4000126063/18/NL/MP
NAVISP-EL3-001

MarRINav
MARITIME RESILIENCE AND INTEGRITY OF NAVIGATION

D3b GNSS Integrity: Maritime Integrity at User Level with EGNOS V3 & M-RAIM

28th February 2020

MarRINav is a project delivered on behalf of the European Space Agency

MarRINav – Maritime Resilience and Integrity in Navigation
Work Package 2 GNSS Integrity (SBAS and Beacons)

Version	Date	Editor	Reason for change
Draft A	14.11.19	G Shaw	Version complete for internal review
1.0	27.11.19	G Shaw	Formal release
2.0	28.02.20	G Shaw	Response to RIDS

Document Information

Client	ESA
Project Title	MarRINav – Maritime Resilience and Integrity in Navigation 4000126063/18/NL/MP NAVISP-EL3-001
Deliverable Number	D3b
Report Title	D3b GNSS Integrity: Maritime Integrity at User Level with EGNOS V3 & M-RAIM
Report Version	v2.0
Report Version Date	28th February 2020
Lead Author(s)	Name: Chris Hargreaves Organisation: General Lighthouse Authorities of UK and Ireland (GLA)
Contributing Author (s)	Names: Dr Paul Williams & George Shaw (Editor) Organisation: GLA The contents of this report have been peer reviewed by a team from Stanford University in the US. Their valuable contribution is acknowledged and included in the report. The Stanford University review team members are: Sam Pullen, Todd Walter, Juan Blanch, Sherman Lo
Project Manager	Richard Greaves richard.greaves@nlaltd.co.uk NLA International Ltd
Circulation	1. Client 2. Project Files
File Name	20 02 28 D3b GNSS Integrity: Maritime Integrity at User Level with EGNOS v3 & M-RAIM v2.0.docx
File Location	Googledrive, Dropbox

Executive Summary

This MarRINav project document is the formal D3b deliverable 'GNSS Integrity' report. It provides a detailed explanation of issues and potential solutions concerning the use of Satellite Based Augmentation Systems (SBAS) and Receiver Autonomous Integrity Monitoring (RAIM) in the provision of user-level integrity and continuity for GNSS-based positioning in the future maritime environment. Those solutions are aimed at the 2025 timescale and beyond, for implementation in a vessel's future Multi-Constellation Multi System Receiver (MSR). As such the contents of this report look beyond the introduction of the 'EGNOS V2 A.1046 maritime service', expected in early 2022, and which provides integrity at system-level only. The user-level integrity solutions envisage a possible future maritime SBAS service that may derive from EGNOS Version 3 (V3) or an alternative SBAS, covering both GPS and Galileo and dual frequency (L1/L5 and E1/E5a) operation.

The proposed solutions draw upon mathematical analyses of two autonomous integrity approaches: Isotropy Based Protection Level (IBPL) and Maritime RAIM (M-RAIM). The derivation of M-RAIM (as an adaptation of aviation Advanced RAIM to maritime) forms part of the GLA' (Trinity House's) background IP declared under the NAVISP Element 3 contract with ESA for the MarRINav project. M-RAIM is considered herein to be the principal effective approach to user-level integrity of maritime navigation and it is included in full detail here with the intention to promote its development, making it freely available to marine receiver manufacturers and other stakeholders without licence constraints or royalties.

The contents of technical analysis, discussion and mathematical derivations in this report have been peer reviewed in detail by experts from Stanford University in the US. This was separately funded by the General Lighthouse Authorities of the UK and Ireland (GLA) as part of a separate review of the background IP for M-RAIM, who acknowledge the substantial value of the Stanford review as a significant contribution to the future safety of maritime navigation. Direct quotes from Stanford feedback are denoted by curly brackets and quotation marks, as {"quote "}, in the text to identify their source.

In Europe, the EC is planning a maritime service based on the existing evolution of EGNOS, known as Version 2 (EGNOS V2). This could possibly be introduced as early 2022. The proposed service would provide warnings to mariners of GPS system faults. It would protect the vessel against errors in position caused by malfunction of GPS satellites or ground processing. This capability is termed "position integrity at system level". Vessels regulated by IMO Safety of Life (SOLAS) resolutions would need to be equipped with new type-approved receivers to benefit from this service.

