_{\varphi}(f) = \begin{cases} \sum(-1)^i \sin \left( \frac{2\pi f \rho_i}{2Kf} \right) \cos(\varphi) \sum(-1)^i \sin \left( \frac{2\pi f \rho_i}{2Kf} \right), & N \text{ even} \\ \sum(-1)^i \sin \left( \frac{2\pi f \rho_i}{2Kf} \right) + \cos(\varphi) \sum(-1)^i \sin \left( \frac{2\pi f \rho_i}{2Kf} \right), & N \text{ odd} \end{cases}$$  \hspace{1cm} (6)

$$\rho' = \begin{cases} \sum(-1)^{i+1} \rho_i + \cos^2(\varphi) \sum(-1)^{i+1} \rho_i, & N \text{ even} \\ \sum(-1)^{i+1} \rho_i + \cos^2(\varphi) \sum(-1)^{i+1} \rho_i, & N \text{ odd} \end{cases}$$  \hspace{1cm} (7)

For comparison, the PSDs and autocorrelation functions of $FHS(1,1,\Theta,\varphi)$ and BOC(1,1) examples are shown in Fig. 3.
and Fig. 4, respectively, where \( \cos(\varphi) = \sqrt{2}/2 \) and \( \Theta = [1/3,1] \).

Fig. 3 shows that the PSD shapes of FHS(1,1,\( \Theta, \varphi \)) and BOC(1,1) are similar. In addition, FHS(1,1,\( \Theta, \varphi \)) modulation can suppress 3\(^{rd}\) harmonic, even 9\(^{th}\) and 15\(^{th}\) harmonics. Fig. 4 depicts the shape of the autocorrelation functions of FHS(1,1,\( \Theta, \varphi \)) and BOC(1,1). As shown, the existing tracking algorithm can be easy used in FHS signal tracking process without modification.

III. PERFORMANCE

The Cramér-Rao lower bound (CRLB) is usually employed to assess the performance of the code tracking errors estimation [5]. The Cramér-Rao lower bounds of the FHS(1,1,\( \Theta, \varphi \)) and BOC(1,1) examples using 1Hz code tracking loop bandwidth and 24 MHz two-sided receiver bandwidth are shown in Fig. 5. As shown, the BOC(1,1) signal has better code tracking performance than FHS(1,1,\( \Theta, \varphi \)). However, Faded Harmonics Subcarrier modulations can suppress any unnecessary harmonics.

Fig. 5. Code tracking errors

Besides the code tracking accuracy, code multipath performance is another argument in the signal design. In this paper, the performance of FHS(1,1,\( \Theta, \varphi \)) and BOC(1,1) examples in terms of multipath using the multipath error envelopes are evaluated. In order to obtain an estimate for the multipath error we will take the distributions of path delays and relative amplitudes into account. Fig. 6 and Fig. 7 show results for the two examples using an early minus late code discriminator spacing of 24.4 ns and a two sided receiver front-end bandwidth of 24 MHz. Assume the multipath signal is -10 dB weaker than the direct signal. Note that the Faded Harmonics Subcarrier signals perform higher multipath errors than BOC(1,1) signal. The reduction of the power of high frequency components on Faded Harmonics Subcarrier signal implied significant deterioration of the multipath performance.

Fig. 6. Multipath error envelopes
In order to make open signals interoperable and maximize benefit to all GNSS users, all GNSS signals must be compatible firstly. A parameter called spectral separation coefficient (SSC) [6] was introduced to distinguish the effects of the interference spectral shape from the effects due to the interfering power. Several SSC results are shown in Table I. These calculations assume long spreading codes, normalized (unit area over infinite bandwidth) power spectrum of each signals and two sided receiver front-end bandwidth using 24 MHz. As shown in Table I, the FHS examples present similar spectral separation performance with BOC(1,1) signal in the L1 band.

<table>
<thead>
<tr>
<th>Signal</th>
<th>BOC(1,1)</th>
<th>FHS(1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS C/A code</td>
<td>-67.88</td>
<td>-67.88</td>
</tr>
<tr>
<td>GPS P(Y) code</td>
<td>-70.56</td>
<td>-70.48</td>
</tr>
<tr>
<td>GPS M code</td>
<td>-83.16</td>
<td>-83.85</td>
</tr>
<tr>
<td>GPS L1C</td>
<td>-65.28</td>
<td>-64.70</td>
</tr>
<tr>
<td>Galileo E1 OS</td>
<td>-65.28</td>
<td>-64.70</td>
</tr>
<tr>
<td>Galileo E1 PRS</td>
<td>-92.78</td>
<td>-105.14</td>
</tr>
</tbody>
</table>

IV. APPLICATION DISCUSSION

According to [7], China is now proposing to move its signal modulation almost entirely into the binary offset carrier (BOC) family. The BeiDou signals at L1 will place a MBOC(6,1,1/11) open service (OS) and a BOC(14,2) authorized service (AS). Galileo is a global navigation satellite system currently being built by the European Union (EU) and European Space Agency (ESA) [8]. According to its signal plan, the whole transmitted Galileo E1 signal consists of E1 OS signal and cosine-phased BOCos(15, 2.5) modulation for its Public Regulated Service (Galileo PRS signal). The spectrum of BeiDou AS signal and Galileo PRS signal will be severely overlapped. Note that Faded Harmonics Subcarrier modulations can suppress any unnecessary harmonics and have good performance similar with BOC modulations. Therefore, Faded Harmonics Subcarrier modulation can provide potential opportunities for solving the BeiDou/Galileo AS frequency overlay issue application.

As mentioned, the modulation parameters play a key role in suppressing the spectral side-lobe of BOC modulation which we want to attenuate it. Note that the amplitude parameter \(\cos(\varphi)\) can control the power of the side-lobe. Table II summarizes the example of harmonics suppression for FHS \((n, n, \Theta, \varphi)\) modulation via different modulation parameters, where chip rate \(f_c\) is equal to subcarrier frequency \(f_s, f_{ref} = 1.023\) Mcps is the reference frequency.

<table>
<thead>
<tr>
<th>Harmonics suppression</th>
<th>Fading parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Theta = [1/3, 1])</td>
<td>(\Theta = [1/5, 3/5, 1])</td>
</tr>
<tr>
<td>(f_c = f_{ref})</td>
<td>3, 9, 15, 21, …</td>
</tr>
<tr>
<td>(f_c = 2f_{ref})</td>
<td>6, 18, 30, 42, …</td>
</tr>
<tr>
<td>(f_c = 3f_{ref})</td>
<td>9, 27, 45, 63, …</td>
</tr>
<tr>
<td>(f_c = 4f_{ref})</td>
<td>12, 36, 60, 84, …</td>
</tr>
</tbody>
</table>

As we can recognize, Faded Harmonics Subcarrier modulations are construct based on five-level waveform. The multiplex scheme will become more complicated in the signal design. In addition, the relative power of the Inter-Modulation product will be increased. Therefore, the spreading symbols and fading parameters of Faded Harmonics Subcarrier modulation must be chosen carefully in the system construction.

V. CONCLUSIONS

This paper presents a class of particularly attractive modulations called Faded Harmonics subcarrier (FHS) modulations, and demonstrates their essential characteristics and performance for GNSS. The general mathematical model of Faded Harmonics modulations was proposed. Examples were presented to illustrate the construction and performance of Faded Harmonics subcarrier modulations for GNSS including spectra and correlation characteristics, code tracking, multipath and spectral separation performance. The Faded Harmonics subcarrier modulations can provide capacity for
compatible with other signals while offering advantages for spectral power suppression, and the opportunity for flexible receiver design.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No.61304230, 51279099, 51379121), the Specialized Research Fund for the Doctoral Program of Higher Education of China (No.20123121120003), the Innovation Program of Shanghai Municipal Education Commission (No.14YZ106, 14YZ121), the Shanghai Municipal Natural Science Foundation (No.12ZR1444700), the development of science and technology of Shanghai Ocean University(A2-0209-14-200063) and the Shanghai Science and Technology Committee (No.12510501800, 13510501600).

REFERENCES

Phase-Phase and Phase-Code Methods Modification for Precise Detecting and Predicting the GPS Cycle Slip Error

A. A. Elashiry, Mohamed A. Yousef, A. M. Abdel hamid

Abstract
There are three famous detecting methods for cycle slip error, which are; Doppler measurement method, Phase - Code differencing method, and Phase - Phase differencing Method. The first method depends on the comparison between observables and the fact that; Doppler measurements are immune to cycle slip error. This method is considered as the most precise method for cycle slip detecting, because it succeed in detecting and predicting the smallest cycle slip size (1 cycle). That in case of the local oscillator has low bias. The second method depends on the comparison between observables (phase and code) and the code measurements are immune to the cycle slip error. But this method can’t detect or predict cycle slip size smaller than 10 cycles, because the code measurements have high bias effect. The third method depends on the comparison between observables (phase 1 and phase 2) and the phases measurements have low noise effect. But this method can’t detect or predict cycle slip size smaller than 5 cycles, because the ionospheric change might be has high variation. For enhancing the precision of the last two methods in detecting the smallest cycle slip size reaches 1 cycle, a new algorithm was deduced in this research to determine the change in the ionospheric values and the code bias from epoch to epoch; that might be done by solving all observables equations by least square technique. This modification on these methods succeed in detecting and predicting cycle slip size 1 cycle.

Keywords: GPS, Cycle slip, Phase-Code differencing, Phase-Phase differencing, Doppler measurements.

Introduction
Global Positioning System (GPS) is a satellite based navigation and surveying system for determining the precise position and time, using radio signals from the satellites, in real-time or post-processing mode (Ren Z., et al. 2011). It consists of a constellation of 32 satellites in six different orbits which give the information of the position of the GPS receiver user. If there are four or more GPS satellites in unobstructed line of sight with the receiver, the precise spatial co-ordinates can be obtained (Dawod, G.M. 1991). Each GPS satellite transmits two carrier phase: L1 = 1575.42 MHz; and L2 = 1227.60 MHz modulated with two types of codes and a navigation message. The L1 signal is modulated with a precise (P) code, known also as the Precise Positioning Service (PPS), and a coarse acquisition (C/A) code, which is known also as the Standard Positioning Service (SPS); the L2 signal is modulated with P code only (Dawod, G.M. 1991 and Raju, P.L.N 2003). This means the GPS carrier phase observations quality play an important role in high precision GPS static or kinematic positioning. However, due to internal tracking problems of GPS receiver or signal interruption of the antenna from the satellite, the continuous original carrier phase observations are destroyed, generating cycle slips and gross errors (Wu, Y. et. al, 2010). The main causes of the cycle-slip are listed below (Kim, D 2002):

1. Obstructions of the satellite signal due to trees, buildings, bridges, mountains, etc.
2. Low signal-to-noise ratio (SNR) or alternatively carrier-to-noise-power-density ratio (C/N0) due to bad ionospheric conditions, multipath, high receiver dynamics, or low satellite elevation angle.
3. Failure in the receiver software which leads to incorrect signal processing.

The cycle slip is a sudden jump in the phase observations (Karaim, M. O., et. al, 2014) as shown in Fig. 1, and it may be as small as one or a few cycles, or contain millions of cycles, which directly affect on the GPS positioning precision. Therefore, the precise detecting and predicting for the gross errors and cycle slips is an important pre-processing step in high precision GPS carrier phase positioning and applications (Wu, Y. et. al, 2010 and Seeber, G., 2003).

A. A. Elashiry
Assistant Lecturer in the Department of Civil Engineering, Faculty of Engineering, Beni-Suef University, Beni-Suef, Egypt.
e-mail: eng_ahmedashiry@yahoo.com

Mohamed A. Yousef
Associate Professor in the Department of Mining and Metallurgical Engineering, Faculty of Engineering, Assiut University, Egypt.

A. M. Abdel hamid
Lecturer in the Department of Civil Engineering, Faculty of Engineering, Beni-Suef University, Beni-Suef, Egypt.
Many cycle-slip detection methods have been proposed since the early 1980s. Goad C at 1985; established the first cycle slip predicting method based on ionospheric residual (PIR) (Seeber, G. 2003). Bastos et. al at 1988; studied Kalman filtering technique for filtering observables data from cycle slip to improve the positioning accuracy (Wu, Y. et. al, 2010). Kleusberg et al. at 1993; established a new approach for cycle slip predicting based on high-order differencing concept (Seeber, G. 2003). Recently, several new methods have been proposed for cycle slip detection based on high-order differencing concept, where Dai, Z., at 2012; studied all GPS measurements (phases and codes) and the estimated doppler shift in the RINEX observations file. From his study, many detected methods were established which are; Phase - Code differencing method, Phase - Phase differencing Method and Doppler measurement method. This study concluded that the Doppler measurement method is the most precise method (Dai, Z., 2012). Cai C. at 2013; developed a new approach for cycle slip detecting and repairing under high ionospheric activity using undifferenced dual-frequency GPS carrier phase observations with new algorithms. These algorithms integration was allowing uniquely detecting and determining for the cycle slips (≥5 cycles); even under high ionospheric activities (Cai, C., et. al. 2013). Banville S. et. al at 2012; studied the decoupled-clock model to improve cycle-slip correction capabilities and use it in single frequency point precise positioning (PPP) to improve the positioning results (Banville, S., et al. 2012). Ren Z. et. al at 2011; implemented Doppler-aided cycle slip detecting and repairing method, using a simplified oscillator model with some modification to avoid the influence of the local oscillator bias. This method with its modification lead to high precision in predicting and repairing the cycle slips at real time kinematics positioning method for size equal 1 cycle (Ren Z., et. al. 2011). Liu Z. at 2010; developed a new automated cycle slip detection and repair method that is based on only one single dual-frequency GPS receiver. This method used the ionospheric total electron contents (TEC) rate (TECR) and Melbourne–Wübbena wide lane (MWWL) linear combination to uniquely determine any cycle slip even under very high level of ionospheric activities and on both L1 and L2 frequencies (Liu, Z. 2010). Wu Y et. al at 2010; established a new approach depended on using three groups of uncorrelated dual-frequency observation. This approach helped on repairing and detecting various real-time cycle slips and gross error under the long sampling condition (Wu, Y. et. al, 2010).

1. Observations Data

The observations data which processed in this research was taken from master thesis work for a demonstrator in Mining and Metallurgy Department in Assiut University. The collected data was for two fixed points called (C and R) at two different places in Assiut city in Egypt Fig. 2; they were observed by GPS receiver (ASHTECH A-12). The observation conditions of the collected data were mask angle = 12°, where it is the optimum value of elevation mask angle (Yousef, M. A. 2004), epoch interval= 1 sec, where it is the optimum interval (Yousef, M. A. et. al. 2014) and occupation period >1 hr. The coordinates of this point are shown in Table1. 

<table>
<thead>
<tr>
<th>ECEF frame (X, Y and Z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point C</td>
</tr>
<tr>
<td>Point R</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geodetic frame (WGS84) (Lat, Long and h_BGN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point C</td>
</tr>
<tr>
<td>Point R</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Old Egypt 1906 frame (E, N and h_BGN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point C</td>
</tr>
<tr>
<td>Point R</td>
</tr>
</tbody>
</table>

Fig. 2: The studied area

2. Single Difference Process

Firstly, cycle slip count and place must be determined in taking sample in our observations to evaluate the performance of the studied methods in detecting it. The theoretical solution is to determine the ambiguity value, but this task isn’t trivial. So the other way depends on removing this unknown ambiguity value by using the Single Difference (SD) between single receiver and single satellite at two constitute epochs Fig. 3, where this technique removes the carrier phase ambiguity term, because this ambiguity term is constant for the entire observation session (Bossler, J. D., et. al. 2002). Also the SD technique reduces the ephemeris, ionospheric, and...
tropospheric errors, if epoch interval is short (less than 5 sec) (Dai, Z. 2012).

\[ \text{Fig. 3: Single Difference technique (one receiver observing one satellite at two epochs)} \]

At Single Difference technique the given data are the phase observations from one satellite (s) to one receiver (r) at two different epochs (t1 and t2) (Ogaja, C. A. 2011). The difference between two consecutive epochs is as follows:

- Phase code equation at epoch t1:
  \[ \Delta \phi_{t1} = \phi_{t1} - \phi_{t1}^{\Delta} + \frac{c}{c_t} \Delta T + \Delta I + \Delta ERT \quad (1) \]

- Phase code equation at epoch t2:
  \[ \Delta \phi_{t2} = \phi_{t2} - \phi_{t2}^{\Delta} + \frac{c}{c_t} \Delta T + \Delta I + \Delta ERT \quad (2) \]

- The difference between (1) and (2) is as follows:
  \[ \Delta \phi_{t2} - \Delta \phi_{t1} = \Delta \phi - (\Delta(c * \delta t) + \Delta T + \Delta I + \Delta ERT) \quad (3) \]

where \( \phi_{t2}^{\Delta} \) is the measured phase of L1 (L1 or L2), cycle; \( r \): is the true geometric range, m; \( c \): is the speed of light, m/s; \( \delta t \): is the receiver clock error, sec; \( \delta t_{\text{sat}} \): is the satellite clock error, sec; T: is the tropospheric delay, m; I: is the ionospheric delay, m; \( \epsilon \): is the earth rotation error, m; \( \Delta \): is the single difference operator, and \( M \text{path} \) is the multipath error, m. But, in this research, the observing must be done in an open area, so ‘Mpath’ error is approximately zero.

The previous equation (3) is a condition equation to identify if the observations have cycle slip error or not; and to test the precision of the proposed detected method in this research.

But firstly the range of each error change between two constitutive epochs must be known, which are \( \Delta t, \Delta T, \Delta I, \) and \( \Delta ERT \); that by estimating the values of these errors. The equations of these errors are listed below:

- The satellite clock error (\( t_s \)) can be estimated from the following equation (Ashby, N. 2003):
  \[ t_s = \alpha f_0 + \alpha f_1(T_T - T_0) + \alpha f_2(T_T - T_0)^2 + \Delta \text{rel} \quad (4) \]
  \[ \Delta \text{rel} = e^\sqrt{h} \text{sin} E_0 \quad (5) \]
  where \( \alpha f_0, \alpha f_1, \alpha f_2 \): coefficients are available in the navigation message file, \( T_T \): transmission time, \( \text{Ephemeris reference time} \), \( \Delta \text{rel} \): relativistic correction, e: Eccentricity, and a: orbital semi-major axis.

- The tropospheric error (T) can be predicted using Hopfield model (Chaib, C., et. al. 2007):
  \[ T = \frac{Kd}{\sin(\delta^2 + 1.94410^{-3})^2} + \frac{Kw}{4(\sin(\delta) + 2 \times 10^{-3})^2} \quad (6) \]
  \[ Kd = \frac{1.55208 \times 10^{-4} \times \text{Pamb} \times 8307.2 \times \frac{\text{Tamb} + 273.16}{\text{Pamb} + \frac{\text{Tamb} + 273.16}{\text{Tamb} + 273.16}}}{\text{Tamb} + 273.16} \quad (7) \]

where Tamb: ambient air temperature, C; Pamb: ambient air pressure, kpa; Pamb: ambient vapor pressure, kpa; El = the satellite vehicle's elevation in radian; and (Kd and Kw): are the dry air wet components respectively.

- The ionospheric error (I) can be predicted using the approximately form of Klobuchar's model (Chaib, C., et. al. 2007 and Klobuchar, J. A. 1987):
  \[ I = t_{\text{ion}} = \left[ 1 + 16 \times (0.53 - E_0)^3 \times 5 \times 10^{-3} \right] \quad (9) \]

- The earth rotation error (ERT) can be estimated from the following equation (Gustavsson, P. 2005):
  \[ \text{ERT} = \frac{\omega E}{c} (X_{\text{sat}} - \text{Y}_{\text{sat}}) \quad (10) \]

where \( \omega E \): The earth turn rate = 7.2921151467 x 10^-5, rad/sec; \( X_{\text{sat}} \) and \( Y_{\text{sat}} \): are the satellite horizontal coordinates, m; \( X_{\text{rec}} \) and \( Y_{\text{rec}} \): are the receiver horizontal coordinates, m. By applying the previous equations on our observations of satellite PRN 6, as an example, at “Point C and Point R”, taking sample size (360 epochs) with observation epoch interval 1 second, the values of \( t_s, T, I, \) and \( \text{ERT} \) can be determined. Then the values of change of each error per 1 sec (error difference), the average value and standard deviation value could be estimated; its values shown in Table 2. The relations of these difference values for each error with time and its histogram are shown in Fig. 4.

| Table 2: The average value (\( \mu \)) and standard deviation (\( \sigma \)) of each error change “\( \Delta t_s, \Delta T, \Delta I, \) and \( \Delta ERT \)” for 1 Sec. |
|---------------------|-----|-----|
| Point C | Point R |
| \( \Delta t_s, \text{m/sec} \) | \( \mu \) | 0.015 | 0.012 |
| | \( \sigma \) | 0.0035 | 0.005 |
| \( \Delta T, \text{m/sec} \) | \( \mu \) | 0.0012 | 0.0016 |
| | \( \sigma \) | 0.0005 | 0.0003 |
| \( \Delta I, \text{m/sec} \) | \( \mu \) | 0.03 | 0.05 |
| | \( \sigma \) | 0.012 | 0.016 |
| \( \Delta ERT, \text{m/sec} \) | \( \mu \) | 0.0075 | 0.006 |
| | \( \sigma \) | 0.005 | 0.004 |
Phase and Phase

-Code method approximately 1 cycle >100000. To clarify, when a cycle slip found in a certain

H. The cycle slip error, place and value, on our testing observations data can be detected easily by applying SD

ments data can be detected easily by applying SD

value of the right part of (3) can be estimated and it would

error is equal the mean value plus or minus three of the
distribution, so the maximum allowable value
differences from each estimated error have a normal

error is equal the mean value plus or minus three of the

Fig. 4: The relations between each error difference/1sec with time and its histogram “Taking Point C as example”

From Fig. 4, it was found that; the resultant error differences from each estimated error have a normal
distribution, so the maximum allowable value for each
error is equal the mean value plus or minus three of the
standard deviation value (Mark 2009). Then the maximum
value of the right part of (3) can be estimated and it would
equals about 15.02 cm (approximately 1 cycle).

The cycle slip error, place and value, on our testing observations data can be detected easily by applying SD
equation (3). Where the receiver coordinate is known and
the precise geometric distance between the receiver and
satellite can be estimated at any epoch by using the precise
ephemeris navigation data. So if the value of the left part
of (3) is higher than 1 cycle; then the cycle slip error is
found.

The following tables show the detected cycle slip error
using Single Difference technique at five observation
minutes (interval 1second), from the phase observations of
satellite PRN 6, at point C and point R; Tables 3-a and 3-b
respectively.

Table 3-a: The detected count of cycle slip error in five
minutes of observation data (interval 1 second) at point C

<table>
<thead>
<tr>
<th>Epoch</th>
<th>Error status</th>
<th>Cycle Slip Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 42</td>
<td>Yes</td>
<td>&gt;100000</td>
</tr>
<tr>
<td>43 - 46</td>
<td>Yes</td>
<td>1500</td>
</tr>
<tr>
<td>47 - 49</td>
<td>Yes</td>
<td>400</td>
</tr>
<tr>
<td>50 - 80</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>81 - 84</td>
<td>Yes</td>
<td>10</td>
</tr>
<tr>
<td>85 - 117</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>118 - 119</td>
<td>Yes</td>
<td>5</td>
</tr>
<tr>
<td>120 - 122</td>
<td>Yes</td>
<td>3</td>
</tr>
<tr>
<td>123 - 128</td>
<td>Yes</td>
<td>2</td>
</tr>
<tr>
<td>129 - 246</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>247 - 250</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>251 - 300</td>
<td>No</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3-b: The detected count of cycle slip error in five
minutes of observation data (interval 1 second) at point R

<table>
<thead>
<tr>
<th>Epoch</th>
<th>Error status</th>
<th>Cycle Slip Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>47 - 48</td>
<td>Yes</td>
<td>2000</td>
</tr>
<tr>
<td>49 - 52</td>
<td>Yes</td>
<td>250</td>
</tr>
<tr>
<td>53 - 92</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>93 - 94</td>
<td>Yes</td>
<td>10</td>
</tr>
<tr>
<td>95 - 123</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>124 - 127</td>
<td>Yes</td>
<td>5</td>
</tr>
<tr>
<td>128 - 193</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>194 - 196</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>197 - 230</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>231 - 235</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>236 - 300</td>
<td>No</td>
<td>0</td>
</tr>
</tbody>
</table>

3. Detected Methods

There are many methods for cycle slip detecting and
predicting, but this research will focus on the famous
methods, which are Phase-Phase and Phase-Code method
and comparing it with the most precise method, which
named Doppler measurements method. A short study on
the Doppler method and its precision in detecting and
predicting the cycle slip error in our data, will be presented
in the following section.

3.1 Doppler Measurement Method

This technique takes the advantage of the fact that, the
Doppler measurements are immune to cycle slips because
they are computed values, where these Doppler
measurements are equal to the expected change at carrier
phase measurement from two adjacent epochs (Silva, P.
2013). To clarify, when a cycle slip found in a certain
epoch, the Doppler measurement doesn’t equal to the
carrier phase time difference between this epoch and its
previous epoch. The carrier phase variation between two
adjacent epochs can be estimated by Doppler data
according to the following equation (Dai, Z. 2012):

\[ \phi(t) - \phi(t-1) = D(t) + D(t-1) \times dt/2 \]  (11)

where: D= Doppler frequency measurements, MHz; and
dt = the sampling interval, sec.

So if this condition equation is not achieved; that
means there is cycle slip error affects the phase
observations. When this method was applied on our phase observations data it was found that; this method succeed in detecting and predicting smallest cycle slip sizes (1 cycle), as shown in (Table 3). The resultant precision of this method is agreed by many researchers (Dai, Z. 2012 and Ren, Z., et. al. 2012), that in case of; the local oscillator has low bias (Ren, Z., et. al. 2011).

3.2 Phase- Code Differencing Method.

This method depends on the comparison between observables (phase and code), where the code measurements are immune to the cycle slip error. At stand-alone GPS receiver, the observations equations for pseudorange code and carrier phase measurements can be formulated as follows (Dai, Z. 2012):

\[ P_i^m(t) = \rho^m(t) + c[\delta r(t) - \delta_s(t)] + T^m(t) + 1^m(t) + \text{ERt}(t) + \text{bias}_i^m \] (12)

\[ \phi_i^m(t) = \rho^m(t) + c[\delta r(t) - \delta_s(t)] + T^m(t) - 1^m(t) + \text{ERt}(t) + \lambda_i N_i + e_i^m \] (13)

By using the differencing technique between two consecutive epochs, free of cycle slip, at time (ti and tj), to eliminate the most affecting errors on the phase and code measurements, equations (12 and 13) can be written as:

\[ \Delta P_i^m = \Delta \rho^m + c[\Delta \delta r - \Delta \delta_s] + \Delta T^m - \Delta 1^m + \Delta \text{ERt} + \Delta \text{bias}_i^m \] (14)

\[ \Delta \phi_i^m = \Delta \rho^m + c[\Delta \delta r - \Delta \delta_s] + \Delta T^m - \Delta 1^m + \Delta \text{ERt} + \lambda_i \Delta N_i + \Delta e_i^m \] (15)

When subtract (14) from (15) most of the errors are removed but the difference of ionospheric error doubled because its sign in phase equation reverse its sign in code equation. But the change in ionospheric delay between adjacent epochs would be very small so it can be neglected, this assumption is right at small sampling interval (less than 5 sec). Also, the time difference between ambiguities (N) is zero in case of no cycle slip; thus, the equation resulted as be as follow :

\[ \Delta \phi_i^m = \lambda_i \Delta \rho^m \] (16)

If this condition equation is not achieved; that means there is a cycle slip error affects the observations. When this method applied on our phase observations data it was found that; this method succeed in detecting and predicting cycle slip size ≥ 5 cycles (Ren, Z., et. al. 2012), that in case of; the local oscillator has high bias level and the ionospheric change might be has high variation (Ren, Z., et. al. 2012 and Karaim, M., et. al. 2014).

3.3 Phase- Phase Differencing Method.

This method depends on the comparison between observables (phase 1 and phase 2), where the phase measurements have low noise effect. The observation equation for carrier phases measurements (13) at the absolute positioning can be formulated for each phase (L1 and L2) as follows (Dai, Z. 2012):

\[ \phi_i^m(t) = \rho^m(t) + c[\delta r(t) - \delta_s(t)] + T^m(t) - 1^m(t) + \text{ERt}(t) + \lambda_i N_i + e_i^m \] (17)

\[ \phi_i^m(t) = \rho^m(t) + c[\delta r(t) - \delta_s(t)] + T^m(t) - \left( \frac{f_1}{f_2} \right)^2 1^m(t) + \text{ERt}(t) + \lambda_{i_2} N_{i_2} + e_{i_2}^m \] (18)

Where: f1 and f2 are the frequencies of phases L₁ and L₂ respectively, MHz.

By using the differencing technique between two consecutive epochs, free of cycle slip, at time (ti and tj), to eliminate the most affecting errors on the phases measurements, equations (17 and 18) can be written as follows:

\[ \Delta \phi_{i_1}^m = \Delta \rho^m + c[\Delta \delta r - \Delta \delta_s] + \Delta T^m - \Delta 1^m + \Delta \text{ERt} + \lambda_{i_1} \Delta N_{i_1} + \Delta e_{i_1}^m \] (19)

\[ \Delta \phi_{i_2}^m = \Delta \rho^m + c[\Delta \delta r - \Delta \delta_s] + \Delta T^m - \left( \frac{f_1}{f_2} \right)^2 1^m(t) + \Delta \text{ERt} + \lambda_{i_2} \Delta N_{i_2} + \Delta e_{i_2}^m \] (20)

The change in errors at (19) and (20) from epoch to adjacent epoch are very small at small sampling interval, and the time difference between ambiguities (N) is zero in case of no cycle slip; thus, the equation resulted as follow:

\[ \Delta \phi_{i_1}^m - \Delta \phi_{i_2}^m = 0 \] (21)

When this method was applied on our phase observations data using this condition equation, it was found that; this method succeed in detecting and predicting cycle slip size ≥ 5 cycles (Ren, Z., et. al. 2012), as shown in (Table 3). Because the ionospheric change might be has high variation.


This modification tried to increase the precise of the Phase-Code and Phase- Phase methods, which they can’t catch small cycle-slip (Ren, Z., et. al. 2012), by deducing a new algorithm to determine the change in the ionospheric values and code bias from epoch to epoch; that by solving all observables equations by least square technique. The linearized observations equations system can be represented using the matrix is as follow (Xu, G. 2007):

\[ AL + v = X \] (22)

where: X: Matrix of the unknowns; A= Matrix of coefficients; and L: Matrix of the observations.

After subtracting each observable equation at certain epoch from the adjacent epoch, the final forms of all observables equations according to the form of basic equation of Least Square technique (22), are as follows:

\[ \Delta \phi_{i_1}^m + v_1 = \Delta \rho^m - \Delta 1^m \] (23)

\[ \Delta \phi_{i_2}^m + v_2 = \Delta \rho^m - \left( \frac{f_1}{f_2} \right)^2 1^m \] (24)

\[ \Delta P_{i_1}^m + v_3 = \Delta \rho^m + \Delta 1^m + \Delta \text{bias}_{i_1}^m \] (25)

\[ \Delta P_{i_2}^m + v_4 = \Delta \rho^m + \left( \frac{f_1}{f_2} \right)^2 1^m + \Delta \text{bias}_{i_2}^m \] (26)

The related least squares normal equation is as follow [24]:

\[ X = (A^T A)^{-1} A^T L \] (27)

Where the matrices of these equation terms can detailed as follows:

\[
\begin{bmatrix}
\Delta \rho^m \\
\Delta 1^m \\
\Delta \text{bias}_{i_1}^m \\
\Delta \text{bias}_{i_2}^m \\
\end{bmatrix}
= \begin{bmatrix}
\Delta \phi_{i_1}^m \\
\Delta \phi_{i_2}^m \\
\Delta P_{i_1}^m \\
\Delta P_{i_2}^m \\
\end{bmatrix}
\]

48
### Table 4: The possibility of the usage methods for detecting cycle slip error

<table>
<thead>
<tr>
<th>Cycle slip count</th>
<th>Prediction method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&gt;100</td>
</tr>
<tr>
<td>Doppler method</td>
<td>D*</td>
</tr>
<tr>
<td>Phase- Code</td>
<td>D</td>
</tr>
<tr>
<td>Phase- Phase</td>
<td>D</td>
</tr>
<tr>
<td>Modified Phase-Code</td>
<td>D</td>
</tr>
<tr>
<td>Modified Phase-Phase</td>
<td>D</td>
</tr>
</tbody>
</table>

*where: D: means cycle slip error detected.

### Conclusion

The deduced algorithm in this research is succeed in increasing the cycle slip error detecting precision for Phase- Code differencing method, and Phase- Phase differencing Method, where their precision become close to, or equal, the precision of Doppler measurement method, where they can detecting and predicting the smallest cycle slip size (1cycle). Because this deduced algorithm here, they succeed in determine the effects error on these methods, which are ionospheric doubling and the code bias affects on the Phase- Code method and the ionospheric high variation affects on the Phase- Phase method.

### Acknowledgements

I wish to express my deepest gratitude to my supervisors Assistant Professor Mohamed A. Youssef, Associate professor of engineering surveying and geodesy in Mining and Metallurgical Engineering Department, Faculty of Engineering, Assiut University, Egypt; and Dr. A.M. Abdel Hamid, lecturer of engineering surveying and geodesy, in Civil Engineering Department, Faculty of Engineering, Beni-Suef University, Beni-Suef, Egypt. I am truly grateful for their continuous support, encouragement, guidance, and great efforts throughout this work.

### References


Session C
Signal processing in navigation systems
Ship’s Heading Integrity using Multi-Compass

Kazumichi OI
Assistant Professor of Navigation Department
Marine Technical College, MTC
Ashiya-city, Japan
k-oi@mail.mtc.ac.jp

Shigeyuki OKUDA
Professor of Navigation Department
Marine Technical College, MTC
Ashiya-city, Japan
okuda@mail.mtc.ac.jp

Masatoshi SHIMPO
Professor of Department of Navigation School
of Marine Science and Technology
Tokai University
Shizuoka-city, Japan
shimpo-masatoshi@tsc.u-tokai.ac.jp

Yasuo ARAI
Professor emeritus
Marine Technical College, MTC
Ashiya-city, Japan
arai@mail.mtc.ac.jp

Yuji OKUTOMI
Assistant Professor of Navigation Department
Marine Technical College, MTC
Ashiya-city, Japan
y-okutomi@mail.mtc.ac.jp

Abstract—In this paper, the concept and algorithm of “Ship’s Heading Integrity” are proposed. Ship’s Heading is one of the most important Navigational Information and it is essential to make a good integrity automatically to execute a safety and efficient navigation. The proposed algorithm is ship’s heading integrity to find the differences between two or more different types of compasses called “multi-compass”, and simulation of compass movement and ship motion were executed to evaluate the performance of integrity. The validity of integrity system monitoring the differences in frequency domain using multi-compass and easy to find the malfunction of compass were confirmed.

Keywords—ship’s heading; integrity; gyrocompass; magnetic compass; satellite compass; multi-compass;

I. INTRODUCTION

Ship’s heading is one of the essential information to navigate, and all of the vessels equip Magnetic compass and the vessels more than 500 GT also equip Gyrocompass. So, the vessels more than 500 GT equip two compasses and a navigator or an Officer Of Watch (OOW) should watch and navigate to make a good course monitoring the difference between two different types of compasses [1]. This is one of the integrity which monitors two compasses and finds errors and malfunction of sensor or compass by the OOW.

Recently, large vessel equips two Gyrocompasses and monitors the differences between two and make a good course without errors and/or malfunction of ship’s heading. The monitoring system only check a difference between readouts of two gyrocompasses and the alarm generates when the difference is greater than threshold level. Using this system, it is difficult to find which compass has malfunction and not to get high accuracy.

The authors proposed the concept and algorithm of ship’s heading integrity using two or more different types of compasses called “multi-compass”. The algorithm is that the deference between two compass readouts in frequency domain because of selecting between “ship’s motion” and “readout motion due to errors and/or malfunctions”.

In this paper, simulations of compass readout motion include the distinctive malfunctions such as a ship’s velocity error, cut off of feedback system, and ship’s motion (use a sample model such as 500GT cargo vessel as an example of higher speed of motion) by MMG (Mathematical Model Group) were executed in case of as follows;

- same type of Gyrocompasses,
- two different types of Gyrocompasses, and
- Magnetic compass and Gyrocompass.

After then we discuss the performance of proposed integrity system to compare the relative merits between “heading only” and “heading and rate of turn (ROT) or temporal differentiation of heading” and the possibility of finding the malfunction automatically.

Finally, it is concluded that the validity of integrity system using multi-compass monitoring in frequency domain and the possibility easy to find the malfunction compass automatically are confirmed.
II. Ship’s Heading Integrity

Now a day, as Ship’s Heading Devices for more than 500GT vessels, Magnetic compass and Gyrocompass are equipped, and a navigator should always monitor these to make a good course and to check a malfunction of compass for safe and efficient navigation. Recently many vessels equip two Gyrocompasses, and double checking is applied to monitor a malfunction of compass. In this method it is applied to check a difference between two readouts, so it is not able to find troubled compass and to change to a correct one automatically.

So, the concept of Ship’s Heading Integrity, the Outline of Compass and the algorithm are described in this chapter.

A. Concept of Ship’s Heading Integrity

Generally a compass is a sensor of the standard direction on earth, and the standard direction is basically meridional direction or North. Onboard the compass direct the meridional direction or North and using a compass card synchronized with compass sensor and indicated by a scale using feedback system, then it is possible to get ship’s heading and/or bearings of targets on time. So, a compass has a direction block and readout block. The basic performance of compass is affected by magnitude of North directing forth and ship’s motion which are in the direction block, and almost of readout error and malfunction occur in the readout block. Gyrocompass and Magnetic compass which all of the vessels more than 500 GT equip are classified into the mentioned compass.

Recently, another type of compass is developed and some types of small boats equip these, which is called Satellite compass or GPS compass. Satellite compass estimates the difference between two compasses in frequency domain to split into the two components which are direction block and readout block. The basic performance of compass is affected by magnitude of North directing forth and ship’s motion which are in the direction block, and almost of readout error and malfunction occur in the readout block. Gyrocompass and Magnetic compass which all of the vessels more than 500 GT equip are classified into the mentioned compass.

The concept of ship’s heading integrity is very simple and that comparing the difference between two compasses in frequency domain to split into the two components which are direction block and readout block, just same as filtering, former block is low frequency part and latter is higher part. The split frequency depends on the ship’s motion which is easy to obtain from the results of sea trial called the turning circle at hard frequency depends on the ship block is low frequency part and latter is higher part. The split direction block and readout block, just same as filtering, former frequency domain to split into the two components which are that comparing the difference between two compasses in adaption.

Antenna. So, in this paper, the satellite compass is not under number of reception satellites and/or effect of structures nearby. Gaussian and the reliability is also poor because of small direction is poorer than gyrocompass, compass error is not ship because of short start-up time but the accuracy of North line. This compass is very convenient for small "difference between heading and the meridional direction compass or GPS compass. Satellite compass estimates the types of small boats equip these, which is called Satellite Magnetic compass which all of the vessels more than 500GT vessels, Magnetic compass and Gyro compass are equipped, and a navigator should always monitor these to make a good course and to check a malfunction of compass for safe and efficient navigation. Recently many vessels equip two Gyrocompasses, and double checking is applied to monitor a malfunction of compass. In this method it is applied to check a difference between two readouts, so it is not able to find troubled compass and to change to a correct one automatically.

So, the concept of Ship’s Heading Integrity, the Outline of Compass and the algorithm are described in this chapter.

A. Concept of Ship’s Heading Integrity

Generally a compass is a sensor of the standard direction on earth, and the standard direction is basically meridional direction or North. Onboard the compass direct the meridional direction or North and using a compass card synchronized with compass sensor and indicated by a scale using feedback system, then it is possible to get ship’s heading and/or bearings of targets on time. So, a compass has a direction block and readout block. The basic performance of compass is affected by magnitude of North directing forth and ship’s motion which are in the direction block, and almost of readout error and malfunction occur in the readout block. Gyrocompass and Magnetic compass which all of the vessels more than 500 GT equip are classified into the mentioned compass.

Recently, another type of compass is developed and some types of small boats equip these, which is called Satellite compass or GPS compass. Satellite compass estimates the difference between two compasses in frequency domain to split into the two components which are direction block and readout block. The basic performance of compass is affected by magnitude of North directing forth and ship’s motion which are in the direction block, and almost of readout error and malfunction occur in the readout block. Gyrocompass and Magnetic compass which all of the vessels more than 500 GT equip are classified into the mentioned compass.

The concept of ship’s heading integrity is very simple and that comparing the difference between two compasses in frequency domain to split into the two components which are direction block and readout block, just same as filtering, former block is low frequency part and latter is higher part. The split frequency depends on the ship’s motion which is easy to obtain from the results of sea trial called the turning circle at hard port/starboard rudder angle. The algorithm will be described later.

B. Outline of Compass

1) Gyrocompass; consists of directing block and readout block. Former one is called gyro sphere and directs to the meridional direction “North” and latter one is readout block which follows the directing North in gyro sphere up with servo system.

In gyro sphere a three degrees of freedom wheel is spinning with high speed revolution, so spinning axis called gyro axis directs to one direction in the absolute space always caused by rotary inertia [2].

\[
\frac{dH}{dt} + \begin{pmatrix} H_x \omega_y - H_y \omega_x \\ H_y \omega_x - H_x \omega_y \\ H_z \omega_x - H_x \omega_y \end{pmatrix} = T_G + T_S
\] (1)

Where, Angular Momentum of Gyroscope (Kg*m^2*rad/sec)

\[
H = \begin{pmatrix} H_x \\ H_y \\ H_z \end{pmatrix} = \begin{pmatrix} 0 \\ I_y \omega_y \\ I_z \omega_z \end{pmatrix}
\] (2)

Torque caused by Gyrocompass Function (Kg*m)

\[
T_G = \begin{pmatrix} 0 \\ K \phi \\ \mu \end{pmatrix}
\] (3)

Torque caused by Ship’s motion (Kg*m)

\[
T_S = \begin{pmatrix} 0 \\ a_\phi \theta \\ a_\theta \end{pmatrix}
\] (4)

XYZ and xyz Cartesian coordinate: there origins are at the center shown in Fig. 1.

Fig. 1. Theory of Gyrocompass (Rotary Inertia)

The Gyrocompass consists of Gyro sphere including gyro spinning wheel with high speed revolution as direction block and follow up system including compass card or readout as readout block shown in Fig. 2.

Fig. 2. Diagram of Typical Gyrocompass

Gyro axis keep to direct to North caused by not only rotary inertia but also damping function which is also in Gyro sphere.
Recently this function is not so different in conventional Gyrocompass. Gyrocompass works to direct North with rotary inertia and damping using the Earth rotation.

\[ \Omega_o = \begin{pmatrix} \Omega_o \cos \varphi \\ 0 \\ -\Omega_o \sin \varphi \end{pmatrix} \]  

Earth Rotation Vector at Latitude \( \varphi \) (rad./sec.)