However, simply delivering "integrity at system level", through EGNOS V2 (or the marine beacon DGPS system), fails to take into account position errors caused by disturbances to the navigation satellite signals local to the vessel. This raises the fundamental question of whether current aviation designs are suitable for maritime service. Perhaps surprisingly, the

accuracy and continuity requirements for maritime port and harbour approach are higher than those for aircraft approaches that can be supported by satellite navigation. In addition, there are, as yet, no extensive databases of satellite signals received on ships, equivalent to those in aviation. Indeed, observations show that maritime receivers experience more satellite signal blocking and reflections, and more radio interference than do those on aircraft. There are also recent reports of possible spoofing of ships receivers. So, while the use of SBAS assures the reliability of the navigation signals transmitted from space, its ability to guarantee the quality of signals received on a ship is limited.

Recent studies, such as the SEASOLAS project of the European Global Navigation Satellite Systems Agency (GSA), have made good progress in establishing operational requirements for a maritime SBAS service. They have identified technical approaches to the maritime use of the next version of EGNOS, V3. These include emerging options for candidate receiver algorithms that estimate position errors and raise alarms when the errors threaten safety of navigation or the success of operations. We term this capability: “integrity at user level”. As yet, no method has been shown to meet both the integrity and continuity standards required to ensure safety of life at sea, on the evidence of satellite signals received aboard ships.

The EGNOS V3 designed for aviation is not optimised to provide the information needed to satisfy user level requirements. EGNOS error overbounds are assessed very conservatively for aviation and applications, whereas maritime integrity may be better accomplished using best-estimate “fault free” error models (and associated fault probabilities) rather than the inflated aviation overbounds. It is necessary to determine a nominal vessel multi-path model, and the associated probability that instantaneous measurements exceed this model (fault probability). This issue could be addressed by changing the system requirements for data parameters provided by EGNOS (and other SBAS around the world), or even by considering a new and separate maritime specific SBAS message. This may involve the integrity bound being broadcast as pairs of mean and standard deviation parameters for each satellite. Potentially, the ideal solution could be for aviation, maritime and other applications each to have its own optimised receiver design to use the SBAS information in the most appropriate way. This is discussed further in section 2.7.

Those developing EGNOS V3 and its complementary receiver software will wish to ensure that not only the signals transmitted from space but also those received on ships meet the high maritime integrity and continuity standards. Success in achieving this will require receivers that can cope with the signal errors caused by multipath and interference. The SEASOLAS project identified two potential solutions to this: IBPL and M-RAIM.

IBPL (Isotropy-Based Protection Level) is a proprietary algorithm developed by the Spanish technology group GMV. It has been designed to allow GNSS receivers to establish their integrity autonomously, especially in urban environments. The technical analysis of IBPL presented in this report, based on the rigorous mathematics described in detail, assesses the capability of IBPL against the maritime performance requirement for continuity. It considers the imbalance between integrity and continuity that would arise from use of IBPL on ships. It is also noted that the fundamental “isotropy” assumption on which IBPL is based is fundamentally untrue in the maritime environment and that IBPL should not be implemented in marine receivers for general maritime navigation.

M-RAIM is an adaptation for maritime conditions of the principal RAIM algorithm now under development for deployment in airborne receivers: Advanced RAIM (ARAIM). M-RAIM was developed by the GLA as an adaptation of ARAIM for maritime. The technical analysis of M-RAIM, based on the detailed mathematics presented in this report, considers the capability of M-RAIM to satisfy both maritime integrity and continuity performance requirements. In particular, the capability of M-RAIM to handle multiple simultaneous GNSS signal faults has been investigated. M-RAIM could work in conjunction with SBAS or be used standalone (especially in locations outside SBAS service coverage).

The report draws the following principal conclusions, with associated recommendations:

1. It would be more difficult and costly to utilise IALA Beacon DGPS than EGNOS V3 (or alternative SBAS) for the future provision of maritime navigation integrity at user-level.