Ship’s Velocities Over the Ground (m/s)

\[ \mathbf{V}_G = \begin{pmatrix} u_G \\ v_G \\ w_G \end{pmatrix} \]

Rotation Vector causing by Ship’s movement (rad./sec.)

\[ \Omega_S = \begin{pmatrix} u_G \cos \psi + v_G \sin \psi \\ u_G \sin \psi - v_G \cos \psi \\ 0 \end{pmatrix} / R_o \]

So, Earth Rotation Vector at Latitude \( \varphi \) (rad./sec.)

\[ \Omega = \Omega_o + \Omega_S \]

(8)

Y axis component of (8) affects gyrocompass error called speed error.

It is possible to resolve the gyrocompass motion using with simultaneous differential equation (1) to (8) in computer simulation, if necessary in precise resolution. If not, the solution will be solved analytically as (9) in [3].

\[ \begin{align*} -B\dot{\Phi} + H(\theta + \Omega(\sin \psi - \Phi \cdot \cos \psi)) &= K\Phi \\ B\dot{\Theta} + H(\Phi + \Omega \cos \psi \cdot \theta) &= \mu \Phi \end{align*} \]

(9)

\( K \) is called Gain to increase the directing forth. If \( K \) is greater, then oscillation of direction will be occurred. \( \mu \) is called Damping factor to deduce the oscillation. If \( \mu \) is greater, then it takes a long time to set the direction. These two parameters are fixed by the type of gyrocompass and parameters which affect to gyrocompass are described as follows (the performance is described in [3]).

Gain : mass and volume of gyro sphere, density of supporting liquid and distance between center of gravity/center of flotation and supporting point, and gravity.

Damping factor : torque around \( y \)-axis which is perpendicular to Gyro axis (\( x \)-axis), and it is generated by liquid flow or servo motor.

2) Magnetic Compass:

Magnetic Compass is traditional compass which is essential to navigation, according to rubabric compass without electricity. It is on force for more than 500 GT vessels to equip Gyrocompass onboard, but for all of the vessels to equip Magnetic compass. The OOW should monitor the Gyrocompas and Magnetic compass to make a good course on his/her watch (the performance is described in [4]).

The characteristics of Magnetic compass are as follows:
- Direct to Magnetic medirian, not True one. (Varidation)
- Deviation should be corrected, even with the correction table to reduce swinging oscillation of compass card.
- Deviation often changes, because it is affected by ship’s hull or cargo material such as iron.

Comparing to another compass such as Gyrocompass, it takes much time to compare such as 3 or 5 minutes because of taking out of effects of swinging oscillation of Magnetic compass card.

3) Satellite Compass:

Satellite Compass is a new heading indicating device developed for Radar Plotting on less than 500 GT vessels not equipped Gyrocompass, called Transmitting Heading Device THD.

This compass is a hybrid system using RTK GNSS and angular velocity sensor, and accuracy with 0.5 degrees rms. To use Heading Integrity, the accuracy is not enough and not white noise for a short time such as 1 minute or so. According to this reason and merchant vessels or more than 500 GT does not equip Satellite compass to use for navigation now. So, it will be near future to use Satellite compass to gain heading integrity.

C. Algorithm

To realize the concept on Ship’s Heading Integrity described at Section A, two ship’s headings from each compass are\( \psi_i(t), i = 1,2 \), and input to FFT, then the power spectrum \( \Psi_i(t), i = 1,2 \) are calculated.

Power spectrum in a part of low frequency is affected by ship’ motion and/or movement of gyro axis, in higher part is affected the malfunction such as hunching and rotation of compass card in readout block using follow-up system.

![Diagram of Ship’s Heading Integrity](image-url)
III. SIMULATION AND DISCUSSION

To survey the effectivity of proposal Ship’s Heading Integrity, the simulation of compass movements and calculation and/or simulation of ship’s movement were executed and discussed.

A. Ship’s Motion

According to the object of this research, the target vessels are decided as more than 500 GT conventional merchant vessels and except high speed vessels more than 30 knots.

The speed of ship’s heading movement is required by MMG simulation [4] and/or of turning circle trial, and the ship which has the highest speed of heading movement is 199 GT cargo ship and her principal particulars shown in TABLE II.

TABLE I. CARGO SHIP’S PRINCIPAL PARTICULARS

<table>
<thead>
<tr>
<th>Gross Tonnage</th>
<th>LPP</th>
<th>Breadth</th>
<th>Depth</th>
<th>Draught</th>
</tr>
</thead>
<tbody>
<tr>
<td>199 tons</td>
<td>54.00 m</td>
<td>9.60 m</td>
<td>5.05 m</td>
<td>2.94 m</td>
</tr>
</tbody>
</table>

According to the result of simulation, her maximum ROT is 3.0 deg./sec. and Time delay $t_o$ [sec] is approximately 8.2 sec. The ship’s motion response of turning circle is just like passing through Low Pass Filter LPF, so the cut-off frequency (Hz) calculated by (10) is 0.38 Hz in [4].

$$f_{cut} = \frac{\pi}{t_o} = \frac{l \times df}{df} = f_{clock}/2N \quad (11)$$

It is possible to divide ship’s navigation into three navigational stages which are (a) mooring, (b) coastal navigation and (c) turning. In (a) and (b), real ship’s heading data were used, which were recorded onboard. She is “T/S Kaigi-maru” 150 GT belong to Marine Technical College, nearly same size as TABLE II.

B. Multi-Compass

Recently, many of large vessels equip two gyrocompasses in steering stand or rear side of bridge and these compasses are same type. A pair of different type of Gyrocompasses is infrequently-encountered, and a pair of Magnetic compass and Gyrocompass is also necessary. Any type of Gyrocompass and Magnetic compass are considered as the feedback system with gain $K$ and damping factor $\mu$.

Using heading Integrity, feedback system in a compass has these parameters and concept is applied to any type of compass. So, difference of pair affects performance of integrity, but it is very difficult to know the parameters directly. In this study, to discuss the basic performance of integrity, simulation will be executed after this section.

1) Same type of Gyrocompasses:

In this case, it is difficult to know parameters in the concrete, but random error and static error of compasses is presented in the specification of this type. Random errors are set same as the specification, but static errors are set each within the range of specification.

2) Different type of Gyrocompasses:

In this case, random error is set in proportion to the gain $K$ and the damping factor $\mu$.

3) Magnetic Compass and Gyrocompas:

In this case, the parameters of Magnetic compass is very poor, so taking a rapid changing course and/or large deviation without correction for a long term cause swinging oscillation of compass card

Once oscillation occurs, it should wait until swinging be settle, it takes approximately three minutes or so, integrity of Magnetic compass and Gyrocompass require a long interval such as 3 minutes or 6 minute.

C. Simulation Result

Set Gyrocompass parameter shown in Table III, and the simulation with typical malfunction “hunching” and “stopping compass card” in coastal navigation, high speed turning and mooring were executed.

TABLE II. GYROCOMPASS PARAMETERS SIMULATED

<table>
<thead>
<tr>
<th>Gyro Co. (1)</th>
<th>Gyro Co. (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random error Deviation</td>
<td>0.2 deg. rms</td>
</tr>
<tr>
<td>0.5 deg.</td>
<td>0.1 deg.</td>
</tr>
</tbody>
</table>

Integrity System has an observation time is set as 25.6 sec. for one term, sampling number is $2N = 1024$ and data refresh clock $f_{clock}$ is 1/40 sec. or 40 Hz.

1) Coastal Navigation:

To survey the performance of detection of Malfunction which is “stopping compass card”, the Gyrocompass data recorded when T/S Kaigi-maru was navigating on Harima-nada in Inland Sea.

In case of Two Compasses are Normal Working

Fig. 4. Time History and Power Spectrum on Coastal Navigation. In case of Two Compasses are Normal working

In case of Malfunction “Stopping Compass Card”

To certificate the performance of malfunction detection, 4 cases (stopping time : 0.5, 1, 2 and 5 sec.) were simulated using coastal navigation data were adapted.

Fig. 5. Time History and Power Spectrum on Coastal Navigation In case of one’s Malfunction “Stopping Compass Card” (Stopping time is 5 sec.)
In case of One’s Malfunction “Hunching Compass Card”

To survey the performance of malfunction detection, 7 cases of hunching Amplitude (10.0, 5.0, 2.0, 1.0, 0.5, 0.25 and 0.125 deg.) with hunching frequency 3.33 Hz were simulated.

In case of Two Compasses are Normal Working

To certificate the performance of malfunction detection, 4 cases (stopping time : 0.5, 1, 2 and 5 sec.) were simulated using mooring data.

D. Performance of Ship’s Heading Integrity

Ship’s heading integrity is “the automatic detecting and recovering malfunction” system because of high reliability and accuracy of gyrocompass except malfunction. So, utility of multi compasses is available to double checking monitor and detecting malfunction.

1) In case of Malfunction of Hunching:

To survey the effect malfunction of hunching, the simulation results shows in III - C - 2). If Malfunction of Hunching is caused, the readout heading information has power spectrum in part of higher than cut of frequency $f_{cut} = 0.38$ Hz.

$f_{cut}$ is set according to the highest ROT, so mainly LPF output contains ship’s movement, but HPF output does not. In case of no malfunction such as hunching except just like as malfunction of stopping compass card, HPF output will be zero or white noise. So, to detect the malfunction of hunching, it is clearer to detect the difference of two that the averages of differences of HPF output power spectrum at each frequency component are used.
The comparing result is shown in TABLE VI, the difference is clearly to separate malfunction Gyrocompass and normal one. In this case, Gyrocompass (2) has a malfunction of hunching for one second, so it shows that even very small hunching amplitude is able to be found because of comparing each frequency component.

The simulation in coastal navigation was executed, and the result is shown in TABLE V. According to this result, more than amplitude 0.25 deg., it is able to detect the effect of hunching. It is affected by LPF output or ship’s movement, so Judgement of detection level should be adjusted considering ship’s motion, etc.

**TABLE III. AVERAGE OF DIFFERENTIAL POWER SPECTRUM AT EACH FREQUENCY COMPONENT IN CASE OF TURNING**

<table>
<thead>
<tr>
<th>Diff. PS (×100 deg.)</th>
<th>Average of Differential Power Spectrum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>Diff. of LPF outputs</td>
<td>0.03</td>
</tr>
<tr>
<td>Diff. of HPF outputs</td>
<td>0.00</td>
</tr>
<tr>
<td>Diff. of ALL outputs</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Hunching time = 1 sec.

**TABLE IV. AVERAGE OF DIFFERENTIAL POWER SPECTRUM AT EACH FREQUENCY COMPONENT IN COASTAL NAVIGATION**

<table>
<thead>
<tr>
<th>Diff. PS (×100 deg.)</th>
<th>Average of Differential Power Spectrum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>Diff. of LPF outputs</td>
<td>-0.02</td>
</tr>
<tr>
<td>Diff. of HPF outputs</td>
<td>0.02</td>
</tr>
<tr>
<td>Diff. of ALL outputs</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Hunching time = 1 sec.

2) In case of Malfunction of Stopping Compass card:

To survey the detection malfunction of stopping compass card, the simulation results shows in III- C - 3). Malfunction of stopping compass affects not to HPF outputs but mainly to LPF output and it is not able to separate ship’s motion.

The phenomenon of stopping card is almost same as mooring to pier. So it is more difficult to detect malfunction in mooring in case of larger ships. Using the mooring data of T/S Kaigi-maru, and simulation of stopping card in time 0, 1, 2, 5 10 and 25.6 sec. and the results of variances of Differential Power Spectrum are shown in TABLE V.

In the results of this simulation, variances of HPF outputs are a little changes but variances increase over the stopping time 10 sec. In mooring, ship’s heading has little changes, so it is of little problem to take more time to detect.

It was surveyed the detection performance of detection time in coastal navigation. The results are “it takes more than 5 sec. It is too long to detect malfunction in navigating. It is possible to solve the problem using ROT data also which is almost gyrocompass outputs.

3) Utility of Ship’s Heading Integrity:

The ship’s Heading Integrity is one application of information processing in frequency domain shown in Fig.3, and also it is possible to solve the average and variances in time domain.

It is possible to make a correction of static compass error and/or magnetic variation using the Ship’s Heading Integrity with electrical compass log and statistical processing.

**TABLE V. VARIANCE OF DIFFERENTIAL POWER SPECTRUM AT LPF AND HPF IN MOORING**

<table>
<thead>
<tr>
<th>Stopping Time (sec.)</th>
<th>Variation of Differential Power Spectrum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Var. of Heading (×100 deg.)</td>
</tr>
<tr>
<td>Diff. of ALL outputs</td>
<td>0.52</td>
</tr>
<tr>
<td>Diff. of HPF outputs</td>
<td>0.53</td>
</tr>
<tr>
<td>Diff. of ALL outputs</td>
<td>0.53</td>
</tr>
</tbody>
</table>

**REFERENCES**


Session D
Marine navigation. Applications of GNSS and other technologies. Special problems.
A Study on Vessel Traffic Management System at Tokyo Bay using AIS

Tomoko Shirai  
Department of Maritime System Engineering  
Tokyo University of Marine Science and Technology  
Tokyo, Japan  
d122009@kaiyodai.ac.jp

Nobuaki Kubo  
Department of Maritime System Engineering  
Tokyo University of Marine Science and Technology  
Tokyo, Japan

Kenji Inada  
Maritime Traffic Research Department  
The Japan Association of Marine Safety  
Tokyo, Japan

Hitomi Tamaru  
Department of Maritime System Engineering  
Tokyo University of Marine Science and Technology  
Tokyo, Japan  
tamaru@kaiyodai.ac.jp

Abstract—With the spread of AIS, it has now become easy for us to obtain the computerized status information of ship. In this study, a basic of the vessel traffic management system using AIS information intended for the vessel underway at Tokyo Bay is suggested.

We established nodes and links from the limited possible navigation sea area between the entrance of Tokyo Bay and the inner part of the bay, and chose the routes from the departure to the destination. In addition, we standardized the navigation route of the vessel based on observed AIS data and performed the extracted route comparison with the route provided by AIS.

We inspected whether we can cope with the prediction of the ship trend from the extracted route and the real route. It was effective for future systems construction although there is a problem in practical use.

Keywords—Port Traffic, Optimal route, AIS

I. INTRODUCTION

Various information for ship operation is provided by AIS (Automatic Identification System). Besides dynamic information such as the position, heading, speed over the ground etc. the necessary information for the vessel traffic management like destinations is included in AIS information. The input method of the destination cord of the AIS was determined with the partial revision of the Act on Port Regulations and the Maritime Traffic Safety Act in Japan from July 1, 2010 [1], and the input was made mandatory. When a port of destination, the course (such as mooring places) in the port, other necessary information (such as passage routes) were coded, and Format of this destination cord was input and performed a traffic analysis from AIS information, the processing with the calculator became easy. It may be possible to improve safety of the vessel traffic in the congested area by using this information effectively.

The distance from the entrance of Tokyo Bay to the inner part of bay is approximately 70km, so Tokyo Bay is a large waters. But the navigation route of the ship especially large is almost decided for the safety in the congested area such as the established route in Tokyo Bay, the management of the traffic by VTS (Vessel Traffic Service), the limit of the navigation area.

II. CHARACTERISTIC OF THE TRAFFIC FLOW NETWORK IN THE TOKYO BAY

An input method of the information about the destination along the method that IMO recommended [2] was made a rule in Japan from July 2010, the input was made mandatory. Vessels carrying AIS enter the destination code, so we can obtain the information of departure and the destination from AIS information. Destination code is used to predict each other's courses, in addition to course display by international maritime signal flags and by the whistle in port. The input methods of the destination cord of the AIS are as follows [3].

- In the case of the ship which will sail on the route to establish in the Maritime Traffic Safety Act (Uraga Suido Traffic Route, Naka-no-se Traffic Route), and the ship which will sail in the port concerned or in the vicinity of its boundary for the purpose of entering the port to which the Act on Port Regulations applies (Keihin port, Chiba port, Kisarazu port, Yokosuka port, Tateyama Port) (Fig.1)
In addition, in Tokyo Bay the area of the sea that can navigate the large ship is limited to and a route and navigation are determined as local rule in detail, so a traffic area is limited by a model of a vessel and it is thought that we can estimate a plan route to some extent. About a plan route, we expressed it by establishing the network using the node and the link in Tokyo Bay, and we decided to use network simulation to reproduce a traffic flow.

We show traffic flow in the Tokyo Bay and network (nodes and links) in Fig.2 [4]. The dark blue lines which showed traffic flow of the Tokyo Bay are trails of the ships (The ships which depart from the entrance of Tokyo Bay for arrival in port or anchorage, and the ships which depart from Keihin port, Chiba Port or the Kisarazu Port for the entrance of Tokyo Bay). We showed a node of the network with a number and showed a link to tie between nodes in yellow, light blue, green line each. The light blue line shows the route of the ship leaving for the entrance of Tokyo Bay from each port, the yellow line shows the route of the ship toward each port from the entrance of Tokyo Bay. The yellow and light blue line is one-way traffic. In contrast, the green line shows the route that is movable in both directions, two-way traffic.
We show the route example of the ship sailing in Tokyo Bay. For example, in the case of a ship which leave the entrance of Tokyo Bay for the Keihin port (Tokyo Ward), the destination cord is “→ JP TYO S”. “TYO” means the ship’s destination is Keihin Port (Tokyo Ward), and “S” means the vessel is heading for the berth in Shinagawa Wharf in the destination port. Orange line in Fig.3 show the trail of the ship which input destination code “→ JP TYO S”. Also, in Fig.3 we superimposed number “1”~“58”, those number mean node of the above-mentioned network. The ship more than 50 meters in length which leave the entrance of Tokyo Bay for the Keihin port (Tokyo Ward) must sail Uraga Suido Traffic Route, so the ship is required to sail “1”, “2” and “3” at a speed through the water of 12 knots or less. In addition, the ship is required to sail “3”, “5”, “6” and “7” because the ship more than 50 meters in length with a draft less than 20 meters must sail Naka-no-se Traffic Route. Next, the ship is required to sail in order of “7”, “14”, “15” and “16”, a vessel of 3,000 gross tons or more which sail toward the Tokyo Wan Aqua Line from the Naka-no-se Traffic Route must sail through the Tokyo Aqua Line East Fairway. Finally, a ship within a circle of 1 mile radius centering around Tokyo Offing Light-buoy (it’s between “17” and “43”) must sail so as to keep the same buoy on their port side. So the ship will sail in order of “16”, “17”, “43”, “Tokyo west passage” and “Keihin Port (Tokyo Ward) Shinagawa Wharf”.

Fig.4 and Table1, 2 shows the traffic route and passage in Tokyo Bay [5] [6].

[Table I: TRAFFIC ROUTES IN TOKYO BAY]

<table>
<thead>
<tr>
<th>Traffic Routes</th>
<th>Compulsory Transit of Traffic Routes</th>
<th>Services Offered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uraga Suido Traffic Route</td>
<td>50 meters and upwards in length (20 meters and upwards of draft is exempted from compulsory usage of the Naka-no-se Traffic Route)</td>
<td>Tokyo Wan Vessel Traffic Service Center (“Tokyo MARTIS”)</td>
</tr>
</tbody>
</table>

[Table II: PASSAGE IN TOKYO BAY]

<table>
<thead>
<tr>
<th>Name of passage</th>
<th>Controlled vessel</th>
<th>Vessel subject to control</th>
<th>Competent authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiba Passage</td>
<td>Over 1,000 GT or over</td>
<td>50 meters or over in length overall (excluding those less than 500 GT)</td>
<td>Chiba Port Traffic Control Office</td>
</tr>
<tr>
<td>Ichihara Passage</td>
<td>Over 1,000 GT or over</td>
<td>50 meters or over in length overall (excluding those less than 500 GT)</td>
<td>Chiba Port Traffic Control Office</td>
</tr>
<tr>
<td>Yokohama Passage</td>
<td>Over 1,000 GT or over</td>
<td>50 meters or over in length overall (excluding those less than 500 GT)</td>
<td>Yokohama Port Traffic Control Office</td>
</tr>
<tr>
<td>Tsurumi Passage</td>
<td>Over 1,000 GT or over</td>
<td>1,000 GT or over</td>
<td>Kawasaki Port Traffic Control Office</td>
</tr>
<tr>
<td>Kawasaki Passage</td>
<td>Over 1,000 GT or over</td>
<td>1,000 GT or over</td>
<td>Kawasaki Port Traffic Control Office</td>
</tr>
<tr>
<td>Tokyo East Passage</td>
<td>Over 1,000 GT or over</td>
<td>50 meters or over in length overall (excluding those less than 500 GT)</td>
<td>Tokyo Port Traffic Control Office</td>
</tr>
<tr>
<td>Tokyo West Passage</td>
<td>Over 1,000 GT or over</td>
<td>100 meters or over in length overall</td>
<td>Tokyo Port Traffic Control Office</td>
</tr>
</tbody>
</table>
III. ESTIMATED ROUTE USING NETWORK DATA

For example, the network where a node and a link were set like Fig.5 was assumed, and was given the distance between each link shown at table 3. In the table, M means disconnected node and 0 mean same node.

Now when I want to find the minimum time route toward "h" from node "a". A minimum distance between node "b", "c" or "d" and "a" is selected. In this case, node "b" is minimum distance, so set flag on node "b". In next step, distance to node "e" and "f" connected node "b" are calculated. Among node "e", "f", "c", "d", a flag is set on node to the minimum distance route. The minimum distance routes from node "a" to all nodes are demanded when I repeat this process.

Using the network where each ports and route to the Tokyo bay entrance, waypoints and junction were set as each node, we obtain a route by a former method.

Fig.6 shown estimated route by above method form entrance to Tokyo West passage off. Using the network where each ports and route to the Tokyo bay entrance, waypoints and junction were set as each node, we obtain a route by a former method.

Fig.6 is shown an estimated route form Tokyo bay entrance to Tokyo West passage off.Fig.7 is shown an estimated route under the condition that could not pass west route.

Actual trajectories are shown in Fig.8. Almost ships pass on West Fairway but, several ships pass on East Fairway. In the case of Tokyo Bay, the route of the ship varies according to a ship type, and size. This means that there are a node and link which I cannot use by a ship type and size.
IV. SIMULATION AND RESULT

In this study, a scenario that includes a departure node, departure time, and arrival node was made. But all ships' speed is assumed at 12 knots.

An estimated route is made by a departure node and arrival node by above method. Ship's positions, heading are calculated on estimated route on a certain time. (Fig.9).

Similarly, we can demand the position of other ships, too and estimate position relations at the certain time and judge the presence of the collision.

A collision risk point is found by CPA. The criterion is DCPA is with \( \frac{L}{2} \times 3.2 \) and less than five minutes at the TCPA.

Fig.10 is shown collision area by this simulation. Blue points is crossing, pink points is head-on vessel, orange points are overtaking. The target ship on simulation was chosen by the next condition (departure place and destination. See Table II).

- From Bay Entrance to Chiba Passage
- From Bay Entrance to Ichihara Passage
- From Bay Entrance to Yokohama
- From Bay Entrance to Kawasaki
- From Bay Entrance to Tokyo East Passage
- From Bay Entrance to Tokyo West Passage
- From Kawasaki to Bay Entrance
- From Yokohama to Bay Entrance

Simulation points a little point but area shown. Fig.11 is shown collision points by red circle. Simulation result is shown the collision risk waters.
When we compare actual trajectories (Fig. 10) and simulation result (Fig. 11), risk point number is different. But an area of collision risk was shown (It shown by Orange circle).

V. CONCLUSION
In this study, the Tokyo Bay traffic network on the seas in consideration of an actual trajectories and regulation was expressed. Using destination code that input into AIS information, we made planned route from network, and a collision points was calculated by planned route in the future.

As a future works, I was able to find some collision place from a simple method, but it is necessary to consider the setting method of the route according to the ship type and speed adjustment.

ACKNOWLEDGMENT
We appreciate the help received from Maritime Traffic Department, Japan Coast Guard.

REFERENCES
[1] Japan Coast Guard, New Rules for Maritime Traffic Safety, Effective as of July 1, 2010
[3] 3rd Regional Coast Guard Headquarters Japan Coast Guard, Entry of information on destination in AIS (for Tokyo bay).
[6] 3rd Regional Coast Guard Headquarters Japan Coast Guard, Introduction of a new port traffic control system.
A Statistical Inference Approach to Extract Principal Fairways of Ship Passages through a Sea-strait*

Jinhai CHEN
Marine Navigation & Pilot Technology Research Center
Navigation College, Jimei University
Xiamen, China
jhc@gmail.com

Feng LU
State Key Laboratory of Resources and Environmental Information System
IGSNRR, CAS
Beijing, China
luf@leis.ac.cn

Abstract—Seamen usually select popular routes according to navigational and hydrographical characteristics when passing through a strait with congested traffic. In order to minimize the possibility of ship-ship collisions, principle fairways (PFs) are commonly used by sailors. It is essential to delineate PFs quantitatively and objectively for designing or refining routing measures. A new vision of PFs is introduced in this paper by examining how habitat use is perceived in the wildlife ecology. PFs in this paper is a concept that attempts to provide basic information about cumulative activity patterns for vessels group in a strait and serve as the basis of quantifying space-use patterns. A space use method found in habitat evaluation of wildlife is applied to extract PFs of ship passages through a strait. Compared with existing methods, the proposed method helps to identify cumulative activity patterns for ship groups derived from mass ship trajectories, and provides a clearer interpretation of shifting space-use patterns within strait corridors. Moreover, it gives a better insight for directional and seasonal factor for PFs in straits. Finally, this novel method is used to extract PFs in western Taiwan Strait and its adjacent sea. The results indicate that the proposed method is helpful to identify gaps between current ship routing system plan and cumulative activity patterns recognized by real ship trajectories.

Keywords—principal fairways; space use; ship trajectories; ship routing system; Taiwan Strait; AIS

I. INTRODUCTION

There were various traditional observation to extract shipping routes due to the application of Voice Radio Communication and Radar in tracking vessel [1]. Thousands of antecedent voyages across the oceans were extracted from ship logbooks in the age of sail [2]; Radar data were implemented to create ship routing system of Dover Strait [3]; Global commercial ship lanes were abstracted by using of locations reported by voluntary observing ships [4]. However, traditional records were just sparse samples and shipping traffic was not captured by these data. Fortunately, nearly all seagoing merchant/passenger ships are required to carry an Automatic Identification System (AIS) according to current mandates released by the International Maritime Organization (IMO). The wide deployment of AIS allows ships be tracked in 24 hours per day. Meanwhile the emerging technology of space-based AIS provides a global coverage of the maritime domain [5]. The ship tracks have increased tremendously, and these advances have been accompanied by the development of new methods, which serve for marine transport planning especially in crowded shipping areas, such as straits.

Straits are shortcuts between two large water bodies. These passageways generally host large amounts of traffic volume with various directions. Thus, the risk of ship-ship colliding is higher in straits than in other waterways [6, 7]. To minimize the possibility of collisions in ships passage straits, states bordering them may create routing systems to separate vessels, control crossing and meeting situations. Developing ship routing schemes in congested straits have become an important issue, especially the straits heavily used for international navigation, such as the Dover Strait [3, 8], the Istanbul Strait [9], and the Malacca Strait [6, 10]. However, only 15 straits throughout the world have implemented Traffic Separation Schemes (TSS) approved by the IMO [11]. Vessels are required to follow certain sea lanes in those TSS areas. There are improvements for other straits on keeping ship traffic flows in good order. Within this context, maritime administrators should consider various factors, including the knowledge of Principal Fairways (PFs), existing navigation aids, the state of hydrographic surveys in the area and accepted standards of routing [12]. There are always attractive routes accessible to sailors with respect to navigational and hydrographical characteristics through a given strait. PFs are the water areas commonly used by sailors and have a large traffic volume. Thus they are empirically significant in recommended routes. In previous work involved ship routing system on straits [9], PFs are subjectively selected based on mariners’ experience, which is not quantitatively repeatable.

Perceiving how mobile objects move about space is a fundamental research question in Geographic Information Science (GISc), and there are many quantitative methods for analyzing movement data [13]. Of these, Origin-Destination (OD) matrices is arguably the most straightforward methods to represent routes in a network [14]. It is also accepted in maritime engineering literature. For example, Kaluza et al. [15] extracted 490 517 routes linking 36 351 distinct pairs of OD in annual global AIS database to represent the global shipping network. Eight main ODs were defined as main routes in the Istanbul Strait [9]. However the method of OD matrices only

*Research supported by the Applied Basic Research Program of Ministry of Transport of the China (2013 329 815 290) and the Natural Science Foundation of Fujian Province of China (2015J01166).
concerns the origin and destination of each movement but ignores the actual trajectories. One of the potential solution is Route Topology Modelling appliance, where routes are abstracted into a topology model composed of discrete legs, junctions, nodes and their associated attributes like urban road network [16]. However, the maritime environment remains an open space where navigation in any direction is allowed. Thus, the shipping network is rather different from the road network, which is known a priori to construct corresponding urban road topology [17].

It is important to keep in mind that in maritime engineering communities, the motive of sea-lanes identification is to preserve navigational safety rather than to port accessibility[14]. This is the reason why a manual strategy is adopted in fairways selection [18, 19]. Dorpa et al. [20] applied an automated filtering process to clean dirty routes and define ideal trajectories in simulator. Tzavella et al. [21] proposed a semi-automatic sea-lane extraction approach with particle filtering. Apart from manual, supervised and semi-automatic approach, there are other various unsupervised strategies of fairway extraction, such as image processing technique [22], shared nearest neighbour algorithm [23], density-based clustering of ships’ waypoints algorithm [24-26]. Those existing methods of ship route extraction just concern waypoints, then give a continuous line or labeled track-lines set. Their results were representative of a sea-lane, but ignore the boundary and utilization distribution of principal fairways. In terms of maritime corridors identification on the whole, it is clear that dedicated efforts should be directed to increase scientific understanding of PFs.

The existing methods only provided low-level contextual information of PFs, such as waypoints through track-lines, and the direction on each leg. It indicates a need for next generation of quantitative method to discover high-level knowledge concerning capturing PFs boundaries and understanding changes in PFs over time. As such, we seek to answer the question: how to delineate the boundary of PFs and recognise the importance of variations in the intensity of space-use within a PF.

This paper is outlined as follows. The background and related work in ship movement tracks analysis are introduced in section 1. Section 2 provides a brief overview of fundamental concepts of space-use, then introduces the material of AIS and bathymetric dataset. Section 3 elaborates the proposed method for delineating PFs. Section 4 illustrates how the space-use technique performance for its application on Taiwan Strait, and shows how to utilise the proposed method to identify gaps between current routing system plan and the extracted PFs. The application is used to guide a practical discussion of the usefulness of the space-use technique for PFs analysis in section 5. Finally, section 6 draws the conclusion and provides some suggestions for future research applications of space-use techniques in maritime engineering.

II. CRITERIA AND MATERIAL

As discussed above, there are few algorithms about quantitative delineating PFs available in the scientific literature. PFs were always subjectively selected in previous routing scheme planning. In order to get a further understanding of PFs, nautical science maybe have a try to embrace other disciplines such as ecology and geography. So we develop a cross-disciplinary application of ecological methods found in habitat use of wild animal.

A. CRITERIA AND MATERIAL

To facilitate cross-disciplinary learning, we take a try to view PFs from the perspective of ecological discipline, especially review how habitat use is perceived in the wildlife ecology and obtain a new knowledge of PFs.

1) The base criteria of habitat use

Historically, animal space-use patterns have been described using a series of concepts, such utilization distribution, home range and core area. Firstly let’s briefly introduce some base criteria from movement ecology as below:

Utilization Distribution (UD): UD was formally defined as “the two-dimensional relative frequency distribution for the point of location of an animal over a period of time” [27]. Thus UD is a concept describing the intensity that an animal uses a certain geographical location in a form of probability density surface. And the intensity could be represented as “the relative amount of time that an animal spends in any place” [28].

Home Range (HR): Wildlife ecology researchers believe that individual animals restrict their movements to finite areas and each wild animal has a HR, which defines the spatial area occupied by an individual during its day-to-day activities[29]. Thus HR is a concept that attempts to describe the spatial context of an animal’s behaviour and its boundaries are delineated by contour lines (i.e. isopleths) based on UD. For density based methods like Kernel Density Estimation (KDE), the 95% Isopleths Volume (IV) of UD is accepted for delineating the boundaries of HR [30-32].

Core Area (CA): Seaman and Powell [33] defined CA as “the core of a HR”. Vander Wal and Rodgers [30] further refined the definition of CA as “the area within which an individual spends a maximum amount of time”, supposed that CA could be “delineated using a time-maximising function derived from kernel analyses”. Harris et al. (1990) concluded that core areas may be “useful in understanding the behaviour of an animal, by providing a clearer interpretation of shifting patterns of use within a home range, and allowing better insight into intra-specific and inter-specific patterns of area use”.

The combination of location data and kernel HR estimators have produced widely accepted and intuitive interpretations of a HR as a probability density surface, depicting the likelihood that any point in space was going to be occupied by the animal of interest at any given time [34]. Although Downs, Horner and Tucker [35] suggested the techniques of animal home range analysis might be useful for analyzing ships tracking data, apparently they are rarely if ever applied in maritime engineering.

2) The derived nautical criteria

The ecological criteria related to space-use are intended for the spatial area occupied by an individual animal. While in
terms of maritime engineering, PF is a concept that attempts to describe cumulative activity patterns for ship groups. A concerted effort should be taken to identify gaps between ecological knowledge and nautical knowledge, then to integrate ecological knowledge into the quantitative approach for delineating PFs. Thus we develop derived novel nautical concepts as below:

**Shipping Utilization Distribution (SUD):** SUD is defined as the two-dimensional relative frequency distribution for the track lines determined by waypoints from all involved ships over a period of time. Such a probabilistic space-use surface could be used to map and quantify the likelihood of strait corridors used by involved vessels group.

**Application Nautical Range (ANR):** ANR is defined as the area of certain convenience for sailors with respect to navigational and hydrographical characteristics. Similar to the HR for an animal, ANR for sailors using given open water could be delineated as the 95% IV from SUD.

**Principal Fairway (PF):** In a quantitative view, we define PF as the area within which the cumulative sailing time for all involved ship groups reaches the maximum. PF is a concept that attempts to provide basic information about cumulative activity patterns for vessels group in a strait and serve as the basis of quantifying space-use patterns.

### B. Study area and relevant datasets

1. **Study area and observing period**

   The study area encompasses the western Taiwan Strait and offshore Fujian Province bounded by the latitudes 23°N and 28°N, longitudes 116.5°E and 121.5°E, as shown on Fig. 1. The Taiwan Strait is a typical busy waterway linking Northeast Asia, Southeast Asia and the Indian Ocean rim.

   The observing period took place from 2011/10/01 to 2012/09/30. Because the East Asian Monsoon has significant wind effects in the Taiwan Strait, summer in this paper is limited to June to August, during which the southwest monsoon occurs, while winter is bounded from November to January[36].

2. **AIS tracks dataset**

   The China Maritime Safety Administration (MSA) owns a terrestrial AIS network system. The original daily log files collected by this system are used to illustrate the proposed PFs extraction method. The dataset covers 39,000 vessels from Oct 2011 to Sep 2012. The routes’ aggregation was visualised and spatially analysed on ArcGIS software. A density map of merchant ship traffic flows is shown in Fig. 1, which illustrates that most commercial ships follow some strait corridors with the northeast/southwest directions.

### III. METHODOLOGY DEVELOPMENT

The proposed quantitative methods for delineating PFs follows a four-step procedure as shown on Fig. 2: converting raw AIS data to labelled route objects with clustering of waypoints in step 1; filtering out uneconomic routes with a geographic cost function in step 2; retrieving SUD with Kernel Density Estimation (KDE) in step 3; and finally constructing a statistical inference model to delineate PFs based on a nonlinear generalised regression model in step 4.

#### A. Extracting transit-passage route

The original ship movement datasets collected by CMSA’s terrestrial AIS network were loaded into the ArcGIS Geodatabase. A 6-minute interval is used to resample these raw datasets. It is essential to divide the contiguous sequence of position reports into a set of routes. 533,800 routes were generated by using the ship route extraction method based on preliminary clustering of waypoints [24]. There are 35,500 transit-passage routes objects for the observing period. Here transit-passage means involved ship undertakes a continuous and expeditious transit of the Taiwan Strait. Only the transit-
passage route objects in summer/winter are used for PFs analysis. We are working on the assumption that those transit-passage ships exclude the handy-size individuals. Thus, 15,200 transit-passage ones are selected as route samples as shown on Table.1.

Table.1 COMPONENTS OF TRANSIT-PASSAGE ROUTE SAMPLES GROUP BY SEASON AND DIRECTION

<table>
<thead>
<tr>
<th>Season/direction</th>
<th>Subset number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter/northbound</td>
<td>2728</td>
</tr>
<tr>
<td>Summer/northbound</td>
<td>4040</td>
</tr>
<tr>
<td>Winter/southbound</td>
<td>3675</td>
</tr>
<tr>
<td>Summer/southbound</td>
<td>4795</td>
</tr>
</tbody>
</table>

B. Filtering out uneconomic routes

To minimize the uncertainty caused by anomaly tracks and filter out the dirty routes objects, we formulated a geographic cost function to evaluate route object’s economic performance. Briefly, for clarity, we suppose ships’ sailing speeds on open waters are almost the same. It implies that the further a ship travels on open waters, the more time it will spend. From the perspective of GISci, given a specified route of OD, the unique Least Cost Path (LCP) can be theoretically generated using the corresponding OD points. There are various numerical cell-based methods used to calculate LCP constrained by bathymetric and topographic obstacles [17, 37-39].

Fig.3 Theoretical LCP cases in the western Taiwan Strait

Given the minimum water depth, Dep_{min}, we can establish a Transition Matrix with $n^2 \times n^2$ elements (for the case of Taiwan Strait, n=300) containing the probability of transition from one cell of the Bathymetric Grid to adjacent cells [37]. The running time complexity of generating the Transition Matrix is $o(n^2)$, and a reasonable value of n to define Bathymetric Grid should be carefully selected. Based on a specific Transition Matrix, the shortest great-circle distance between pairs of OD sites, avoiding bathymetric and topographic obstacles, can be calculated, which is noted as $D_{LCP}$. For the case of the Taiwan Strait, Fig.3 depicts the corresponding theoretical LCP of typical routes predefined by certain OD under the condition of water-depth threshold (i.e. $Dep_{min}=3m$ and $Dep_{max}=20m$). Theoretical LCP research in the northern part of the study aligns with the real ship traffic density map (Fig. 1) better than the southern the other part does. This discrepancy occurs because seamen prefer the route with less turning waypoints. In this paper, the theoretical LCP with a 20-meter water-depth threshold is used as the benchmark for evaluating the geographic cost of route, because this research is intend to identify strait corridors in open water.

A simple geographic cost function of route can be developed as follows:

$$Cost = \frac{D_{traj}}{D_{LCP}} \tag{1}$$

where $D_{traj}$ is the real length of the specified route $D_{LCP}$ defined by the corresponding OD sites.

Because there are many transits of ships sharing the same OD elements from Transition Matrix, the efficiency of algorithm for computing Formula (1) can be improved by setting a cache of the adjacency matrix of OD. By calculating the Cost of each route object, the uneconomic routes can be identified by tuning a reasonable threshold of Cost.

The uneconomic routes are filtered using Formula (1). Table.2 shows the statistics for the 15,200 transit samples' Geographic Cost to highlight the economic performance of routes in different seasons: Firstly, the Standard Deviation (SD) of northbound traffic was larger than that of southbound traffic for the dispersion of geo-cost for opposing streams of traffic; Secondly, the economic performance of northbound traffic during the winter was the worst. These variables demonstrate that a seasonal effect is significant for opposing streams of traffic. Such a seasonal effect may be attributed to the fact that the northeast monsoon during winter is stronger than that of the southwest monsoon in summer.

Table. 2 DISTRIBUTION OF TRANSIT SAMPLES’ GEO-COST GROUPED, BY DIRECTION AND SEASON

<table>
<thead>
<tr>
<th>Season/Direction</th>
<th>1st Quartile</th>
<th>2nd Quartile</th>
<th>3rd Quartile</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter/northbound</td>
<td>1.001</td>
<td>1.006</td>
<td>1.012</td>
<td>0.012</td>
</tr>
<tr>
<td>Summer/northbound</td>
<td>0.997</td>
<td>1.002</td>
<td>1.008</td>
<td>0.012</td>
</tr>
<tr>
<td>Winter/southbound</td>
<td>0.997</td>
<td>1.001</td>
<td>1.006</td>
<td>0.008</td>
</tr>
<tr>
<td>Summer/southbound</td>
<td>0.997</td>
<td>1.002</td>
<td>1.008</td>
<td>0.008</td>
</tr>
</tbody>
</table>

To examine route samples ranked by geo-cost, we used winter values and conducted a spatial correlation analysis of seasonal differences and similarities with ArcGIS software. The analysis resulted in the division of the spatial features of route subsets by quartile from the ranked set of geo-cost. Fig. 4 shows selected route objects with respect to the top and bottom 5% of geo-cost.

The worst economic subset with sharp turns or twists is in the set of geo-costs of the 3rd quartile or above. Most of the turns or twists from the subset occurred in the northern region of the research area, as shown on Fig 4. Thus, an uneconomic route subset should be filtered out. The other subsets depicted in Fig.5 were selected as clean routes for delineating the principal fairways.
C. KDE estimator for the retrieval of SUD

Based on the lined-based kernel density estimator for the retrieval of UD from animal movement tracks [40], we applied Kernel Density Estimation (KDE) to estimate the SUD of transit-passage routes. The KDE method begins by centring a bivariate probability function with unit volume, namely a kernel, over each route. A regular grid with a user-defined cell size is then superimposed on route samples. Each line segment \( l_{i,i+1} \) (where \( i=1 \ldots n \)) of route samples is rasterised individually, and the raster \( r_{i,i+1} \) is calculated using the KDE. A probability density estimate is calculated at each grid-cell by summing the overlapping volumes. A two-dimensional probability surface over the entire grid is then generated. Thus the retrieval of SUD is represented by a grid of regular cells so as to measure the hard boundary delineating the edge of polygon containing the area used by transit-passage routes. Namely, the isopleths from SUD delineate different portions of Application Nautical Range (ANR) volumes, as shown on Fig 6.

D. Nonlinear generalised regression model for PFs

Similar to the individual-based quantitative model for delineating CA of animal space use [28, 30], an exponential curve of SUD could be drafted by a statistical inference approach (i.e. Nonlinear generalised regression model), as shown on the left side of Fig. 6. The SUD area is the coverage of the involved ships’ Utilization Distribution, shown on the right side of Fig. 6. In other words, given a certain IV of \( p \) from SUD, there is a Specific Nautical Range (SNR) representing a unique portion of ANR. So the IV is indicative of the likelihood of the cumulative sailing time for all involved ships in ANR: the greater the volume, the more cumulative sailing time spent. The curve in the left side Fig. 6 is to interpret the relationship between the IV (x-axe) and the area of its corresponding SNR. For the sake of congruent axes, the area of SNR is standardised proportional to the total area covered by the Application Nautical Range (ANR) and displayed as a percentage, which is called Percent SNR Area. Especially, the IV of \( p=0.95 \) is selected to delineate the ANR. The ratio is obtained as below:

\[
R_{psa} = \frac{S_p}{S_{0.95}}
\]

where \( R_{psa} \) is the Percent SNR Area, \( S_{0.95} \) is the area of ANR and \( S_p \) is the area of SNR.

Because SUD is estimated by KDE, the relationship between \( p \) and \( R_{psa} \) fits an Exponential Regression equation as below:

\[
R_{psa} = a \cdot e^{b \cdot p} + c
\]

where \( a \) is the y-intercept, \( b \) is the constant of the Exponential Regression equation, and \( c \) is the surplus constant. \( a, b \) and \( c \) can be calculated using statistical inference procedures.

Vander Wal el al.[30] revealed that there is a unique point at which the slope is equal to 1 that can be found in the monotonically increasing and asymptotically fitted ER curve by differentiation. The superficial mathematical meaning of the unique point is that \( R_{psa} \) begins to increase more rapidly than IV. Additionally, there is an underlying nautical significance that the involved ships’ sailing time spent in the SNR are maximised relative to their movement on ANR. Therefore, the explicit empirical threshold for delineating PFs can be obtained as below:

Calculate the first derivative of the ER curve and assign the slope of the curve equation a value of 1:

\[
\frac{\partial R_{psa}}{\partial p} = a \cdot b \cdot e^{b \cdot p} = 1
\]

A unique \( P_{pf} \), which determines the contour of DR, can be solved as follows:

\[
P_{pf} = \frac{-\ln(a \cdot b)}{b}
\]

The more closely the result approaches zero, the stronger the spatial heterogeneity of the related sample becomes, which means the corresponding PF is used more heavily.
A nonlinear generalized regression method to delineate PFs

Fig. 7 Economic voyages' PFs in the western Taiwan Strait by direction and season

Table 3: Regression results of the extracted PFs in the western Taiwan Strait (grouped by season and direction)

<table>
<thead>
<tr>
<th>Season/direction</th>
<th>(a)</th>
<th>(b)</th>
<th>(P_{\rho} = -\ln(a*b)/b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(2.5%</td>
<td>97.50%</td>
<td>LCE)</td>
</tr>
<tr>
<td>Winter/Southbound</td>
<td>0.019</td>
<td>0.024</td>
<td>0.022</td>
</tr>
<tr>
<td>Winter/Northbound</td>
<td>0.093</td>
<td>0.105</td>
<td>0.099</td>
</tr>
<tr>
<td>Summer/Southbound</td>
<td>0.038</td>
<td>0.044</td>
<td>0.041</td>
</tr>
<tr>
<td>Summer/Northbound</td>
<td>0.044</td>
<td>0.051</td>
<td>0.048</td>
</tr>
</tbody>
</table>
IV. APPLICATION

A. Examining the directional and seasonal factor for Taiwan Strait’s PFs

In order to highlight the seasonal variation between opposing streams of traffic in the Strait, the PFs of ship traffic are extracted from the cleaned routes with the proposed algorithm. Table 3 shows the confidence intervals for the two parameters (a, b) in the Exponential Regression Model and IV, designating the PFs for each group route subset divided by the direction and season. The results indicate that each subset's IV fits the exponential curve well.

The minimum and maximum IV among the subsets occurred in winter. The PFs of northbound traffic flow in winter shows the strongest convergence, but that of southbound traffic is the most spatially homogeneous. The result shows that the spatial heterogeneity of the PFs is stronger for northbound vessels than southbound ones both in winter and summer. And southbound traffic is more convergent in summer than in winter, whereas northbound traffic shows the opposite trend. To achieve a more accurate assessment of spatiotemporal characteristics for opposing traffic streams. We used ArcGIS to visualise each boundary of the PFs associated with the IV items listed in Table 3. The result shown in Fig. 7, presents the extracted principal fairways crossing the Taiwan Strait in northeast/southwest directions. It shows that an extra strait-corridor in offshore regions is used for northbound traffic in winter besides the other commonly used sea-lanes.

B. Identifying gaps between the extracted PFs and current routing system plan

![Image](Image)

Fig. 8 Current ship routing system plan issued by CMSA in the western Taiwan Strait

To Identify gaps between the extracted PFs and current routing system plan, a comparison is made with the China’s Coastal Ships Routing System Plan (CCSRSP) issued by CMSA [41], which involves recommended coastal routes in the western Taiwan Strait, Precautionary Areas (PA) and TSSs on open sea, as shown on Fig. 8. Recommended coastal routes of CCSRSP are intended for handy-size vessels conducting short sea shipping trips. Precautionary areas (PAs) are specific zones where ships must keep cautious because of frequent crossing traffic. Opposing streams of traffic using open sea routes are managed and controlled by the TSS, which is generally composed of a separation zone, borderline and traffic lane. Because there is an extra corridor in offshore regions derived from the northbound-winter PFs, the layer of CCSRSP is then superimposed on the PFs as shown on Fig. 8.

Fig. 8 shows the following two gaps between cumulative activity patterns of transit passage ships and current NCSRSP. The first is that the lines linking current No.2-3 PAs and corresponding TSSs (i.e. No.1-3 TSSs) deviate from the axis-lines of the extracted PFs. A concerted effort should be taken to further refined routing system of the strait. The second gap lies in that existing recommended coastal routes wriggle their way in an ‘S’ shape. There are so many sharp turns or twists that they are unsuitable for merchant ships to pass the strait efficiently. This indicates that new routes should be recommended to supplement the current CCSRSP and some PAs.

V. CONCLUSION

Methods of HR or CA are routinely applied to analyse movement of wildlife, while the techniques are equally applicable for ship movement tracks. Although the idea of the nonlinear generalised regression approach for delineating core areas of animal space use was already introduced [30], and the line-based KDE for the retrieval of UD from animal movement tracks was proposed [40], the quantitative approach for delineating PFs presented in this manuscript makes a number of potentially important contributions in the field of maritime engineering.

First, the nonlinear generalised regression model for delineating PFs provides a rigorous mathematical framework for accurately modelling maritime corridors than cluster-based counterpart [22-26], as it increases understanding PFs boundaries; In the case of the Taiwan Strait, this research shows a superior ability to delineate strait corridors commonly used by transit-passage merchant ships. Meanwhile, when coupled with current routing system plan, PFs boundaries can be useful to identify gaps between current TSSs and cumulative activity patterns for vessels group derived from real AIS tracks.

Second, the line-based KDE method for geo-visualising changes of shipping space-use patterns over time is also useful. Not only is it easy to implement in existing GIS software packages, but geo-visualising shipping space-use patterns under different conditions are also helpful to view the effect of seasonal and directional factors for maritime corridors. There is no information about the temporal dimension of strait corridors and space-use patterns for traditional TSS analysis [3, 9, 42]. While our proposed method is useful not only for mapping the total area over the whole observing period, but also for exploring the changes of most intense regions used by involved vessels over a broad time period such as season.
References


METHODOLOGY OF CREATION SIMULATION BASIN BASED ON THE CHANNEL THROUGH THE VISTULA SPiT

Piotr Zwolan
Polish Naval Academy
Institute of Navigation and Hydrography
Gdynia, Poland
p.zwolan@amw.gdynia.pl

Krzysztof Czaplewski
Gdynia Maritime University
Department of Geodesy and Oceanography
Gdynia, Poland
krzysztof@czaplewski.wp.pl

The biggest problem in the process of implementation of the new sea areas project or aids to navigation systems is to check the assumptions without compromising security on real waters. Today, digital models are available for easy and inexpensive replacement of the research methods used so far. For this purpose the navigational and maneuvering simulators are perfect.

Simulators as a research tools allow you to define a more realistic and accurate data for the study and evaluation sea areas. The purpose of the simulation is to identify and reduce the risk for seafarers during navigation on the waterways, canals and port areas. These include quantitative and qualitative assessment of the channels and fairways construction. The main requirement for a simulation system is to have a multi-task simulation software, which includes effective tools for repositioning and designing safe waterways and port infrastructure.

The Institute of Navigation and Hydrography Polish Naval Academy since the eighties of the twentieth century is carried out projects in the field of navigational - hydrographic support of human activities at sea. This is possible thanks to the extensive set of simulation tools. Planning this tasks in a simulators software environment allow not only to design of virtual counterparts of real sea areas, but also to evaluate the quality of manufactured parts before their actual exposure to sea areas.

The paper presents the possibility of simulators on the example of the navigable channel construction project through the Vistula Spit, which is planned for implementation in Polish maritime areas. The paper presents the algorithm for the basin creation and maneuvering test for maximum vessel provided in the newly formed basin.

Keywords—navigation, simulation

I. INTRODUCTION

Due to the current political situation, access to the Vistula Lagoon is limited as it is either periodically closed or European Union vessels are not allowed to enter its waters from the Baltic Sea through the Pilawa Strait. This situation compels the Polish government to work on delineating a new, alternative international waterway that would connect the Baltic Sea with the port of Elblag and smaller ports that are located on the Vistula Lagoon. The project is primarily of economic significance since the creation of a navigable canal that would go through the Vistula Spît is a prerequisite for a more effective exploitation of the potential of the port of Elblag, which was modernised in 2003, and for a significant increase in the lagoon’s importance to tourist and recreational sailing. Four alternative locations where a cut could be made through the spit were suggested, i.e. near the following localities: Skowronki, Nowy Świat, Przebrno, and Piaski. Then three different designs of the cut were proposed; these variants mainly differ in terms of the arrangement of wharves and the location of the entrance to the cut. This paper shows how the navigational limitations of the cut that is planned be made through the Vistula Spît can be assessed by using a navigation and manoeuvring simulator. The study was conducted based on analysing the behaviour of a vessel with the maximum dimensions for manoeuvring in the canal with the aim of identifying the best design variant. A navigation and manoeuvring simulator produced by Transas Marine, which is used by the Polish Naval Academy, served as a research tool. A wide range of vessel models, with different manoeuvring characteristics, are available in this simulator. The Navi-Trainer 5000 software, based on which the simulator operates, also includes the Model Wizard, a technologically advanced application which makes it possible to create new bodies of water for training and edit the existing ones [1]. By using this application, a training water body that encompassed a cut through the Vistula Spît was created in accordance with the data that had been obtained from the Maritime Office in Gdynia. Then, manoeuvring tests were conducted in different hydrometeorological conditions for the three variants of the entrance to the cut. Due to the diverse skills and experience of the research team’s members, the simulation experiments were reliable and demonstrated any navigational dangers associated with the cut.
II. GENERAL CHARACTERISTICS OF THE DESIGN OF THE CUT THROUGH THE VISTULA SPIT

There are four alternative locations of the navigable canal that is planned to be constructed through the Vistula Spit, from which the one that will be the best in economic and environmental terms will be chosen. Experts decided that the best locations for the project will be the following:

- ALTERNATIVE 1 – Skowronki in the Sztutowo Commune;
- ALTERNATIVE 2 – Nowy Świat in the Sztutowo Commune;
- ALTERNATIVE 3 – Przebrno in the Krynica Morska Commune;
- ALTERNATIVE 4 – Piaski in the Krynica Morska Commune.

Regardless of which of the alternatives will be selected, the newly built body of water will be equipped, for example, with a canal lock, storm gates, a berth, two drawbridges or a permanent bridge (about 17 m high) or a tunnel (13 m deep), and other facilities, with the following approximate parameters:

- a length, depending on the location, of: 1150 m – "Skowronki", 1260 m – "Nowy Świat", 1650 m – "Przebrno" and 750 m – "Piaski";
- a basic width of 60 m, which will be increased to 100 m at a distance of 200 m from the beginning of the canal;
- a depth of about 5 m;
- a cubic volume of the cut, depending on the location, of: 1.26 million m³ – "Skowronki", 1.74 million m³ – "Nowy Świat", 1.96 million m³ – "Przebrno" and 0.79 million m³ – "Piaski";
- a canal lock with the following dimensions: a length of about 200 m, a width of about 25 m and a depth of about 5 m.

A general design concept of the canal was adopted based on the “Skowronki” alternative, which is presented in Fig 2. It includes two alternative entrances, i.e. on the west and east sides of the canal. In the variant that features two drawbridges, which are located above the lock heads, one of the bridges would always be passable [2].

The following maximum dimensions of vessels that could safely manoeuvre in the cut were adopted:

- a length of up to 100 m;
- a width of up to 20 m;
- a draught of up to 4 m.

III. PROCESS OF CREATING A BODY OF WATER IN A SIMULATED ENVIRONMENT

The process of creating a water body in the environment of the Navi-Trainer Pro 5000 simulator, which involves using the specialist software Model Wizard [3], is complex and time-consuming. This process requires preparing a large amount of essential data and knowing how to use specialist software. The time that is needed for this depends on the accessibility of information sources, the shoreline’s complexity and the number of additional objects that are necessary for creating a given body of water. This process occurred in accordance with the algorithm below.

- Selecting an S-57 ENC cell or raster chart showing a particular area

In order to generate a water body with a high degree of detail and position it in space as accurately as possible, one must use official charts. An S-57 ENC p4map41 cell, which had been created by the Navy Hydrographic Office, was used as a basis for developing a body of water representing a cut through the Vistula Spit. Satellite images that have been
superimposed on a chart can be used to verify the correctness of the simulated shoreline and compare it against the actual shoreline. Having selected the appropriate chart cell, one has to import it into the program. When an S-57 cell has been implemented, it becomes transformed into the TX97 format, which is used by the Model Wizard software. While changing formats, the program automatically imports the existing shoreline together with the navigation marks, bathymetric characteristics and heights.

- Editing the existing shoreline and creating a new one

After the basic shoreline has been generated, the editing process begins. During the transformation of the shoreline, parameters such as altitude, type of shoreline (e.g. a shoreline designed for mooring) and construction material used on the shoreline (e.g. sand and stones) are edited.

Apart from correcting the existing shoreline, it is also possible to create a new one based on construction plans in the form of a raster chart or any kind of bitmap in order to precisely draw a given area. Bitmaps and the dimensions that had been used in a feasibility study were employed to create three variants of a virtual model of the entrance to the cut. Accurate mapping of the shoreline will allow one to later generate a cell of an electronic navigational chart which will show the specific area.

- Editing bathymetric information

Since it is possible that the depth data that are shown on the charts might not be up-to-date, it may be necessary to correct them. In order to do this, one must change the shape of the isobaths or add new ones. One can also map particular depth points and altitudes onto the chart. When working on the cut’s model, it was necessary to create a new depth grid for the selected area. After all the essential corrections to the bathymetric data have been made, one should verify their accuracy. The Model Wizard program has a function that allows one to generate a preview of a given area’s topography (the Sensor Viewer function). When analysing a 3D visualisation that was generated by the Sensor Viewer, one can notice random individual points representing wrong depth values. The figure below shows a three-dimensional model of the bottom of the cut.

- Editing the terrain layer

The area of the Vistula Spit that is located in close proximity to the cut does not have very diverse surface features as it is mostly covered by trees, grass and sand. One can use a library of objects that are included in the program to generate a land cover grid. The flat layers that are adjacent to the canal and the entrance to the port are created by superimposing the appropriate layers which are located at different heights above the water surface. These layers were covered by materials with various structures.
Creating a navigational infrastructure with navigation marks

After generating the topography and accurately mapping the shoreline together with the port basin and navigable canal, one can begin the process of creating the navigational infrastructure. Certain navigation marks and elements of port infrastructure that are included in the program’s library of objects were used to create the body of water representing the cut through the Vistula Spit, along with newly generated objects. A canal lock is an example of the latter. In order to generate a lock, one must additionally use a program for creating 3D objects. It is essential that an object that has been created have the .3do extension since only files with this extension can be imported to the Model Wizard. Among these programs are AutoCad, 3D Studio Max and the extended version of Google SketchUp. The process of designing and creating 3D objects can be divided into four basic stages, which are presented below.

**STAGE 1 – collecting data about the objects that are being created**

This is the preliminary stage of generating 3D objects, which involves developing a database of photographic data and completing all the necessary information about the structures of the objects that are being created. Each of the elements is very important because when properly made, photographs that are to serve as textures allow one to realistically represent an object in a simulated environment. However, it is the data on the actual dimensions of particular elements of the navigational infrastructure and on their spatial location that are the most important. It was necessary to develop a model of the canal lock and the bridges, which were not available in the program’s database of objects, in order to create a virtual model of the cut.

**STAGE 2 – creating a grid, i.e. the structure of an object**

After collecting all the data that are necessary to generate the elements of the navigational infrastructure, one can proceed to the next stage, which involves creating a grid. Previously prepared textures will later be placed on this grid. The figure below shows an example of the canal lock’s structure.

**STAGE 3 – connecting the points (closing the shape) and filling the area**

The third stage involves connecting all of the loose lines and points of the grid and verifying the connectivity between all the planes that form a closed solid.

**STAGE 4 – superimposing a texture on the object that has been created**

At the last stage of the process of generating an object, a texture is superimposed on that object’s surface area. When preparing the photographs or graphics that are intended to serve as a texture, one should remember to select the appropriate resolutions and sizes. A photograph must provide an en face image of an object so as to avoid distortions.

**Importing the generated 3D objects and selecting objects from the program’s database**

The next stage of creating a training body of water involves placing the last layer, which is constituted by 3D objects. New elements should be imported to the Model Wizard program and the existing ones are selected from the library of objects. A satellite image of the area can be used as a background in order to properly locate the real-life objects. A canal lock and bridges are the new structures which were generated for the purposes of the cut and positioned in accordance with the design.
Verifying the model's coherence by generating a 3D visualisation

After completing the process of designing a body of water, one can verify its correctness. In order to do this, one should use the Visual Viewer function, which generates a 3D preview of the body water that has been created and makes it possible to preliminarily look at the designed landscape from any point, height and angle of view. This function allows one to generate the appropriate hydrometeorological conditions and find out what influence they have on the area in question. It is also possible to see what the situation looks like at night in order to check whether the navigation lights function properly.

IV. SIMULATION TESTS

In order to assess a body of water that is being designed in navigational terms, one should carry out a range of simulation tests under changeable hydrometeorological conditions. These tests were conducted for all of the proposed design variants.

Simulations of a vessel going through the canal were carried out for winds coming from the four main directions: the north, south, west and east. The wind’s speed corresponded to forces 4 and 6 on the Beaufort scale. After the simulation, points at which the risk of collision was the highest were determined and then the influence of a given wind force on the likelihood of a collision was described. Also, the threshold value of wind force was determined, at which it would be impossible to safely manoeuvre a vessel. Additionally, tests were conducted concerning a vessel passing through the lock after being unmoored from a dolphin. The analysed variants only differed in terms of the location of the entrance to the cut, and apart from that, the structure of the canal was identical for all the alternatives. Therefore, there was no need to simulate a vessel’s passage through the canal lock for each variant separately.

The model of a vessel that is used for research purposes is equipped both with two propellers and a bow thruster. Using torques and a thruster makes it possible to significantly reduce the turning circle diameter and even the turning diameter for a vessel that turns on the spot.

However, it should be assumed that not all vessels are equipped in this way, which is why neither a thruster nor torques were used during tests. The tests were conducted by navy officers with different levels of professional experience.

The figure below presents partial results for one instance of a vessel’s passage for the design variant that is shown in Figure 12. This structure allows a vessel to enter the cut from the east.

Analysis of the rudder blade deflection shows that the rudder must be deflected to the left at the maximum angle when a vessel is executing a turning manoeuvre, even under perfect meteorological conditions. A significant decrease in a vessel’s speed during a turning manoeuvre is another important parameter, which can limit that vessel’s manoeuvring capabilities if it has reached its minimum manoeuvring speed. When analysing a vessel’s wake while it is moving under...
neutral conditions, one can notice that when passing the heads, a vessel goes much closer to the port head of the entrance. This is because it is forced to begin a turning manoeuvre, which ends right before the entrance to the canal, faster. In order to set a vessel on a course to enter the lock, one must manoeuvre in the canal, which proves that there may be a significant risk of collision in adverse meteorological conditions.

The results of the simulation tests allow one to draw conclusions which make it possible to select the best variant from among those that are analysed as regards navigational safety. The table below provides a simplified presentation of the test results for both variant entrances.

**TABLE I. RESULTS OF THE TESTS FOR THE FIRST VARIANT (SOURCE: OWN WORK).**

<table>
<thead>
<tr>
<th>Wind</th>
<th>North</th>
<th>South</th>
<th>East</th>
<th>West</th>
</tr>
</thead>
<tbody>
<tr>
<td>4°B</td>
<td>successful</td>
<td>successful</td>
<td>successful</td>
<td>successful</td>
</tr>
<tr>
<td>6°B</td>
<td>successful</td>
<td>successful</td>
<td>successful</td>
<td>failed</td>
</tr>
</tbody>
</table>

**TABLE II. RESULTS OF THE TESTS FOR THE SECOND VARIANT (SOURCE: OWN WORK).**

<table>
<thead>
<tr>
<th>Wind</th>
<th>North</th>
<th>South</th>
<th>East</th>
<th>West</th>
</tr>
</thead>
<tbody>
<tr>
<td>4°B</td>
<td>successful</td>
<td>successful</td>
<td>successful</td>
<td>successful</td>
</tr>
<tr>
<td>6°B</td>
<td>failed</td>
<td>successful</td>
<td>failed</td>
<td>failed</td>
</tr>
</tbody>
</table>

Analysis of the tables that summarise the results of particular simulations shows that the first design variant failed one manoeuvring test, whereas the second design variant failed as many as three tests. This demonstrates that the first arrangement of wharves makes it possible to enter the area of the cut under much more difficult weather conditions. However, the failed attempt to enter the cut with a west wind of force 6 on the Beaufort scale shows that this structure is not perfect and requires certain adjustments. The first variant has a considerable advantage over the second one in terms of the distance that a vessel must travel while executing a turning manoeuvre. The need to make a turn of 20° degrees more makes it much more difficult to perform a manoeuvre when considering wharves with dimensions similar to the given ones. This is especially visible when the wind is strong. Three of the four tests that were conducted for the first variant were passed, while only one manoeuvre was successful when the second variant was used.

**CONCLUSIONS**

The tests that were carried out for the purpose of this study were aimed at collecting data which would allow one to select the more advantageous design variant or which would be a part of the process of making this choice. Therefore, the conclusions that can be drawn from this study are directed solely towards assessing these variants with regard to the influence of the wind’s force and direction on navigational safety.

After examining both variants, one can state that the aim of this study has been achieved as it is possible to identify the better design alternative without major problems by analysing the simulation tests. The design with an east side entrance proved to be better. However, this structure is not perfect, and one of the conclusions that have been drawn from the tests is that it is necessary to move the entrance to the canal to the right. This will eliminate the negative influence from an east wind, which pushes a vessel and thus increases the turning circle radius. It can also be stated that this change will eliminate the frequent error that results from beginning a turning manoeuvre too late. These tests showed that this error is one of the most often repeated errors that are made during manoeuvres when they are performed by a human being.

The results allow one to state that the simulator that is used by the Polish Naval Academy is a good tool which can aid in preparing and verifying marine traffic engineering designs. Since being economical is very important these days, simulators, which are relatively inexpensive to use, have an additional advantage over real-world experiments.

**REFERENCES**

Ship monitoring and location estimation based on spaceborne GNSS reflections

*Wei Liu, Shengzheng Wang, Jing Peng, Chao Jian Shi, Bao Jia Xiao
Merchant Marine College
Shanghai Maritime University
Shanghai, China
*E-mail: liuwei@navisagere.com

Abstract—This paper studies the ship detection and location estimation method based on space-borne GNSS reflections. Firstly, the applicability of the GNSS-R concept to ship detection and positioning is discussed. A rapid prediction approach of GNSS-R Specular Reflection Point will be presented for ship detection based on spaceborne GNSS-R receiver. Experiments have been performed using the proposed rapid prediction approach. The positioning algorithms based on GNSS-R signal time delay and Doppler shift information will be proposed. Two positioning constellation configurations are discussed, including three GNSS satellites and a LEO satellite, and three LEO satellites and a GNSS satellite.

Keywords—Ship monitoring, GNSS-R, Location estimation

I. INTRODUCTION

Nowadays, there is an increasing interest in remote sensing applications which can be helpful to maritime monitoring from space including traffic surveillance, maritime security, fisheries and border control. Reflectometry of opportunity signals such as Global Navigation Satellite System (GNSS-R) was originally proposed for mesoscale altimetry, but it has yielded many promising results: from altimetry applications, to soil moisture determination, ice characterization, or sea state retrieval. Experiments have shown that L-band Global Positioning System (GPS) signals scattered from the ocean surface can be detected by aircraft and space-borne [1-3]. In recent years, ship detection based on GNSS-R was also proposed for maritime monitoring [4].

This paper studies the ship detection and location estimation method based on space-borne GNSS reflections. Firstly, we discuss the applicability of the GNSS-R concept to ship detection and positioning. A rapid prediction approach of GNSS-R Specular Reflection Point (SP) is presented for ship detection based on spaceborne GNSS-R receiver. Experiments have been performed. The positioning algorithm based on GNSS-R signal time delay and Doppler shift information is proposed. Two positioning constellation configurations are discussed including three GNSS satellites and a LEO satellite, and three LEO satellites and a GNSS satellite.

II. SHIP DETECTION METHOD

A theoretical model that describes the power of a scattered GNSS signal as a function of geometrical and environmental parameters has been developed by Valery U. Zavorotny and Alexander G. Voronovich (Z-V) [5]. This proposed model is based on a bistatic radar equation derived using the geometric optics limit of the Kirchhoff approximation. The most complete measurement of a GNSS-R instrument in the spaceborne receiver is the Delay Doppler Map (DDM). Conventionally, Delay Doppler Maps depict the power distribution of the signals scattered from the glistening zone in the DD domain using the GNSS Z-V scattering model, and it is expressed as [5]:

\[
\left| Y(\Delta \tau, \Delta f) \right|^2 = \frac{2^3 T^2}{(4 \pi)^2} \frac{P G}{R_g^2 R_s^2} \chi^2 (\Delta \tau, \Delta f) \sigma_d dS
\]

where \(T\) is the coherent integration time, \(P G\) is the transmitter power, \(G\) is the receiver antenna gain pattern, \(R_g\) and \(R_s\) are distances between the nominal specular point and the transmitter/receiver, respectively. \(\Delta \tau = \tau (\hat{s}) - \tau\) is the relative delay, in which \(\tau\) and \(\tau (\hat{s})\) are the GNSS code delay of specular point (SP) and the observed surface point, respectively. \(\Delta f = f_d (\hat{s}) - f_d\) is the relative Doppler shift where \(f_d\) and \(f_d (\hat{s})\) are the Doppler frequency of the SP and
the observed surface point, respectively. $\sigma_0$ is the normalized bistatic radar cross section (BRCS) of the rough surface. $\chi^2(\Delta r, \Delta f)$ is the Woodward Ambiguity Function (WAF), which describes the range and Doppler selectivity of the coherent radar [6].

The scattering coefficient $\sigma_0$ is expressed as

$$\sigma_0 = \pi \left| \mathbf{p} \right|^2 \left( \left| \frac{\mathbf{q}}{q_z} \right| \right)^4 p\left( \frac{-\hat{q}_z}{q_z} \right)$$  \hspace{1cm} (2)

where $\left| \mathbf{p} \right|^2$ is the Fresnel reflection coefficient, determined by the polarization, the complex dielectric constant of sea water, and the local elevation angle [7]. The scattering vector $\mathbf{q}$ can be obtained with the locations of the transmitter, receiver and corresponding surface point. $\frac{-\hat{q}_z}{q_z}$ is the ocean surface slope.

$P\left( \frac{-\hat{q}_z}{q_z} \right)$ is the probability density function (PDF) of the sea surface slope.

The WAF in the expression (1) can be approximated by the square product of two functions: the triangularly shaped (the correlation function $\Lambda(\Delta r)$) and the sinc-shaped function $S(\Delta f)$:

$$WAF = \chi^2(\Delta r, \Delta f) = \Lambda^2(\Delta r)|S(\Delta f)|^2$$  \hspace{1cm} (3)

where $\Lambda(\Delta r)$ is the GNSS PRN code correlation function defined as $\Lambda(\Delta r) = 1 - |\Lambda(\Delta r)|/r_c$ if $|\Lambda(\Delta r)| \leq r_c$ and zero elsewhere, with $r_c$ being the length of a chip of the GPS C/A code defined as 1ms/1023. $S(\Delta f)$ is the sinc-shaped function defined as $S(\Delta f) = \sin(\pi \Delta f) / \pi \Delta f$.

Fig.1 shows an idealized case of the space-borne receiver flying at 600-km altitude, in the same plane as the GNSS transmitter [8]. Note that the mapping of the $xy$ domain into the DD domain is ambiguous. As shown in Fig.1, each DD point corresponds to two physical $xy$ points. To solve for this ambiguity, some solutions were proposed in [9].

According to the expression (1), the GNSS PRN code delay of specular point (SP) $\tau$ and Doppler frequency of the SP $f_d$ are the required parameters for spaceborne ship detection. Therefore, the prediction of the specular reflection point becomes one of the key issues that has to be addressed for ship detection based on GNSS-R. In this paper, a rapid prediction approach of GNSS-R Specular Reflection Point (SP) will be presented for ship detection based on spaceborne GNSS-R receiver.

Fig.2 shows the geometrical relationship of GNSS-R. $O$ is the geocenter, $R$ is receiver location, $T$ is satellite location, $S$ is the specular reflection point location, $r$ is radius of the earth. $U$ is the mirror point of $R$ respect to $MS$, $C$ is the mirror point of $M$ respect to $RU$. $H_r$ is the height of receiver, $H_s$ is the height of satellite.

In order to find the point on the Earth’s surface that satisfies the three conditions, a function of the unknown specular point location is represented as the signal path magnitude.

$$P(\mathbf{S}) = \left| \mathbf{TS} + \mathbf{RS} \right|$$  \hspace{1cm} (4)

Normally, the receiver location is known from the standard navigation output from the GNSS receiver, and the transmitter location is either calculated during this same navigation solution or afterwards using data from the International GNSS Service. As this equation is non-linear, an iterative method based on an initial guess will be used. In order to minimize this path we first take the partial derivatives of the specular point $\mathbf{S}$ with respect to $x$, $y$ and $z$. The partial derivative with respect to $S_x$ is shown below with the results being identical with respect to $S_y$ and $S_z$ [11].

$$\partial_x P(\mathbf{S}) = \frac{(T_x - S_x)}{\sqrt{(T_x - S_x)^2 + (T_y - S_y)^2 + (T_z - S_z)^2}}$$

It can be noted that the denominators above are the incoming and reflected vector magnitudes, respectively. Simplifying the above equation and expanding to include three dimensions follows as [11],

$$d\mathbf{S} = \partial_{S_x} P(\mathbf{S}) = \frac{\mathbf{TS}}{|\mathbf{TS}|} + \frac{\mathbf{RS}}{|\mathbf{RS}|}$$  \hspace{1cm} (5)
Iterating on $\overline{S}$ will then result in a convergence to the minimum path. The point on the Earth’s surface that satisfies the three conditions listed above is then solved for iteratively, using expressions (7), (8) and (9) below.

$$\overline{S}_{\text{tmp}} = \overline{S} + K \frac{d\overline{S}}{d\overline{S}}$$  \hspace{1cm} (7)

$$K = |RS|$$  \hspace{1cm} (8)

where, $K$ is correction gain to quicken the convergence. From the initial experimental results, when $K = |RS|$, the calculation time is very quickly.

This intermediate value is then converted to a unit vector and scaled by the Earth radius, giving us the new estimate for $\overline{S}$ to be used during the next iteration.

$$\overline{S}_{\text{new}} = r \frac{\overline{S}_{\text{tmp}}}{|\overline{S}_{\text{tmp}}|}$$  \hspace{1cm} (9)

The specular point can be considered found when the difference between the old and new values of $\overline{S}$ falls below a specified tolerance after several iterations. Finally, as a last sanity check that the value of $\overline{S}$ is correct, test the third condition listed above, that Snell’s law is satisfied with respect to the incoming and reflected wave directions.

Experiments have been performed. The radius $r$ of the earth is derived from the WGS84 model for GPS. To test the specular point, incoming wave and reflected wave are calculated in the end, it must satisfy Snell’s law. The results are shown in Fig. 3 which shows location of the specular reflection point.

![Fig. 3. Receiver, transmitter and specular point locations](image)

**III. SHIP POSITIONING ALGORITHM BASED ON GNSS-R**

According to [5], the GNSS code delay refers to the time delay of signals reflected from different grid elements of sea surface as given by

$$\tau(\overline{s}) = \frac{f}{c} (R_c + T_c)$$  \hspace{1cm} (10)

where $R_c$ and $T_c$ are the distances from the grid element of sea surface to the receiver and transmitter, respectively. $f = 1.023 \text{MHz}$ is the frequency of the GNSS code code, e.g GPS C/A code. For the SP, the code delay is

$$\tau = \frac{f}{c} |\overline{p} - \overline{T}| + |\overline{R} - \overline{p}|$$  \hspace{1cm} (11)

The Doppler frequency shift $f_d$ is caused by the relative motion between the transmitter, receiver and grid elements of sea surface. The Doppler frequency shift can be expressed as [5,12]

$$f_d = f_0 \left( \overline{V}_r - \overline{V}_i \right) \hat{m} + \left( \overline{V}_r - \overline{V}_i \right) \cdot \hat{n}$$  \hspace{1cm} (12)

where $\overline{V}_r$, $\overline{V}_i$, and $\overline{V}_i$ represent the velocity of transmitter, receiver and a grid element of sea surface, respectively. $f_0$ is the GNSS carrier frequency, e.g. GPS L1 carrier frequency = 1575.42 MHz. $\hat{n}$ is the unit vector of the scattered wave and $\hat{m}$ is the unit vector of the incident wave.

For the SP, the Doppler frequency shift is

$$f_d = f_0 \left( \overline{V}_r - \overline{V}_i \right) \hat{m} + \left( \overline{V}_r - \overline{V}_i \right) \cdot \hat{n}$$  \hspace{1cm} (13)

For the stationary or slower ship, the velocity of sea surface $\overline{V}_i$ can be assumed as zero since the magnitude of the ship velocity is much smaller than the speed of the transmitter and receiver.

$$f_d = f_0 \left( \overline{V}_r - \overline{V}_i \right) \hat{m} + \left( \overline{V}_r - \overline{V}_i \right) \cdot \hat{n}$$  \hspace{1cm} (14)

Using the expressions (11) and (14), the joint equations for determine the location of ship target $(x,y,0)$ can be written as

$$\tau = \frac{f}{c} \left( \frac{(x_{\text{GNS}} - x)^2 + (y_{\text{GNS}} - y)^2 + (z_{\text{GNS}})^2}{(x_{\text{LEO}} - x)^2 + (y_{\text{LEO}} - y)^2 + (z_{\text{LEO}})^2} \right)$$  \hspace{1cm} (15)

$$f_d = f_0 \left( \overline{V}_r - \overline{V}_i \right) \hat{m} + \left( \overline{V}_r - \overline{V}_i \right) \cdot \hat{n}$$  \hspace{1cm} (16)

According to the expressions (15) and (16), the location and velocity of GNSS and LEO satellites are known, while the ship target is unknown. Note that four satellites (normal navigation) can be used to determine three position dimensions and time. Therefore, two positioning constellation configurations can be used in the space-borne ship detection system. The
configurations consist of three GNSS satellites and a LEO satellite, and three LEO satellites and a GNSS satellite. When using three LEO satellites and a GNSS satellite configuration, we need more LEO satellites. This will increase the cost of the ship detection system. However, Space-based Automatic Identification System (SAT-AIS) is a promising solution to overcome terrestrial coverage limitations with the potential to provide AIS service for any given area on Earth. With the development of SAT-AIS, we can implement the GNSS-R receiver in the SAT-AIS satellite payload.

IV. CONCLUSIONS

In this paper, the ship detection and location estimation method based on space-borne GNSS reflections was presented. Firstly, we discussed the applicability of the GNSS-R concept to ship detection and positioning. Then, a rapid prediction approach of GNSS-R Specular Reflection Point was presented for ship detection based on space-borne GNSS-R receiver. Experiments have been performed. The positioning algorithm based on GNSS-R signal time delay and Doppler shift information was proposed. Two positioning constellation configurations were discussed including three GNSS satellites and a LEO satellite, and three LEO satellites and a GNSS satellite.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No.61304230, 51279099, 51379121), the Specialized Research Fund for the Doctoral Program of Higher Education of China (No.20123121120003), the Innovation Program of Shanghai Municipal Education Commission (No.14YZ106, 14YZ121), the Shanghai Municipal Natural Science Foundation (No.12ZR1444700), the development of science and technology of Shanghai Ocean University(A2-0209-14-200063) and the Shanghai Science and Technology Committee (No.12510501800, 13510501600).

REFERENCES

Session E
Aviation and terrestrial navigation.
Applications of GNSS and other technologies.
Sensors and their data fuse. Special problems.
Advanced-Receiver Autonomous Integrity Monitoring for Horizontal Civil Aviation Applications

Mikael Mabilleau, Jonathan Vuillaume
System Definition and Performance / Navigation Unit
Egis Avia
Toulouse, France

Pr. Igor Nikiforov
Professor
Université de Technologie de Troyes
Troyes, France

Gerhard Berz
Focal Point Navigation Infrastructure
EUROCONTROL
Brussels, Belgium

O.F. Bleeker
Consultant
ADSE
(Retired from Rockwell Collins R&D Director position)
Brussels, Belgium

This paper analyzes ARAIM performance to support civil aviation horizontal only operations subject to different constellation constraints. Results in downgraded/partial GPS and Galileo cases indicate that previous availability limitations can be mitigated when assuming specific probabilities of constellation failure. Guidelines on civil aviation needs in terms of satellite and constellation failure probability commitment can be derived, with the aim to speed up the receiver standardization process. Practical implementation issues are discussed with the idea to limit the complexity and cost of ARAIM introduction.

Keywords—GNSS, ARAIM, Civil Aviation

1. INTRODUCTION

Because receiver autonomous integrity monitoring (RAIM) techniques are based on measurement redundancy, multi-constellation signals will have a positive impact on the integrity monitoring performance of RAIM algorithms. Increasing reliability and resilience of next generation global navigation satellite system (GNSS) services is an agreed objective of the Global Air Navigation Plan agreed by International Civil Aviation Organization (ICAO) member states. Service robustness is also a key performance indicator as introduced in the GPS-Galileo Concept of Operation developed by the EUROCAE WG62. This includes supporting approaches with lateral as well as vertical guidance to achieve an increased level of safety and other operational and environmental benefits. However, current integrity-critical assumptions being used for RAIM based on single frequency GPS no longer hold true in a multi-constellation architecture. Consequently, this paper investigates both the updates required to implement future multi-constellation RAIM algorithms and evaluates associated system implications and performance.

Multi-constellation and multi-frequency technology (MCMF) introduces challenges concerning avionics processing power, complexity of operational modes and the number of core constellations to be used. Furthermore, current investigations on MCMF Advanced RAIM for vertical approach guidance have introduced a ground segment or Integrity Support Message (ISM) which increases cost and complexity and may impact implementation timescales. While the ISM does provide flexibility which has the potential to reduce the cost of future upgrades and facilitate service introduction, it is not sure if and in which form it is required for horizontal guidance and positioning applications. Previous work [1] has evaluated these trade-offs focusing in particular on horizontal positioning and associated implementation issues. This includes an evaluation of operational requirements in light of how positioning performance impacts aircraft guidance quality.

Based on these identified system trade-offs, the achievable level of performance for various navigation and surveillance applications is analyzed in several operational contexts. The paper includes an evaluation of the availability of the targeted accuracy and integrity requirements. Additional analyses have been performed to assess the impact of Galileo and GPS L5 full operational capability (FOC) declaration. The objective of these evaluations is to derive suitable and beneficial performance targets for horizontal advanced RAIM to sustain robust operations in civil aviation. The paper will provide recommendations on how to develop and implement Advanced RAIM to support horizontal positioning applications in multi-constellation, dual frequency receivers in a 2020 timeframe. The analysis performed in the paper will assess civil aviation needs in terms of core constellation performance commitments to sustain robust horizontal only operations. Recommendations for core constellation service providers will be derived considering the latest information made available from current civil aviation MFMC standardization activities.
Section 2 introduces the navigation and surveillance operational applications targeted in this paper and a synthesis of the initial results presented in previous work [1]. The methodology used to perform the simulation will also be described in this section. New operational scenarios will be presented in section 3 with the description of the assumptions retained in the sensitivity analyses. Simulation results are shown in section 4. The results are used in section 5 to address practical implementation issues associated to H-ARAIM for short term benefits and further action for MCMF GNSS receiver standardization activities. A conclusion synthesizes the main outcomes of the paper and suggests directions for further investigations such as the characterization of the minimum detectable bias in a multiple failure environment.

II. HORIZONTAL ARAIM SERVICES

A. Operational targets

ARAIM is being designed to globally provide localizer performance with vertical guidance (LPV) approaches. It also includes a horizontal only guidance function. The development of such techniques aims to overcome limitations of the current RAIM function already standardized for non-precision approach. RAIM algorithms operating using one constellation (i.e. GPS), have been built on the assumption of a single satellite failure, considering a GPS constellation-wide failure to be negligible. Considering the safety criticality of the operation served by GPS + RAIM equipment, those assumptions were acceptable. In a multiple constellation environment supporting more advanced operations, this will not be the case because of the increased complexity and criticality. Especially when considering the number of satellite signals available at the receiver antenna in a multiple constellation environment, the assumption of only a single satellite fault at a given time becomes difficult to justify based on current core constellation standards or observation [2] [3]. Due to this, standardization bodies and civil aviation institutions wish to develop next generation of advanced RAIM algorithms which provide increased robustness without undue complexity for all potential applications of interest, which may or may not include precision approach operations.

Currently proposed ARAIM concepts [4] introduces an ISM component that should provide to the user information from ground observations to help the user algorithm meeting LPV requirements while taking into account the potential evolution in core constellation service quality. The current work on ISM architecture and means of dissemination show that this component may not be fully validated in a 2020 timeframe. Indeed, some questions linked to the certification, feasibility and cost-benefit have been asked and are not trivial to answer. Consequently, in order to define a first step in MCMF GNSS Receivers providing initial ARAIM services in a 2020 timeframe while paving the way for a full implementation with more advanced capabilities, it is of interest to investigate if a solution can be described which does not a dynamically updatable ISM component for horizontal-only based operations.

As described in [1], the main navigation operations of interest for horizontal only ARAIM (H-ARAIM) analysis are LNAV and RNP 0.1 applications. A 50/50 split of the total system error assumption has been used for RNP 0.1 navigation following the EU-US WG-C assumption [5]. While these values have not been standardized in an ICAO navigation specification, they nonetheless represent a useful service target for H-ARAIM. It has been also introduced in [1] that ADS-B mandates need to be considered for H-ARAIM as current GPS equipment without SBAS may not provide 100 % availability compliance. The next table summarized the operational requirements retained for the analyses.