   Recommendation: The dual frequency multi constellation (DFMC) capability of EGNOS V3 (or alternative SBAS), supported by the ship's Multi System Receiver (MSR), should be used in the development of position integrity for vessels rather than modifying the beacon system.

2. SBAS (EGNOS V3) alone will be insufficient to address user-level integrity for general maritime navigation due to the local GNSS signal reception environment (noise, interference, multipath and non-line-of-sight reception) on vessels.

   Recommendation: Receiver algorithms for receiver autonomous integrity monitoring (RAIM) should be designed, and an appropriate IEC test specification produced to ensure future type approved receivers adequately protect the user from potentially misleading GNSS errors caused by effects local to the vessel.

3. Maritime RAIM (M-RAIM) is a method that shows considerable promise as a candidate form of RAIM for inclusion in the maritime user-level integrity solution.

   Recommendation: M-RAIM should be researched further and evaluated for implementation in future maritime receivers when used either in combination with SBAS (e.g. EGNOS V3) or standalone (for locations outside SBAS coverage).

4. The fundamental assumption of IBPL autonomous integrity monitoring is not generally valid in maritime operations and IBPL cannot be relied upon to provide user-level integrity and continuity for vessels.

   Recommendation: IBPL should not be implemented in receivers for general maritime navigation.

5. Protection levels derived from SBAS and RAIM may be overly conservative if they are driven by “worst-case” fault scenarios and a “specific risk” integrity design.

   Recommendation: Consideration should be given to “specific” vs. “average” risk; a “fault-averaged risk” approach would provide some degree of probabilistic averaging over the prior probabilities of faults. It is noted that M-RAIM adopts a “fault-averaged risk” approach based on a-priori fault probabilities.

6. For M-RAIM, it is necessary to determine a nominal vessel multi-path model, and the associated probability with which instantaneous measurements exceed this model (fault probability).

   Recommendation: More information should be gathered from real-world measurements in the maritime environment (including how the environment varies under different operational conditions) to establish a multipath model, and the associated probability with which instantaneous measurements exceed this model (fault probability).

Contents

Table of Contents

List of Figures

List of Tables

Glossary

ARAIM	Advanced RAIM (under development for aviation)
CDF	Cumulative Distribution Function
CNMP	code noise and multi-path
DGPS	Differential GPS (provided by IALA Marine Beacons)
DFMC	Dual Frequency Multi Constellation
DFRE(I)	Dual Frequency Range Error (index)
DGPS	Differential GPS (Global Positioning System)
EC	European Commission
EEZ	Exclusive Economic Zone
EGNOS	European Geostationary Navigation Overlay Service
ESA	European Space Agency
FD(E)	Fault Detection (and Exclusion)
GBAS	Ground Based Augmentation System
GIVE	Grid Ionosphere Vertical Error
GLA	General Lighthouse Authorities of UK and Ireland
GNSS	Global Navigation Satellite System
HAL	Horizontal Alert Limit
HDOP	Horizontal Dilution of Precision
HMI	Hazardously Misleading Indormation
HPL	Horizontal Protection Level
IALA	International Association of Lighthouse Authorities
IBPL	Isotropy Based Protection Level
ICAO	International Civil Aviation Organization
IEC	International Electrotechnical Commission
IMO	International Maritime Organisation
IP	Intellectual Property
M-RAIM	Maritime RAIM
MDB	mimimum detectable bias probability of false alert
MHSS	multiple-hypothesis solution-separation
MOPS	(Aviation) Minimum Operational Performance Standards
MSI	Maritime Safety Information
MSR	(Multi Constellation) Multi System Receiver
NLOS	Non-Line-of-Sight (signal reception)
PDF	Probability Distribution Function
PFA	Probability of False Alarm (or Alert)
PL	Protection Level
PMD	Probability of Missed Detection
PMI	Probability of Misleading Information
PNT	Position, Navigation, & Timing
PPP	Precise Point Positioning
PVT	Position, Velocity & Time

RAIM	Receiver Autonomous Integrity Monitoring
RFI	Radio Frequency Interference
RIMS	receiver-integrity-monitor-stations
RNP	Required Navigation Parameters
RSS	Root Sum Squared
RTCM	Radio Technical Commission for Maritime Services
SBAS	Satellite Based Augmentation System
SDD	System Definition Document
SOLAS	Safety Of Life At Sea
TSO RAIM	Technical Standard Order (Aviation Regulations)
UDRE(I)	User Differential Range Error (Index)
UIRE	User Ionospheric Range Error
WAAS	Wide Area Augmentation System
WP	Work Package

Introduction

This report is the main deliverable D3 completing WP2 of MarRINav project. It examines the use of the European Geostationary Navigation Overlay Service (EGNOS), the Satellite Based Augmentation System (SBAS) in Europe, to support the integrity and continuity of navigation data used by the mariner. Dependence on GNSS at sea is increasing. The maritime risk of GNSS loss or degradation through interference and spoofing is becoming ever more apparent. Demands for positioning resilience, integrity and continuity are growing more urgent to support the evolving e-Navigation concept and the growth in autonomy for shipping. Current maritime electronic position-fixing capability requirements for integrity at only the system-level are inadequate to achieve this. Consolidation of user-level performance requirements is needed, leading to integrity at system and user level within a vessel's integrated navigation system.

This report should be read in conjunction with MarRINav Deliverable D6 ('MSR Integrity Report') [1] which describes the balance that must be struck between the integrity and continuity of positioning information for maritime navigation.

The report recognizes that the maritime user environment can be substantially different than the "open sky" environment assumed by aircraft at 200 feet and higher above the ground. Reflections from the ship structure and surrounding obstacles (buildings, bridges, cranes, etc.) can block direct signals and/or create significant unwanted multipath. Without the ability to exclude the possibility of such obstacles, maritime users will need the capability to detect such misleading inputs and either prevent them from affecting the positioning or to flag their potential impact on the expected positioning accuracy. As such, the report advocates a combination of SBAS and some form of receiver autonomous integrity monitoring (RAIM). SBAS is there to improve accuracy and eliminate the contribution of signal in space errors, while RAIM identifies the presence of local threats through consistency checks on the received signals' information.

The report evaluates two different consistency checking routines: The Isotropy-Based Protection Level (IBPL) and a multiple fault tolerant variant of RAIM that is labelled as Maritime RAIM (M-RAIM). IBPL is quite simple to implement but its performance raises many concerns that are highlighted in the report. Chief among these concerns is the simplifying assumption that the N-dimensional measurement error vector (for N satellites combined in the navigation solution) is equally likely to be pointing in any direction. This *isotropy* assumption may be approximately true under nominal conditions if all satellites have comparable error distributions. But under faulted conditions, where most typically only a subset of the satellites are affected and have very different error distributions, this assumption can be very far from true. Abnormal multipath affecting multiple (but not all) satellites in different ways is an example of these conditions to which maritime users in ports and harbours are susceptible. Other issues are IBPL's high degree of variability in its protection levels (due to their proportionality to the residual vector magnitude) and its inability to recognize weak satellite geometries that are critically dependent on a single potentially faulty satellite.

In contrast to IBPL, and as an improvement on "traditional" RAIM, the authors of the Stanford University peer review support the use of Maritime RAIM (M-RAIM) as explained in the report. While M-RAIM is indeed dependent on assumptions regarding prior probabilities of failures and nominal error distributions, these assumptions are much more realistic and easier to validate than the isotropy assumption required of IBPL. The Stanford University review authors do not believe that IBPL should be used as part of a high-integrity solution for maritime or other GNSS applications. The isotropy assumption that is central to the performance and simplicity of IBPL is simply too far from being true. Further, as pointed out in this report, IBPL needs extensive modification if it is to provide continuity and availability. These modifications remove the advantage of simplicity and may not be applied rigorously. Approaches based on the principles of ARAIM, as adopted for M-RAIM, remove these difficulties and are strongly recommended by the Stanford review authors.