### TABLE I. NAVIGATION POSITION DETERMINATION AND SURVEILLANCE REQUIREMENTS

<table>
<thead>
<tr>
<th></th>
<th>Accuracy (m)</th>
<th>Alert Limit (m)</th>
<th>Integrity Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNAV</td>
<td>220</td>
<td>556</td>
<td>10^{-7}hour</td>
</tr>
<tr>
<td>RNP 0.1</td>
<td>92</td>
<td>184</td>
<td>10^{-7}hour</td>
</tr>
<tr>
<td>European ADS-B mandate</td>
<td>185</td>
<td>1111</td>
<td>10^{-7}hour</td>
</tr>
<tr>
<td>US ADS-B mandate</td>
<td>92</td>
<td>370</td>
<td>10^{-7}hour</td>
</tr>
</tbody>
</table>

B. ARAIM simulation methodology

The ARAIM simulations are using a platform implementing an ARAIM algorithm introduced in [6]. Table 2 provides the settings of the simulation platform configuration selected for this paper. Additional information needs to be provided to complete the operational scenarios such as:

- **The GNSS environment**: It refers to the constellations and signals used as well as the number of satellites being operational in each constellation.

- **The GNSS measurement error model**: The model used to characterize the GNSS nominal error budget. It gathers the contribution of the satellite clock and ephemeris error budget, the troposphere residual error budget, the multipath/receiver noise error budget and possibly the ionosphere residual error budget in single frequency mode simulations. In addition, a nominal bias due to signal deformation and satellite/receiver antenna group delay is also defined.

- **The GNSS threat model**: The description of the faulted cases for the different constellations used in the operational scenarios. It is based on the values assumed for the probability of satellite and constellation failure for each constellation. It also includes the integrity and continuity risk requirements derived from the targeted operation to compute the probability of false alarm and probability of missed detection.

A complete definition of all the elements described above provides an operational scenario that can be tested on the ARAIM simulation platform. As outcomes, maps and statistics of the 99.9% horizontal protection level (HPL) are provided. The characteristics of the HPL are analyzed further in post-processing.
C. Initial H-ARAIM performance analysis

Initial operational scenarios have been described in [1] taking into account the potential nominal and fall back modes of future Dual Frequency Multiple Constellation (DFMC) receivers. They provide the inputs needed by the ARAIM algorithm as defined by the ISM in the ARAIM concept of operation [5], without requiring a particular source for the ISM (e.g., they could be either hard-coded or dynamic, but for a given simulation run they are static). Four categories of scenarios have been characterized:

- Constellation/Satellite geometry impact scenario
- Single Frequency impact scenario
- Constellation/Satellite failure impact scenario
- Satellite clock and ephemeris sensitivity scenario

The results show that ARAIM is robust to the loss of one frequency and can provide 100 % availability for Lateral Navigation (LNAV), RNAV, RNAV operation, and ADS-B mandates based on conservative assumptions for URA, Psat, and Pconst. In nominal mode of operation, based on nominal 24 + 24 satellites constellations, ARAIM can sustain HPL requirements of 20 m. It provides promising technical performance that may accommodate various avionics innovations as the Navigation System Error (NSE) will not be a driver. In case of non-nominal Galileo constellation distribution and with some values of Psat and Pconst, the algorithm is not capable to provide a protection level. An increased number of satellites doesn’t necessarily bring additional performance benefits. It appears that some combinations of Psat/Pconst (10^7/10^5 for instance) together with downgraded constellations may lead to worse operational results than current GPS/RAIM equipment. This specific outcome has led to more in depth analysis of H-ARAIM robustness using downgraded constellation and various Psat/Pconst assumptions.

III. H-ARAIM PERFORMANCE SENSITIVITY ANALYSIS

The next generation of aeronautical receiver standards may still be under development while the Galileo constellation may also take some time to reach the final operation capability of 24 satellites and demonstrate its actual performance. The most recent declaration from the European Commission [8] indicates that Final Operation Capability (FOC) should be reached around 2019. Nevertheless, the sensitivity of H-ARAIM performance with respect to Galileo FOC provides valuable information on H-ARAIM sensitivity toward the Galileo FOC declaration. Recent work [7] provides some arguments justifying a 10-8/h constellation failure for the GPS constellation. Considering that the ARAIM algorithm is supporting operations with a 10-7/h integrity integrity, this assumption on GPS Pconst leads to a simplification through not having to assess subsets with GPS out only, based on probability checks. This new setting for GPS Pconst which is consistent with current RAIM assumptions has been implemented alongside a downgraded Galileo constellation in order to challenge previous results of less than 60 % of availability over the globe for LNAV.

The United States have recently published the 2014 Federal Radio Navigation Plan [9] indicating that the GPS L5 FOC will be postponed to 2024. This delay also questions the timeline of H-ARAIM potential operational introduction as 24 L5 operational GPS may not be available in the beginning of the 2020’s. Further analyses are conducted to assess H-ARAIM sensitivity to the GPS L5 FOC milestone.

Finally, although GPS users are provided with more operational GPS satellites than the committed number provided in the GPS SPS [2], observations of the location of space vehicles within their orbits have indicated that they were not always located in the primary slot but nearby. An operational scenario has been set up with a nominal number of GPS satellites but not all placed in a primary slot in order to assess H-ARAIM sensitivity to the compliance of satellites with their nominal slot.

In summary, the following classes of operational scenarios have been introduced and will be further described in the next sub-section:

- Downgraded or partial Galileo constellation design scenarios
- Downgraded or partial GPS constellation design scenarios
- Non-nominal GPS constellation design scenarios

The nominal GNSS environment used for the assessment of ARAIM performance is a dual frequency multi-constellation system (DFMC) that is expected to be available in a 2020 timeframe. Multi-constellation refers to GPS and Galileo in this article. Multi-frequency refers to L1/L5 frequencies in the GPS case and E1/E5a frequencies in the Galileo case. If there is no mention of a different setting, the baseline number of GPS and Galileo satellites used is 24 each.

It is assumed that future receivers will be capable to provide sufficient channels to be capable to track all satellites in view, even if maybe not all of them will be processed. An “all-in view” approach is adopted with a receiver mask angle of 5° for GPS and Galileo.

A. Downgraded or partial Galileo constellation design

The Galileo constellation specification has recently been changed from a 27 satellites constellation to a 24 satellites constellation [10]. The European commission adopted a step
wise approach for the deployment of the Galileo system leading to 2 major milestones: FOC1 (or IOC) and FOC [10]. FOC1 configuration is defined with 18 satellites:

- Plane A / Slots 2, 3, 5, 8
- Plane B / Slots 2, 3, 4, 5, 6, 8
- Plane C / Slots 1, 2, 3, 4, 5, 6, 7, 8

FOC configuration is based on 24 satellites (every 8 slots in the 3 orbital planes) plus 6 spare satellites. Almanac information on the Keplerian coordinates associated with this updated satellite configuration is provided in [10]. The FOC configuration represents a priori a minimum configuration of the constellation as 2 spares per orbit will be present to replace any faulty satellite.

Using the new hypothesis on GPS Pconst, different assumptions have been retained for Galileo Psat and Pconst. Indeed, current Galileo documentation doesn’t provide such information while standardization activities at RTCA SC 159 and EUROCAE WG62 depend on this information for the further development of civil aviation receiver standards. An objective of this paper is to evaluate different scenarios with a range of value for Psat and Pconst in order to identify civil aviation needs for robust H-ARAIM operation, hopefully leading to useful initial service targets or commitments.

The following scenarios have been retained:

- Simulation 1.1: Optimist Psat Case
  - 18 Galileo
  - PconstGPS / PsatGPS : $10^{-8}$ / $10^{-5}$
  - PconstGAL / PsatGAL : $10^{-3}$ / $10^{-5}$

- Simulation 1.2: Medium Psat Case
  - 18 Galileo
  - PconstGPS / PsatGPS : $10^{-8}$ / $10^{-5}$
  - PconstGAL / PsatGAL : $10^{-3}$ / $10^{-4}$

- Simulation 1.3: Worst Psat case
  - 18 Galileo
  - PconstGPS / PsatGPS : $10^{-8}$ / $10^{-5}$
  - PconstGAL / PsatGAL : $10^{-3}$ / $10^{-3}$

B. Downgraded or partial GPS constellation design

The GPS constellation is simulated according to the GPS SPS [2] leading to a nominal case of 24 satellites or an extended case of 27 satellites. The design of this class of simulation scenarios has been derived from the downgraded Galileo constellation case. An 18 GPS constellation almanac has been created based on the nominal 24 GPS case but removing A1, B2, C3, D4, E1 and F2.

It was assumed that GPS failure probabilities will be in line with current GPS performances leading to a sensitivity analysis for Galileo Psat and Pconst in order to see how results are linked to GPS FOC. The number of Galileo satellites is 24, corresponding to a situation that could exist between 2020 and 2024.

The following scenarios have been retained:

- Simulation 2.1: Optimist Psat Case
  - 18 GPS
  - PconstGPS / PsatGPS : $10^{-8}$ / $10^{-5}$
  - PconstGAL / PsatGAL : $10^{-3}$ / $10^{-8}$

- Simulation 2.2: Medium Psat Case
  - 18 GPS
  - PconstGPS / PsatGPS : $10^{-8}$ / $10^{-5}$
  - PconstGAL / PsatGAL : $10^{-3}$ / $10^{-4}$

- Simulation 2.3: Worst Psat case
  - 18 GPS
  - PconstGPS / PsatGPS : $10^{-8}$ / $10^{-5}$
  - PconstGAL / PsatGAL : $10^{-3}$ / $10^{-3}$

C. Non-nominal GPS constellation design

For this scenario, the GPS satellites in B1, D2 and F2 slots have been placed to slots B1A, D2A and F2A as defined in [2] for the GPS extended configuration. The idea was to keep a constellation of 24 satellites with an unbalanced distribution of satellites in certain orbits.

The following scenarios have been simulated:

- Simulation 3.1: Reference case
  - Nominal 24 GPS constellation

- Simulation 3.2: Unbalanced distribution case
  - 24 GPS unbalanced on 3 orbits

IV. SIMULATION RESULTS AND ANALYSIS

The simulation results are provided in the core article section through the tables below including the 99.9% horizontal protection level (HPL) figure and availability of ARAIM to a user as a function of the operational requirements used in this study. The 99.9% HPL represents the maximum of the 99.9 percentile of the HPL distribution at each user point tested over the globe. It is anticipated that such information can support further discussions of the Flight Technical Error (FTE) / Navigation System Error (NSE) requirement allocation in standards, since this can be seen as a robust NSE achievable with good availability for a given operational scenario.

Maps of the HPL and HPL distribution figures are provided in Annex 2.

The results have been realized based on the baseline simulation settings provided in Annex 1. Changed parameters are indicated in the results tables when appropriate.
A. Downgraded or partial Galileo constellation design analysis

Table 3 presents the results for a partial Galileo scenario. In this scenario, the GPS constellation is set at 24 with well performing assumptions for Psat and Pconst.

The targeted applications are available at 100% for all cases tested. The high level of performance in terms of only limited GPS faults compensates the low number of Galileo satellites and the high fault probabilities used in the settings of the scenario. It can be concluded that a nominal high performing constellation mixed with a constellation with a limited service record could bring operational benefits by overcoming current ABAS/RAIM limitations of operations based GPS L1 signals only. Indeed the drop of performance observed in [1] with an incomplete Galileo constellation is fully mitigated by the optimized Psat case with the settings used for GPS. Even with more pessimistic assumptions on Galileo (the medium and worst case introduced above), the availability is not harmed. Only a slight augmentation of the HPL is observed.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>99.9 % HPL (m)</th>
<th>RNP 0.1</th>
<th>LNAV</th>
<th>US ADS-B</th>
<th>EU ADS-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 GAL + Optimistic Psat Case</td>
<td>26.87</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>18 GAL + Medium Psat Case</td>
<td>27.44</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>18 GAL + Worst Psat Case</td>
<td>27.48</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

B. Downgraded or partial GPS constellation design analysis

The first operational scenarios have high-lightened the interest of the introduction of Galileo for H-ARAIM targeted civil aviation operations even in a short term timeframe with pessimistic assumptions on the number of satellites and the associated fault probabilities. The next results presented in the following table provide additional information on the sensitivity of the GPS L5 calendar considering the same high level hypothesis: robust GPS for Psat and Pconst but with a limited number of satellite for earlier introduction of H-ARAIM services. The Galileo constellation used is based on 24 satellites.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>99.9 % HPL (m)</th>
<th>RNP 0.1</th>
<th>LNAV</th>
<th>US ADS-B</th>
<th>EU ADS-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 GPS + Optimistic Psat Case</td>
<td>Infinite</td>
<td>47.39%</td>
<td>56.65%</td>
<td>53.35%</td>
<td>63.42%</td>
</tr>
<tr>
<td>18 GPS + Medium Psat Case</td>
<td>Infinite</td>
<td>47.35%</td>
<td>56.57%</td>
<td>52.83%</td>
<td>63%</td>
</tr>
<tr>
<td>18 GPS + Worst Psat Case</td>
<td>Infinite</td>
<td>47.27%</td>
<td>56.49%</td>
<td>52.79%</td>
<td>62.86%</td>
</tr>
</tbody>
</table>

A very low level of availability is obtained mainly due to the fact that not enough satellites where available in some areas when the Galileo constellation and one GPS satellite are out in a particular separation solution subset due to their respective Pconst and Psat. Comparing the results with the previous scenario, the same number of satellites is simulated but the constellation with the better failure probability needs to be well populated to sustain a user availability of anything near 99%. Additional simulations have indicated that a 99% availability target can be achieved with 21 GPS satellites. Considering that today GPS is the core constellation with the longest service history and user confidence, GPS will certainly be the most robust constellations among those ones available. The H-ARAIM achievable performance will be therefore linked to the availability of the GPS L1/L5 satellites in the future.

C. Non-nominal GPS constellation analysis

The previous operational scenario has illustrated the importance of a well-populated constellation with a good quality of service in terms of Psat and Pconst. The next table will presents the sensitivity of the H-ARAIM performance with respect to the alignment of those satellites within their orbit.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>99.9 % HPL (m)</th>
<th>RNP 0.1</th>
<th>LNAV</th>
<th>US ADS-B</th>
<th>EU ADS-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal 24 GPS</td>
<td>20.26</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>24 unbalanced GPS</td>
<td>78.07</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

The availability of the targeted operations has not been harmed because of the high level of horizontal alert limit which can be tolerated. Nevertheless, it is of interest to observe the increase of 280% of the 99.9 % HPL which is significant with respect to other simulation results presented in [1] and [11].

V. H-ARAIM PERFORMANCE ANALYSIS AND PRACTICAL IMPLEMENTATION

Initial analyses made in [1] are still valid. The information provided by the Integrity Support Message (ISM) will not drastically change the performance of H-ARAIM except if the Pconst goes below 10−7 for one of the two constellations used. Indeed such modifications will be similar to the scenario in section IV.A providing 100% availability even with a low number of satellites from the second constellation. It seems that some values of Psat and Pconst have the most significant impact on the final user performance but not all. The final architecture and dissemination means of the ISM should also take this into account in order to reduce complexity and cost. This can be done assuming different states of constellation quality of service with a fixed number of possible satellite failure probability sets having no significant impact on user performance. This solution may avoid the requirement to implement a communication channel between the ISM service provider and the users (as has been assumed for ARAIM with vertical guidance, e.g., it may become optional for horizontal ARAIM).
Other parameters have an impact and need to be carefully monitored such as the number of satellites (especially for the high performing constellation such as GPS in this paper) and their slot alignment within the orbit. Those parameters are not particularly monitored or accounted for in the current ISM design but have an impact on user performance and/or metrics such as the protection levels. The recent delay of the L5 FOC may produce a situation where Galileo reaches dual frequency FOC at 24 satellites earlier than GPS. In the case that Galileo cannot ensure or demonstrate probabilities of satellite/constellation in line with the assumptions used for GPS, H-ARAIM users may suffer availability issues until GPS L1/L5 FOC is achieved. The availability of at least 21 GPS L1/L5 satellites appears to be a key milestone for the robustness of H-ARAIM operation.

The results provided in section IV.A also provide minimum operational requirements for civil aviation interests in H-ARAIM to be considered by a new core constellation service provider (CSP). Current discussion during RTCA and EUROCAE meetings indicate that obtaining such formal commitments on Psat and Pconst from the CSP may be difficult, however, such information is required in order to move forward on receiver standards and enable implementation and adoption. For newcomers such as Galileo with very little service history, this paper provides some values that could be of interest for a first useful commitment even if it may not be so challenging from a system perspective. Hopefully a commitment to such less challenging values can provide the required information to start the development of DFMC receiver standards and accelerate implementation which will include H-ARAIM as one of the modes of operation that can be used when other integrity monitoring schemes are not available.

VI. CONCLUSIONS

Additional analyses complement the promising results depicted in [1]. The downgraded/partial constellation cases have indicated that a robust well populated constellation can mitigate the lack of service history of a new constellation which will force the assumption of high failure probabilities due to the limited observation period. E.g., even when assuming such initial high failure probabilities, it will be possible to provide 100% availability of H-ARAIM horizontal service. Such results may accelerate the standardization process as CSP may have a more easily achievable target of initial constellation commitments to meet civil aviation needs to sustain H-ARAIM service. The results have also indicated the sensitivity of H-ARAIM performance toward GPS being the robust constellation in this paper. Any loss of satellite from the nominal constellation may have a direct impact on operations.

It appears that some parameters values are performance drivers such as the $10^{-4}$ to $10^{-5}$ probability of constellation failure or 21+ satellites available within the strong constellation. The sensitivity of H-ARAIM performance toward variation of those parameters is not linear. Some drops in availability have been observed which means that not the entire threat space needs to be covered by the ISM but only a part of it. As an example, the fact that a constellation has a Pconst below $10^{-8}$ is dimensioning the operational availability. However, a Pconst evolving from $10^{-4}$ to $10^{-5}$ is not impacting horizontal applications performance. The ISM can take credit for such results by identifying states of the threat space that have an operational impact. It may simplify the H-ARAIM implementation by relaxing the need for dynamic ISM updates.

The number of satellites, the failure probability of a satellite and the constellation are parameters which have already been identified as drivers for ARAIM user performance. This is one of the reasons why they are included in the ISM message. This paper has also identified that the alignment of the space vehicles within their orbit is also important. No impact on availability has been observed but the significant increase of HPL will be also seen in the vertical domain for applications with more stringent alarm limits. This is a parameter to be monitored as well in the future.

It may be of interest to further characterize in a next step the detection performance of ARAIM algorithm in presence of multiple failures. Specific event analysis has been conducted in [7] but a characterization of the ARAIM minimum detectable bias will be a solid contribution to the on-going availability assessment and performance sensitivity analysis on ARAIM.

ACKNOWLEDGMENT

This work has been supported in part by SESAR project WP2.7 partners Thales and Honeywell, as well as Juan Blanch from Stanford University.

REFERENCES

ANNEX 1 – BASELINE SIMULATION PARAMETERS

The following table provides the baseline simulation parameters settings concerning part of the measurement error model and threat model for the simulation conducted.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constellation</td>
<td>24 + 24</td>
</tr>
<tr>
<td>Signals</td>
<td>L1/L5 and E1/E5a</td>
</tr>
<tr>
<td>URA / URE</td>
<td>1 / 0.5</td>
</tr>
<tr>
<td>SISA / SISE</td>
<td>1.5 / 0.75</td>
</tr>
<tr>
<td>Bnom</td>
<td>0.75</td>
</tr>
<tr>
<td>PnomGAL / PnomGPS</td>
<td>10^5 / 10^-5</td>
</tr>
<tr>
<td>PnomGAL / PnomGPS</td>
<td>10^5 / 10^-5</td>
</tr>
</tbody>
</table>

ANNEX 2 – DETAILED SIMULATION RESULTS

Detailed maps and figures for the simulation described in this paper are provided in the next table.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>HPL Distribution</th>
<th>HPL 99.9% maps (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1.1</td>
<td>18 GAL – 24 GPS Optimist GAL Psat Case</td>
<td><img src="image1" alt="Histogram" /> Scale 0m to 35m  Scale 9.20m to 26.87m</td>
</tr>
<tr>
<td>Scenario 1.2</td>
<td>18 GAL – 24 GPS Medium GAL Psat Case</td>
<td><img src="image2" alt="Histogram" /> Scale 0m to 35m  Scale 9.35m to 27.43m</td>
</tr>
<tr>
<td>Scenario 1.3</td>
<td>18 GAL – 24 GPS Worst GAL Psat Case</td>
<td><img src="image3" alt="Histogram" /> Scale 0m to 35m  Scale 9.47m to 27.47m</td>
</tr>
</tbody>
</table>

Scenario 2.1
24 GAL – 18 GPS Optimist GAL Psat Case
This histogram is unavailable because the maximum of HPL values is infinite.

Scenario 2.2
24 GAL – 18 GPS Medium GAL Psat Case
This histogram is unavailable because the maximum of HPL values is infinite.

Scenario 2.3
24 GAL – 18 GPS Worst GAL Psat Case
This histogram is unavailable because the maximum of HPL values is infinite.

Scenario 3.1
Nominal 24 GPS
![Histogram](image4) Scale 0m to 20m  Scale 8m to 20m

Scenario 3.2
Non Nominal 24 GPS
![Histogram](image5) Scale 0m to 60m  Scale 8.17m to 78.07m

2015 International Association of Institutes of Navigation World Congress
Prague, Czech Republic, 20–23 October 2015
Enhancing the EGNOS NOTAM Proposals: increased automation and state-of-the-art prediction engine

J. Morán, J. Vázquez, M.A. Sánchez
ESSP SAS (European Satellite Services Provider)
Madrid, Spain

Jorge.Moran@essp-sas.eu, Juan.Vazquez@essp-sas.eu, Miguel-Angel.Sanchez@essp-sas.eu

I. INTRODUCTION

NOTAM stands for NOTice to AirMen. NOTAMs contain any important information that could affect the safe flight of an aircraft or information that is required by the pilot to conduct safe operations. NOTAMs cover a wide variety of aspects, including navaids availability but also runways and even meteorological conditions. A NOTAM must be originated and issued promptly whenever the information to be distributed is of a temporary nature and of short duration or when operationally significant permanent changes or temporary changes of long duration are made at short notice, except for extensive text and/or graphics. Before a flight, pilots request a printout of any NOTAMs that are relevant to the route, destinations and alternate for the flight. A NOTAM service is essential in order to approve a flight procedure.

In Europe, the first analyses about the NOTAM concept for GNSS started in 2006 when EUROCONTROL launched a contract to analyse several questions and actions opened within the scope of the Area Navigation (RNAV) Approach Task Force (RATF). The main results of those analyses were published in the “Requirements for a European GNSS NOTAM system” [1]. After that, the EURONOTAM project was launched by Eurocontrol in 2007 with the objective to develop a prototype of a system based on the requirements defined in the previous GNSS NOTAM project. Finally, in 2009, the GISE project was activated in order to develop the first version of the operational platform (called “EURONOTAM”) that would be provided to ESSP, as the EGNOS Services Provider, in order to generate and distribute EGNOS NOTAM proposals to the corresponding NOTAM Offices (NOF) for their validation and issuance. Also, all these efforts resulted in the definition of a harmonized approach at European level for the provision of GNSS NOTAM which is defined in the European Concept for GNSS NOTAM document [2].

A. Applicable Regulation

The need for a NOTAM service when implementing SBAS based approach procedures is stated by the ICAO SARPs [6]: “Before publishing procedures based on SBAS signals, a State is expected to provide a status monitoring and NOTAM system. To determine the effect of a system element failure on service, a mathematical service volume model is to be used.

[...] The system element status data (current and forecast) required for the model could be obtained via a bilateral arrangement with the SBAS service provider, or via connection to a real time “broadcast” of the data if the SBAS service provider chooses to provide data this way.”

The EGNOS NOTAM Proposals Provision is operational since the March 2nd 2011, when the EGNOS Safety of Life (SoL) service was declared available for aviation users. The first EGNOS-based approach procedure was published in Pau-Pyrénées Airport (France) on March 17th 2011, hence becoming the first airport subscribed to the EGNOS NOTAM Proposals Provision. This publication was only possible after the signature of an EGNOS Working Agreement (EWA) between the ESSP and DSNA (Air Traffic Services provider in France). This standard agreement between ESSP and the ATS provider is required by the SES (Single European Sky) regulation prior to the publication of any EGNOS-based procedure and includes the EGNOS NOTAM Proposals Provision as one of the required enablers for the EGNOS-based approach procedures implementation.

Apart from establishing the NOTAM service as a key element in the implementation of SBAS-based approach procedures, the ICAO SARPs [6] also lay down the recommendations for this kind of service, in terms of notification timeliness:

“For scheduled events, notification should be given to the NOTAM authority at least 72 hours prior to the event. For unscheduled events, notification to the NOTAM authority should be given within 15 minutes. Notification should be given for events of 15 minutes, or longer, duration.”

As stated above, the final responsible for the provision of a NOTAM service is the State publishing SBAS-based approach procedures. Hence, the ICAO recommendations are not directly applicable to ESSP (which acts as data originator in the EGNOS NOTAM generation chain) but to the ANSPs having published EGNOS-based approach procedures and to which ESSP is providing EGNOS NOTAM proposals, in line with the EGNOS Working Agreement provisions. Nevertheless, ESSP intends to bring the EGNOS NOTAM Proposals Provision to a level enabling its Customers (EWA signatories) to fully comply with the ICAO recommendations for a NOTAM service.
B. The European concept for GNSS NOTAM

The GNSS NOTAM concept in Europe has been defined by EUROCONTROL [2], on the basis of a Stakeholders’ consultation, including discussions with the RNAV Approach Task Force (RATF, now transformed into the joint ICAO PBN Task Force and Eurocontrol RNAV Approach Implementation Subgroup –RAISG-) and the Aeronautical Information Operations Sub-group (AI Operations), working arrangements of EUROCONTROL Navigation and AIM Domains. The document describes the types of GNSS NOTAM to be distributed to final users, the generation and transmission process of the NOTAMs, the roles of each ANSP/State and International NOTAM Office (NOF) as well as the format to be used.

This concept foresees two types of NOTAM, GPS RAIM NOTAM and EGNOS NOTAM; with respect to the EGNOS NOTAM, it is stated that:

- EGNOS NOTAM will indicate periods of time at an airport when EGNOS is predicted not to be available to support RNAV (GNSS) approach to LPV minima: “EGNOS IS NOT AVAILABLE FOR LPV”.
- The criteria for EGNOS to support LPV apply for EGNOS to support LNAV/VNAV. Consequently, in the case the local supervisory authority allows the use of EGNOS for approaches to LNAV/VNAV minima, the EGNOS NOTAM could indicate “EGNOS IS NOT AVAILABLE FOR LNAV/VNAV”.

The EGNOS NOTAMs are formatted according to the aeronautical information applicable requirements from ICAO Annex 15 [3], ICAO Doc 8126 [4] and the EUROCONTROL Operating Procedures for AIS Dynamic Data (OPADD) [5].

Also, and for the case of EGNOS NOTAM, it should be highlighted that in case the National Supervisory Authority (NSA) of the corresponding EU State approves the use of EGNOS for vertical guidance to fly APV Baro procedures, EGNOS NOTAMs are provided not only for EGNOS based approach procedure to LPV minima but also for LNAV/VNAV minima. The reason for this is that, the lateral performance required for LNAV/VNAV is much lower than the one required for LPV (i.e. lower than APV I lateral performance) and vertical performance required for LNAV/VNAV is similar to the one required for LPV (similar to APV I vertical performance). The loss of vertical performance is most likely to happen first whenever navigation performances degrade. In conclusion, and considering that the loss of vertical guidance results in the loss of the selected approach capability in total, the monitoring of EGNOS performance against APV-I criteria illustrates the capability of EGNOS to support both LPV and LNAV/VNAV approaches.

C. EGNOS NOTAM generation chain

Focusing on the EGNOS NOTAM, the following steps take place since the occurrence of an event causing a service unavailability (at one of the airports with published EGNOS-based procedures) until such situation is communicated to final users:

1. An event at EGNOS and/or GPS system level is detected.
2. The impact of the event on the user performances is predicted. In case service unavailability at a relevant location (airport with EGNOS based approach procedures published) is confirmed, it is formatted in a notification compliant with the ICAO specifications (EGNOS NOTAM Proposal).
3. The EGNOS NOTAM proposal is distributed to the concerned NOTAM Office (NOF). Transmission can be done directly via AFTN or through the European Aeronautical Database (EAD), managed by EUROCONTROL. It is up to the concerned ANSP to decide which transmission channel better suits its needs.
4. The NOTAM Office validates the EGNOS NOTAM proposal format correctness, and if no issue is detected, creates the EGNOS NOTAM (only the country sequence number needs to be changed in order to obtain the final NOTAM from the NOTAM proposal).
5. The EGNOS NOTAM is published and made available to final users.

In the above chain, ESSP is fully responsible for steps 1 and 2. In case the concerned ANSP requests a direct transmission of the NOTAM proposals to the NOF (step 3), ESSP is also fully in charge of this transmission. Otherwise (ANSP requests transmission through EAD), ESSP is responsible for the transmission up to EAD, and EUROCONTROL takes over the responsibility to redirect the information to the NOF.

II. EGNOS-BASED APPROACH PROCEDURES IN EUROPE

Since ESSP started providing EGNOS NOTAM Proposals for Pau Airport in France (March 17th 2011), the EGNOS NOTAM Proposals Provision has grown both in terms of countries subscribed (following the signature of an EGNOS Working Agreement between the ESSP and the corresponding ANSP) and operational airports. At the time of writing this paper (AIRAC cycle 1507 - June 25th 2015) thirty one ANSPs were subscribed to the EGNOS NOTAM Proposals Provision.

The EWA negotiation status for each European country at the time of writing can be seen in Fig. 1. In addition, the evolution of the number of published APV SBAS (LPV) procedures is also depicted:
The full list of airports having at least one EGNOS-based approach procedure, including the type and number of procedures available per airport, can be found at EGNOS user support website ([http://egnos-user-support.essp-sas.eu/](http://egnos-user-support.essp-sas.eu/)). The aforementioned list is regularly updated by ESSP. To sum up, at the time of writing (AIRAC cycle 1507 - June 25th 2015) ESSP was providing EGNOS NOTAM proposals to 17 countries and 150 airports for a total of 261 EGNOS-based approach procedures (192 LPVs and 69 APV Baro).

## III. EURONOTAM TOOL: FUNCIONALIDADES AND INTERFACES

### A. Introduction to the EURONOTAM tool

The EURONOTAM tool implements a mathematical service volume model and is in charge of the EGNOS NOTAM Proposal generation considering GPS and EGNOS GEO satellites, and EGNOS stations (RIMS) outages. The service volume, after propagating the orbits of the satellites based on the GPS almanacs, computes the satellites in view for each airport under study. Once the satellite visibility is computed, the prediction module estimates the UDRE and GIVE values for each Line Of Sight (LoS), taking into account the number of RIMS in view (for the UDRE computation) and the number of IPPs (Ionospheric Pierce Points) surrounding the IGP (for the GIVE estimation). After estimating these two variances, the protection levels for each airport are computed as defined by MOPS [9][9]. Finally, the prediction module assesses the level of service available at each aerodrome and in case that an unavailability of the EGNOS SoL service is predicted, the EURONOTAM tool generates airport-specific EGNOS NOTAM proposals in ICAO format. Finally, the NOTAM proposals are delivered to the concerned NOF through the AFTN network, either directly or via EAD (European Aeronautical Database).

### B. EURONOTAM Tool Interfaces

As represented in Fig. 2, the EURONOTAM tool uses as input the Almanacs and NANUs in order to predict the EGNOS service availability. This information is read from a local directory and from the database, populated by the operator through the HMI (Human Machine Interface).

On the other hand, the EURONOTAM tool is connected to a dedicated interface with the EGNOS system (GNSS Asset Status interface) from which the tool retrieves the following data sets:

- a) GPS PRN List
- b) GEO PRN List
- c) RIMS List
- d) RIMS Positions
- e) GEO satellite and signal unexpected unavailability notice
f) RIMS unexpected unavailability notice.
g) GPS satellite unexpected unavailability notice (GPS satellite not monitored by EGNOS).

Every time the EURONOTAM tool connects to the GNSS Asset Status interface, the latest available information is received (all the above data sets). Then, upon change, updates are delivered.

Should a change in any of these assets occur, a new prediction is launched by the EURONOTAM tool in order to assess the impact of this event on the EGNOS performance. The information about the EGNOS unavailabilities identified is processed to generate the contents of the corresponding NOTAM proposals in the appropriate format. These NOTAM contents are sent through the AFTN network directly to the NOFs for its validation and distribution.

C. UDRE and GIVE computation

At first, the prediction module computes the orbit of the satellites based on the GPS almanacs following the algorithm as stated in GPS ICD [8]. Once the orbits of the satellites have been obtained, the prediction function computes the satellites that are in view by at least three EGNOS RIMS, using the stations locations \(r_{\text{rims}}\) and the propagated satellite position \(r_{\text{sat}}\).

After computing the satellite visibility, the prediction engine estimates the UDRE and GIVE values. The UDRE for each satellite in a particular moment in time depends on the number of stations that see that satellite with an elevation higher or equal to a configurable RIMS mask angle. The computation of the visible RIMS for each satellite is done using the fixed RIMS position \(r_{\text{rims}}\) and the propagated satellite position \(r_{\text{sat}}\), taking into account the applicable RIMS unavailabilities. The RIMS visibility angle for UDRE computation is also a parameter that can be configured through the HMI. When the number of available RIMS has been calculated for each of the satellites, a table containing the UDRE estimated values as a function of the number of RIMS is used.

The GIVE value at one IGP depends on the number of RIMS surrounding its position. More exactly, it depends on the number of surrounding IPPs determined by the RIMS and the corresponding satellites. To select the surrounding IPPs, the system measures an angular distance (configurable through the HMI) between the IGP location and the RIMS IPPs.

The computation of the number of IPPs is done following an algorithm with the following inputs:

- RIMS position \(r_{\text{rims}}\).
- GPS satellite position \(r_{\text{sat}}\).
- IGP position \(r_{\text{IGP}}\).

Once the surrounding RIMS IPPs have been localized and counted, a table calculated by averaging the resulting GIVEs of several EGNOS simulations as a function of the number of IPPs is used to compute the GIVE value at each IGP.

D. EGNOS PLs computation

Before the PL computation begins, the prediction function calculates the UIVE value for each satellite-airport pair, which requires localizing the corresponding IPP. Once the IPPs positions have been identified, the prediction function calculates the distance to the different IGPs and selects the closest ones, based on the IPP position \(r_{\text{IPP}}\), the IGP position \(r_{\text{IGP}}\) and the distance from the IPP to the IGP \(r_{\text{IGP-IPP}}\).

Using the GIVE values of the selected IGPs, it interpolates the desired UIVEs at each of the airport IPPs. The IGP selection process, the interpolation algorithm that the prediction function follows and the expressions to obtain the GIVE value based on the UIVEs at the different airports, are fully described in [9].

The next step is to calculate other error corrections for each visible satellite, following MOPS guidelines [9]:

- Tropospheric variance \(\sigma_{\text{tropo}}\).
- Standard deviation for fast-long term corrections (residual error) \(\sigma_{\text{fl}}\).
- Airborne receiver standard deviation \(\sigma_{\text{air}}\).

Then the prediction function works out the complete deviation for each of the visible satellites using the next expression:

\[
\sigma^2_i = \sigma^2_{i,\text{fl}} + \sigma^2_{i,\text{URB}} + \sigma^2_{i,\text{air}} + \sigma^2_{i,\text{tropo}}
\]

Using this information, the PLs are computed as stated in [9].

Finally, the prediction module assesses the level of service available. The service is considered ‘LPV available’ at a given location if predicted HPLSBAS < HALLPV and predicted VPLSBAS < VALLPV.

IV. CALIBRATION ACTIVITY

A. Calibration Methodology

In order to align the performance predicted by the EURONOTAM tool with respect to the current EGNOS
service performance, an exhaustive campaign has been carried out to calibrate the prediction model.
As explained before, one of the main drivers of the EGNOS performances predictions are the UDRE and GIVE tables, relating the number of visible RIMS and IPPs from a satellite with the value of UDRE and GIVE. Taking this into account, one of the main objectives of the calibration campaigns was to calibrate the UDRE and GIVE tables in order to get performances closer to reality.
The methodology for this analysis, which will be deeply explained in the following section, is briefly introduced hereafter:

- First, a statistical data analysis has been performed using EGNOS real data (EGNOS v2.3.2) in order to assess:
  - The UDRE evolution as a function of the type of the GPS satellites (II-A, II-R, IIR-M or IIF) and the number of RIMS in view.
  - The GIVE evolution as a function of the IGP location, the hour of day (night and day comparison) and the number of IPPs.
- Then, an iterative calibration process (by modifying the UDRE and GIVE models, as well as other relevant parameters such as the RIMS mask angle or the radius used to compute the number of IPPs surrounding an IGP) has been carried out, comparing the Protection Levels predicted by the service volume with respect to the real ones, in a set of points evenly distributed throughout the EGNOS service area.

Finally, it is important to remark that although the target of this first calibration activity was to accurately model the performances provided by EGNOS System Release 2.3.2 (operational version at the time of writing), the methodology has been designed so that it will remain valid for future EGNOS system releases.

B. Calibration Results

1) EGNOS real data analysis

On a first step, an exhaustive data campaign has been done in order to model the UDRE and GIVE evolution as a function of the number of RIMS and IPPs in view. For that purpose, one month of EGNOS data (from 2014-07-26 to 2014-08-24) has been analysed using eclayr. This tool (developed by GMV), apart from other functions, is able to automatically collect and process EGNOS data to generate comprehensive performance assessment reports.

As part of these reports, eclayr provides for each epoch and satellite, the UDRE, broadcasted by EGNOS and the number of RIMS in view. In the same way, for each epoch and each IGP, the GIVE, transmitted by EGNOS and the number of IPPs in view is reported.

a) UDRE Analysis

The mean UDRE, for all the GPS satellites during a month of data has been computed as a function of the RIMS in view. This information, depicted in Fig. 4, has been used to calibrate the UDRE table configured in the EURONOTAM tool.

![Mean UDREi vs Number of RIMS](image1)

Fig. 4. Mean UDREi vs Number of RIMS

![Mean UDREi vs Number of RIMS for each Satellite Block](image2)

Fig. 5. Mean UDREi vs Number of RIMS for each Satellite Block

Additional analyses have been done aimed at identifying potential evolutions for the prediction engine of the tool. For instance, the correlation between the GPS satellite block and the UDRE, has been studied. However, as shown in Fig. 5, no relevant correlation has been found between the GPS satellite block and the UDRE, evolution as a function of the RIMS in view.

b) GIVE Analysis

Also, an analysis of the GIVE index as a function of the IPPs surrounding the IGP has been done for each epoch (during a month of data) and each IGP. The mean GIVE, depicted in Fig. 6, has been used to calibrate the GIVE table configured in the EURONOTAM tool.

![Mean GIVEi vs Number of RIMS](image3)

Fig. 6. Mean GIVEi vs Number of RIMS for each IGP
Again, in order to identify potential evolutions of the prediction engine of the tool, several analyses have been performed. For instance, the correlation between the GIVE$_i$ and the hour of day (night and day comparison) has been studied by computing the mean GIVE$_i$ at 4 hours intervals (see Fig. 7). As it was expected, for the same number of IPPs (surrounding the IGP), the GIVE$_i$ is greater in the central hours of the day (i.e. 12-16h) when the ionospheric activity is higher.

On the other hand, the evolution of the EGNOS ionospheric corrections has been analysed depending on the IGP location. As shown in Fig. 8, the comparison of the mean GIVE$_i$ evolution for the IGPs located in the three latitudinal bands defined shows that there is an important difference, especially when the number of IPPs surrounding the IGP is reduced. Based on this information, a potential evolution of the tool could consist on upgrading EURONOTAM to manage multiple GIVE tables (currently it is only possible to configure one table) depending on the IGP location and/or the time of the day.

### Iterative Process

In order to evaluate the improvement obtained with the new UDRE and GIVE models, the EURONOTAM tool configuration was updated to use the tables derived from the analysis described in the previous section. After applying these changes, the availability performances predicted by the tool were compared with respect to the real performances of EGNOS. For this comparison, two days with nominal performances were analysed (August 8th and 10th 2014).

As depicted in Fig. 10, the availability performances predicted by EURONOTAM before the calibration activity were mostly in line with the real performance in the central area of Europe (where most of the airports with EGNOS LPV procedures are located). However, focussing on the locations placed in the borders of the commitment area, several differences may be observed, especially on the Eastern areas of Europe where the predicted availability is more pessimistic. For instance, for the days under analysis, the EGNOS APV-I availability is above 99.9% throughout Poland but the predicted performance is below 99% in the east part of the country. Something similar occurs in Finland, where the EFJ0 (Joenssu) airport - with LPV procedures published – is located. In fact, due to these differences between the real and predicted performances, several NOTAM proposals have been generated for this airport, notifying false periods of time when EGNOS was predicted not available. These differences reflect the improvement of the EGNOS services performance as new releases are deployed.

In order to minimize these discrepancies and improve the prediction capabilities of the tool, an iterative process was carried out modifying several configuration parameters of the tool. On a first step, the UDRE and GIVE models derived from the previous analysis were configured in the tool and the availability performances were predicted for the two days under analysis. Although some improvements were obtained...
with these new tables, still some discrepancies were observed on the border areas of Europe.

Focussing on these differences, it was detected that most of them were due to the fact that the EURONOTAM prediction module was pessimistic when selecting the IPPs to be used for the GIVE computation. As shown in Fig. 12, the EURONOTAM tool selected less IPPs than EGNOS. Hence, for a given IGP, less IPP were available for computing the ionospheric correction (penalising the service performance).

In order to improve the predicted availability performances, the IGP visibility criterion has been modified by increasing the maximum distance IGP-IPP for the GIVE computation in EURONOTAM. This modification improved the monitoring of IGP located at the ECAC border and therefore the availability performance on these areas.

After implementing all the changes derived from the calibration analysis, a significant improvement has been obtained (especially on the Eastern areas of Europe), and a very good level of alignment between the availability performance predicted by EURONOTAM and the real performance of EGNOS (see Fig. 9 and Fig. 11) on nominal conditions.

V. WHAT’S NEXT?

In the coming months, the EGNOS NOTAM Proposals provision is expected to progressively evolve in order to support new EGNOS based operations.

For instance, as detailed in the excerpt of the ICAO PBN Manual [7] included hereafter, EGNOS could be a key enabler for the publication of RNP 0.3 routes:

“Operators relying on GNSS are required to have the means to predict the availability of GNSS fault detection (e.g. ABAS RAIM) to support operations along the RNP 0.3 ATS route […]. This prediction will not be required where the navigation equipment can make use of SBAS augmentation […]. Should the State permit the operator of an SBAS-equipped aircraft to disregard the requirement for a RAIM prediction when the RNP 0.3 operation occurs in an SBAS service area, then it is recommended the State consider establishing a requirement for that operator to check SBAS NOTAMS prior to the flight to ensure the availability of the SBAS SIS”

In order to support the publication of RNP 0.3 routes based on EGNOS, the EURONOTAM prediction engine responsible for assessing the level of service available at each aerodrome, will be able to predict the EGNOS service availability not only in airport-specific locations but also along the waypoints defining a RNP 0.3 route.