The work undertaken for this report describes an approach and associated mathematical methods that support the development of a potential future maritime service of a European SBAS, possibly using EGNOS V3 or a separate derivative system, which aims to provide integrity and continuity of positioning at user level (sometimes referred to as 'integrity in the positioning domain'). It recognises the value of a combination of SBAS and a form of Receiver Autonomous Integrity Monitoring (RAIM) working together to protect maritime safety-of-life operations in the challenging environment of GNSS receivers on ships. Only a method of RAIM can combat the potentially high levels of multipath, non-line-of-sight signal reception, jamming and natural interference that onboard GNSS receivers may experience. Although a number of EU studies and data measurement activities towards a future maritime SBAS service with positioning integrity have been conducted or are in progress, this study introduces possible requirements for SBAS data parameters and proposes M-RAIM as an approach to meet navigation performance standards for vessels regulated by the International Maritime Organization (IMO) and governed by the Safety of Life at Sea (SOLAS) Convention.

SBAS could eventually provide not only system-level integrity but also information used within M-RAIM to determine user-level integrity. The MSR's M-RAIM would process redundant measurements from all satellites to determine the risk of an erroneous position fix being sufficiently small for safe operation. Ultimately, it is possible that the standalone capability of M-RAIM, without the additional support of SBAS but with sufficient multi-constellation satellites, could deliver the required integrity at user level to assure future maritime safety, efficient operations (including e-Navigation services and autonomous vessels) and enhanced marine environmental protection.

The document is organised as follows:

- Section 1 provides an overview of the user-level integrity approach.
- Section 2 contains a technical analysis and discussion of the solution approach.
- Section 3 presents mathematical descriptions of IBPL and M-RAIM.
- Section 4 presents the conclusions and recommendations.

- Appendices (provided by Stanford as part of their review:
- Appendix A discusses 'specific' and 'average' risk.
- Appendix B provides mathematical clarification of IBPL.
- Appendix C provides mathematical clarification of M-RAIM.

Background

Navigation integrity is a well-described and mature concept within the aviation community, but the same cannot be said for the maritime domain. The aviation required navigation parameters (RNP) for instrument approach procedures demand certain levels of performance from the aircraft's on-board navigation system. Inherent in these requirements is the concept of user-level integrity: the user's navigation equipment is tasked with providing an integrity warning to the pilot if ever the error on the derived position-fix is likely to exceed the given threshold for safe operation. Currently, mariners use GNSS as their primary aid-to-navigation, with no navigation integrity at user-level and with only limited coverage of system-level integrity (which protects against faults in signal transmission but not those in signal reception).

For clarity, the terms system-level integrity and user-level integrity are explained here:

- **System-level integrity** is a guarantee that the GNSS system itself is operating fault-free. Marine Beacon DGPS and SBAS both provide this service by guaranteeing that each GNSS satellite's transmitted *signal in space* is within acceptable parameters and is considered safe to use. This guarantee does not extend to the quality of reception of the signal and individual pseudo-range measurements made by the user's receiver, nor does is guarantee that the receiver itself is operating correctly.

- **User-level integrity** is a guarantee that the derived *navigation solution* is safe to use. Inherent in this service is a guarantee that both the GNSS signal-in-space is safe to use, and that the individual measurements made by the user's receiver are sufficiently accurate. Because the augmentation system (Differential GPS Beacons or SBAS) cannot know about the local environment of the user, or the operation of their receiver, user-level integrity can only be achieved by the receiver itself (with co-operation from the augmentation service, or any other systems which may aid the integrity-monitoring function).

An EGNOS V2 'maritime 1046 service' is expected to become operational in early 2022 to provide system-level integrity of the GPS-based position solution for vessels equipped with appropriate type approved receivers. The service will alert mariners to faults occurring with the GPS ground or space segments, effectively advising of the non-usability of signals broadcast from one or more GPS satellites. Whilst this is an important step in the provision of integrity for maritime navigation, it falls short of alerting mariners to adverse effects local to a vessel (e.g. the noise & interference environment and signal obscuration or multipath) that could introduce unacceptable position errors. The conditions for GNSS signal reception on vessels vary with many factors, from ship to ship, with the cargo load, the phase of voyage and the associated surrounding infrastructure, while overall being dependent on the quality of the GNSS antenna installation. Achieving adequate user-level integrity under this wide spread of operational conditions presents a significant technical challenge.

SBAS is designed to provide accurate corrections for the signal-in-space (originating from the satellite or in the ionosphere or troposphere) as well as associated confidence bounds when applying those corrections. It cannot monitor the extremely localized error conditions affecting the user measurements such as multipath, RF interference or receiver faults. The same limitations affect GBAS, DGPS, and any other external augmentation. Aviation users are fortunate to operate in open sky environments where these local effects can be presumed to be bounded by relatively small uncertainty values. The responsibility of meeting these localized bounds falls upon the user, not the SBAS provider. The maritime community could take a similar approach; however, their operating environment may be considerably more cluttered such that an upper bound on potential locally-generated errors may be prohibitively large. The maritime community is likely better served by having local error detection (e.g. RAIM) to detect large local errors (such as multipath) rather than assuming that large errors may always be present. This report suggests optimizations that may allow for combinations of SBAS and RAIM that support user-level integrity with better availability and continuity.

RAIM is traditionally implemented to only handle the presence of a single faulty measurement. However, different versions of RAIM have been developed to handle multiple faults. It is important to remember that RAIM is primarily a measure of measurement inconsistency. It is possible to construct an entirely consistent set of measurements that results in a false position (e.g. spoofing or when all direct signals are blocked and only reflected signals are received).

1 SBAS and Maritime Navigation Integrity

1.1 EGNOS Provision of System-Level Integrity

It has long been accepted that satellite navigation systems such as GPS or Galileo are unable on their own to provide the reliability required for safety-of-life use. For decades, GPS has been supported by augmentation systems, such as maritime radiobeacon differential GPS (DGPS); here ground stations monitor the satellite transmissions and send warning messages to users should the GPS signals become unreliable.

The civil aviation sector has chosen to adopt Satellite Based Augmentation Systems (SBAS) such as WAAS (Wide Area Augmentation System) and its European counterpart EGNOS (European Geostationary Navigation Overlay Service). The navigation satellites' signals are monitored on a continental scale by a network of ground stations. Any malfunction detected causes a warning to be communicated to users' receivers via geostationary satellites. The messages generated at the SBAS ground stations also include continuous corrections for smaller position errors, enabling the receivers on-board the aircraft to deliver more accurate position fixes.

In the course of developing these aviation SBAS systems, large numbers of measurements were made on satellite signals received aboard aircraft. On the basis of statistical, analytical and physical analysis, SBAS and receiver software was then developed that ensured that airborne users would enjoy extremely high levels of "integrity"; that is, confidence that the aircraft's position was correct and prompt warnings should there be a failure. Another vital goal was that "continuity" should be protected: that is, an aircraft embarking on a certified airport approach procedure would have a very high probability of completing it successfully without system failure. SBAS have been designed to meet the needs of the civil aviation community for most, but not all, operations, using four basic performance parameters: accuracy, availability, integrity and continuity.

There are now proposals for the formal use of SBAS on ships. In Europe, the EC is planning to introduce a 'maritime 1046 service' in early 2022, based on the existing evolution of EGNOS, known as Version 2 (EGNOS V2). The proposed service would respond to the requirements of IMO Resolution A.1046 [2] to provide warnings to mariners of GPS system faults. It would protect the vessel against errors in position caused by malfunction of GPS satellites or ground processing. This capability is termed "position integrity at the system level". Vessels regulated by IMO Safety of Life (SOLAS) resolutions would need to be equipped with new type-approved receivers to benefit from this service.

Existing maritime DGPS beacons already provide suitably equipped vessels with position integrity at the system level, provided they are within the coverage of a beacon. The proposed EGNOS V2 service has the clear advantage over the beacon system of covering a much wider geographical region. Many DGPS beacons in Europe are approaching the end of their operational service life and face obsolescence of components. Some maritime authorities have elected to upgrade their systems using EGNOS information as the source of GPS fault and error information. Others, including the GLA in UK and Ireland, have yet to announce plans for the future of their beacons. It is also evident that the cost to the maritime community of the maintenance and upgrading of the beacon infrastructure would be avoided if EGNOS V2 could replace some or all of the beacon DGPS systems in Europe.