Additionally, the next version of EGNOS system will provide LPV-200 capability, enabling the publication of RNP APCH procedures down to 200 feet (considered as SBAS CAT I Precision Approach according to the new ICAO Approach Classification scheme). In line with this, the EGNOS NOTAM Proposals provision will support the publication of SBAS CAT I operations.

VI. CONCLUSIONS

The information presented in this paper has shown that, as the European aviation community continues with the progressive adoption of EGNOS, the EGNOS NOTAM Proposals
Provision is also improving. At the time of writing (AIRAC cycle 1507 - June 25th 2015) ESSP was providing EGNOS NOTAM proposals to 150 airports for a total of 261 EGNOS-based approach procedures. This process is expected to continue speeding up in the coming years due to the wider consensus in the aviation community with regards to the added value that GNSS can bring to all phases of flight, including approach.

Among all the continuous activities to provide the best quality of service, this paper has focused on the adjustments that are regularly performed to the EURONOTAM tool prediction engine, in order to adapt it to the progressive improvement of the EGNOS performance. The outcomes of the calibration analysis carried out during the last year have been outstanding, providing very good level of alignment between the availability performance predicted by EURONOTAM and the real performance of EGNOS as well as some ideas for improvements. These potential improvements will be assessed and if necessary injected in future evolutions of the tool. Finally, it should be noted that this calibration activity will be exercised again for future EGNOS system releases in order to maintain aligned the EURONOTAM performances with EGNOS’. In line with the progressive adoption of EGNOS and the growing interest from the aviation community in using EGNOS for all phases of flight, the EGNOS NOTAM Proposals provision will evolved in the coming months in order to support new EGNOS based operations (e.g. RNP 0.3 and LPV-200).

To conclude, from the authors’ perspective, the results presented in this paper illustrate the commitment from the EGNOS programme and its stakeholders (GSA, ESSP, industry) towards the aviation user community and the work that has and will continue to be done to maximize the EGNOS adoption and EGNOS users’ satisfaction.

ACKNOWLEDGMENTS

The authors would like to thank the European Commission and the European GNSS Agency (GSA) for their continuous support and close cooperation that has been key in order to make the EGNOS NOTAM Proposals Provision a reality. The authors would like to acknowledge the outstanding contribution made by GMV (special mention to Jorge Humberto and Román Rodriguez) in the calibration activity presented in this paper and the important support provided for the successful performance of the service. Finally, the authors would like to thank their colleagues Francisco Cantos and Pedro Gómez, which have provided invaluable support to this paper.

REFERENCES

[8] GPS Interface Control Document, ICD-GPS-200, Revision D.
[9] MOPS for GPS/WAAS airborne equipment, RTCA D0229D.
EGNOS empowers tracking and tracing for logistics in Europe and around the Mediterranean

A. Di Fazio; D. Bettinelli
Telespazio S.p.A.
Rome, Italy
antonella.difazio@telespazio.com; danielle.bettinelli@telespazio.com

U. Celestino
European Commission
Brussels, Belgium
ugo.celestino@ec.europa.eu

C. Gracia
Novacom Services
Toulouse, France
cesar.gracia@novacom-services.com

Abstract—Satellite navigation technology is a building block for numerous applications, for example, it is widely adopted for freight transport and logistics applications.

EGNOS (European Geostationary Navigation Overlay Service), which is operational since 2009, and Galileo, presently under deployment, constitute the European Global Navigation Satellite System (EGNSS).

EGNOS improves the accuracy of the current GPS signal and delivers integrity information, making it suitable for applications requiring accurate and reliable positioning. EGNOS provides the Open Service (OS) free to users from the satellites’ open signal, and the EGNOS Data Access Service (EDAS) delivered through a server via terrestrial networks and enabling to further increase the performances of the GPS.

Thanks to its capability to provide more precise and trustworthy information on the position with respect to GPS-only, EGNOS enables to empower tracking and tracing systems, and it allows to enhance today’s operational systems based on GPS in Europe and in countries outside the European boundaries. Galileo will provide further improvements on a global scale when it will become operational.

In the European project CONTAIN (CONtainer security Advanced Information Networking) a system for tracking and tracing intermodal containers has been developed and validated in life operations for monitoring the shipment of goods along road and rail corridors.

In the European project MEDUSA (MEDiterranean follow-Up for EGNOS Adoption), the system developed in CONTAIN has been used in a life demonstration in Jordan, with the involvement of the Royal Jordanian Customs and the aid of the Ministry of Transport of Jordan, to track and trace containers shipped across the Mediterranean sea and vehicles transporting goods through the country’s roads. This life demonstration represents the first real experience of EGNOS services’ usage in far-flung neighbouring regions.

This paper presents the results of the tests conducted in CONTAIN and of the life demonstration carried out in MEDUSA. In the paper, the analysis of the collected data is reported and discussed, showing the added value of EGNOS for the freight transport and logistics applications, that can be extended also to customs applications, and that has the potential to generate benefits in Europe and in the Euromed region.

Keywords - EGNOS; Galileo; freight transport, tracking and tracing; accurate and reliable positioning; customs applications; Euromed region

I. EGNOS AND GALILEO

Satellite navigation technology is an increasingly common component of innovative applications in different sectors. GNSS (Global Navigation Satellite System) solutions are already used in the road industry, aviation, surveying, agriculture, the maritime environment and rail for instance. In the logistics sector today the satellite positioning is widely adopted. Besides, the possibility to extend its use for institutional purposes, to support and facilitate the customs procedures and the control of the freight traffic flow across the entire supply chain, is presently under consideration.

EGNSS (European Global Navigation Satellite System) is the Europe’s initiative for a state-of-the-art GNSS, and it includes EGNOS (European Geostationary Navigation Overlay Service) and Galileo.

EGNOS is a regional system and it is the first pan-European satellite navigation system. EGNOS is a Satellite-Based Augmentation System (SBAS) that augments the GPS (Global Positioning System) by providing correction data that enable to improve GPS position accuracy, and integrity information about the GPS system. EGNOS is operational and it is offering the following services:

- EGNOS Open Service (OS), launched in 2009, is delivered free of charge for use by anyone with an EGNOS-enabled or SBAS-compatible receiver. Being based on GPS frequencies and complementary
signal design, the EGNOS signal does not require major changes for receivers. Today, many mass-market receivers available on the market are also EGNOS-enabled. EGNOS OS is particularly suitable for mass market and some applications like surveying.

- EGNOS Safety-of-Life Service (SoL) is authorized for European civil aviation and operational since March 2011. EGNOS SoL delivers the integrity message verifying the usability of the GPS system and providing timely warnings (within six seconds) when the system or its data should not be used for navigation.

- EGNOS Data Access Service (EDAS) launched in 2012, delivers a terrestrial commercial data service that is particularly suitable for professionals. It consists of a server that gets data directly from the EGNOS system and disseminates them via terrestrial networks in real time, with a guaranteed maximum delay, security, and performance. Software solutions connect to EDAS and use its provided data to implement products and value-added services built on them. For example, service providers can deliver EGNOS data via different telecommunication means, augment EGNOS OS performances to improve its availability and GPS position accuracy, or qualify and guarantee the GPS position information by exploiting EGNOS’s integrity feature.

Galileo will provide highly accurate, guaranteed global positioning service specifically designed for civilian purposes, to its users worldwide. Additionally, while providing autonomous navigation and positioning services, Galileo will assure interoperability with the GPS - USA and Global Navigation Satellite System (GLONASS) - Russia, the two other already deployed global satellite navigation systems as well as with Beidou, the global Chinese GNSS system. The fully deployed system established under the Galileo programme will consist of 30 satellites and the associated ground infrastructure. Eight satellites of the final operational constellation are already in their orbits.

II. EGNSS FOR LOGISTICS APPLICATIONS

Though conceived for aviation, EGNOS has concrete perspectives of utilization also in other markets. Through its services, EGNOS gives opportunities for users to have more accurate and reliable positioning, for enhancing existing applications and developing new ones. Because of their capability to provide more precise and trustable information on the position with respect to GPS-only, EGNOS OS and EDAS are particularly suitable for transport applications (such as road, freight transport and logistics) requiring accurate and reliable positioning, and for which the positioning data are linked to legal and economic liabilities.

In the freight transport and logistics sector, the possibility to extend the use of satellite positioning also for institutional purposes, to support and facilitate the customs procedures and the control of the freight traffic flow across the entire supply chain, is presently under consideration. In this respect, EGNOS, by enabling a robust and reliable traceability of the freight, helps to define liability schemes among the various stakeholders and support law enforcement. In particular, the use of EDAS enhances the real-time positioning and traceability (for example in the case of goods transport), by making them more robust and reliable, thanks to improved performance and data with greater added value than those obtained through EGNOS OS only.

Over the past decade, the European Commission (EC) has incentivized the development of solutions and applications based on EGNSS for commercial and professional markets, including freight transport, such as tracking and tracing of professional and regulated fleets, intermodal (road/rail/maritime) asset tracking (containers and tankers), city logistics, monitoring of vehicle transiting/movements in limited traffic areas.

The European project CONTAIN (CONtainer securiTy Advanced Information Networking) [1], recently ended, had the objective to create tools and methods that can be applied to increase the security in container transport. To this end, over the past three years, CONTAIN developed and proved in real operational cases, innovative and advanced technologies, including those using satellite navigation, for empowering the transport and the security of the intermodal containers traffic.

Within CONTAIN, Telespazio and Novacom Services, partners of the project, have developed a system based on EGNOS to provide value added tracking and tracing services. The system has been extensively proved in real business cases in Europe, in cooperation with Interporto Bologna, for monitoring the shipment of intermodal containers along road and rail corridors.

III. EUROMED GNSS II/MEDUSA

EGNSS can provide benefits not only to European Union (EU) member-states but also to non-EU countries.

EGNOS regional coverage could be extended to areas adjacent to the EU by deploying limited additional elements of the ground infrastructure. Additionally, the EU Regulation No 1285/2013 on the implementation and exploitation of EGNSS explicitly considers the extension of EGNOS system to other regions of the world, particularly to the territories of candidate countries and of third countries belonging to the European Neighborhood Policy.

Over the past decade, the EC has launched a series of projects under the umbrella of the Euromed GNSS programme to support EGNOS service extension over the EU neighbouring countries in southern Europe, in Africa and in part of the Middle East (i.e. the so called Euromed region that includes Algeria, Egypt, Israel, Jordan, Lebanon, Libya, Morocco, Palestine, Syria, and Tunisia). Through the Euromed GNSS programme ([2] and [3]), from the originally envisaged coverage over European countries, the EGNOS coverage is today extended in countries of North Africa and of the Middle East.
The MEDUSA (MEDiterranean follow-Up for EGNSOS Adoption) project [3] is part of this programme, and it is aimed to assist the countries of the Euromed region to achieve optimal use and adoption of EGNSOS services and preparing them for Galileo. Through MEDUSA, EC is implementing capacity-building and technological transfer in the Euromed region. For this purpose, MEDUSA has set-up a cooperation and operation structure named GEMCO (Galileo EuroMed Cooperation Office) located in Tunisia (Tunisia) and acting as a catalyzer of initiatives related to EGNSS in the Euromed region.

Running since January 2012, the project has been implementing technical assistance actions in the countries, for training, technical and feasibility assessment, support to decision-making process for the actual service introduction.

IV. MEDUSA’S EGNSOS SERVICES DEMONSTRATIONS IN JORDAN

In the frame of one of MEDUSA’s technical assistance actions, life demonstrations have been conducted in Jordan to prove EGNSOS OS and EDAS for tracking and tracing containers shipped across the Mediterranean Sea and vehicles transporting goods by road. Conducted with the involvement of the Jordan Customs and the aid of the Ministry of Transport of Jordan, these MEDUSA’s life demonstrations represent the first real experience of EGNSOS services’ usage in far-flung neighboring regions.

In the Euromed region, Jordan presents an excellent and concrete scenario for the use of EGNSOS in freight transport for professional/commercial and law enforcement purposes.

From a geographical perspective, the strategic position of Jordan makes it the centre for the movement of goods between Europe, Africa and the Arabian Gulf region. From an economic viewpoint, transit of free goods is a strategic market for the country, with more than one million trucks crossing Jordan yearly. In terms of security, the transiting trucks are among the most risky means of smuggling goods and dangerous materials. The density and irregular flow of trucks from neighbouring countries has led to the increase of traffic jams at customs houses. Some trucks are delayed for many hours between the shifts of convoys and often have to stay overnight at border customs centres. The traditional customs escorting process (convoys) causes traffic jams on highways, disturbs the smooth movements of passenger cars, and delays the movement of goods across the country.

In 2008 the Jordan Customs adopted an electronic tracking and tracing system based on GPS, to monitor the vehicles and the goods in transit across the country’s territory and to ensure the integrity and the security of consignments while facilitating transit traffic. The Jordan’s system uses GPS, GSM-GPRS (Global System for Mobile Communications-General Packet Radio Service), and electronic seals based on Radio Frequency IDentification (RFID) technology.

Within MEDUSA, Jordan Customs have proved the system developed in the CONTAIN project and consisting of:

- a tracking device installed on the container and able to provide the relevant position by means of GPS+EGNOS. The tracking device integrates a GPS/EGNOS chipset and it is configured to use EGNSOS OS and EDAS. Detailed information about the characteristics of the device and how it can be mounted on a container is given and shown below.

- a software solution named LCS (LoCation Server). LCS enables the tracking and tracing system to use EDAS and to provide value-added services by exploiting the relevant features of EGNSOS (i.e., corrections and integrity information).

- a monitoring and localization platform, which enables tracking and tracing, statistical analysis, and alarm management functions via Internet access.

The architecture of the system is presented in Fig. 1.

LCS is a “plug-in” navigation solution that easy retrofits GPS tracking and tracing systems and allows them to deliver value-added positioning services based on EDAS, such as the improvement of the quality and the reliability of the position of the tracked asset. LCS consists of software modules running on a server connected to the tracking and tracing system and to EDAS. LCS obtains data and corrections from EDAS and the data from the tracking devices installed on the assets being tracked (e.g., vehicles, containers, tankers, wagons, trucks).

The added value of EGNSOS data distributed by EDAS arises not only from obtaining position fixes that are more accurate than standalone GPS but also from metrics that qualify a measured position. The so-called “protection level” expresses the system’s “confidence” in the reliability of a position.

In addition to remote localization, tracking and tracing, a web-based application performs such functions as alarming, reporting and geofencing.

Fig. 1. MEDUSA GPS+EGNOS-based system for tracking and tracing containers in Jordan.

EDAS-augmented position information accessed through LCS consists of latitude and longitude (as for a GPS-only

1 Source: Jordan Customs, presentation during the MEDUSA national workshop in Jordan, Amman 11 December 2013
position, but more accurate), time of the measured position, and the “protection level”.

The tracking device (whose key specifications are reported in Table 1) integrates a GPS/EGNOS chipset that uses EGNOS OS and EDAS by means of LCS.

The device’s design and dimensions are driven by the need to cope with stringent requirements and the harsh operating conditions posed by the shipment of freight through different environments (inland and overseas) and being unpowered for long periods of time. For this reason, the unit is enclosed in a waterproof package that also provides robust protection against dust. Additionally, the enclosure’s dimensions allow it to be mounted in the grooves of intermodal containers’ sidewalks (as shown in Fig. 2).

The tracking device can be installed and removed by means of strong magnets to avoid the need for screws and holes to attach the unit. Long-life rechargeable batteries support up to six months of operation without maintenance, accommodating the absence of any external power source in containers during long-distance hauls. The batteries can be recharged by plugging the tracking device into a standard electrical outlet.

The tracking device periodically sends information on its position, with the transmission rate that can be remotely programmed. The device also integrates motion sensors, so that a more frequent position update rate is automatically set when the container is stopped (e.g., in ports or intermodal terminals). Otherwise, one position report per day is sent when the container is in motion (indicating, for example, that the container is on a vessel).

![Fig. 2. Example of a GPS+EGNOS tracking device able to use EGNOS OS and EDAS mounted on a container.](image)

**TABLE I.**

<table>
<thead>
<tr>
<th>Device name</th>
<th>MEDUSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions</td>
<td>838 mm x 116 mm x 25 mm</td>
</tr>
<tr>
<td>Weight</td>
<td>2.17 Kg</td>
</tr>
<tr>
<td>Battery</td>
<td>Rechargeable batteries (12 units 3,7V/2900 mAh type Li-Ion)</td>
</tr>
<tr>
<td>Battery life</td>
<td>6 months (minimum) autonomy at polling rate of 1 fix/day</td>
</tr>
<tr>
<td>Mounting</td>
<td>Exterior mounting with magnets (or screws - optional)</td>
</tr>
<tr>
<td>Packaging</td>
<td>One single box - Rugged Packaging and IP68 enclosure</td>
</tr>
<tr>
<td>Communication</td>
<td>Dual communication: GPRS/Sat Com INMARSAT</td>
</tr>
<tr>
<td>Transmission mode</td>
<td>- Default transmission mode GPRS</td>
</tr>
<tr>
<td></td>
<td>- Satellite transmission/automatic switch-over in case of GPRS unavailability</td>
</tr>
<tr>
<td>Satellite navigation</td>
<td>Telit Jupiter JF2/Sirf Star IV</td>
</tr>
<tr>
<td></td>
<td>GPS + EGNOS</td>
</tr>
<tr>
<td></td>
<td>EGNOS OS</td>
</tr>
<tr>
<td></td>
<td>EDAS powered by LCS (CWA 16390 compliant)</td>
</tr>
<tr>
<td>GPS Chipset</td>
<td>Sirf IV</td>
</tr>
<tr>
<td>Polling rates</td>
<td>- Remotely configurable</td>
</tr>
<tr>
<td></td>
<td>- Two default polling rates: In-Motion (1 fix/day) and Stationary (1 fix/hour)</td>
</tr>
<tr>
<td></td>
<td>- Minimum polling rate: 1 fix/10 minutes</td>
</tr>
<tr>
<td>Sensors</td>
<td>Motion sensor</td>
</tr>
</tbody>
</table>

Fig. 2. Example of a GPS+EGNOS tracking device able to use EGNOS OS and EDAS mounted on a container.

To ensure a robust positioning, the tracking device transmits positions via terrestrial GSM-GPRS network, and it
is able to switch to satellite INMARSAT communications in the case of unavailability of the terrestrial GSM-GPRS network. Data transmission has been optimized for cost-effectiveness and reduced power consumption.

Containers equipped with GPS/EGNOS tracking device were shipped in an operational traffic route from Italy to Jordan (Fig. 3), crossing geographic areas with different EGNOS service coverage. This enabled to appreciate the EGNOS added value in those areas where EGNOS is currently available with respect to GPS-only. Some containers were shipped from the port of Ravenna in Italy via sea to the port of Aqaba, Jordan, and others from Brescia, Italy, by rail to the port of Genoa and then via sea to the port of Aqaba.

Fig. 3. Screenshot of the web-based application - Mediterranean Sea path.

Fig. 4 presents a screenshot of the web-based application showing that EDAS is used: EGNOS corrections are applied to the position and the “protection level” in the horizontal dimension (HPL – horizontal protection level) is provided resulting in a user’s higher confidence in a container’s reported position.

The screenshot in Fig. 5 represents a condition in which only standalone GPS is used without the benefit of EDAS. EGNOS corrections are not applied to the position, making it impossible to calculate a “protection level”; thus, a value is not given for the HPL.

In Jordan, the GPS/EGNOS tracking devices were installed on a Jordan Customs patrol vehicle (Fig. 6) and proved from the port of Aqaba to Amman (Fig. 7). Static and dynamic tests were conducted in various urban and extra-urban environments, such as the route Aqaba-Amman and in desert areas (dynamic tests), and in the city of Amman (static and dynamic tests).
The screenshot below (Fig. 8) is related to the test in Jordan. It shows that in some cases the EGNOS corrections are applied to the position, meaning that the EGNOS performance is suitable, and in these cases the "protection level"/HPL value is given. While in some cases the EGNOS corrections are not applied to the position meaning that the EGNOS performance is not suitable, and thus GPS-only is used and the "protection level"/HPL valid value is not given.

V. RESULTS OF THE MEDUSA’S EGNOS SERVICES DEMONSTRATIONS IN JORDAN

Table 2 presents the results of a comparison between the precision of the positions obtained by four tracking devices installed in various locations (under similar GNSS environmental conditions), in the cases of GPS-only and GPS corrected with EDAS using LCS. The table reports the values (in meters) of the horizontal and vertical distance root-mean-square (DRMS), being the root mean square of the distances (projected on the horizontal plane and vertical dimension respectively) from the average 3D position to the observed positions obtained from a number of trials. The values show that LCS improves the position.

Fig. 9 and 10 show the availability of the “protection level” in the horizontal dimension and the relevant statistical distribution, for the positions of containers in Jordan and in Europe (Italy and France). As expected, these plots show an EGNOS performance degradation in Jordan with respect to Europe.

For Jordan, the availability of an HPL is about 25 percent, and in most of the cases the HPL’s values are between 12 and 14 meters. On numerous occasions the HPL’s values are even higher. At locations in Europe, the availability of HPL is about 78 percent, and in most cases the HPL values are between 8 and 10 meters.

As long as EGNOS service coverage will be extended into the Euromed countries, the situation in Jordan will improve and produce similar results as in Europe.

### Table II

<table>
<thead>
<tr>
<th>Device (Location)</th>
<th>GPS only</th>
<th>GPS+EDAS (LCS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of samples</strong></td>
<td>26</td>
<td>93</td>
</tr>
<tr>
<td><strong>Horizontal DRMS (m)</strong></td>
<td>3.04</td>
<td>2.26</td>
</tr>
<tr>
<td><strong>Vertical DRMS (m)</strong></td>
<td>3.95</td>
<td>3.65</td>
</tr>
<tr>
<td><strong>Device 2 (Toulouse-France)</strong></td>
<td>44</td>
<td>501</td>
</tr>
<tr>
<td><strong>Horizontal DRMS (m)</strong></td>
<td>7.32</td>
<td>2.61</td>
</tr>
<tr>
<td><strong>Vertical DRMS (m)</strong></td>
<td>8.30</td>
<td>2.6</td>
</tr>
<tr>
<td><strong>Device 3 (Toulouse-France)</strong></td>
<td>33</td>
<td>432</td>
</tr>
<tr>
<td><strong>Horizontal DRMS (m)</strong></td>
<td>9.10</td>
<td>1.72</td>
</tr>
<tr>
<td><strong>Vertical DRMS (m)</strong></td>
<td>7.63</td>
<td>2.13</td>
</tr>
<tr>
<td><strong>Device 4 (Toulouse-France)</strong></td>
<td>37</td>
<td>538</td>
</tr>
<tr>
<td><strong>Horizontal DRMS (m)</strong></td>
<td>7.95</td>
<td>1.55</td>
</tr>
<tr>
<td><strong>Vertical DRMS (m)</strong></td>
<td>11.07</td>
<td>1.74</td>
</tr>
</tbody>
</table>

Performance analysis/comparison between GPS-only and GPS+EDAS via LCS.
Moreover, the MEDUSA Jordan “case study” also proves that EGNOS and Galileo can provide benefits not only to European countries. EGNOS OS and EDAS services are ready for use, though with degraded performances, far beyond the European borders. In this respect, the MEDUSA’s demonstration in Jordan, understood as a proof of concept, is a valuable “case study” for other countries outside Europe interested to use EGNOS for freight and road transport applications. The role of projects like MEDUSA is to give the opportunity to the Euronmed countries to share European experiences and have advices and guidance to implement their own activities, such as the development of research projects, national and regional initiatives. During the “All day long Think Tank: Exploiting the potential of E-GNSS in non-EU countries for road regulated applications” [4], organized by MEDUSA in Tunis on 19 May 2015, the Jordan Customs appreciated the added value of EGNOS with respect to GPS-only for tracking and tracing systems. Particularly, EGNOS’s ability to provide users with a more accurate position of the freights and goods movement, and also a measure of the reliability of the position information was considered a useful tool for enhancing the security of operations, the enforcement and the control.

VI. CONCLUSIONS

The availability of EGNOS over Europe, as a precursor of Galileo over the world, enables reliable positioning for robust tracking and tracing services. EGNOS added value as compared to GPS relies in the capability to provide high accuracy, combined with continuity and “protection level” information, enabling the involved commercial stakeholders and authorities to use GNSS to support their operations. Thanks to EGNOS, tracking and tracing systems provide users with a more accurate position of the freights and goods movement, and enable also a measure of the reliability of the position information. The resulting benefits are in a higher robustness, and thus positive impacts on security of operations, monitoring, planning and optimisation. Thanks to this, EGNOS based tracking and tracing systems could be useful tools for enforcement and control. This is not the case for tracking and tracing systems based on GPS-only. Therefore, in applications where reliability is a key feature, the use of EGNOS OS and EDAS could be of interest. This is the case of the freight and asset tracking and tracing applications, especially when security concerns are involved. Suitable technological solutions and products are available on the markets for fruition in operation. The most relevant benefits are social, thanks to an enhanced security; however also commercial interests are implied.

The MEDUSA Jordan “case study” shows the potential that the use of EGNOS can bring in freight transport tracking and tracing and road applications, thanks to its added value with respect to GPS-only.

REFERENCES

[1] CONTAIN project www.containproject.com
[4] All day long Think Tank: Exploiting the potential of E-GNSS in non-EU countries for road regulated applications http://galileo.cs.telespazio.it/medusa/public/All%20day%20long%20Think%20Tank%20road%20regulated%20applications%20Tunis%2019%20May
Session F
Spectrum management.
Jamming and spoofing of navigation systems.
Abstract—Receiver autonomous integrity monitoring (RAIM) is designed for fault detection and exclusion of a single satellite failure. It determines the failure by checking the consistency between satellite range measurements. Therefore, it is not able to detect the spoofing attack that affects all the measurements to fake the user position, as the consistency maintains even though the position is faked. Thus, RAIM cannot be used as a reliable spoofing detection method on its own. In this paper, a spoofing detection method based on the combined use of single authentic ranging source and RAIM algorithm is proposed. By using the additional authentic signal as a reference, the proposed RAIM can detect a spoofing attack. Experimental spoofing tests using multiple RF GPS simulator have been carried out to verify the feasibility of the proposed algorithm. The results show that, contrary to a conventional GPS receiver, the proposed algorithm can detect a spoofing attack.

Keywords—anti-spoofing; spoofing detection; authentic signal; RAIM; eLoran; pseudolite; GNSS

1. INTRODUCTION

As the location information is getting increasingly important, there is a growing reliance on Global Satellite Navigation System (GNSS) in many applications. Nowadays, GNSS has a broad range of applications, ranging from the location-based service of a personal smartphone to aerospace and military applications [1]. As the dependence on GNSS is getting higher, the vulnerability of GNSS caused by jamming and spoofing is becoming a real threat to GNSS users. Unlike the jamming that disables access to satellite navigation signal, the spoofing may result in a greater risk, since the user could be led to a wrong position without being aware of it [2]. Therefore, many recent studies on anti-spoofing technique have been actively conducted [3]. This paper deals with a spoofing detection technique that does not require an additional hardware modification on a conventional GNSS receiver and, thus, is applicable to a variety of GNSS applications.

The conventional anti-spoofing techniques known to be most effective are the consistency check with inertial sensor, angle-of-arrival discrimination using multiple antennae, and cryptography authentication [3]. In addition, recent studies show that the multiple receiver consistency check method is effective as well [4]. From a user receiver perspective, the cryptography authentication is considered to be the best among these methods. While other methods require additional hardware equipment, such as multiple antennae, a communication link, and a RF collecting device, there is no additional hardware modifications on a receiver and, therefore, no shape and size constraints in the cryptography authentication method. However, since this method requires encrypted ranging sources, such as GPS P(Y) code, most civil users cannot use it.

In this paper, a spoofing detection method based on the combined use of single authentic ranging source and receiver autonomous integrity monitoring (RAIM) is proposed. As is well known, RAIM has originally been designed for fault detection and exclusion of a single satellite failure, not for the detection of a spoofing attack [5]. It uses redundant pseudorange residuals to check consistency with the computed receiver position and, when the residual exceeds an expected value, indicates a fault of the associated satellite. This is why it cannot raise an alarm against a spoofing attack that misleads the position solution itself and thus maintains the consistency between measurements and position by deceiving all the pseudorange measurements. However, if only a single authentic ranging signal can be available even when all the other signals are deceived, RAIM can detect an inconsistency between the authentic measurement and the deceived receiver position. This is the key idea of the proposed method that utilizes the combined use of a single authentic signal and RAIM. To provide the authentic signal, an eLoran, a pseudolite or a geostationary satellite could be used to transmit the encrypted ranging signal.

For validation of the proposed algorithm, we implemented a spoofing test environment using multiple synchronized GPS simulator modules operated on a same scenario. Two simulator modules were used to generate GPS and spoofing signals, respectively. Furthermore, a software GPS receiver was used to monitor the spoofing procedure. Under the test environment, it was confirmed that, unlike a commercial off-
the-shelf (COTS) GNSS receiver, the proposed method is able to effectively detect a spoofing attack.

This paper starts with a brief review of the RAIM algorithm (Section II). In Section III, the proposed scheme is described introducing the overall system architecture and the detailed algorithm. In Section IV, the configuration of an experimental spoofing test using multiple RF GPS simulator is addressed, which is followed by the presentation of the results (Section V). Finally implementation issues are briefly addressed (Section VI) and the conclusions of the present study are drawn (Section VII).

II. BRIEF REVIEW OF THE RAIM ALGORITHM

Before getting into the main subject, the conventional RAIM algorithm will be briefly reviewed in this section. Since the proposed method is based on the conventional RAIM, the formulas provided below will be used to derive the method proposed in the next section.

At first, the linearized measurement equation of satellite navigation system is given as follows:

\[ y = H\bar{x} + \xi \]  \hspace{1cm} (1)

where

\[ \bar{x} \] is the 4×1 linearized user state vector

\[ y \] is the n×1 linearized pseudorange measurement vector of n visible satellites

\[ \xi \] is the n×1 measurement error vector

\[ H \] is the n×4 matrix whose elements are three columns of direction cosines from user to each satellite and a value 1 as a fourth column

The RAIM algorithm cannot use the user position error directly as a test statistic, as both the user position estimation and integrity verification are to be performed simultaneously [6]. Instead, the range residual vector \( w \), the difference between measurement \( y \) and the estimated linearized measurement \( \hat{y}_{\text{LS}} \) derived from the least squares estimate for \( \hat{x} \), is used to make a scalar test statistic.

The range residual vector is derived as follows:

\[ \hat{\xi}_{\text{LS}} = (H^T H)^{-1} H^T y \]

\[ \hat{y}_{\text{LS}} = H \cdot \hat{\xi}_{\text{LS}} \]

\[ w = y - \hat{y}_{\text{LS}} \]  \hspace{1cm} (2)

The square sum of error (SSE) and the test statistic D, the square root of SSE, are defined as follows:

\[ SSE = w^T \cdot w \]

\[ D = \sqrt{SSE} \]  \hspace{1cm} (3)

More detail derivation and other approaches can be seen in [5] and [6].

III. SPOOFING DETECTION WITH A SINGLE AUTHENTIC SIGNAL

As described in the previous section, RAIM can detect and exclude the failure on a single satellite. As RAIM is based on the consistency check between satellite ranging measurements and the estimated user position, it could cope with a very simple spoofing attack that affects a minority of satellites. However, when a spoofing attack is successful on all the visible satellites or a majority of them, RAIM cannot raise an alarm against such a high-level attack. This is why RAIM is not an effective anti-spoofing measure in itself.

This paper proposes a new RAIM scheme that can detect spoofing attacks by utilizing an additional authentic signal. The new algorithm consists of two parts. One is an additional ranging signal source transmitting encrypted signal that cannot be regenerated by a spoofer and thus can be used as a reference by a user receiver. The other is a modified RAIM algorithm implemented in a user receiver to make the use of the authentic signal for spoofing detection. The key idea of the proposed algorithm is to check the consistency between the received GNSS spoofing signals and the authentic one. As the authentic signal is encrypted and thus cannot be deceived by a spoofing attack, it can be used as a reference to authenticate other signals. Assuming that all visible GNSS signals are deceived except for the authentic one, conventional RAIM would detect the authentic signal as a fault. However, the receiver already knows that the authentic signal is the only one that is trustworthy and thus it can determine other signals as faulty. This is a very simple but effective approach of the proposed method.

![Fig. 1. Conceptual view of GNSS spoofing environment and the additional ranging source of the proposed scheme](image)
Fig. 1 shows the conceptual view of GNSS spoofing environment and the additional ranging signal source to show how the algorithm works. The pseudorange measurements of the user receiver from N visible GNSS satellites are \( \rho_{s,N}^{u} \). Simultaneously, the spoofer, transmitting spoofing signals, sends out deception signals for all visible satellites and the pseudorange measurements of the user receiver are \( \rho_{s,N}^{e} \). Also, it may use a pseudolite or a geostationary satellite in order to transmit an additional encrypted ranging signal, which is a component of the proposed algorithm. The pseudorange measurement of the user receiver for the encrypted signal is \( \rho_{e}^{u} \). In order for the user to utilize the encrypted signal as a reference, it must be synchronized with GNSS. This issue will be covered later.

Under the spoofing scenario described above, let us assume that the user receiver has already been deceived by the spoofing signals and, therefore, observable pseudorange measurements are all spoofing signals, not GPS excepting the encrypted one, which is unable to be regenerated by the spoofer. Therefore, the pseudorange measurements of the receiver are \( \rho_{s,N}^{e}, \rho_{s,N}^{e}, \ldots, \rho_{e}^{u} \). Now, the proposed algorithm in the user receiver will be explained in two steps.

**A. Step 1 : Conventional RAIM without the authentic signal**

The first step is to apply the conventional RAIM to the received pseudorange measurements except for the encrypted one received from the additional transmitter. As we assumed above, when all observable measurements are from the faked signals by a perfectly successful spoofing, the conventional RAIM could not detect any failure. However, in the real world, considering the body effect and uncertainty of the location information of the victim user, a spoofing attack is likely to succeed only on some satellites, rather than on the entire visible satellites. Assuming that the majority of satellites are spoofed, the expected response of the conventional RAIM is one of the two. One is to raise an alarm notifying an integrity failure. This is a successful detection and there is no need to continue with the second step. The other response is to detect true GNSS signals as faulty and exclude them, and then keep calculating the user position by using the deceived signals only. This is missed detection, the limitation of the conventional RAIM against a spoofing attack. When all the visible satellites are spoofed, the first step could not raise an alarm at all.

**B. Step 2 : Spoofing detection RAIM with the authentic signal**

The measurements which have passed the first step are all faked signals and the additional encrypted one. Now, assuming that all the visible satellites are deceived, the pseudorange measurements are as follows:

\[
\begin{bmatrix}
\hat{\mathbf{y}}^* & \hat{\mathbf{y}}^* & \hat{\mathbf{y}}^* & \hat{\mathbf{y}}^* & \hat{\mathbf{y}}^* & \hat{\mathbf{y}}^* & \hat{\mathbf{y}}^*
\end{bmatrix}
\]  

(4)

The geometry matrix is reconstructed by adding a row representing the direction cosines from the user to the additional transmitter.

\[
H^* = \begin{bmatrix}
\mathbf{e}_{s,1} & \mathbf{H} & \mathbf{e}_{s,1} & \mathbf{e}_{s,1}
\end{bmatrix}
\]  

(5)

where

\[
\mathbf{e}_{s,1}, \mathbf{e}_{s,1}, \mathbf{e}_{s,1}
\]

are the direction cosines from the user to the additional transmitter.

Now, the least squares user solution, estimated measurement vector, and range residuals are calculated again with the measurement vector \( \hat{\mathbf{y}}^* \) and geometry matrix \( H^* \).

\[
\hat{\mathbf{x}}_{1S}^* = (H^* H^*)^{-1} H^* \hat{\mathbf{y}}^*
\]

\[
\hat{\mathbf{y}}_{1S}^* = H^* \cdot \hat{\mathbf{x}}_{1S}^*
\]

\[
\mathbf{w}^* = \hat{\mathbf{y}}^* - \hat{\mathbf{y}}_{1S}^*
\]

Finally, SSE and the test statistic \( D \) of the second step are defined as follows:

\[
\text{SSE}^* = \mathbf{w}^* \cdot \mathbf{w}^*
\]

\[
D^* = \sqrt{\text{SSE}^*}
\]

(7)

The test statistic in (7) is calculated from (4) and, therefore, it can pick out \( \rho_{e}^{u} \) as a fault because it breaks the consistency from the majority of faked signals. However, the receiver already knows that \( \rho_{e}^{u} \) is the only trustworthy because of the encryption. It means that the step 2 can detect a spoofing attack even when the majority of visible signals have been faked.

**IV. CONFIGURATION OF EXPERIMENTAL SPOOFING TEST**

In order to test and verify the proposed algorithm, spoofing test environment has been implemented using a multiple RF GPS simulator. It can simultaneously generate GPS signals for multiple users under the same satellite environment with an accurate timing synchronization. In general, it is used to test an anti-jamming GPS receiver with controlled reception pattern antenna (CRPA). Since the simulator can control multiple GPS simulating module using single control software, it can be used to simulate a GPS spoofing scenario as well. Fig. 2 shows the conceptual view of an experimental spoofing test set-up using two simulator modules. Each module generates GPS and spoofing signal, respectively. They are combined to be provided to the receiver using a RF combiner. The advantage of this set-up is that it is possible to provide spoofing environment for any commercial GPS receiver through the RF input to test its anti-spoofing performance. Fig. 3 shows the actual configuration of the spoofing test that consists of multiple RF GPS simulators, a RF combiner, a RF front-end, and a software GPS receiver.
Fig. 2. Conceptual view of the experimental test set-up for simulation of spoofing scenario using multiple GPS simulator modules and GPS receiver

Fig. 4 shows the response of the IFEN SX-NSR software GPS receiver under the experimental spoofing test using the configuration shown in Fig. 3. The scenario is to deceive a static user into moving. First, it can be seen that the spoofing signal, as well as GPS signal, is successfully generated (see the autocorrelation function in the upper left part of Fig. 4). As the spoofing signal is getting close to the original GPS signal, the tracking loop of the receiver loses its lock on the GPS signal and then starts tracking the spoofing signal (see Steps 2 and 3 in Fig. 4). Finally, the receiver is deceived and its position solution is led to a faked location. The response of the software GPS receiver, autocorrelation function, and positioning result against a simulated spoofing attack can confirm the validity of the implemented test set-up.

V. EXPERIMENTAL TEST RESULTS

The proposed algorithm has been tested using the same configuration as described above, except for the fact the victim GPS receiver is a COTS, not the software receiver. A Novatel FlexPak6 GNSS receiver, based on OEM628, has been used to get comparison results and to collect raw measurements. The proposed algorithm has been implemented in MATLAB to post-process the collected measurement. The testing has been carried out in two types. First, we performed a spoofing test on the Novatel receiver and monitored its response. The results of this test could be regarded as a level of the spoofing detection capability of legacy RAIM algorithm of a COTS GNSS receiver. Secondly, we tested the proposed algorithm by post-processing the raw measurement collected from Novatel receiver.

A. Spoofing tests on a conventional RAIM implemented on a COTS GNSS receiver

Three case studies were carried out with respect to the number of spoofed satellites. Cases 1, 2, and 3 were as follows: only one satellite was spoofed, four of seven visible satellites were spoofed, and all visible satellites were spoofed, respectively. Fig. 5 is the screen shot of the monitoring program of Novatel receiver showing the response of the receiver for each case. The top figure for each case shows the number of tracked satellites and solution status, indicating the normal condition or integrity warning. The figure in the middle illustrates the received signal strength for each channel. During the test, all GPS signals were set at the same signal strength and that of the spoofing signals was set to be 3dB stronger than GPS. Therefore, checking the signal power level makes it possible to establish which channel has been spoofed.
For case 1, when only one satellite was spoofed, the receiver can detect the spoofed satellite and exclude it to provide an appropriate navigation solution. For case 2, when most visible satellites were spoofed, the receiver raised an alarm of integrity warning and did not provide any solutions. The RAIM must have detected inconsistency between measurements. These two cases can be considered to demonstrate that the conventional RAIM functioned well. For case 3, all satellites were spoofed, the receiver could not detect the spoofing and kept calculating the misled position solution without alarm. It shows the vulnerability of the conventional RAIM against a spoofing attack. This is the result from the geodetic grade Novatel receiver. We have tested several kinds of COTS receivers under the same scenarios, most of them could not raise an alarm and keep providing the misled position, even for case 2.

B. Spoofing test on the proposed algorithm with post-processing

The proposed algorithm was tested using the measurements from the Novatel receiver with post-processing. As described in Section III, the proposed algorithm requires an additional encrypted ranging source. To simulate the additional signal, PRN 4 in Fig. 6 has been assumed as the encrypted ranging signal by letting the spoofing simulator not to generate spoofing signal for PRN 4 and setting the receiver not to use the measurement to calculate the position solution. Therefore, the response of the Novatel is similar to the results shown in Fig. 5. Fig. 6 shows the sky plot of the test scenario. Although the receiver did not use PRN 4, the only authentic signal, for calculating position solution, it tracked the signal and stored the pseudorange measurement to be used for the post-processing of the proposed algorithm. 

In this paper, we would like to address the results of the proposed algorithm for two cases. One is for the four spoofed signals of six visible satellites, the case when most of the visible satellites was spoofed. The other is the case when all visible satellites were spoofed. For both cases, PRN 4 was used as an additional authentic signal for the proposed algorithm. However, the conventional algorithm was set not to use it. Fig. 7 and 8 show the results of two cases, respectively. For the comparison, the conventional RAIM was implemented with post-processing as well.

As can be seen in Fig. 7, both conventional and the proposed RAIM can detect spoofing. The red line is the threshold to raise an alarm and the blue line is the test statistic described in (3) and (7). The spoofing signal started to affect the tracking loop at about the epoch time 250s. Before 250s, for both RAIMs, the test statistic is below the threshold. Then, the spoofing was completed at about the epoch time 380s. After that, the test statistic for both RAIMs exceeded the threshold and could raise an alarm for the spoofing attack. As shown in case 2 of Fig. 5, the conventional RAIM could cope with the spoofing attack performed only on the part of visible satellites.

Fig. 8 shows the result when all the visible satellites were spoofed. In this test, the spoofing started at about 300s and was completed at about 490s. As can be seen in the left figure of Fig. 8, after the completion of the spoofing attack, the test statistic is still below the threshold, meaning missed detection.
This is identical to the result of the Novatel receiver shown in case 3 of Fig. 5. However, the test statistic of the proposed algorithm exceeds the threshold after about 490s, meaning that it can raise an alarm even when all satellites were deceived.

VI. IMPLEMENTATION ISSUES

The proposed scheme has the advantage as compared to the other anti-spoofing techniques. Specifically, the user receiver does not require additional hardware equipment and, therefore, can be applied to a variety of applications. On the other hand, well-known effective anti-spoofing methods, such as angle-of-arrival discrimination method using array antenna and GPS P(Y) codeless correlation method between multiple receivers, have constraints in shape and size [7].

However, the need for an additional ranging signal is a disadvantage of the proposed scheme, from the perspective of service provider. Despite this limitation, the proposed scheme can be a good measure to provide anti-spoofing capability to many users by simply adding a single transmitter. An eLoran, a pseudolite or a geostationary satellite could be used for that purpose with synchronization with GNSS for ranging. Especially, the eLoran, being considered as an alternative navigation system, could be a good solution [8]. Furthermore, the requirement for the accuracy of the timing synchronization and the accuracy of the location of the transmitter does not need to be tight, since it is not used for calculating the position solution. The detection threshold could be set in accordance with the available ranging accuracy.

VII. CONCLUSIONS

In this paper, a very simple but effective spoofing detection method has been proposed. The proposed method uses an additional encrypted ranging signal to check the consistency between GNSS spoofing signals and the encrypted authentic one. We implemented a RF spoofing test configuration using multiple a RF GPS simulator and it has been used to check the validity of the proposed algorithm. First, to check the anti-spoofing performance of the conventional RAIM, we performed the test on a COTS GNSS receiver. The results showed the limitation of the RAIM against a spoofing attack. The proposed algorithm was implemented in MATLAB for post-processing. Using the raw measurement collected from the Novatel receiver, its spoofing detection performance has been confirmed even when all the visible satellites were spoofed, whereas the conventional RAIM could not cope with the task.

As is well known, the best solution for a spoofing attack is using an encrypted ranging signal that cannot be regenerated. However, the accessibility to GPS P(Y) code signal is very limited and Galileo commercial service is not yet available. At this moment, the proposed algorithm can be a viable alternative by simply adding a single transmitter. As mentioned above, except for a simple software implementation, the receiver does not require any hardware modification or additional equipment.

Our future work is to implement the algorithm on a real-time operable receiver with a pseudolite or an eLoran as an additional signal transmitter. Furthermore, the recovery algorithm that can distinguish the spoofing signals and control the receiver to keep tracking the authentic signal will be studied.

REFERENCES

Session H
Radar, electronic maps, special sensors.
Characteristics of Echo Fluctuation and Echo Paint from the Small FRP Boat on the Marine Radar Screen

Tatsuto Yamada
Maritime Science and Technology
Japan Coast Guard Academy
Kure, Japan
yamada@jcga.ac.jp

Shinji Mizui
Department of Maritime Technology
National Institute of Technology, Hiroshima College
Oosakikamisima-cho, Japan

Akihiko Tsukisaka
Department of Waterfront and Urban Planning
Chuden Engineering Consultants Co., Ltd
Hiroshima, Japan

Ruri Shoji
Faculty of Marine Technology
Tokyo University of Marine Science and Technology
Tokyo, Japan

In this study, authors examine the characteristics of echo fluctuation and echo paint of the point target such as the small FRP boat. In echo fluctuation, we discuss mainly the probability density of RCS by comparing with the distribution of actual data and the Swerling target model. In echo paint, we discuss an echo size and shape on the raster-scan synthetic display which is presented by the cartesian co-ordinates. The evaluated models for the echo size and shape on the marine radar screen are proposed.

Radar, Radar Cross Section, Echo Paint, Radar signal processing

I. INTRODUCTION

AIS is the very variable nautical instrument to get ship’s information around given ship. But it is hard to install AIS on all small ship such as pleasure boats or fishing boats under 5 tons. From this viewpoint, radar is one of the most important tools to detect these boats and navigators are required to use radar perfectly. In particular, under poor visibility conditions such as night, a foggy day and so on, it is necessary to get appropriate information by radar for safety navigation. However, it is difficult to detect these boats on the radar screen at a greater distance since the Radar Cross Section (RCS) of these boats is so low. So, it is important for the navigator to grasp the characteristics of echo strength and echo fluctuation of these echoes on the radar screen. Additionally, the radar image of these boats are overlapped with other targets since these boats are often located closely each other or other targets, namely these are related to the size and shape of radar echo on the radar screen, which are called “echo paint”. Due to this reason, it is also important to realize the characteristics of echo paint. In order to bring out these characteristics, authors carried out an observation experiment to get radar echo of a small FRP boat at the front sea area of the Japan Coast Guard Academy (JCGA) using a system for sampling the radar echo signals in high frequencies by A/D conversion. We explore the characteristics of echo strength, echo fluctuation and echo paint based on these data.

At the First, in chapter II, we describe about the outline of experiments and echo strength which was measured from the reception signal intensity of small FRP boat. In chapter III, echo fluctuation as the fluctuation of RCS of small FRP boat is discussed. The results of analysis with regard to echo paint, namely the size and shape of echo, is discussed in chapter IV. Finally, we present our conclusions in chapter V.

This work was supported by JSPS KAKENHI Grant Number 23510216.
II. OUTLINE OF THE OBSERVATION AND ECHO STRENGTH

The observation experiment to get radar echoes from the small FRP boat was carried out at the front sea area of the JCGA [1]. The weather was clear with light wind. Fig.1 shows the image of the observation area and measurement. TABLE 1 shows the dimension of the small FRP boat which is named “L2” owned by the JCGA. TABLE 2 shows the specification of the radar in use. We got data on the situation with the reflector and without reflector to compare with echo strength each situations and “L2” was positioned at two different ranges from the radar. Radar reflector was installed in the height of about 2.5 meters on the sea level. In this observation experiment, we prepared two type radar reflectors which are type approval by the Ministry of Land, Infrastructure, Transport and Tourism in Japan [2]. Main characteristics of two type reflectors from each catalogues are shown in TABLE 3. We took care so that the boat’s heading had been kept 090 or 000 degree as much as possible during measurement. Therefore the radar reflective signal from the abeam or the bow of the boat was always received. These situations were observed by the radar which is installed on the training ship “Hikari” owned by National Institute of Technology, Hiroshima College. Since “Hikari” was moored at the pontoon, the height of radar antenna is 3.6 meters regardless of the tide. Fig.2 shows “L2” with the reflector and the training ship “Hikari”. We measured 40 times in each situation. The sampling system achieved 1.5 meter of the range cell size by the sampling frequencies of 100 megahertz and was able to measure the radar video signal in voltage as 12-bit digital values. As the pulse repetition frequency was 2,000 hertz and the revolution per minute of antenna was 24, the angle of one sweep was 0.072 degree of the bearing cell size. So the video signal is stored on the memory according to its range from the origin and bearing from the heading marker. After measurement, the video signal was converted to the signal intensity in decibel-milliwatt by calibration data. Fig.3 shows the relation between the signal intensity of the radar echoes from the FRP boat and the distance in each situation. In this figure, the horizontal axis is the distance in meter, the vertical axis is the signal intensity in decibel-milliwatt, open marks mean the case of abeam-aspect (i.e. aspect angle is about 90 degree) and other marks mean the cases of bow-aspect (i.e. aspect angle is about 0 degree). And circles are the signal intensity on echoes with no reflector, squares and x marks are that with reflector A, and triangles and + marks are that with reflector B. Since the noise level of this radar receiver showed about -85 (dBm) we analyzed the reflective signal of the level exceeding -80 (dBm). Several echoes were not taken out the radar image around 1,500 meters range from radar because the signal intensity of echo was low under -80 (dBm). The signal intensity of all situations is shown the fluctuation in the range of 20 (dB) and more. The fluctuation of cases of bow-aspect is larger than that of abeam-aspect.

III. ECHO FLUCTUATION

A. Basic Theory

The received signal power from the target (p_r in milliwatt) is calculated as follows [3]
where $p_t$ is the transmitted power in milliwatt, $G$ is the antenna gain, $\lambda$ is the wavelength in meter, $\sigma$ is the RCS in square meter, $d$ is the distance between antenna and target in meter, $h_1$ is the height of antenna and $h_2$ is the height of target in meter. The unit of radar signal power is decibel relative to decibel-milliwatt (dBm), which is defined in (2)

$$P_r = 10 \times \log (p_r)$$

The received signal power is dependent upon the prime characteristics which are aspect, surface texture, material, shape and size of the target. According to (1), influences of such the prime characteristics appear the amount of RCS. Especially, the RCS of complex targets such as ships are complicated functions of viewing aspect. A complex target may be considered as comprising a large number of independent objects that scatter energy in all direct ion. The relative phases and amplitudes of the echo signals from the individual scattering objects as measured at the radar receiver determine the total cross section. The phases and amplitudes of the individual signals might add to give a large total cross section, or the relationships with one another might result in cancellation [4]. Therefore RCS is presented by the probability density function. A popular method for presenting echo fluctuation of targets is the four statistical models described by Peter Swerling [5]. It is said that the RCS of echo from the small complex target is presented by the case 3 in these models [4]. The probability density function $(P(\sigma))$ is given by

$$P(\sigma) = \frac{4 \times \sigma}{\sigma_m} \times \exp \left( - \frac{2 \times \sigma}{\sigma_m} \right) \quad \sigma \geq 0$$

where $\sigma$ is the target RCS and $\sigma_m$ is the average value of target RCS in square meter. We examine and discuss RCS by the comparison of (3) and the actual observation data.

### B. Comparison with the Actual Data

By judging on the actual condition, we assumed that the effective reflection height without reflector was 1.5m and that with reflector was 2.5m. The average value of RCS ($\sigma_m$) is estimated by adapting the theoretical carve to the average signal intensity. These results are shown in TABLE 4.

Each calculated RCS are classified into nine groups per unit interval which is 10 to half power, and are totaled for every class. Since we estimated the average value of RCS ($\sigma_m$), we get the probability density using (3). Fig.4 shows the probability density of calculated RCS in the case of abeam-aspect. In this figure, the horizontal axis is the RCS in square meter by logarithm, the vertical axis is the probability density, thin line is presented the probability density of Swerling target model, thick line with solid circles is the probability density without reflector, thick line with open squares is that with reflector A, thick line with open triangles is that with reflector B, thick line with star is that with reflector C.

**TABLE 4. Estimated value of average RCS**

<table>
<thead>
<tr>
<th></th>
<th>RCS(m²) (Aspect 90°)</th>
<th>RCS(m²) (Aspect 0°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>without reflector</td>
<td>3.5</td>
<td>1.5</td>
</tr>
<tr>
<td>with reflector A</td>
<td>8.5</td>
<td>5.5</td>
</tr>
<tr>
<td>with reflector B</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

We can calculate the RCS using (4) in each situation on Fig.3. Each calculated RCS are classified into nine groups per unit interval which is 10 to half power, and are totaled for every class. Since we estimated the average value of RCS ($\sigma_m$), we get the probability density using (3). Fig.4 shows the probability density of calculated RCS in the case of abeam-aspect. In this figure, the horizontal axis is the RCS in square meter by logarithm, the vertical axis is the probability density, thin line is presented the probability density of Swerling target model, thick line with solid circles is the probability density without reflector, thick line with open squares is that with reflector A, thick line with open triangles is that with reflector B, thick line with star is that with reflector C.
reflector A and dotted line with open triangles is that with reflector B. The probability density of RCS of the small FRP boat with reflector almost adapt to the Swerling target model. But, the probability density of the condition without reflector does not quite adapt to this model. In the case of bow-aspect, the same tendency was indicated. For example, from this Figure, if the RCS will be under 1 square meter, the probability density of RCS in the case without reflector and with reflector are about 0.2 and about 0.1, respectively. Thus echo fluctuation is evaluated as the probability density of RCS.

IV. ECHO PAINT

A. THEORETICAL CONCEPT AND ACTUAL DATA

Fig.5 shows a situation in which a radar beam sweeps across a point target and illustrates how the echo paint is built up [6]. As the beam sweeps across the point target, it will be struck at regular intervals (the pulse repetition rate) by a number of pulses. Thus the echo paint is built up of a number of the radial elements each of which is a radial length. The angular width of the paint of the point target depends on the horizontal beam width. This is illustrated by Fig.5 (b). The length of the arc which subtends the angular beam width increases with the range of the target. Fig.5 (c) shows the effect of pulse length. Each pulse which strikes the point target and is successfully detected produces the paint on screen which has a radial extent related pulse length. The most popular explanation is that the aggregate of the each effect composes the total echo paint which is shown by Fig.5 (d). However, the angular width and the radial length of echo paint on the actual radar screen have close relation to the signal intensity, the setting sensitivity, the beam shape and the pulse shape [7].
In this analysis, the threshold level was set -70 (dBM). We measured the angular width and the radial length of the echo data. Fig.6 shows the relation between the angular width of echo and the distance in each situation. The horizontal axis is the distance in meter and the vertical axis is the angular width in degree. Open marks mean the case with reflector and solid marks mean the cases without reflector. Circles and triangles are the case of bow-aspect, and squares and diamonds are the cases of abeam-aspect. From this figure, the variation of the angular width shows a same tendency, independent of the aspect angle. On the other hand, measured value in the situation with reflector is bigger than that without reflector. In the case of the radial length of echo, the same tendency was indicated. It follows from this that the angular width and the radial length of echo paint depend on the echo strength; it seems that the distribution of the signal intensity of echo on the screen is related to the beam shape and the pulse shape.

B. Evaluated Model of echo paint

Fig.7 shows the relation between the angular width and the radial length of echo and the echo strength in each situation. In Fig.7 (a) and (b), the vertical axis is the angular width in degree and the radial length in meter, respectively. The horizontal axis is the relative strength which is the differential value between the echo strength and the threshold level in decibel. The respective lines in these figures are determined by the regression analysis as shown (5) and (6).

\[
\alpha = 0.1324 \times dP + 0.1194 \quad (5)
\]

\[
\beta = 0.519 \times dP + 14.97 \quad (6)
\]

where \(\alpha\) is the angular width in degree, \(\beta\) is the radial length in meter and \(dP\) is the relative strength in decibel. As each correlation coefficients are more than 0.8, we can conclude that the angular width and the radial length have a significantly correlation with the relative strength. And from these equations, we can estimate the angular width and the radial length of echo when the echo strength and the threshold level are assumed.

By the way, taking into account the actual radar display, it is necessary to consider the effect of the raster-scan synthetic display which is presented by the cartesian co-ordinates. Fig.8 shows the relation between the polar co-ordinates and the cartesian co-ordinates. Fig.8 (a) shows that a target locate at a distance \(D\) in miles and a bearing \(\theta\) in degree. Such storage is said to be in term of the polar co-ordinates. In the raster-scan synthetic display, the co-ordinate conversion is carried out by a digital scan conversion. For any given polar location specified by the range and the bearing, the rectangular co-ordinates will be given by \(x\) and \(y\) in Fig8 (b). Such storage is said to be in term of the cartesian co-ordinates. So we should analyze the characteristics of echo paint by which convert the image from the polar co-ordinates to the cartesian co-ordinates. The caetesian location \(x\) and \(y\) correspond to memory cells according to screen size. Commonly available cell size is 512, 768, 1024 or 1360 per horizontal or vertical dimension [5]. The echo paint on the cartesian co-ordinates have a close relation to a setting range scale and a memory cell size, except for the angular width and the radial length of echo. The cell size of echo on the cartesian co-ordinates are calculated as follows.

\[
dx = \frac{N \times d \times \sin(\alpha)}{2 \times R \times 1852} \quad (7)
\]

\[
dy = \frac{N \times \beta}{2 \times R \times 1852} \quad (8)
\]

where \(dx\) is the horizontal cell size of echo, \(dy\) is the vertical cell size of echo, \(N\) is the maximum cell size on the screen, \(R\) is the range scale in mile and \(d\) is the distance from antenna to target in meter. \(dx\) and \(dy\) are positive integers. An echo size and an echo shape can be evaluated as the aggregate the horizontal cell size and the vertical cell size, namely we propose that the echo size and the echo shape are represented by the product of \(dx\) and \(dy\), and the ratio of the \(dx\) and \(dy\), respectively. Fig.9 shows these relations. The echo size and the echo shape are calculated as follows.

\[
\text{echo size} = dx \times dy \quad (9)
\]

Fig.9. The image of cell size of target echo on the cartesian co-ordinates
\[ \text{echo shape} = \frac{dx}{dy} \quad (10) \]

From (10), when the value of echo shape ratio is smaller than 1, echo shape stretches to the radial direction, and when the value of that is larger than 1, echo shape stretches the angular direction. Substituting (5) into (7) and (6) into (8) give the modeled cell size of echo in this case. Fig.10 shows the echo size and the echo shape which are calculated by using (1), (5) through (10) and the value defined in TABLE 2 and TABLE 5. In this figure, the horizontal axis is distance in meter. The left vertical axis is the ratio which is indicated echo shape and the right vertical axis is the number of cells which is indicated echo size. From this figure, the value of echo shape ratio is gradually small by more than 1500 meters and the echo size is gradually small by more than 1000 meters. In general, echo shape stretches to the radial direction and echo size is smaller at a shorter distance, and echo shape stretches to the angular direction and echo size is bigger at a greater distance. These qualitative characteristic is shown in Fig.5 (d). On the other hand, as the signal intensity becomes weak when the distance between the target and the antenna is far, the relative strength (the differential value between the echo strength and the threshold level) falls. From Fig.7, when the relative strength is small, the angular width and the radial length of echo become small. So, we should consider that the aggregate of effect by beam shape, pulse shape, signal intensity and setting sensitivity composes the actual echo paint on the marine radar screen. And we can simulate and evaluate quantitative the echo paint of the small boat on the raster-scan synthetic display by (9) and (10) if we assume some calculation conditions.

V. CONCLUSIONS

To recognize radar images correctly, the navigator is particularly required to grasp the relation between the echo strength and the threshold level, the echo fluctuation which influence the target detection, and the distortion effects which influence the angular width and the radial length of echo, on the actual radar screen. In this paper, we analyzed mainly the echo fluctuation and the echo paint of the small FRP boat by using the data which was obtained by the observation experiment. The echo fluctuation were evaluated as the average value of RCS and the probability density of RCS. The echo paint was evaluated as the aggregate the horizontal cell size and the vertical cell size on the raster-scan synthetic display which is presented by the cartesian co-ordinates. This research gives the fundamental data and method for evaluating the echo fluctuation and the echo paint of the small boat. However, the further investigation is also required about these characteristics.

ACKNOWLEDGMENT

We would like to thank Capt. Akimoto and other crew members of “Hikari” (National Institute of Technology, Hiroshima College), Mr. Nakagawa (the SEAMEC Co.,Ltd.) and Mr. Kajitani (the Chuden Engineering Consultants Co.,Ltd.), for their assistance and cooperation in correcting the data. Furthermore, we would like to thank the Nippon Foundation and the Japan Coast Guard Foundation for support.

REFERENCES

Availability of Navigational Fusion System for Small Domestic Vessels

Yoshifumi Ichikawa
Professor of Navigation Department
Marine Technical College
Ashiya, Japan
ichikawa@mail.mtc.ac.jp

Shigeyuki Okuda
Professor of Navigation Department
Marine Technical College
Ashiya, Japan
okuda@mail.mtc.ac.jp

Akihiko Hori
Director of Educational & Administrative Planning Department
Independent Administrative Institution Marine Technical Education Agency
Shizuoka, Japan
hori-a@mtea.ac.jp

Yasuo Arai
Professor emeritus
Marine Technical College
Ashiya, Japan
arai@mail.mtc.ac.jp

Masatoshi Shimpo
Department of Navigation School of Marine Science and Technology
Tokai University
Shizuoka, Japan
shimpo-masatoshi@tsc.u-tokai.ac.jp

Abstract—Since AIS coming into being onboard, it becomes more convenient to get the navigational information. It will be required that the fusion of navigational information shall be proceeded to increase the reliability, according to satisfaction both of safety and efficiency of marine navigation. In a past study, authors concluded for the fusion of navigational information which are ARPA and AIS and/or GNSS should be essential to keep and to enhance the safety navigation. In this paper, we proposed new method for the association of ARPA and AIS as part of the fusion system. Furthermore, we developed new algorithm which includes the association and automates it and improves the reliability of data.

Keywords—ARPA (TT), AIS, Fusion, Association, Navigational Information, Domestic Vessels

I. INTRODUCTION

Since AIS coming into being onboard, it becomes more convenient to get the navigational information and also in the near future it is expected that a lot of navigational information will be fulfill as the development of e-Navigation will be proceeding, and it will be required that the fusion of navigational information shall be proceeded to increase the reliability, according to satisfaction both of safety and efficiency of marine navigation. Authors presented the characteristics and performances on the ARPA information such as true vector, heading and COG (Course Over the Ground) as the first step of proceeding the fusion of nautical information, and concluded for the fusion of navigational information which are ARPA data and AIS data or GNSS data to avoid collision should be essential to keep and to enhance the safety navigation in [7].

Authors precede the study on fusion between ARPA and AIS and/or GNSS which is one of examples. After this, the system which fuses plural navigational information is needed to increase the precision of obtained information. So this system is named “Navigational Fusion System”, but there are many problems to realize this system. Especially, the vessels of which number is maximum sailing along coast are domestic vessels, and the performance of individual systems in domestic vessels is inferior to that of large ships and/or ocean-going vessels. Although the domestic vessels while operating similar systems as large vessels, an environmental condition is different or conditionality comparing with large ships. Small domestic vessels less than 500 GT in the SOLAS, have no carriage requirement of gyrocompass, but can support for carriage requirement of THD. However, as for the domestic vessels, there is performance degradation by lower antenna height of AIS and/or RADAR, and their performance is inferior to the performance said in general in [4] and [5]. Although the almost results of marine accidents are caused by human errors, it is essential not only to educate and train the seafarers but also to grade up performance of navigational systems for small domestic vessels. So, it should be necessary to validate and achieve that system performance for the domestic vessels should be able to improve by fusion of this navigational information.
In the previous paper [1], the establishment of ARPA information was discussed systematically to fuse between ARPA and AIS. Furthermore, we proposed to devise a countermeasure which decrease ARPA information noise according to check of the time rate of range, bearing and their combination. RADAR which has been installed after resolution of MSC.192 (79) [2], added a function of Association between ARPA target and AIS target. However, association function on present ARPA is to set limitation value of difference of range, bearing, etc. by manual. Moreover, as the method of advanced association, it has been proposed a method to compare with radar raw echo and expected radar echo of corresponded ship hull. In case of large ship, there are few problems because radar echo from large ship becomes to be comparatively ideal image. On the other hand, in case of small vessels, it does not often become ideal shape, target image get behind large ship or other structure and become the complicated shape by multiple reflections [6]. So, it becomes to be source of error for DCPA and TCPA because plotting point which is assumed as target ship’s center.

In this paper, it establishes plotting point using AIS information as association method and compares with ARPA information. Moreover, it calculates DCPA and TCPA and compares with their calculated on ARPA. So, there is possibility to be able to judge it which is near to the truth.

II. REVIEW OF THE PERFORMANCE OF ARPA INFORMATION

In the previous paper [7], ARPA information was organized by systematic approach because of discussion for the performance of ARPA information. In this paper we use same coordinate, symbols and marks to be used in the previous paper for discussion of association between ARPA and AIS. So, it is described below.

ARPA information is processed from RADAR information, ship’s speed and ship’s heading. To proceed the development of fusion between ARPA and AIS, the systematic approach is necessary to do. That means to analyze not only ship’s speed but also ship’s Double Axes Velocities (DA Velocities) which are longitudinal and lateral velocities.

\[ \theta \] is ship’s heading (degree)

\[ \mathbf{S}_{OG} = \begin{bmatrix} u_g \\ v_g \end{bmatrix} \] (1)

\[ \mathbf{S}_{TW} = \begin{bmatrix} u_w \\ v_w \end{bmatrix} \] (2)

\( \varphi \) is called as COG (Course Over the Ground) when Velocities are Over the Ground.

A. Relative Position

Relative Position from own ship to target ship is measured by RADAR described in Fig. 2, and ship’s heading is obtained by gyrocompass. So, the relationship between relative position and ship’s DA OG velocities are as follows.

\[ \mathbf{r}_t(0) = \mathbf{r}_o - \mathbf{r}_g \] (4)

The distance between own ship and target ship \( r \) [nm.] and bearing \( \beta \) [deg.] are measured by RADAR and bearing will be changed to true bearing by the heading or compass data.

B. Relative / True Vector

1) Relative Vector

It is possible to get the target ship’s motion from RADAR information, and Relative Vector is shown as (5) to calculate derivative of the relative position by time.

\[ \frac{d\mathbf{r}}{dt} = \mathbf{\dot{r}} = \begin{bmatrix} r \cos \beta - r \sin \beta \varphi \\ r \sin \beta + r \cos \beta \varphi \end{bmatrix} \] (5)

2) True Vector

Before calculating the true vector, the DA OG / TW Velocities of own ship and target ship are shown as (6) and (7). These Velocities are derived from (1) and (2) which show the longitudinal and lateral velocities onboard according to easy derive true motion.

\( \mathbf{T}_o \) and \( \mathbf{T}_t \) mean the disturbance effect such as current and waves at own ship and target ship, \( \mathbf{L}_o \) and \( \mathbf{L}_t \) mean the disturbance effect such as wind, etc. at own ship and target ship.
\[ S_{OG} = S_{TW} + T_0 = (H_{TW} + L_0) + T_0 \]  
\[ S_{OG} = S_{TW} + T_t = (H_{TW} + L_t) + T_t \]

Where, \( H_{TW}\) and \( H_{TW}\) mean the longitudinal velocities TW on own ship and target ship, called ship’s speed in traditional technical words.

It is possible to detect the relationship shown as (8) according to the Figure 2.

\[ \hat{r} = S_{OG} - S_{OG} \]

So, the true vector of target ship \( S_{OG} \) is derived by (9). Here relative vector \( \hat{r} \) is obtained by RADAR, and \( S_{OG} \) is input from gyrocompass as heading and SDME.

\[ S_{OG} = \hat{r} + S_{OG} \]  
\[ S_{OG} = \hat{r} + S_{OG} \]

3) Ship’s heading

Using Equation (6), (7) and (9), it is possible to derive the target ship’s heading which is nearly equal to the direction of the longitudinal velocity on the target ship in the case of no difference of wind effect and disturbance effect between own and target ship.

\[ H_{TW} = \hat{r} + H_{TW} - (L_t - L_0) - (T_t - T_0) \]

C. DCPO and TCPA

In ARPA, it is possible to detect a possible collision to monitor the CPA (Closest Point of Approach) using (11) and (12).

\[ DCPO = |r| \cdot \sin \alpha \]  
\[ TCPA = |r| \cdot \cos \alpha / |r| \]

Where, \( \alpha = \angle |r| - \beta - \pi \)

In AIS, relative position \( r \) is driven by (4) and relative vector \( \hat{r} \) is derived by (8) using own data \( S_{OG} \) and target signal reception data \( S_{OG} \) at the time of signal reception, so AIS also provide the CPA using (11) and (12).

III. PROPOSED FUSION ALGORITHM

Fig. 3 was shown in the previous paper [7] and they are data of time series in which difference between ARPA and AIS information appeared. It is essential to be carried out averaging method in order to obtain more accurate ARPA information. Consequently, we focused that ARPA information includes time delay.

As comparison with ARPA and AIS information, AIS information is more superior to ARPA one in term of the accuracy and the time response except the time interval of information reception. In Fusion system the plotting point should be obtained with precise because of execution of correct identification and no swapping data or missing. The Information should be interpolated during one and next of AIS reception using Kalman Filter, and weighting the parameter will be set. So, the new algorithm consists of two parts: (a) plotting point detection and (b) AIS or ARPA information extrapolation.

IV. ASSOCIATION WITH ARPA AND AIS TARGET

According to MSC.192 (79) adopted in 2004 [2], RADAR which is equipped on board after July 1st 2008 was mandatory to display AIS information on radarscope. When it is considered that the tracking target by ARPA and the reception target by AIS are same one, it has been added ARPA function that ARPA can display the association symbol not to be displayed 2 symbols of ARPA and AIS target respectively. In
present ARPA, it is assumed to be the same target when
differences of ARPA and AIS information such as range,
bearing etc. are smaller than setting value by user. With
respect to Japanese manufacture, setting items to judge the
association with ARPA and AIS target are difference of range,
bearing, course and speed in case of A company and adds
difference of distance between ARPA and AIS target in case of
B company.

Under the present circumstances, all of setting is by manual
and there is neither the value that should be recommended not
the guidance. Still more, user selects to give priority which
data of ARPA and AIS. So, not a thing judging the accuracy of
the guidance. Still more, user selects to give priority which
and there is neither the value that should be recommended not
the guidance. Still more, user selects to give priority which
data of ARPA and AIS. So, not a thing judging the accuracy of
the guidance. Still more, user selects to give priority which
data of ARPA and AIS. So, not a thing judging the accuracy of
the guidance. Still more, user selects to give priority which
and there is neither the value that should be recommended not

V. PROPOSAL OF ASSOCIATION ALGORITHM

The conventional Association is the function of ARPA that it
is assumed to be one target which of ARPA and AIS are
considered to be same one. However there is little time when
the association is established because various parameters have
some error. For this in this paper, we propose the new
association algorithm that provides not only the judgement of
the association but also high reliable data which a re course,

Position (1) ~ (5) are

\[
P_{k_i} = P_{o_i} + \left(\sin \theta_i - \cos \theta_i \right) \begin{pmatrix} x_{k_i} \\ y_{k_i} \end{pmatrix}
\]

(16)

Where \( k = 1 \sim 5 \), and each position is

\[
\begin{align*}
(x_{1i}) &= \frac{L}{2}, & (x_{2i}) &= \frac{(L-B)/2}{2}, & (x_{3i}) &= -\frac{L}{2}, \\
(y_{1i}) &= 0, & (y_{2i}) &= -\frac{B}{2}, & (y_{3i}) &= \frac{B}{2}.
\end{align*}
\]

(17)

Where \( L \) is Ship’s Length and \( B \) is Breadth.

Range and bearing of target which are obtained on ARPA
are made by RADAR’s raw data. Range and bearing of target
which are obtained from AIS reception data are made by GPS
position information of target ship. Namely, in fact the
association on ARPA compares with different data source.
Range and bearing of ARPA target are plotting point for
tracking. In order to compare with ARPA and AIS in the same
standard, using AIS data, it is estimated the plotting point
which is decided on ARPA, so it should be compared with
range and bearing of ARPA information and estimated the
plotting point which is calculated by AIS data.

Position of AIS reception data is antenna position of target
corresponding \( P_{ANT_i} \) in Fig. 5. If the plotting point on ARPA
represents the center of target ship, it should be found \( P_{o_i} \).
However, finding \( P_{o_i} \) is impossible because \( P_{o_i} \) does not
coincide the plotting point which is decided by analysis of
radar echo in [1], [3] and [8]. Consequently, in actual ARPA
it is guessed that the plotting point gets as the center of target
or minimum Difference of Range and Bearing, DCPA
and TCPA distance and the center in azimuth which means
range of forward edge of target echo and the center in azimuth
direction of it. In this paper, we carried out numerical
simulation that the plotting point is decided as minimum
distance and the center in azimuth which is able to estimate by
using AIS information.
First of all, it is estimated type of target ship by obtained from AIS information shown in Fig. 6. In this paper, because it was aimed for constructing algorithm, it was considered the model of simplified target ship shown in Fig. 6.

At feature point from $P_{ij}$ to $P_{ij}$, it is extracted points to generate radar echo, and it decides distance, and minimum and maximum degree of azimuth direction. Finally, $P_i$ is estimated as the plotting point shown in Fig. 4. Fig. 6 shows differences between Calculated by AIS and ARPA.

$\Delta R$ is the difference between AIS and Radar Position. $\Delta B$ is the difference between AIS and Radar Position.

In comparing DCPA and TCPA, (plot $p'$ t) means calculated by AIS data, and (ARPA) means measured by ARPA.
of range and bearing, DCPA and TCPA of $P_F$ and ARPA information at the same time of Fig. 3.

Accordingly, we suggest the identification algorithm.

1) To find AIS target in which it uses window to acquire and/or to track.
2) To calculate difference between estimated plotting point using AIS data and the plotting point by ARPA.
3) If its difference is within a certain value, it is considered that they are the same target.
4) If target made by using AIS data and radar echo exist in the same window, it is also considered that they are the same target.
5) To execute 1) to 4) at the end of ARPA calculation.
6) Finding a difference using AIS data and/or dead reckoning calculation, if its difference is less than a certain value, to calculate and display CPA etc. using AIS data.
7) To calculate and display CPA etc. using AIS data established shortly before when ARPA can not receive AIS data, if prediction error become to be big, AIS data will be lost and it will be used only ARPA data.
8) If prediction error is within a certain value, it is carried out 6).

Results of numerical simulation are shown in Fig. 6. At about 8 minutes, it was occurred big variation of ARPA information. At about 18 minutes, time delay of ARPA data occurred. In both case, DCPA and TCPA using AIS data are reasonable.

VI. CONCLUSION

As described above, we proposed new method for the association of ARPA and AIS as part of the fusion system. Furthermore, we developed new algorithm which includes the association and automates it and improves the reliability of data

Finally, we describe below some future problems.

Used target data exists near 10 nm. With respect to analysis of ARPA plotting point, we understand that the plotting point of closed target changes drastically. On the other hand, range and bearing data of long distance target are not constant relation to echo resolution. It will be necessary to validate that proposed fusion system can resolve for these target situation.

REFERENCES

3D ECDIS Implementation Base on Area Level of Detail Rendering Techniques

Shengzheng Wang, Chaojian Shi, Baojia Xiao, Jing Peng
Merchant Marine College
Shanghai Maritime University
Shanghai, China
szwang@shmtu.edu.cn

Abstract—3D ECDIS has aroused widely interest in the research community of the state of the art marine navigation equipments since it can provide an effective technique platform for developing e-Navigation. However, there is still not an effective solution to implement 3D ECDIS due to the display problem of the sea-bed terrain. In order to implement 3D ECDIS, an area-oriented level of detail (ALOD) rendering algorithm is proposed to optimize the rendering of the sea-bed terrain dataset. Namely, the navigational safe levels of the waterways around the ship and the real-time navigational speed of the ship are utilized to construct ALOD rendering model, which is used to control GPU tessellation program, and thus the sea-bed terrain can be fast rendered with high qualities since the number of 3D vertexes is largely reduced on the premise of preserving the geometric accuracy. The effectiveness of the proposed method is demonstrated by performing quantitative and qualitative analysis.

Keywords—Area Level of Detail; 3D ECDIS; Marine Navigation; E-Navigation

I. INTRODUCTION

Since 2012, Electronic Chart Display and Information System (ECDIS) has been one of the official navigation equipments with which ships were equipped, the most important contribution is that it has implemented the transition from the traditional paper charts into the electrical navigation charts and thus brought about very important significance to ship navigation and safety. The application of ECDIS are very useful, however, from the view of human cognition, ECDIS based on two-dimensional vector chart still needs the crew to understand the meaning of various data by combining the priority knowledge with their experience and cognition. For example, for soundings, buoys etc., crew must use the brain cognition and knowledge to understand the internal meanings of the data and then translate them into the useful navigational information. However, misunderstanding would make a mistake and even cause fatal accidents. Thus, Ford et., proposed firstly the ideas and the framework of three dimensional electronic chart display and information system (3D ECDIS)\[1\], and they found that the 3D visualization of a chart is an effective means to help the crew quickly understand the navigation information and make a proper decision so as to reduce the risks of ship navigation.

Furthermore, 3D ECDIS usually depends on the 3D computer graphics to display space geographic information. That is to say, the real world is described more accurately, which would be useful for the crew to more intuitively read the geographic information rather than translate the data into the recognized information by the brain cognition model. Recently, Ray[2] who referred to the design concept of 3D ECDIS proposed the basic framework of the implementation of 3D ECDIS based on the idea of the 3D GIS\[3\].

But in the implementation of the 3D ECDIS, one of the key problems is how to implement effectively real-time display for the large ranges of the sea-bed terrain. In the past two decades, many research efforts have been dedicated to resolving this problem, in the most of the methods, the soundings and the depth contours were extracted from the electrical navigation chart (ENC) and were used to reconstruct the terrain by employing the surface fitting algorithms\[4-8\]. The focus of these algorithms is to look for a robust method to reconstruct the grid surface so as to render the sea-bed terrain. Although the techniques of the level of detail (Level of Detail, LOD) were developed to simplify the meshes of the surface, the balance between the performance and accuracy has been a common issue for these methods. But if keeping real-time display performance, the accuracy and the effects become very poor. Therefore, the existing methods are limited to be applied in the practical applications.

Based on the object features of ENC and the navigational requirements, we propose an area-oriented level of detail rendering algorithm and implement the real-time dynamic display of the 3D ECDIS on the premise of maintaining the precision of sea-bed topography by employing the latest graphics hardware tessellation technology. Firstly, the pseudo-normal of each vertex is calculated by applying the angle and distance weighted pseudo-normal algorithm, and the dense vertexes are simplified by adaptively controlling the change rates of the normal vectors. Thus the sparse vertexes are obtained and the density distributions of the vertexes are relatively uniform.

Then, the region around the ship is divided into different sub-regions in terms of the navigational features the chart object. In addition, the different safe levels are specified in terms of the ship navigational speed. ALOD display model is built by integrating these two factors.

Finally based on ALOD, the different accurate meshes are rendered by using the graphics hardware GPU.
tessellation technology. The experimental results show that the proposed method not only saves the large quantity of redundant information, but also eliminates a large amount of redundant information. Thus it is superior to the traditional 3D terrain display methods in display performance and efficiency.

II. 3D ECDIS SYSTEM OVERVIEW

As shown in figure 1, the framework of 3D ECDIS system mainly includes four stages: 1) Data preprocessing, including the extraction of the dense depth point cloud dataset, using the proposed pseudo-normal algorithm based on the included angles and the distances of adjacent triangles to calculate the normal vectors of the point cloud dataset; 2) Simplification of point cloud dataset, the density vertexes are simplified by adaptively controlling the change rates of the normal vectors. 3) Meshing, the proposed ALOD is used to dynamically control the quantity of the meshes, and GPU grid processing techniques are employed to advance the rendering efficiency of the geometry of the sea-bed terrain. 4) 3D ECDIS display, the geometry of 3D terrain are dynamically rendered and updated with the change of the ship position, various navigation functions are integrated into the rendered 3D scene.

In order to maintain the geometric topology, we propose an efficient algorithm to calculate vertex normal vectors based on the included angles and barycentric coordinates of adjacent triangles. As shown in figure 2 (a), the vertex \( p \) is required to calculate the normal vector, the nearest vertices (usually takes 5 ~ 8) with \( p \) are selected and are used to construct a group of adjacent triangles. Figure 2(b) shows the calculation method of the normal vector of each triangle, namely, for each triangle, the cross product between vectors is calculated and is written as follow.

\[
n_j = v_j \times v_j'
\]

Thus, the calculation method of the normal vector of the vertex is \( p \) as follows:

\[
N = \sum \frac{n_i (\theta_i / S_i)}{\sum n_i (\theta_i / S_i)}
\]

where \( n_i \) represents the normal vector of the triangle \( T_i \) at the point \( p \) that form \( k \) triangles sequence \( \{ T_i \}_{i=1}^k \), \( \theta_i \) is corresponding to the interior angle of the triangle, \( S_i \) is corresponding to the distance from triangles barycentric coordinates to point \( p \).

B. Simplification of the depth contour and sounding data

Electronic Navigation Chart(ENC) consists of two kinds of depth data, namely sounding and depth contour. The soundings are composed of the non-uniform distributed 3D points, and the depth contour is a contour line that is composed of some adjacent equal depth points. If both the dense sounding and the depth contour are directly used to reconstruct the triangulated irregular network of the sea-bed terrain, the number of the triangles is very large so that the rendering efficiency of the geometry is very low and the visual effects could not be satisfied with the requirements of the terrain. Therefore, it is necessary to simplify the dense contour and sounding data. There are many existing methods that are used to process the point cloud data in the computer graphics, but most of methods were applied to process the point cloud data of the laser scanning in which the key issue is to eliminate the noise point cloud dataset.

For 3D sounding dataset of 3D ECDIS, the most important task is to keep the geometric accuracy when the dense depth contour and sounding are simplified instead of eliminating the noise point cloud. Based on that, an adaptive simplification method is proposed to eliminate the redundant vertices based on the change rates of the normal vectors of the sounding vertices.

Fig. 2. The angle weighted pseudo-normal \( N \).

(a) the pseudo-normal, calculation, (b) the calculation of the normal vector of each triangle.

Fig. 1. 3D ECDIS system implementation framework
IV. 3D ECDIS SEA-BED TERRAIN RECONSTRUCTION

A. The weighted model based on area-oriented level of detail (ALOD)

As shown in figure 4, in the ECDIS system, the bottom of the sea areas can be divided into four layers which are shallow water area, non-safe area, safe area and deep water area. If the general terrain rendering algorithms in computer graphics are used to render the sea-bed terrain, the uniform level of detail is applied no matter what are navigable waters or non-navigable waters. Although LOD is employed to optimize the rendering of the sea-bed terrain, the large quantity of redundant data would affect the efficiency. Even so, for some particular waterways, the geometric accuracy still cannot meet the requirements of the navigation safe.

In general, in terms of the characteristics of the sea-bed terrain object and the navigation tasks, the waterway can be divided into two layers consisting of navigable area and non-navigable area. Therefore, for shallow area and deep area, the geometric accuracy is not required to be considered because the vessels cannot enter the shallow area and the ship is safe in the deep area. Thus, the focus is not the geometric accuracy but the rendering effects. On the contrary, there are some risks in the non-safe area and the safe area because the ship would pass these areas where there are some dangerous wrecks. Therefore, the geometric accuracy in these areas is significant for ship’s safety. Thus, for these areas, both the visual effects and the geometric accuracy are important for the sea-bed rendering.

Figure 5 shows the different safe areas for ECDIS, but in the actual situation, the topography of the sea-bed do not fully follow the rules that the depths are linearly changed from shallow to deep. As shown in figure 5, there are some wrecks in the navigable areas that are called as dangerous areas, which would have potential risks for ship’s safe. Therefore, in order to guarantee the ship’s safe, we should put emphasis on some special areas, such as wrecks, isolated obstructions etc., besides the surrounding areas of the ship.
is within the triangle respectively represent the shallow water can be calculated as follow two kinds of control parameters, the weighted control defined as the weighted parameters of the tessellation. By using two kinds of control parameters, the weighted control parameters can be computed as follow:

\[ d^{(-1)} = a \cdot c_j \]  

**B. Local tessellation algorithm based on the Phong model**

The traditional methods were proposed to first using point cloud data as the control nodes, the Bezier and spline surface algorithms are employed to generate the uniform grids. Then the tessellation algorithm is applied to refine each grid [16-19]. These methods can be implemented easily, and may generate the smooth surface, but the generated surface only coarsely approximates the random point cloud data, which shows the poor accuracy. For 3D ECDIS, the high accuracy of the sea-bed terrain is required. Thus, the irregular triangles are first generated directly by using the point cloud dataset. As shown in figure 6, for a triangle , 3 vertex are denoted as \( p_1, p_2, p_3 \in \mathbb{R}^3 \), the normal vectors of 3 vertices are \( N_1, N_2, N_3 \in \mathbb{R}^3 \), respectively, and 3 projection planes are formed in terms of 3 vertices and 3 normal vectors. Thus, any point \( q \) in the triangle \( t \) can be defined by the gravity weighted coordinate system, written as follows

\[ q(u, v) = [u, v, w] [p_1, p_2, p_3] \]  

\( (u, v, w) \) are the weighted parameters of the point within a triangle, which meets the following equation. \( u + v + w = 1 \) and \( u, v \in [0,1] \).