1.2 Potential for EGNOS Support to User-Level Integrity

Simply delivering "integrity at system level", through EGNOS V2 and the DGPS system, fails to take into account position errors caused by disturbances to the navigation satellite signals local to the vessel. Limited observations, such as those obtained in 2018 during the SEASOLAS project of the European Global Navigation Satellite Systems Agency (GSA), indicate that maritime receivers experience more satellite signal blocking and reflections, and more radio interference than do those on aircraft. There are, as yet, no extensive databases of satellite signals received on ships, equivalent to those in aviation. The ESA MARGOT project ('Multipath & Interference Error Mitigation Techniques for Future Maritime e-NAV Services') within the NAVISP Element 1 programme is currently investigating GNSS signal reception on vessels. There are also recent likely reports of actual spoofing of ships receivers, corroborated by many independent observers and traced back to a source location. Whilst there are user-level bounds on SBAS suitable for aviation in the flight environment (e.g. bounds on multipath errors from aircraft reflecting surfaces), the principal maritime challenge is perhaps that SBAS consideration of integrity at the user level is probably not yet adequate to fully address the natural and man-made threats encountered in maritime operations. So, while the use of SBAS assures the reliability of the navigation signals transmitted from space, its ability to guarantee the quality of signals received on a ship is limited.

The peer review of this report by Stanford University in the US commented:

{"The maritime user environment can be substantially different than the 'open sky' environment assumed by aircraft at 200 feet and higher above the ground. Reflections from the ship structure and surrounding obstacles (buildings, bridges, cranes, etc.) can block direct signals and/or create significant unwanted multipath. Without the ability to exclude the possibility of such obstacles, maritime users will need the capability to detect such misleading inputs and either prevent them from affecting the positioning or to flag their potential impact on the expected positioning accuracy. As such, the GLA paper advocates a combination of SBAS and some form of RAIM. SBAS is there to improve accuracy and eliminate the contribution of signal in space errors, while RAIM identifies the presence of local threats.

SBAS is designed to provide accurate corrections on the signal-in-space (originating on the satellite or in the ionosphere or troposphere) as well as associated confidence bounds when applying those corrections. It cannot monitor the extremely localized error conditions affecting the user measurements such as multipath, RFI, or receiver faults. The same limitations affect GBAS, DGPS, and any other external augmentation. Aviation users are fortunate to operate in open sky environments where these local effects can be presumed to be bounded by relatively small uncertainty values. The responsibility of meeting these localized bounds falls upon the user, not the SBAS provider. The maritime community could take a similar approach; however, their operating environment may be *considerably more cluttered such that an upper bound on potential locally-generated errors may be prohibitively large. The maritime community is likely better served by having local error detection (e.g. RAIM) to detect large local errors (such as multipath) rather than assuming that large errors may always be present. It is not so much that SBAS is not designed to meet maritime needs as it is that the maritime community is better served by having localized consistency checks as part of its overall integrity monitoring."}*

Recent studies, such as the SEASOLAS project, have made good progress in establishing operational requirements for a maritime SBAS service. They have identified technical approaches to the potential maritime use of the next version of EGNOS, V3. These include emerging options for candidate receiver algorithms for the provision of "integrity at user level", that estimate position errors and raise alarms when the errors threaten safety of navigation or the success of operations. However, as yet, no method has been shown to meet both the integrity and continuity standards required to ensure safety of life at sea, on the evidence of satellite signals received aboard ships.

The EGNOS V3 designed for aviation is not optimised to provide the information that a maritime user would require. This shortcoming could be addressed by changing the data parameters provided by EGNOS (and other SBAS around the world), or even by considering a new and separate maritime specific SBAS message, if either of these were shown to be cost-effective. Potentially, the ideal solution could be for aviation, maritime and other applications each to have its own optimised receiver design to use the SBAS information in the most appropriate way.