When refines the grids, vertex \( q \) within the triangle \( t \) is projected onto 3 projection planes which are formed by the triangle vertices and the corresponding normal vectors, and three projection vertices \( \{ \pi \}_{i=1}^3 \) are calculated and are written as:

\[ \pi_i(q) = q - ((q - p_i)^T \cdot n_i) n_i \]  

The local vertex coordinate \( q' \) can be calculated by using the triangle formed by 3 projection vertices and the barycenter coordinate \( (u, v, w) \) of vertex \( q \) and is written as

\[ q'(u, v) = [u, v, w] \begin{bmatrix} \pi_1(q(u, v)) \\ \pi_2(q(u, v)) \\ \pi_3(q(u, v)) \end{bmatrix} \]  

Similarly, according to Phong model, a new normal vector \( \mathbf{N}' \) of vertex \( q' \) can be calculated as follow

\[ \mathbf{N}'(u, v) = [u, v, w] [N_1, N_2, N_3] \]  

The normalized vector can be represented as follows

\[ \mathbf{N}'(u, v) = \mathbf{N}' \| \mathbf{N}' \| \]  

The task of the refinement of the grids actually is to divide a grid into many sub-grids so as to accurately represent the geometry of the sea-bed terrain. The tessellation of the grid can be calculated by traditional CPUs. However, with the increase of the number of meshes, much more RAM is required, but the performance decreases rapidly. Therefore, GPU tessellation is employed to implement the refinement of the grids so as to resolve the performance problem, and the formula (3) is used to compute the tessellation levels which are applied to control the GPU grid refining program.

**V. EXPERIMENTS AND RESULTS**

In order to validate the effectiveness of the proposed method, the rendering program of 3D ECDIS was implemented by using Visual C++ 2010 and OpenGL, and
the proposed algorithm was evaluated by using the quantitative and qualitative validation analysis.

The proposed algorithm was performed on the workstation with Intel®Core™ i7-3537U CPU@2.00 GHz processor, 8G of memory, NVIDIA GeForce GT 630M video card, memory 1G.

A. Data preprocessing

Data preprocessing mainly consists of the calculation of the normal vectors and the simplification of the point cloud data. Figure 7(a) and 7(b) show the results of the calculation of the angle and distance weighted pseudo-normal. In addition, in order to validate the advantages of the normal vector calculation method proposed in this paper, we compared it with the traditional method based on the average normal vector of the adjacent triangles, figure 7(c) and 7(d) show the results of the traditional average normal vectors and the rendering results respectively. As seen from figure 7, for the vectors of the vertices (1) ~ (5), the proposed method outperforms the traditional method by analyzing the geometric topology and the rendering results of the triangles. In addition, as seen from the rendering results of the grids, in the traditional methods, only lights are smoothly interpolated, but the geometric topology cannot be correctly maintained after the mesh refinement so that there are many non-smooth vertices. However, the proposed algorithm not only can optimize the calculation of the normal vector, but also guarantee the smooth geometric structures.

Figure 8 shows the simplification results of the point cloud dataset based on the control of the change rates of the normal vectors, which can obtain the sparse point cloud dataset with the premise of without changing the geometric topology and accuracy.

B. The sea-bed terrain rendering based on area level of detail model

Figure 9(a) shows the triangular irregular network (TIN) rendering effects of the sparse point cloud dataset which is obtained through preprocessing. Figure 9(b) shows the improved rendering results based on ALOD. As seen from figure 9, the proposed ALOD method would not refine the triangles of the land areas and shallow areas, but the important waterways would be refined in terms of the ship’s speed and the safe levels and the more detail geometric meshes are displayed in the wreck areas.
In order to validate the effectiveness of the 3D ECDIS rendering method, we tested the geometry errors, and compared it with the traditional B-spline surface fitting method. Table 1 shows the geometry errors in different situations by simplifying the 6236 random soundings and depth contour dataset with different simplification factors. As seen from table 1, the proposed method outperforms the traditional surface fitting methods about the geometry errors. The average errors are less than 1m, which is significant for navigation safety.

<table>
<thead>
<tr>
<th>C_1(%)</th>
<th>C_2</th>
<th>C_3</th>
<th>C_4/%</th>
<th>C_5/m</th>
<th>C_6/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6236</td>
<td>2728</td>
<td>56.25</td>
<td>0.75</td>
<td>2.12</td>
</tr>
<tr>
<td>10</td>
<td>6236</td>
<td>1626</td>
<td>73.92</td>
<td>0.68</td>
<td>2.14</td>
</tr>
<tr>
<td>15</td>
<td>6236</td>
<td>1140</td>
<td>81.71</td>
<td>0.67</td>
<td>2.14</td>
</tr>
<tr>
<td>20</td>
<td>6236</td>
<td>929</td>
<td>85.1</td>
<td>0.66</td>
<td>2.13</td>
</tr>
<tr>
<td>25</td>
<td>6236</td>
<td>821</td>
<td>86.83</td>
<td>0.70</td>
<td>2.13</td>
</tr>
<tr>
<td>30</td>
<td>6236</td>
<td>781</td>
<td>87.47%</td>
<td>0.71</td>
<td>2.13</td>
</tr>
</tbody>
</table>

*C_1*: Change rates of normal vectors  
*C_2*: Quantity of Original Sounding  
*C_3*: Quantity of Sounding after Simplification  
*C_4*: Simplification rates/%  
*C_5*: Average Errors of the proposed method /m  
*C_6*: Average Errors of B-Spline /m

Figure 11 shows the rendering results of 3D ECDIS, figure 11 (a) shows the overview visual effects of a sea-port, figure 11 (b) is the local rendering effects of the waterway.

**VI. CONCLUSION**

3D ECDIS technology would be very significant for enhancing the navigation safe. In order to solve effectively some key problems in large range 3D rendering, we proposed ALOD algorithm to optimize the rendering performance and effects of the sea-bed terrain, namely, both the navigation safe levels around the ship and the real-time navigational speed are used to build ALOD rendering model, which is the key difficulty of the 3D sea-bed rendering. The proposed algorithm provided an effective solution for the implementation of 3D ECDIS. This approach not only presented an effective preprocessing method to simplify the point cloud dataset through controlling the change rates of the normal vectors, which maintains the geometric topology of the 3D ECDIS, but also fully used the characteristics of the ship navigation to divide the sea area into different safe levels. Then the safe levels are regarded as the weighted values, and the graphics hardware technology GPU is employed to implement the area level of detail algorithm, which is useful to implement rapidly highly accurate 3D ECDIS display. The proposed approach can promote significantly the development of 3D ECDIS technology.

**REFERENCE**


Navigation with ECDIS: A Novel Approach to Navigational Calculations

Adam Weintrit
The Faculty of Navigation
Gdynia Maritime University
Polish Navigation Forum, The Royal Institute of Navigation, The Nautical Institute,
Gdynia, Poland
weintrit@am.gdynia.pl

Abstract—The article presents base issues concerning navigation in maritime transport with use of electronic chart systems. The Author presents a novel approach to navigational calculations and computational algorithms implemented in the software applied in marine navigation electronic devices and systems, such as GNSS (GPS, GLONASS, Beidou, Galileo), AIS, ECDIS/ECS, and other marine GIS that deserve to be collected together. The question he asks affects the range and point in applying the loxodrome (rhumb line) in case the ECDIS equipped with the great circle (great ellipse) approximation algorithms of given accuracy replaces the traditional nautical charts based on Mercator projection. Moreover, the navigation based on geodesic lines and connected software of the ship’s devices (electronic chart, positioning and steering systems) gives a strong argument to research and use geodesic-based methods for calculations instead of the loxodromic trajectories in general. Further research activities are pointed out, together with planned practical actions in raising awareness regarding navigation with ECDIS.

Keywords—Geodesy, Rhumb Line (Loxodrome), Great Circle (Orthodrome), Geodesic Line, ECDIS, GNSS, Marine Navigation, Route Planning Calculations, Great Elliptic Sailing (GES)

I. INTRODUCTION

Almost from the first days of the development of the basic navigational software built into satellite navigational receivers, it has been noted that for the sake of simplicity and a number of other reasons, this navigational software is often based on the simple methods of limited accuracy. It is surprising that even nowadays the use of navigational software, even in ECDIS and more sophisticated INS/IBS systems, is still used in a loose manner, sometimes ignoring basic computational principles and adopting oversimplified assumptions and errors such as the wrong combination of planar, spherical and ellipsoidal calculations in different steps of the solution of a particular sailing problem. The lack of official IMO and IHO standardization on both the “accuracy required” and the equivalent “methods employed”, in conjunction to the “black box solutions” provided by GNSS navigational receivers and navigational systems (ECDIS and ECS) suggest the necessity of a thorough examination of the issue of sailing calculations for navigational systems and GNSS receivers.

The paper was written taking into account the views of various readers, ranging from practicing navigators, navigational equipment manufacturers, to researchers and theoretical analysts. It is also the author’s goal to present a current and uniform approaches to sailing calculations highlighting recent developments. Much insight may be gained by considering the examples that have recently proliferated in the literature reviewed. The additional aim is to demonstrate to the manufacturers of ECDIS systems and GNSS receivers wide spectrum of solutions which may be used in the implemented computational algorithms ranging from simple (on the flat or the sphere) to sophisticated and computationally complex (on the surface of ellipsoid) solutions.

The author compares algorithms for rhumb line sailing (RLS), great circle sailing (GCS) and great elliptic sailing (GES) calculations used for route planning and route monitoring in the context of marine navigational applications. The existence of approximating and basically differing implemented computational procedures leads to a significantly different results of distance and angle calculations which play the essential role in theoretical solutions and applications requiring high accuracy and precision. Additionally, compact vector formulas are presented for the great circle and rhumb line on the sphere and rhumb line, great circle, great ellipse and geodesic line on the spheroid providing vector solutions to both the forward and the inverse problems and conversions of longitude and latitude. The solutions incorporate a closed form for the azimuth and the derivation of the corresponding equation is illustrated. The paper also shows that a computer algebra system is a powerful tool to solve some mathematical derivations that may be useful in navigation, geodesy and cartography.

The paper provides an overview of the computational algorithms, which can be applied in the navigational software.

II. ECDIS APPROACH

As part of the curriculum for ECDIS navigation it became apparent that the standard navigation manuals were perpetuating a flawed method of calculating rhumb lines on the Earth considered as an oblate spheroid. On further investigation it became apparent that these incorrect methods were being used in programming a number of marine calculators, computers and satellite navigation receivers. Although the discrepancies were not large, it was disquieting to compare the results of the same rhumb line calculations from a number of
such devices and find variations of a few miles when the output was given, and therefore purported to be accurate, to a tenth of a mile in distance and/or a tenth of a minute of arc in position. The problem was highlighted in the past and the references at the end of this paper show that a number of methods have been proposed for the amelioration of this problem [Pallikaris ed.al, 2009a, b and 2010].

This paper presents and recommends the guidelines that should be applied for the accurate solutions. Most of these may be found in standard geodetic text books, such as, but also provided are new formulae and schemes of solution which are suitable for use with computers or tables. The paper also takes into account situations when a near-indeterminate solution may arise. Some examples are provided which demonstrate the methods. The data for these problems do not refer to actual terrestrial situations but have been selected for illustrative purposes. Practising ships' navigators will find the methods described in detail in this paper to be directly applicable to their work and they also should find ready acceptance because they are similar to current practice. In almost none of the references cited at the end of this paper has been addressed the practical task of calculating, using either a computer or tabular techniques.

In papers [Weintrit, Kopacz, 2011 and 2012] the authors present the review of different approaches to contact formulae for the computation of the position, the distance, and the azimuth along a great ellipse. The proposed alternative formulae are to be primarily used for accurate sailing calculations on the ellipsoid in a GIS environment as in ECDIS and other ECS. Among the ECDIS requirements is the need for a continuous system with a level of accuracy consistent with the requirements of safe navigation. At present, this requirement is best fulfilled by the Global Positioning System (GPS). The GPS system is referenced to World Geodetic System 1984 Datum (WGS 84). Using the ellipsoid model instead of the spherical model attains more accurate calculation of sailing on the Earth. Therefore, we aim to construct a computational procedure for solving the length of the arc of a great ellipse, the way-points and azimuths along a great ellipse. We announced our aspiration to provide the straightforward formulae involving the great elliptic sailing based on two scenarios. The first is that the departure point and the destination point are known. The second is that the departure point and the initial azimuth are given (direct and inverse geodetic problems on reference ellipsoids).

A. ECDIS Calculations

ECDIS system must be able to perform as a minimum the following calculations and conversions [Weintrit, 2009], [IMO, 2006]:
- geographical coordinates to display coordinates, and display coordinates to geographical coordinates;
- transformation from local datum to WGS-84;
- true distance and azimuth between two geographical positions;
- geographic position from a known position given distance and azimuth (course);
- projection calculations such as great circle and rhumb line courses and distances;
- “RL-GC” difference between the rhumb line and great circle in sailing along the great circle (or great ellipse?).

B. Route Planning Calculations

When planning a route the ECDIS allows the navigator to create waypoints and routes including setting limits of approach and other cautionary limitations. Both rhumb line and great circle routes can be defined. Routes can be freely exchanged between the ECDIS and GNSS or ARPA. Route checking facility allows the intended route to be automatically checked for safety against limits of depth and distance as defined by the navigator.

The mariner can calculate and display both a rhumb line and a great circle line and verify that no visible distortion exists between these lines and the chart data. But is it really rhumb line and great circle? And perhaps this is geodesic lines?

The Author predicts the early end of the era of the rhumb line. This line in the natural way will go out of use. Nobody after all will be putting the navigational triangle to the screen of the ECDIS. Our planned route is not having to be a straight line on the screen. So, why keep this line still in the use? Each ship’s position plotted on the chart can be the starting point of new updated great circle GC, or saying more closely, great ellipse GE or geodesic line! But now a model of the Earth is no longer a sphere, but an ellipsoid WGS-84. So how many orthodromes (great circles) can we plot on the surface of the ellipsoid? Only one – an equator. So, what do we really compute?

C. Most Important Questions

It is an important question whether in the ECDIS time still Mercator projection is essential for marine navigation. Do we really need it? And what about loxodrome? Also not? So, let start navigation based on geodesics. It is high time to forget the rhumb line navigation and great circle navigation, too. But the first we need a clearly established methods, algorithms and formulas for sailing calculations. But it is already indicating the real revolution in navigation - total revolution. We will be forced to make the revision of such fundamental notions as the course, the heading and the bearing.

And another very important question: do you really know what kind of algorithms and formulae are used in your GNSS/GPS receiver and your ECS/ECDIS systems for calculations mentioned in section II A? I am sure, your answer is negative. So, we have a problem – a serious problem.

III. NEW MERIDIAN ARC FORMULAS FOR SAILING CALCULATIONS IN NAVIGATIONAL GIS

A. Introduction

In traditional navigation, the calculation of the elements of the shortest navigation path between two points on the surface of the Earth is usually conducted by the use of spherical model of the Earth and the assumption that one minute of arc of any great circle is equal to one international nautical mile. It is well known that more accurate results can be obtained by the adoption of an ellipsoidal model of the Earth and the calculation of geodesic distances and azimuths.

The discrepancies between the results on the spherical and the ellipsoidal model of the Earth are in the order of 0.27% according to Tobler [Tobler, 1964], and in the order of 0.5% according to Earle [Earle, 2006]. In reality these discrepancies can exceed 10 nautical miles for a number of common
navigational routes. These matters became especially important with the appearance of the first satellite positioning system Transit and were discussed a new approach to the theory of geodesics on an ellipsoid [Holmstrom, 1976], [Williams and Phytian, 1996]. An example of such a discrepancy is shown through the calculation of the shortest navigational distance from a departure location in the west coast of USA such as the entrance of San Francisco Bay (φ=37º45.047'N, λ=122º42.023'W) to a destination point in Japan such as the approaches to Yokohama Harbour (φ=34º26.178'N, λ=139º51.139'E) [Pallikaris et al, 2010]. This calculation on the spherical Earth model using spherical trigonometry and the classical assumption that 1 minute of a great circle arc is equal to the international nautical mile (1852 metres) yields a distance of 4489.9 nautical miles. The calculation of this distance on the WGS-84 ellipsoid, using very accurate methods for the calculation of long geodesics, as the method of Vincenty [Vincenty, 1975], yields 4502.9 nautical miles. For this example the difference in calculated distances on the spherical model from those on the ellipsoid is 13 nautical miles (~24 km). But the discrepancies between RLS calculations on the sphere and the ellipsoid for very long sailing distances may exceed even 19 nautical miles (~35 km) [Pallikaris et al, 2010].

Despite these differences the use of the spherical model in traditional navigation for most practical purposes is considered satisfactory. Nevertheless for the case of sailing computations in GIS navigational systems such as ECDIS and other ECS systems the computations must be conducted on the ellipsoid in order to eliminate these errors but without seeking the submeter accuracies pursued in other geodetic applications. Seeking extremely high accuracy for marine navigation purposes does not offer any real benefit and requires more computing power and processing time. For these reasons and before proceeding with the adoption of any geodetic computational method on the ellipsoid for sailing calculations it is required to adopt realistic accuracy standards in order not only to eliminate the significant errors of the spherical model but also to avoid the exaggerated and unrealistic requirements of sub meter accuracy.

Section B of this chapter addresses the topic of accuracy requirements of sailing calculations in GIS, such as ECDIS and other ECS. Section III.C explains the relation of the meridian arc distance formulas with the process of calculating sailing routes. Section III.D overviews the general geodetic methods and formulas with their main variations used for the calculation of the length of the arc of the meridian on the ellipsoid. The proposed new equations are presented in section III.E. Section III.F presents the results of a comparative study of selected methods and formulas in terms of accuracy achieved and CPU time required, which was conducted in order to evaluate the proposed new formulas. The results of this comparative study can also be employed for the selection of the proper computing method according to the requirements of any, other than sailing computations, application. Section III.G concludes the subchapter. The basic benefits of the proposed new equations (see section III.E) are that:

- They are much simpler and faster than traditional geodetic methods of the same accuracy.
- They provide extremely high accuracies for the requirements of sailing calculations on the ellipsoid.

B. Accuracy Requirements for Sailing Calculations in GIS

The IMO performance standards for ECDIS [IMO 2006] do not provide specific accuracy standards for sailing calculations, except for the following general requirements:

"It should be possible to carry out route planning and route monitoring in a simple and reliable way".

"The accuracy of all calculations performed by ECDIS should be independent of the characteristics of the output device and should be consistent with the SENC accuracy".

Setting accuracy requirements in relation to SENC, depends directly on the category of the Electronic Navigational Charts (ENCs) used in ECDIS. This is a reasonable requirement for calculations relating to real time positions that affect the safety of navigation when using ECDIS. This safety is assured through the installation of the proper ENCs in the SENC. Nevertheless these standards, when applied to set the accuracy of sailing calculations for route planning may result in vague, ambiguous and sometimes unreasonable standards due to their indirect dependency on the installed ENCs. This deficiency is illustrated in the attempt to apply this general ECDIS accuracy requirement for consistency with SENC accuracy in sailing calculations. Taking into consideration that the SENC contains ENCs of various categories, the average compilation scale of each category and considering SENC accuracy equivalent to 0.5 mm at the compilation scale of the contained ENCs, we obtain accuracy requirements ranging from 5 metres to more than 1250 metres (even to 5.000 metres for “category 1” ENCs compiled from 1/10.000.000 paper charts). For the above mentioned reasons the study for the development of more realistic formulas for the computation of the length of the arc of the meridian has been based on the requirements of Table 1 rather than on the IMO general ECDIS accuracy requirements.

Accuracy requirements for sailing calculations [Pallikaris et al, 2009b]

<table>
<thead>
<tr>
<th>Calculated Distance</th>
<th>Maximum Acceptable Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>up to 250 nautical miles</td>
<td>0.1 nautical miles</td>
</tr>
<tr>
<td>between 250 and 500 nautical miles</td>
<td>0.2 nautical miles</td>
</tr>
<tr>
<td>between 500 and 2000 nautical miles</td>
<td>0.3 nautical miles</td>
</tr>
<tr>
<td>greater than 2000 nautical miles</td>
<td>0.5 nautical miles</td>
</tr>
</tbody>
</table>

C. The Length of the Meridian Arc in Sailing Calculations

The calculation of the length of the arc of the meridian is a basic prerequisite for many accurate sailing calculation methods on the ellipsoid concerning both Rhumb Line Sailing (RLS) and shortest sailings on the ellipsoid such as Great Elliptic Sailing (GES). A lot of specific papers present in detail the advantages and benefits of these methods [Benet 1996], [Earle 2000] and [Pallikaris et. al. 2009a, b, 2010].

It is noted though that in certain sailing calculation methods it is not necessary to calculate the length of the meridian arc. Typical examples of these methods concern:

(i) RLS calculations by the employment of the general formulas of the Mercator projection [Snyder, 1987] and isometric latitude [Bowring, 1985];
(ii) calculation of shortest sailings paths on the ellipsoid by a geodetic inverse method such as the Andoyer-Lambert method proposed by the Admiralty manual of Navigation [AMN, 1987].

The Rhumb Line Sailing (RLS) calculations employing direct formulas on the ellipsoid, which require the calculation of the length of the arc of the meridian, [AMN, 1987], [Bennet 1996], are simpler than those employing the Mercator projection formulas and isometric latitude. In addition, the formulas on the ellipsoid provide more flexibility for the solution of the direct problem for the calculation of the geodetic coordinates of an unlimited number of intermediate points for the purpose of the display of RLS routes on the electronic navigational chart of the ECDIS and ECS.

If we consider the great ellipse as a slanted version of the meridian ellipse, it is possible to calculate the great elliptic arc (sailing distance) in a similar manner to that used for the calculation of the meridian arc. Various numerical tests and comparisons show that discrepancies in the computed distances between the “geodesic” and the “great elliptic arc” are practically negligible for marine navigation [Williams, 1996], [Earle, 2000]. [Palliarkis & Latsas, 2009]. Moreover GES calculations are much simpler and straightforward and can be easily implemented in navigational software. They provide the same and in some cases, higher accuracy than other methods and formulas for sailing calculations on the ellipsoid. An example is that GES calculations provide more accurate results than the Geodesic inverse solutions with the Lambert method. GES calculations can be also used for the precise calculation of the geodetic coordinates of an unlimited number of intermediate points along the great elliptic arc, and thus be implemented in GIS navigational systems (ECDIS and ECS) for the display of navigational paths on the electronic chart. The purpose of this paper is to present new simpler and faster formulas for meridian arc computations that can be immediately implemented in various sailing calculation methods that require the calculation of the meridian arc. The detailed presentation of these sailing calculations can be found in the relevant bibliographic references [Palliarkis et al, 2009a, b and 2010].

D. Geodetic Formulas for the Meridian Arc Length

The methods and formulas used to calculate the meridian arc length to precise calculations on the ellipsoid sailing, such as "rhumbowel sailing line", "big elliptical Sea" and "surveying sailing" are simplified forms of the formulas used in applications geodetic surveying

The methods, algorithms and formulas used to calculate the meridian arc length for precise sailings calculations on the ellipsoid, such as “rhumb-line sailing”, “great elliptic sailing” and “geodesic sailing” are simplified forms of general geodetic formulas used in geodetic applications. In this section an overview of the most important geodetic formulas along with general comments and remarks on their use is carried out. For consistency purposes and in order to avoid confusion in certain formulas the symbolization has been changed from that of the original sources. The fundamental equation for the calculation of the length of the arc of the meridian on the ellipsoid $M_\psi^0$ (Figure 1), is:

$$ M_\psi^0 = \int_0^\psi R_M d\psi $$ (1)

In (1), $R_M$ is the radius of curvature of the meridian given by (2).

$$ R_M = \frac{a(1-e^2)}{(1-e^2 \sin \psi)^{3/2}} $$ (2)

In (2), a is the semi-major axis and e is the eccentricity of the ellipsoid.

![Figure 1. The length of the arc of the meridian [Palliarkis et al, 2009b]](image)

Replacing the value of $R_M$ from (2) in (1), we obtain:

$$ M_\psi^0 = \int_0^\psi \frac{a(1-e^2)}{(1-e^2 \sin \psi)^{3/2}} d\psi $$ (3)

Equation (3) can be transformed to an elliptic integral of the second type, which cannot be evaluated in a “closed” form. The calculation can be performed either by numerical integration methods, such as Simpson’s rule, or by the binomial expansion of the denominator to rapidly converging series, retention of a few terms of these series and further integration by parts. According to Snyder [Snyder, 1987] and Torge [Torge, 2001], Simpson’s numerical integration does not provide satisfactory results and consequently the standard computation methods are based on the use of series expansion formulas. Expanding the denominator of (3) by the binomial theorem yields:

$$ M_\psi^0 = a \cdot (1-e^2) \int_0^\psi \left(1 + \frac{3}{2} e^2 \sin^2 \psi + \frac{15}{2} e^4 \sin^4 \psi + \frac{35}{16} e^6 \sin^6 \psi \right) dx $$ (4)

Since the values of powers of $e$ are very small, equation (4) is a rapidly converging series. Integrating (4) by parts we obtain:
\[ M_0^\psi = a(1 - e^2) \left( 1 + \frac{3}{4} e^2 + \ldots \right) \varphi - \frac{3}{8} e^2 + \frac{15}{32} e^4 + \ldots \sin 2\varphi + \frac{15}{256} e^4 + \frac{105}{1024} e^6 + \ldots \sin 4\varphi + \ldots \]  
\[ (5) \]

Equation (5) is the standard geodetic formula for the accurate calculation of the meridian arc length, which is proposed in a number of textbooks such as in Torge’s “Geodesy” using up to \( \sin(2\varphi) \) terms, [Torge, 2001] and in Veis’ “Higher Geodesy” using up to \( \sin(6\varphi) \) terms [Veis, 1992]. A rigorous derivation of (4) for terms up to \( \sin(6\varphi) \) is presented in [Pearson, 1990].

Equation (4) can be written in the form of equation (5) provided by Veis [Veis, 1992]
\[ M_0^\psi = a(1 - e^2)(M_0\varphi - M_2\psi + M_4\varphi - M_6\varphi + M_8\varphi + \ldots) \]  
\[ (6) \]
\[ M_0 = 1 + \frac{3}{4} e^2 + \frac{45}{64} e^4 + \frac{175}{256} e^6 + \frac{11025}{16384} e^8 + \ldots \]
\[ M_2 = \frac{3}{8} e^2 + \frac{15}{32} e^4 + \frac{525}{1024} e^6 + \frac{2205}{4096} e^8 + \ldots \]
\[ M_4 = \frac{15}{256} e^4 + \frac{105}{1024} e^6 + \frac{2205}{8820} e^8 + \ldots \]
\[ M_6 = \frac{35}{3072} e^6 + \frac{315}{12288} e^8 + \ldots \]
\[ M_8 = \frac{315}{130784} e^8 + \ldots \]

Equation (7) is derived directly from equation (6) for the direct calculation of the length of the arc between two points (A and B) with latitudes \( \varphi_A \) and \( \varphi_B \). In the numerical tests for the assessment of the relevant errors of selected alternative formulas, we will refer to equations (6) and (7) as the “Veis-Torge” formulas.

\[ M_{\psi A}^\varphi = a(1 - e^2)(M_0\varphi - M_2\psi_B + M_3\varphi_B - M_4\psi_B - \sin 2\psi_B + M_4\varphi_B - M_6\psi_B + M_6\varphi_B - \sin 6\psi_B + M_6\varphi_B - \sin 6\varphi_B) \]  
\[ (7) \]

Equations (6) and (7) are the basic series expansion formulas used for the calculation of the meridian arc. They are rapidly converging since the value of the powers of \( e \) is very small. Very accurate results are obtained by formula (6) and the retention of terms up to \( \sin(6\varphi) \) and \( \sin(4\varphi) \) and \( 8^{th} \) or \( 10^{th} \) powers of \( e \). For sailing calculations on the ellipsoid it is adequate to retain only up to \( \sin(2\varphi) \) terms, whereas for other geodetic applications it is adequate to retain up to \( \sin(4\varphi) \) or \( \sin(6\varphi) \) terms. The basic formulas (6) and (7) can be further manipulated and transformed to other forms. The most common of these forms is formula (8). Simplified versions of 8 (retaining up to \( A_6 \) and \( e^6 \) terms only) are proposed in texts such as Bombord’s “Geodesy” [Bombord, 1985], and in the “Admiralty Manual of Navigation” [AMN, 1987].

\[ M_0^\psi = a(A_0\psi - A_2\psi + A_4\psi - A_6\psi + A_8\psi \ldots) \]  
\[ (8) \]
\[ M_0 = 1 - \frac{1}{4} e^2 - \frac{3}{64} e^4 - \frac{5}{256} e^6 - \frac{175}{16384} e^8 \ldots \]
\[ A_2 = \frac{3}{8} \left( e^2 + \frac{1}{4} e^4 + \frac{1}{15} e^6 + \frac{35}{512} e^8 \ldots \right) \]
\[ A_4 = \frac{15}{256} \left( e^4 + \frac{3}{4} e^6 + \frac{35}{64} e^8 \ldots \right) \]
\[ A_6 = \frac{35}{3072} e^6 + \frac{175}{12288} e^8 \ldots \]
\[ A_8 = \frac{315}{131072} e^8 \ldots \]

Another formula for the meridian arc length is equation (9), which is used by Bowring [Bowring, 1983] as the reference for the derivation of other formulas, employing polar coordinates and complex numbers. The basic difference of formula (9) from (6), (7) and (8) is that (9) uses the ellipsoid parameters \( a, b \), instead of the parameters \( a, e \) which are used in formulas (6), (7) and (8).

\[ M_0^\psi = A_1 \left( \psi - B_1 n \sin 2\psi - \frac{15}{16} n^2 \sin 4\psi + \frac{35}{48} n^3 \sin 6\psi - \frac{315}{512} n^4 \sin 8\psi + \ldots \right) \]  
\[ (9) \]
\[ A_1 = \frac{a(1 + \frac{1}{8} n^2)^2}{1 + n} \]
\[ B_1 = 1 - \frac{3}{8} n^2 \]
\[ n = \frac{a - b}{a + b} \]

Bowring [Bowring, 1985] proposed also formula (10) for precise rhumb-line (loxodrome) sailing calculations. This formula calculates the meridian arc as a function of the mean latitude \( \varphi_m \) and the latitude difference \( \Delta \varphi \) of the two points defining the arc on the meridian.

\[ M_{\psi A}^\varphi = a(A_0\Delta \varphi - A_2 \cos(2\varphi_m) \sin(\Delta \varphi)) + A_4 \cos(4\varphi_m) \sin(2\Delta \varphi)) - A_6 \cos(6\varphi_m) \sin(3\Delta \varphi)) + A_8 \cos(8\varphi_m) \sin(4\Delta \varphi)) \]  
\[ (10) \]
\[ \Delta M = k_0 \Delta \varphi - k_2 \cos(2\varphi_m) \sin(\Delta \varphi) + k_4 \cos(4\varphi_m) \sin(2\Delta \varphi)) - k_6 \cos(6\varphi_m) \sin(3\Delta \varphi)) + k_8 \cos(8\varphi_m) \sin(4\Delta \varphi)) \]  
\[ (11) \]
\[ In (10), the coefficients \( A_0, A_2, A_4, A_6, \) and \( A_8 \) are the same as in (8). Equation (10) has the general form of equation (11). \]

\[ In (11), the coefficients \( k_0, k_2, k_4, k_6, k_8 \) are: k_0= a A_0, k_2= a A_2, k_4= a A_4, k_6= a A_6, k_8= a A_8 \]
E. The Proposed New Formulas by Pallikaris, Tsoulos and Paradissis Length

The proposed new formulas for the calculation of the length of the meridian in sailing calculations on the WGS-84 ellipsoid in meters and international nautical miles are (53) and (54), respectively [Pallikaris, et al, 2009 and 2010].

\[
M_{\varphi_0} = 111132.95251 \cdot \Delta \varphi - 16038.50861 \cdot 
\left( \sin \left( \frac{\varphi_B - \pi}{90} \right) - \sin \left( \frac{\varphi_A - \pi}{90} \right) \right) \tag{12}
\]

\[
M_{\varphi_\Lambda} = 60.006994 \cdot \Delta \varphi - 8.660102 \cdot 
\left( \sin \left( \frac{\varphi_B - \pi}{90} \right) - \sin \left( \frac{\varphi_A - \pi}{90} \right) \right) \tag{13}
\]

In both formulas (12) and (13) the values of geodetic latitudes \( \varphi_A \) and \( \varphi_B \) are in degrees and the calculated meridian arc length in meters and international nautical miles respectively. Formulas (13) and (12) have been derived from (7) for the WGS-84, since the geodetic datum employed in Electronic Chart Display and Information Systems is WGS-84. The derivation of the proposed formulas is based on the calculation of the \( M_0 \) and \( M_2 \) terms of (7) using up to the 8th power of \( \epsilon \). This is equivalent to the accuracy provided by (8) using \( A_0 \) and \( A_2 \) terms with subsequent \( e \) terms extended up to the 10th power since in formula (48) the terms \( M_0, M_2, M_4 \ldots \) are multiplied by \( (1-e^2) \). According to the numerical tests carried out, which are presented in the next section, the proposed formulas have the following advantages:

- they are much simpler than and more than twice as fast as traditional geodetic methods of the same accuracy.
- they provide extremely high accuracies for the requirements of sailing calculations on the ellipsoid.

F. The Author’s Modification

Taking into account that the polar distance for WGS-84 is 10001965.7293127 m [Weintrit, 2013] the author proposes some modification to the formula (12) presented by Pallikaris, Tsoulos and Paradissis:

\[
M_{\varphi_0} = 6367449.1458234 \cdot \Delta \varphi - 
16038.50862 \cdot \left( \sin (2\varphi_B) - \sin (2\varphi_A) \right) \tag{14}
\]

with \( \varphi \) in radians, and result in meters).

This formula will be a little bit more accurate than formula (12).

The proposed new formula by Pallikaris, Tsoulos and Paradissis [Pallikaris, et al, 2009 and 2010] for the calculation of the meridian arc are sufficiently precise for sailing calculations on the ellipsoid. The author strongly supports that approach and highly recommends it for use in practice.

Higher sub metre accuracies can be obtained by the use of more complete equations with additional higher order terms.

Seeking this higher accuracy for sailing calculations does not have any practical value for marine navigation and simply adds more complexity to the calculations only. In other than navigation applications, where higher sub metre accuracy is required, the Bowring formulas showed to be approximately two times faster than alternative geodetic formulas of similar accuracy.

CONCLUSIONS

An analytical method, algorithm and formulas for great elliptic sailing (GES) calculations was presented. The method solves the complete GES problem calculating not only the great elliptic arc distance, but also other elements of the sailing such as the geodetic coordinates of intermediate points along the great elliptic arc. The proposed formulas provide extremely high accuracies and are straightforward to be exploited immediately in the development of navigational software, without the requirement to use advanced numerical methods. Their validity and effectiveness have been verified with numerical tests and comparisons to extremely accurate geodetic methods for the direct and inverse geodetic problem [Pallikaris & Latas, 2009].

Comparisons and numerical tests show that discrepancies in the computed distances between the “geodesic” and the “great elliptic arc” are practically negligible for marine navigation. On the other hand the discrepancies between the results of shortest navigational path calculations on the spherical model of the earth as great circle arcs and on the ellipsoidal model as great elliptic arcs, or geodesics, may in some cases exceed 15 nautical miles (~28.5 km).

The discrepancies between RLS calculations on the sphere and the ellipsoid for very long sailing distances may exceed 19 nautical miles (~35 km).

For short sailing distances (smaller than 600 nautical miles) the discrepancies between the calculations on the spherical and the ellipsoidal model of the earth practically can be considered negligible for navigation.

The proposed formulas for both RLS and GES calculations are straightforward and can be easily implemented in navigational software. They provide the same and in some cases, higher accuracy than other methods and formulas for sailing calculations on the ellipsoid. They can also be used in programmable pocket calculators for the solution of the inverse RLS and GES problems.

The proposed algorithms for the direct RLS and GES problems are used for the calculation of the geodetic coordinates of an unlimited number of intermediate points along the rhumb line and the great elliptic arc. Thus they can be easily implemented in GIS navigational systems (ECDIS and ECS) for the display of navigational paths on the electronic chart.

The proposed formulas provide a more realistic balance between accuracy and computing power required for the sailing calculations in a GIS environment and particularly in ECDIS, in compliance with the performance standards of the International Maritime Organization (IMO). These formulas can be immediately used not only for the development of new
algorithms for sailing calculations, but also for the simplification of existing algorithms without degrading the accuracies required for precise navigation. The simplicity of the proposed method allows for its easy implementation on pocket calculators for the execution of accurate sailing calculations on the ellipsoid [Pallikaris, 2009b].

Original contribution affects and verifies established views based on approximated computational procedures used in the software of marine navigational systems and devices. Current stage of knowledge enables to implement geodesics based computations which present higher accuracy. It also lets to assess the quality of contemporary algorithms used in practical marine applications. It should be noted that an important step in the solution is simplification by the omission of the expansion part into power series of mathematical solutions, previously known from the literature, i.e., [Torge, 2001] and [Veis, 1992], and reliance in the explanatory memorandum of application, in particular, on the amount of the available processing power of modern calculating machine (processor). In the author’s opinion this criterion is relevant from a practical point of view, but temporary, given the growth and availability of computing power, including GIS [Pallikaris et al., 2009a, b, 2010], [Weintrit & Kopacz, 2011 and 2012], [Weintrit, 2013].

Professional scientific workshop employed to solve the problem makes use of various tools, i.e. of differential geometry, marine geodesy (marine navigation), analysis of measurement error, approximation theory and problems of modelling and computational complexity, mathematical and descriptive statistics, mathematical cartography. Geometrical problems are important aspect of the tested models which are used as the basis of calculations and solutions implemented in contemporary navigational devices and modern electronic chart systems.

The algorithms applied for navigational purposes, in particular in ECDIS, should inform the user on actually used mathematical model and its limitations. The shortest distance (geodesics) between the points depends on the type of metric we use on the considered surface in general navigation. It is also important to know how the distance between two points on considered structure is determined.

The marine navigation based on geodesic lines and connected software of the ship’s devices (electronic chart, positioning and steering systems) gives a strong argument to research and use geodesic-based methods for calculations instead of the loxodromic trajectories in general. The theory is developing as well what may be found in the books on geometry and topology. This should motivate us to discuss the subject and research the components of the algorithm of calculations for navigational purposes.

REFERENCES

Session I
Special problems, other topics.
Low-Cost GPS/INS/OBD-II Integration
On an Embedded Linux Platform

Joonhoo Lim, Je Young Lee, Kwang Ho Choi, Hyung Keun Lee
Korea Aerospace University
School of Electronics, Telecommunication & Computer Engineering
Goyang-si, Korea
limjh@kau.ac.kr, jeylee@kau.ac.kr, sahnhara@kau.ac.kr, hyknlee@kau.ac.kr

Abstract—As widely known, GPS (Global Positioning System) alone cannot provide reliable and continuous positioning results under complex signal environments in urban area such as street canyons, tunnels, underpasses, and overpasses. For the reason, integration of GPS with other complementing sensors has been investigated actively. The integration of GPS with INS (Inertial Navigation System) has been one of the most well-known methods. However, positioning errors of GPS/INS systems increase gradually depending on inertial sensor grades. Therefore, if low-cost MEMS-IMUs (Micro Electro Mechanical Systems-Inertial Measurement Unit) were utilized, the GPS/INS error would increase quite rapidly.

For reliable and continuous land vehicle positioning in GPS-denied environments with low-cost sensors, this paper proposes an efficient integration method combining GPS, MEMS-IMU, and OBD-II (On-Board Diagnostic-II) device. In the proposed method, a BBB (BeagleBone Black) board was utilized as the integration platform for the three low-cost sensors. The embedded Linux platform performs collection of GPS, MEMS-IMU and OBD-II measurements and extraction of time synchronization information.

The proposed method corrects the INS velocity information continuously by utilizing OBD-II speed even when GPS measurements are not available. Therefore, the proposed integration method can provide more continuous position estimates with less error growth. To evaluate the feasibility of the proposed method, experiments were performed with real measurements including GPS-denied environments. The comparison results with the conventional GPS/MEMS-IMU integration method are presented.

Keywords—GPS, INS, MEMS, OBD-II, Embedded Linux, GPS-denied environments

INTRODUCTION

As widely known, GPS (Global Positioning System) alone cannot provide reliable and continuous positioning of the vehicle under complex signal environments in urban area such as street canyons, tunnels, underpasses, and overpasses incur many problems [1, 2]. For the reason, integration of GPS with other complementing sensors has been investigated actively. The integration of GPS with INS (Inertial Navigation System) has been one of the most well-known methods to provide more accurate positioning results with higher update rate. However, positioning errors of GPS/INS systems increase gradually depending on inertial sensor grades [3, 4]. If low-cost MEMS-IMUs (Micro Electro Mechanical Systems-Inertial Measurement Unit) were utilized, the GPS/INS errors would increase quite rapidly. Therefore, the GPS/INS integration method require more augmentation with other aiding sensors to complement the weakness of MEMS inertial sensors.

Recently, integrated navigation systems augmenting GPS/INS with other aiding sensors have gained many attentions to improve navigation accuracy and continuity. A representative integrated navigation system combining GPS, INS, odometer, and OV (Omni-directional Vision) sensor was introduced in [5]. It showed that the OV sensor plays a key role to reduce the growth of INS coasting errors by providing incremental heading angle between two successive images. In addition, a tightly-coupled GPS/INS/Vision system was proposed for navigation in urban environments where traffic lights were exploited to correct the lateral position in urban area [6]. To use traffic lights as aiding information, it was assumed in [6] that the coordinates of traffic lights are known in advance.

The OBD-II (On-Board Diagnostic-II) standard is purposed to diagnose all types of land vehicles in the near future [7]. The OBD-II standard provides unified hardware interface to the vehicles from various manufacturers. By OBD-II, running condition of a vehicle can be monitored [7-9]. From the view point of positioning and navigation, OBD-II is attractive since it can provide vehicle speed information continuously. With a simple interface, vehicle speed can always be obtained independently from the vehicle's surrounding environments.

For reliable and continuous land vehicle positioning in GPS-denied environments with low-cost sensors, this paper proposes an efficient low-cost GPS/MEMS-IMU/OBD-II...
integration method on an embedded Linux platform. The OBD-II device is utilized to complement the weakness of low-cost and low-grade MEMS inertial sensors. For flexible design and implementation of navigation software, this study utilizes a BBB (BeagleBone Black) board as the embedded Linux platform.

To integrate a single frequency GPS receiver, a MEMS-IMU, and an OBD-II device, the proposed embedded Linux platform consists of four functional parts. The first part processes GPS measurements and extracts information for the time synchronization, the second part collects inertial measurements from the MEMS-IMU, the third part extracts the speed of a vehicle from the OBD-II device, and the last part merges all the outputs from the former three parts based on the time synchronization information.

This paper organized as follows. At first, the embedded Linux platform is introduced. Secondly, the OBD-II standard is described. Thirdly, the integration of GPS/MEMS-IMU/OBD-II sensors is explained. Fourthly, experiment results are presented. Finally, concluding remarks are given.

EMBEDDED LINUX PLATFORM

This paper utilizes a BBB [10, 11] to collect GPS, MEMS-IMU, and OBD-II measurements with the time synchronization information. Fig. 1 shows the components of the BBB. It can support Linux and Android operating systems. Fig. 2 depicts the configuration of synchronized measurement sampling. A flash memory is basic data storage for the BBB board. However, external memory can also be utilized through micro-SD slot.

GPS receiver and OBD-II are connected via USB and IMU is connected via GPIO-pin. To merge all sensor measurements, the time synchronization information from the GPS receiver is utilized. The time synchronization information is utilized for different sampling rates of the three sensors. Sampling rates of GPS receiver, OBD-II and MEMS-IMU are 1Hz, 2Hz and 20Hz, respectively. In this way, the proposed method can collect the time synchronized measurements from the three different sensors.

ON-BOARD DIAGNOSTICS-II

A. OBD-II

OBD-II is the standard anticipated to diagnose the conditions of all types of land vehicles in the near future [7]. The OBD-II specification provides unified hardware interface to the vehicles from various manufacturers. By OBD-II, states of the vehicle can be checked by MIL (Malfunction indicator Lamp).

B. ELM327

To obtain vehicle speed, an ELM327 device was utilized. The ELM327 is a low power CMOS (Complementary Metal-Oxide Semiconductor) device that supports high speed communications. It can be programmed using AT (Attention) commands. In this study, the ELM327 operates as the interface between RS232 and OBD-II standards and

<table>
<thead>
<tr>
<th>PID (hex)</th>
<th>Data bytes returned</th>
<th>Description</th>
<th>Min value</th>
<th>Max value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>0D</td>
<td>1</td>
<td>Vehicle speed</td>
<td>0</td>
<td>255</td>
<td>km/h</td>
</tr>
</tbody>
</table>

Typically, a user can obtain the messages with a scan tool connected to the vehicle’s OBD-II connector. The procedure is summarized by the following steps [12].

1) The user enters the PID (Parameter ID)
2) The scan tool sends the PID to the vehicle’s CAN (controller–area network) bus, VPW, PWM, ISO, KWP. (After 2008, CAN only)
3) The device on the CAN bus recognizes the PID and returns the response to the PID
4) The scan tool reads the response message and displays it to the technician

As summarized in Table I, vehicle speed in unit of km/h can be obtained by the PID 0D (hex).


Fig. 1. Components of BBB [11]

Fig. 2. Configuration of synchronized measurement sampling
supports all the PIDs messages. It automatically identifies and interprets the messages. All data exchange with the ELM327 device is in hexadecimal format [9].

**GPS/MEMS-IMU/OBD-II INTEGRATION**

For the efficient integration of heterogeneous measurements from three different sensors, the well-known Kalman filter is utilized as the estimator. In the integrated system, MEMS-IMU measurements are utilized as the fusion center since it provides measurements on high-resolution vehicle movements and robust to external interference, jamming, or disturbance. Thus, the time propagation of the navigation variables such as position, velocity, and attitude of the vehicle is performed solely by the IMU measurements.

Fig. 3 depicts the configuration of the proposed GPS/MEMS-IMU/OBD-II integration method. When sufficiently many satellites are available, the proposed method corrects INS position and velocity with GPS and OBD-II measurements. When GPS measurements are not available, INS velocity is compared with the OBD-II speed. Therefore, the proposed integration method provide more continuous position estimates with less error growth.

For the time propagation of the integration Kalman filter, the following equation is utilized [13].

$$X_{k+1} = F_k X_k + W_k$$ \hspace{1cm} (1)

where,

- $F_k$: system matrix
- $X_k = [\delta L, \delta \theta, \delta h, \delta \nu, \delta \nu, \delta \phi, \phi, \phi, \nu, \nu, \nu, \nu, \phi, \phi, \phi]^T$: system state vector
- $\delta L, \delta \theta, \delta h$: position error
- $\delta \nu$: velocity error vector
- $\phi$: attitude error vector
- $\nu$: accelerometers bias vector
- $\varepsilon$: gyro drift vector
- $W_k \sim N(0, Q)$: white Gaussian propagation noise
- $N(x, Q)$: Gaussian distribution of mean $x$ and covariance $Q$

For the measurement update of the integration Kalman filter, OBD-II speed information is converted to the velocity measurement as follows.

$$\tilde{V}_{OBD} = C_k^o \begin{bmatrix} \tilde{V}_X \\ 0 \\ 0 \end{bmatrix}$$ \hspace{1cm} (2)

where, $\tilde{V}_{OBD}$ is the velocity vector with respect to the locally-level navigation frame [1] and $\tilde{V}_X$ is the scalar vehicle speed by obtained OBD-II. $C_k^o$ denotes transformation matrix form the body frame to the navigation frame.

When GPS position and velocity measurements are available, the indirect measurement vector is formed as follows by utilizing both GPS and OBD-II measurements;

$$Z_k = H_k \delta \hat{X}_k = \begin{bmatrix} V_{GPS} \\ V_{GVEL} \\ V_{OBD} \end{bmatrix}$$ \hspace{1cm} (3)

where,

$$Z_k = \begin{bmatrix} \hat{P}_{INS} - \hat{P}_{GPS} \\ \tilde{V}_{INS} - \tilde{V}_{GPS} \\ \tilde{V}_{INS} - \tilde{V}_{OBD} \end{bmatrix}$$

$$H_k = \begin{bmatrix} I_{3x3} & 0_{3x3} & 0_{3x3} & 0_{3x3} \\ 0_{3x3} & I_{3x3} & 0_{3x3} & 0_{3x3} \\ 0_{3x3} & I_{3x3} & -\tilde{V}_{OBD} & 0_{3x3} \end{bmatrix}$$

$\hat{P}_{GPS}, \tilde{V}_{GPS}$: GPS position, velocity

$\hat{P}_{INS}, \tilde{V}_{INS}$: INS position, velocity
\( \dot{V}_{OBD} \) : OBD velocity  
\( v_{GPS} \) : GPS position measurement error  
\( v_{GVEL} \) : GPS velocity measurement error  
\( v_{OBD} \) : OBD velocity measurement error  
\( \langle x \rangle \) : skew-symmetric matrix of vector \( x \)

When GPS position and velocity measurements are not available, the indirect measurement vector is formed utilizing only the OBD-II measurement.

\[
Z = \dot{V}_{INS} - \dot{V}_{OBD} = \delta V_{INS} = \langle \dot{V}_{OBD} \rangle \phi - C^b \nu_{OBD} \quad (4)
\]

**EXPERIMENT**

To evaluate the feasibility of the proposed method, two experiments were performed. The first experiment contains the simulated signal blockages and the second experiment contains real signal blockages by an underpass and a tunnel. For the experiments, three sensors were utilized; an u-blox AEK-4T [15] with a patch antenna as the single frequency GPS receiver, an MPU-6050 [16] as the MEMS-IMU, and an ELM327 as the OBD-II device. A BBB board as the embedded Linux platform. Fig. 4 shows the installation of sensors in the rear of the vehicle. The MEMS-IMU provides 3-axis angular rates and 3-axis specific forces. To mitigate the effects of atmospheric errors by differential GPS, a Septentrio PolaRX2e dual frequency receiver with a Choke-Ring antenna was utilized as the stationary reference station.

The first experiment was conducted in an area with good satellite visibility. The simulated GPS signal blockages were generated as described in Table II. In the experiment, three trajectories were generated by GPS, GPS/IMU, GPS/IMU/OBD measurements, respectively. The GPS-only trajectory was generated by the RTK software GAFAS (GNSS Algorithm For Accuracy and Safety) [18] not applying the simulated signal blockages. Thus, it can be considered as the reference trajectory with cm-level accuracy. The second and third trajectories were generated by applying simulated signal blockages.

Fig. 5 compares the three trajectories generated by the three different methods in the first experiment. Fig. 5-(a) shows the overall trajectories and Fig. 5-(b) shows the magnified for the area A-1. In Fig. 5, the gray squares correspond to the GPS-only solutions, the gray circles correspond to the GPS/IMU integration method, and the black crosses correspond to the proposed GPS/IMU/OBD integration method. The horizontal positioning errors of the GPS/IMU and GPS/IMU/OBD methods during the simulated GPS outages are compared in Table II. It can be seen that the proposed GPS/IMU/OBD integration method improves accuracy by 85.5\% than the conventional GPS/IMU integration method.

The second experiment was performed in an urban area where GPS signal reception is affected by an underpass and a tunnel. Fig. 6 shows the second experiment area by Google Earth [17] where yellow dots correspond to the experiment trajectory. The experiment area is nearby Seoul World Cup Stadium, Korea. In Fig. 6, encircled areas B-1 and B-2 correspond to the underpass and the tunnel, respectively. As shown in Fig. 6, GPS cannot provide the position solutions while the vehicle is in beneath underpass and tunnel. For the reason, integrated systems are required for continuous positioning.
Fig. 6. The trajectory generated by GPS-only method

Fig. 7. Comparison of trajectories generated by different positioning systems

Fig. 8. The trajectory generated by the proposed method

TABLE III. HORIZONTAL ERROR FOR EXPERIMENTAL TRAJECTORY

<table>
<thead>
<tr>
<th>Area</th>
<th>Outage Duration (s)</th>
<th>Outage Distance (m)</th>
<th>Horizontal Error (m)</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-1</td>
<td>IMU</td>
<td>20</td>
<td>398.05</td>
<td>183.36</td>
</tr>
<tr>
<td></td>
<td>IMU/OBD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-1</td>
<td>IMU</td>
<td>11</td>
<td>202.45</td>
<td>33.65</td>
</tr>
<tr>
<td></td>
<td>IMU/OBD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-2</td>
<td>GPS/IMU</td>
<td>23</td>
<td>325.15</td>
<td>106.01</td>
</tr>
<tr>
<td></td>
<td>IMU/OBD</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSIONS

This paper proposed an efficient low-cost GPS/MEMS-IMU/OBD-II integration method against GPS-denied environments such as tunnels and underpasses. Since the propagation of accurate position requires accurate velocity, the proposed method utilized the OBD-II speed continuously. By the result, the velocity accuracy degradation could be mitigated even when GPS measurements were not available.

In the proposed integration method, a BBB board was utilized as the embedded Linux platform to interface three low-cost sensors; a single frequency GPS receiver with a patch antenna, a MEMS-IMU, and an OBD-II device. In order to merge all the sensor measurements and synchronize their sampling rates, the time synchronization information provided by the GPS receiver was utilized.

To evaluate the feasibility of the proposed integration method, two experiments were performed. The first experiment contained the simulated signal blockages and the second experiment contained real signal blockages by an underpass and a tunnel.
By the results of both experiments, it was shown that the proposed GPS/MEMS-IMU/OBD-II integration method is more feasible to provide accurate and reliable positioning results than the conventional GPS/MEMS-IMU integration method. It was also shown that the proposed method is possible to improve the positioning accuracy about 80% compared with GPS/INS integration method. More experiments are required to obtain accuracy and reliability improvement statistics.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program (2010-0010899) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology, Korea. The first author was also supported by Expert Education Program of Maritime Transportation Technology (GNSS Area), Ministry of Oceans and Fisheries of Korean government.

REFERENCES


[17] https://www.google.com/earth/

Performance Analysis of Loosely Coupled RTK-GNSS/IMU/Vehicle Speed Sensors in Urban Environment

Ren Kikuchi, Nobuaki Kubo
Tokyo University of Marine Science and Technology
Tokyo, Japan
lnkikut1@gmail.com

Shigeki Kawai, Ichiro Kato, Nobuyuki Taguchi
DENSO CORPORATION
Japan

Abstract—Many sensors are now being mounted in vehicles to improve safety, comfort, and efficiency of driving, to realize applications such as automatic braking to avoid or minimize conflict or a system to follow the preceding vehicle. For intelligent transportation system (ITS) applications such as lane recognition/keeping, precise absolute position is vital. The Real-Time Kinematic-Global Navigation Satellite System (RTK-GNSS) is one candidate to obtain precise position and is becoming widely available in the world. By using multi-GNSS constellation in an urban environment, both reliability and availability of RTK-GNSS are guaranteed with a relatively high level because of the use of many satellites. However, visibility of the RTK-GNSS satellites in dense urban areas is still not sufficient and the performance of RTK-GNSS is significantly degraded under tunnels or elevated roads. Our present goal is to obtain 100% availability and within 1.5 m horizontal accuracy even in dense urban areas. To achieve these objectives, we propose loosely coupled integration of RTK-GNSS, inertial measurement units (IMUs), and vehicle speed sensors. We consider the outputs of RTK-GNSS to be highly reliable. The second priority of our solution is GNSS Doppler frequency-based velocity. Without any GNSS solutions, only IMU-based direction and speed sensors are used to deduce the position. The popular Kalman filter is used to integrate gyro information with GNSS-based direction to calculate direction. To check the performance of our proposed method, an experiment was performed in 2014 in the urban areas of Nagoya city, Japan. Compared with only GNSS-based direction, loosely coupled direction by using gyro information was improved from 1.5° to 0.6° in the standard deviation. We used dual-frequency geodetic-grade GNSS receiver and antenna. The IMU and speed sensor used in this test were originally equipped in a car, which resulted in a very low cost of the sensors. The experiment results show that the availability was improved to 100% and the maximum absolute horizontal errors were within 2-3 m in all tests.

Keywords—RTK-GNSS, IMU, speed sensor

I. INTRODUCTION

Recently, advanced driver assistance systems (ADASs) to realize such features as lane change assist and automatic braking in automotive applications have experienced rapid growth. For a more advanced operating system implementation, improvement of the vehicle location accuracy is desired. It is possible to calculate a highly accurate absolute position by using the Real-Time Kinematic-Global Navigation Satellite System (RTK-GNSS). However, large errors arise in urban areas because of shielding and multi-path of the GNSS signal when there are many high-rise buildings in the area. Sometimes, positioning might not be possible at all. It is important to integrate GNSS with other sensors in order to produce robust precise position estimation in such an environment. A method using a gyro sensor and speed sensor has been widely used to estimate vehicle position [7]-[8]. These sensors can estimate position regardless of the surrounding environment with high accuracy in a short period of time. However, in general the measured values of internal measurement units (IMUs) have a bias error. If these sensors are used for long periods, the direction bias increases. On the other hand, error of the direction bias can be corrected at any time as long as the direction obtained from the GNSS Doppler frequency is available. In this study, we assume the use of very low-cost automotive sensors and propose a highly accurate and robust trajectory estimation method by integrating GNSS and IMU, with a target horizontal accuracy of less than 1.5 m.

II. INTEGRATION METHOD OF RTK-GNSS, IMU, VEHICLE SPEED SENSOR

First, each sensor type was evaluated for use in an urban environment, as summarized in Table I. In the table, the symbol $\Delta$ indicates that the sensor cannot be used for the indicated purpose. For GNSS, the accuracy depends on the surrounding environment. For IMU, as previously described, the bias error is cumulative. In order to use IMU to compensate for GNSS shortcomings, the IMU bias error must be occasionally corrected by using the observed value of the GNSS Doppler. However, the azimuth accuracy for determining the GNSS Doppler observations depends on the surrounding environment, which may degrade the precision because of factors such as satellite signal interruption and multipath. Therefore, a quality validation must be performed to eliminate the bias error, in order to correct the IMU bias error.
An overview of our GNSS/IMU integration algorithm is shown in Fig. 1.

In the proposed method, position is calculated in this order of priority: RTK-GNSS positioning solutions, position estimation by GNSS speed, and heading estimation by IMU and speed sensor.

A. RTK-GNSS

The overview of RTK-GNSS for a single epoch is introduced here. In the satellite selection, the mask angle was set to 15°. The maximum HDOP was set to 10. For RTK-GNSS positioning, the required number of satellites changes according to the satellite constellation. For the GPS or GPS/QZSS constellation, the minimum number of satellites is five. This number was chosen because the reliability of RTK deteriorates significantly when using only four visible satellites, as reported in past studies [10]. The QZSS constellation is treated in the same way as GPS satellites in this study because the GPS/QZSS ambiguity resolution has already been evaluated. This choice does not cause a problem because the QZSS timescale is also controlled with respect to the offset from the GPST, similar to the other GPS satellites. Furthermore, common frequencies and common coordinates are used by GPS and QZSS. On the other hand, there are no common frequencies between the BeiDou B1/B2/B3 and GPS L1/L2/L5 carrier frequencies. Therefore, the respective reference satellites are selected for the BeiDou and GPS systems when constructing the double-difference (DD) observation equation, so as to retain the integer property of the DD ambiguities. For simultaneous measurements over short baselines (< 10 km), the receiver- and satellite-related errors are completely eliminated; the DD troposphere and ionosphere errors can also be neglected. For the GPS/QZSS/BeiDou constellation, the minimum number of total satellites is six and at least two satellites in both GPS/QZS and BeiDou are required because we use the DD technique in this work. For the GPS/QZSS/BeiDou/GLONASS constellation, the minimum number of total satellites is seven and at least two satellites in each of GPS/QZS, BeiDou, and GLONASS are required (in the same manner). European Galileo satellites were not used in this work because the number of Galileo satellites available was insufficient in all tests. Regarding the ambiguity resolution of GLONASS satellites, there is an additional difficulty in terms of inter-frequency biases (IFBs) because GLONASS uses frequency division multiplexing (FDMA) to differentiate the signals from individual satellites [12]. GLONASS RTK positioning is similar to GPS RTK as long as a priori corrections to the carrier phase IFBs are applied. Fortunately, recent geodetic-level GNSS receivers calibrate the IFBs before shipment. Therefore, we do not need to calibrate these biases if we use suitable GNSS receivers. However, IFB calibration strongly depends on the manufacturer and the firmware.

B. Speed Estimation

The GNSS receiver can detect Doppler shift because of the relative speed between the GNSS receiver and the GNSS satellite. An effective method is to use the carrier frequency to obtain a highly accurate speed. If the carrier frequency is not tracked, it is possible to use the Doppler frequency. The carrier frequency is the integrated value of the Doppler frequency. The Doppler observation value is obtained by frequency-locked loop (FLL), and has tolerance to the effects of noise and multipath compared with the pseudo-range measurement value obtained by the delay-locked loop (DLL). The speed of a moving object can be calculated as follows by using the satellite velocity and Doppler measurements. First, Doppler measurements are defined as

\[ \Delta f = f \rho' / c \]  

(1)

where \( f \) is the GNSS carrier frequency, \( \Delta f \) is the Doppler measurement, \( c \) is the speed of light, and \( \rho' \) is the relative velocity between the GNSS satellite and the moving object.
In Fig. 2, \( I_i \) is the unit vector of the line of sight to each satellite, \( V_i \) is the velocity vector of each satellite, \( \rho_i' \) is the true radial velocity, \( V \) is the velocity vector of the GNSS receiver, and

\[
(V_i - V)S_i = \rho_i'
\]

\( \rho_{obs}' \) is the observed value of the line-of-sight velocity of the \( I_i \)-th satellite, and the measurement error of the line-of-sight velocity is \( \Delta \rho_i' \); therefore,

\[
(V_i - V)S_i = \rho_{obs}' - \Delta \rho_i'
\]

\( V_i \) can be computed from the known orbit information of the satellite. There are four unknown values in (3), so the velocity vector of the receiver \( V \) and \( \Delta \rho_i' \) in the expression, if observing using four satellites, it can be solved. In addition, by using GNSS Doppler observations it is possible to estimate the velocity on the order of cm/s.

### C. Direction Estimation

The accuracy of direction estimated from the GNSS Doppler observed value depends on the speed of the moving platform. In addition, it depends on the satellite constellation at that time. When the moving platform stops, it cannot be directly used as direction because the velocity information becomes noise. Therefore, in the case where the speed of the moving platform is very low, the reliability of the direction estimated from the GNSS Doppler decreases. Thus, in this study, the estimate of the azimuth angle is divided into very low speed (under 1 m/s) or normal speed (over 1 m/s). In the case of very low speed, only the IMU output value is used, without GNSS-estimated direction. In the case of normal speed, GNSS-based orientation is used if HDOP is also less than 5. For all other cases, we use the IMU output value in the direction estimation, using a Kalman filter or moving average, if necessary.

### III. TEST AND RESULTS

#### A. Experiment Outline

On November 19, 2014, an experiment was conducted in the urban areas of Nagoya city in Aichi Prefecture, Japan. We used Trimble Co. NetR9 and SPS855 geodetic-grade GNSS receivers. The NetR9 was used as a reference station, and was placed in a nearby test track. The SPS855 and geodetic grade antenna were installed on a vehicle roof. In this experiment, the correction data were not received in real time and we assumed real-time observation data post-processing to implement the baseline analysis. A vehicle-mounted IMU (Analog Devices Co. MEMS gyro) was used to acquire the vehicle speed information from a speed sensor, also mounted on the vehicle. This experiment was intended to obtain GNSS observation data at 10 Hz. Speed sensor data were obtained at 100 Hz, and IMU data were obtained at 20 Hz. Speed sensor and IMU data were linearly interpolated to be synchronized to GPSTIME. In addition, a POS/LV (manufactured by Applanix Inc.) was installed as a reference. The horizontal position accuracy of this system is guaranteed to approximately 20–30 cm. The test course is shown in Fig. 3, where the origin is the data acquisition start position. We traversed this course twice.

<table>
<thead>
<tr>
<th>Test</th>
<th>Test duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11/19 2014 2:00-2:30 (UTC)</td>
</tr>
<tr>
<td>2</td>
<td>11/19 2014 4:10-4:45 (UTC)</td>
</tr>
</tbody>
</table>

Analysis conditions were as follows.

1) Mask angle 15°
2) HDOP < 10
3) Lowest signal strength 20 dB/Hz
4) Using the system GPS/QZS/BDS/(GLONASS)
5) Using dual-frequency observations
6) GNSS speed information used for FLOAT solution
7) AR is LAMBDA method + Ratio test

#### B. Experiment Results

Table III shows the results for RTK-GNSS. NVS is short for the number of visible satellites. Fig. 4 shows the result of the horizontal plot of RTK-GNSS fixed solutions in Test 1. Fig. 5 shows temporal horizontal errors of RTK-GNSS in Test 1. In Test 1, there were no large errors for RTK-GNSS, which means there were no fixed solutions with large errors. All horizontal results were below 1 m. In contrast, in Test 2, there were several wrong fixed solutions over 1 m, with a maximum error of approximately 3 m. The overall results for Test 1 are...
presented here in detail, whereas the results for Test 2 were similar to the results of Test 1.

### TABLE III. RESULTS FOR RTK-GNSS

<table>
<thead>
<tr>
<th>Test</th>
<th>FIX Rate [%]</th>
<th>NVS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>57.4</td>
<td>9.2</td>
</tr>
<tr>
<td>2</td>
<td>71.0</td>
<td>9.7</td>
</tr>
</tbody>
</table>

C. Speed Information Estimated from GNSS Doppler Observations

The vehicle speed accuracy obtained from GNSS Doppler observed value and speed information of the POS/LV was compared. Fig. 7 shows the GNSS velocity error, and Table IV shows the standard deviation of the GNSS velocity error. The accumulation of errors is not reduced even by integrating the velocity information because its errors are noise-like.

### TABLE IV. GNSS VELOCITY ERRORS

<table>
<thead>
<tr>
<th></th>
<th>Latitudinal [cm/s]</th>
<th>Longitudinal [cm/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>12.3</td>
<td>12.4</td>
</tr>
<tr>
<td>Test 2</td>
<td>9.6</td>
<td>8.4</td>
</tr>
</tbody>
</table>

D. Output of IMU Data and Speed Sensor

Fig. 8 shows the time series results of the IMU in Test 1, and Fig. 9 shows the IMU direction error. The angle is set to 0° to 360°, and the first epoch direction was the direction deduced from GNSS. To remove the IMU bias error, the temporal bias was corrected by the slope information when the
vehicle was stopped. The accumulated error of the IMU is approximately 10° for 30 min.

\[ \text{Fig. 8. Direction comparison in Test 1.} \]

\[ \text{Fig. 9. IMU direction error of Test 1.} \]

The errors of the speed sensor are shown in Fig. 10, which indicates that the speed sensor cannot output the data when the speed is very low, under approximately 35 cm/s, because of the limited number of pulses. The standard deviation of the speed error for this sensor was approximately 5 cm/s.

\[ \text{Fig. 10. Speed sensor error of Test 1.} \]

**E. Integrating RTK-GNSS and Sensors**

The position was estimated from the GNSS velocity information when there was no RTK-GNSS solution. When the reliability of the GNSS results was low, they were complemented by IMU and speed sensor data. In Test 1, the direction of IMU was estimated by integrating gyro data with the GNSS-based direction using a moving average. Specifically, the IMU-based direction was subtracted from the GNSS-based direction when the GNSS direction output had high reliability, which means that the speed was higher than 5 m/s and HDOP was lower than 5. Then, we stored these differences and calculated an average value after 20–30 epochs. Finally, the IMU bias error was evaluated and compensated. The standard deviation of the direction error for total period was 0.35°. The temporal direction error is shown in Fig. 11. We used a Kalman filter for the direction estimation in Test 2, where the standard deviation of the direction error was 0.57°. Fig. 12 and Fig. 13 show the horizontal absolute error of Test 1 and Test 2, respectively. In addition, Table V shows the percentages of each solution used in each test, including RTK, GNSS velocity-based position, and sensor-based position. The maximum horizontal error in Test 1 was 2.6 m and 3.2 m in Test 2. This indicates that the integration of dual-frequency RTK-GNSS with low-cost IMU/speed sensors can output a navigation solution within 2–3 m in horizontal position, even in urban areas.

\[ \text{Fig. 11. Temporal estimated direction error of Test 1.} \]

\[ \text{Fig. 12. Horizontal absolute error of Test 1.} \]

\[ \text{Fig. 13. Horizontal absolute error of Test 2.} \]
IV. CONCLUSION

This paper described an enhanced RTK-GNSS with an IMU and vehicle sensors for application in an urban environment. First, the performance of the RTK-GNSS was evaluated. Second, the integrated performance of the RTK with IMU/speed sensors was evaluated. The experimental data were obtained in an urban environment in Nagoya city and compared to the data of a very precise and reliable reference system. The results show that the fix rate of RTK was increased significantly by adding other navigation satellite systems to the existing GPS constellation.

By integrating the RTK solutions with IMU/speed sensor information, the availability was increased to 100% with a maximum horizontal error of 2–3 m in the two tests. GNSS Doppler frequency-based velocity was useful not only for determining the positions but also for correcting the direction of the vehicle.

Because a geodetic-grade GNSS receiver was used in this study, a low-cost GNSS receiver will have to be evaluated for use in our proposed integration method.

REFERENCES

[1] DAVID M. BEVLY, STEWART COBB, GNSS for Vehicle Control, 2010
[8] L. Yand, Y. Li, Y. Wu, C. Rizos, An enhanced MEMS-INS/GNSS integrated system with fault detection and exclusion capability for land vehicle navigation in urban areas, GPS Solutions, Vol. 18, Issue 4, pp 593-603
North Finding System with Real Time Attitude Determination for a Mortar

Jin Seung Lee  
The 3rd R&D Institute-4  
Agency for Defense Development  
Daejeon, Korea  
butler@add.re.kr

Suk Won Jang  
The 3rd R&D Institute-4  
Agency for Defense Development  
Daejeon, Korea  
feldspar1@add.re.kr

Jae Gun Choi  
The 3rd R&D Institute-4  
Agency for Defense Development  
Daejeon, Korea  
jgchoi@add.re.kr

Tae Gyoo Lee  
The 3rd R&D Institute-4  
Agency for Defense Development  
Daejeon, Korea  
tglee@add.re.kr

Abstract—North identification is simply accomplished through the use of digital magnetic compass. However the accuracy of magnetic sensor is easily degraded by spatial and temporal distortions due to ferrous material or electromagnetic interference. Moreover GPS-based north finding system (NFS) can accurately determine the azimuth by two separated GPS antennas. But it still has practical limitation in a jamming or indoor environment. Therefore inertial NFS is developed to replace the classic equipment used in estimating the fire direction of hand-carried projectile weapons such as a mortar. To improve the precision of seeking north with the middle grade gyro, multi-position method is implemented. With the minimum number of inertial sensors on a manual rotary platform, the NFS can be compact, light-weight, low-cost and robust for man-portable mortar. Using additional auxiliary information which can be obtained by magnetic sensors, the sun, the stars or the geographic features, the necessary time for north finding can be dramatically reduced. The test results compared with the navigation grade inertial navigation system shows that the single digit mrad heading accuracy can be achieved. The operability of NFS mounted beside the mortar was tested in the field in which the reference positions are acquired by DGPS. To insure the durability in gun-fire shock, the drop tower test was performed based on the shock profile measured in real fire of mortar.

Keywords—gyrocompassing, north-finding, NFS, azimuth seeking, heading

I. INTRODUCTION

Heading accuracy of single digit milliradian (mrad) or better is required for setting the initial orientation in the applications such as targeting, pointing, and inertial guidance. The north finding techniques are mostly by means of sensitive devices, detecting and calculating the azimuth angle, thus determining the “north” detection. North finding is important part of inertial technology field, is widely used in military and civil applications.

North identification is accomplished through the use of digital magnetic compasses (DMCs) which are compact, low-cost instruments capable of several mrad heading precision. However the accuracy of magnetic compasses is easily degraded by spatial and temporal distortions due to ferrous materials or electromagnetic interference [1]. Moreover GPS-based north finding system (NFS) is a high accuracy, easily deployable pointing system which can determine the azimuth by passively measuring phase differences between two separate GPS antennas [2]. Although GPS-NFS is not affected by magnetic interference, it still has practical limitations in jamming or indoor environment.

Currently the inertial NFS are widely used, which can calculate the azimuth with the earth rotation rate. Gyroless north finder utilizes the Coriolis effect with a rotating linear accelerometer [3]. Nevertheless accelerometers on a PLL controlled DC motor to measure the earth spin rate is still complex and bulky. The main device of inertial NFS is gyroscope such as fiber optic gyro, ring laser gyro and hemispherical resonator gyro [4], [5]. Single digit mrad north finding and tracking over dynamic environment is often assumed unattainable by MEMS gyro in spite of light weight, low power consumption, and batch-fabrication.

II. NORTH SEEKING PRINCIPLE

Generally the inertial navigation system (INS) composed of 3 gyros and 3 accelerometers can easily calculate the azimuth. The high precision gyrocompassing also needs a high-grade expensive gyroscope such as rotor gyro or ring laser gyro. However the NFS should be compact, light-weight, and low-cost for the man-portable mortar. To improve the precision of seeking north with a middle-grade gyro, “carouseling” or multi-position method can be implemented [6].
In the conventional methods, the multi-position method was proposed to find true north with FOG. And some specified cases like four-position or three position method are widely used in practice. It is also commonly referred to as a two-position gyrocompassing by the 180° turning of the gyro sensitive axis. The two-position method is effective for the constant gyro bias through differential azimuth measurement. The main shortage of the multi-position method is that it can’t effectively suppress the bias drift of the gyro especially in a full temperature environment [7].

In the NFS, the FOG is installed on the rotation table and the sensitive axis of the FOG is parallel to the plane of the rotation table. The average measurement rotation rate \( \Omega \) at position 1 can be given as

\[
\Omega_1 = \Omega_e \cos(\phi) \cos(H + \theta) + B + \frac{\sigma_{(RW)}}{\sqrt{T_e}} \tag{1}
\]

where the subscript i denotes the number of the measurement position, \( \Omega_e \) is the earth rotation rate with a value 15.041°/h, \( \phi \) is the terrestrial latitude, \( H \) is the azimuth angle from the true North and \( \theta \) is the angle of the measurement position \( i \) from the start position. \( B \) is the gyro bias, \( \sigma_{(RW)} \) is the standard deviation of the gyro random walk and \( T_e \) is the averaging time. As the measurement time increases, the error caused by angular random walk decreases whereas the error caused by bias drift increases. Therefore, the suppression of angular random walk and bias drift are conflicted with each other, and there exists a minimized error and a corresponding optimized north finding time.

Since the random walks in the average measurement rotation rates are probabilistic values even in the same gyroscope, it may have different values every time. However the probabilistic standard deviations are identical to each other. If the average measurement rotation rates in two positions are combined with each other to obtain the gyro bias \( B \), the following equation is obtained.

\[
\Omega_1 + \Omega_2 = \Omega_e \cos(H) \cos(H + \theta) + 2B + \frac{\sqrt{2} \sigma_{(RW)}}{\sqrt{T_e}} \tag{2}
\]

Where \( \Omega_e \) is \( \Omega_e \cos(\phi) \). In a random walk term that is the last term of equation (2), the error is increased by \( \sqrt{2} \) times because the random walks in the two positions do not have a correlation. An estimated bias value \( B \) can be obtained by the following equation.

\[
B = \frac{\Omega_1 + \Omega_2}{2} - \cos\left(\frac{\theta}{2}\right) \Omega_e \cos(H + \theta) \tag{3}
\]

\[
= B + \frac{\sigma_{(RW)}}{\sqrt{2T_e}}
\]

If we choose \( \theta = \pm 180° \), there is no need for the information about the azimuth angle for the estimated bias value. With the average measurement rotation rates in two positions we can eliminated the gyroscopic bias error.

\[
\Omega_1 - \Omega_2 = \Omega_e \cos(H) - \cos(H + \theta) + \sqrt{2} \sigma_{(RW)} \sqrt{T_e} \tag{4}
\]

The random walk error in equation (4) is increased by \( \sqrt{2} \) times regardless of addition or subtraction. If we know the latitude, the approximated azimuth angle \( \hat{H} \) can be obtained using the following equation.

\[
\hat{H} = \sin^{-1} \frac{\Omega_1 - \Omega_2}{2 \Omega_e \sin(\theta/2)} = \frac{\theta}{2} \tag{5}
\]

\[
= \sin^{-1} \frac{\cos(H) - \cos(H + \theta)}{2 \Omega_e \sin(\theta/2)} \frac{\sqrt{2} \sigma_{(RW)}}{\sqrt{T_e}} \tag{6}
\]

The error of the estimated azimuth angle \( \delta \hat{H} \) is represented by following equation.

\[
\delta \hat{H} = \frac{1}{\cos(H + \theta/2) \Omega_e \sin(\theta/2)} \sigma_{(RW)} \sqrt{2T_e} \tag{7}
\]

To minimize the error of the estimated azimuth angle, the following conditions must be satisfied.

\[
\cos(H + \theta/2) = \sin(\theta/2) = 1
\]

For \( \theta = \pm 180° \) and \( H = 90° \), the upper equation can be satisfied. So using the additional auxiliary information such as from magnetic sensors, we can make \( H = H_0 + \theta_{\text{initial}} = 90° \) with the initial turning angle, \( \theta_{\text{initial}} \) of the gyroscopic sensing axis on the rotational platform as shown in figure 1. Where \( H_0 \) means the real azimuth angle which is aligned with the fire direction of the mortar. In the case that the approximated azimuth angle \( \hat{H} \) is not within 45°~135°, it means that there is large error in the estimated azimuth angle. In that case to improve the accuracy of the estimated azimuth angle, we have to perform
the second estimation process at \( H = H_H + \theta_{\text{cmd}} \) where \( \theta_{\text{cmd}} = 90^{\circ} - \hat{H} \).

\[
H_{\text{H}} + \theta_{\text{cmd}} = 90^{\circ} - \hat{H}.
\]

Fig. 1. Mortar fire direction and initial turning angle of the gyro sensing axis

Using the auxiliary information, the gyroscope sensing axis can located around the east and the problem of the many-valued function can be solved. However, there are still many scenarios in which we cannot get any auxiliary information. In that case, we need the 3-position method, and the accurate azimuth angle can be determined.

III. DEVELOPMENT OF NFS

The developed NFS is composed of the inertial sensor package, 2-axis manual rotary table and CDU. The schematic view of the inertial sensor package in figure 2 shows a single fiber optic gyroscope, three MEMS accelerometers, 3-axis magnetic sensor, ADC and some signal processing electronics. Among the sensors, the x-axis accelerometer doesn’t need for our north seeking principle. To improve the usability, the ISP data can be selectively transmitted to CDU by RS-422 or wirelessly Bluetooth. MEMS accelerometers for leveling the inertial sensor package on the rotary table are Colibrys MS9005.D series. And Honeywell HMC5883L is used to inform the initial turning angle to coarsely align the gyro sensing axis to the east. To measure the earth rotation rate, the fiber optical gyroscope used in the inertial sensor package is Kvh DSP-1750 equipped with magnetic shielding. The angle random walk of it is less than \( 0.013^\circ/\sqrt{\text{hr}} \) and the bias stability \( 0.05^\circ/\text{hr}, 1\sigma \) by Allan Variance method. Using these sensors, the best performance of NFS is theoretically better than 3mil heading accuracy for 5 minutes azimuth estimation time at the 36º latitude.

Fig. 2. Schematic of the inertial sensor package

Besides the performance limitation of the inertial sensors, there are several mechanical error sources in the inertial sensor package in figure 3 and 2-axis manual rotary table in figure 4. Using the combination of 64 align-holes and 4 ball-plungers between the turn table and the main frame, the inertial sensor package can be exactly aimed at every 100mil. This ball-plungers can dramatically reduce the turning error and improve the estimated heading accuracy.

CDU window in figure 5 presents the outputs of all sensors, indicates the procedure to estimate azimuth angle and displays the real-time attitude determinations with the estimated mortar heading.

Fig. 3. Inertial sensor package for NFS

Fig. 4. 2-axis manual rotary table for NFS

Fig. 5. CDU for the NFS
The developed NFS mounted beside the Korean 81mm mortar instead of a conventional sight unit. Using the dovetail slot in the bipod assembly, the NFS can be easily and firmly assembled and dissembled from the mortar. Because the mortar barrel is composed of iron and nickel, the magnetic compensation must be performed. In spite of several degree compass heading accuracy of HMC5883L, the hard-iron distortion by the mortar barrel can degrade the compass heading accuracy to several tens degree. By rotating the mortar with the NFS, magnetic compensation can be done. The hard and soft iron compensations can guarantee the minimum required compass heading accuracy for our NFS.

Table I. Residual error of NFS after calibration using INS

<table>
<thead>
<tr>
<th>$H_{\text{ref}}$ [deg]</th>
<th>$H_{\text{NFS}}$ [deg]</th>
<th>Residual error [mil]</th>
</tr>
</thead>
<tbody>
<tr>
<td>44.523</td>
<td>46.492</td>
<td>-3.465</td>
</tr>
<tr>
<td>46.319</td>
<td>-0.153</td>
<td></td>
</tr>
<tr>
<td>46.266</td>
<td>2.010</td>
<td></td>
</tr>
<tr>
<td>46.193</td>
<td>2.260</td>
<td></td>
</tr>
<tr>
<td>46.439</td>
<td>-2.451</td>
<td></td>
</tr>
<tr>
<td>60.814</td>
<td>-2.931</td>
<td></td>
</tr>
<tr>
<td>61.180</td>
<td>-9.766</td>
<td></td>
</tr>
<tr>
<td>59.862</td>
<td>14.868</td>
<td></td>
</tr>
<tr>
<td>60.081</td>
<td>10.771</td>
<td></td>
</tr>
<tr>
<td>60.759</td>
<td>-1.904</td>
<td></td>
</tr>
<tr>
<td>75.459</td>
<td>-3.929</td>
<td></td>
</tr>
<tr>
<td>75.595</td>
<td>-6.296</td>
<td></td>
</tr>
<tr>
<td>75.062</td>
<td>3.414</td>
<td></td>
</tr>
<tr>
<td>75.293</td>
<td>-0.795</td>
<td></td>
</tr>
<tr>
<td>75.142</td>
<td>1.956</td>
<td></td>
</tr>
<tr>
<td>Ave</td>
<td>0.247</td>
<td></td>
</tr>
<tr>
<td>STD</td>
<td>6.218</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6. The NFS accuracy test comparing with reference INS

Figure 6 shows the NFS heading accuracy test setup for comparing with the reference INS which has the sub-mil performance. The raw data which have every five measurements for each 5 minutes azimuth estimation time are obtained in 6 different heading directions. The heading accuracy test results in figure 7 verify the error of the estimation azimuth angle in equation 6. To ensure the heading accuracy, the developed NFS would be better to estimate $\hat{H}$ within ±45º from the east. Within this region, we can get the error model from figure 7 and calibrate the NFS. There are the residual errors after the calibration in table 1. After the calibration of our NFS, the heading accuracy is dominantly affected by repeatability and our NFS shows the 7mil (1σ) grade heading accuracy performance. When we start the azimuth estimation more close to east, the heading accuracy performance can be directly improved.

To test the operability of the developed NFS, we build a test field which have several reference points and azimuth angles. The two reference base points are acquired by DGPS. From these points, the several target azimuth points are marked using theodolite. In the test field, the mortar heading is aligned to the known reference heading angle using the monocular such as in figure 8. After alignment, the monocular is replaced by our NFS and the operability test is performed as shown in figure 9. After compensating $\delta H_{\text{setup}}$, the misalignment angle between the monocular and the mortar, the total repeatability error in the operability field test is 5.477mil (1σ) in table 2. From the performance comparison test in the laboratory and operability test in the field, the heading accuracy of our NFS is better than single digit mrad.

Fig. 8. The mortar alignment by the monocular in the operability test field with reference positions acquired by DGPS and theodolite
The developed NFS should have the durability for the real gun fire. From the real gun fire, the amplitude and the duration time of shock profile is measured to be 250g, 2msec. With this shock profile, the drop tower test of the inertial sensor package in figure 10 is performed three times at all directions. And there is no damages and no changes before and after drop test.

**V. DISCUSSION AND CONCLUSION**

In this study, a novel approach was developed to design a NFS using a single gyro and two accelerometers optionally with auxiliary sensors. North seeking principle with a manual rotary table was studied as the key techniques for the development of compact, light-weight and low-cost NFS. To improve the precision of seeking north with a middle-grade gyro, multi-position method on the manual rotary table was implemented. The error equations of estimated azimuth angle by 2-position methods were derived and verified. Our NFS has theoretically 3mil heading accuracy for 5min azimuth estimation time at 36° latitude. After calibrating the NFS using reference INS, the residual error shows 7mil in case of initial estimation heading angle within ±45° from east. And the single digit heading accuracy was verified in the operability field test. The durability for gun-fire shock was insured by drop tower test. Besides man-portable mortar, the proposed NFS can also be applied to the various kinds of artillery system which doesn’t have an attitude heading reference system (AHRS) for itself. The developed NFS will ensure that weapon system can beat the target quickly, mobile, and precisely in modern war.

**REFERENCES**