1.3 The Need for a form of RAIM

The development of any future maritime EGNOS V3 service and its complementary receiver software will wish to ensure that not only the signals transmitted from space but also those received on ships meet the high maritime integrity and continuity standards. Success in achieving this will require receivers that can cope with the signal errors caused by multipath and interference, through the inclusion of a form of Receiver Autonomous Integrity Monitoring (RAIM).

RAIM is traditionally implemented to only handle the presence of a single faulty measurement. However, different versions or RAIM have been developed to handle multiple faults. It is important to remember that RAIM is primarily a measure of measurement inconsistency. It is possible to construct an entirely consistent set of measurements that results in a false position (e.g. spoofing or when all direct signals are blocked and only reflected signals are received).

This report evaluates two different consistency checking routines: The Isotropy-Based Protection Level (IBPL) and a multiple fault tolerant variant of RAIM that is labelled as "M-RAIM" (Maritime RAIM). IBPL and M-RAIM were identified as two potential solutions within the SEASOLAS project.

1.3.1 IBPL

IBPL is a proprietary algorithm of the Spanish technology group GMV. It has been designed to allow GNSS receivers to establish their integrity autonomously, especially in urban environments.

IBPL is quite simple to implement but has many concerns that are highlighted in this report. Chief among these concerns is the simplifying assumption that the N-dimensional measurement error vector (for N satellites in solution) is equally likely to be pointing in any direction. This *isotropy* assumption may be approximately true under nominal conditions if all satellites have comparable error distributions. But under faulted conditions, where most typically only a subset of the satellites are affected and have very different error distributions, this assumption can be very far from true. Abnormal multipath affecting multiple (but not all) satellites in different ways is an example of these conditions to which maritime users in ports and harbours are susceptible. Other issues are IBPL's high degree of variability in its protection levels (due to their proportionality to the residual vector magnitude) and its inability to recognize weak geometries that are critically dependent on a single potentially faulty satellite. Maritime users in ports and harbours may experience signal obstructions that weaken satellite geometry and make such "critical satellites" more likely.

The technical analysis of IBPL presented in Section 2.4, based on the rigorous mathematics in section 3.2, assesses the capability of IBPL against the maritime performance requirements. It concludes that the rapid variations that may occur in the integrity values computed by IBPL on ships may present an unacceptable continuity risk. It is also noted that the fundamental "isotropy" assumption on which IBPL is based is demonstrably false. The assumption does not hold against satellite faults which are likely to affect only one satellite at a time. It also does not hold against large reflections and blocked direct signals, which will only affect a subset of satellites.

The peer review of this report, by Stanford University in the US, highlighted a key recommendation:

{"The authors do not believe that IBPL should be used as part of a high-integrity solution for maritime or other GNSS applications. The isotropy assumption that is central to the performance and simplicity of IBPL is simply too far from being valid. Further, as pointed out in the GLA report, IBPL needs extensive modification to be able to protect continuity and availability. These modifications remove the advantage of simplicity and are not strictly correct in places. Approaches based on the principles of ARAIM and M-RAIM remove these difficulties and are strongly recommended by the authors."}

1.3.2 M-RAIM

Maritime Receiver Autonomous Integrity Monitoring (M-RAIM) is an adaptation for maritime conditions of one of the principal RAIM algorithms now emerging for deployment in airborne receivers: Advanced RAIM (ARAIM). ARAIM standards for aviation are being developed and harmonized between many international stakeholders with deployment aimed for the mid-2020s. Based on the principles of ARAIM, M-RAIM has been developed by the General Lighthouse Authorities of the UK and Ireland (GLA) and forms the background IP provided in

Section 3.3. The technical analysis of M-RAIM presented in Section 2.6, based on the rigorous mathematics of Section 3.3, considers the capability of M-RAIM to satisfy both maritime integrity and continuity performance requirements. In particular, the capability of M-RAIM to handle multiple simultaneous GNSS signal faults without imposing an impractical data processing load remains under investigation.

The peer review of the contents of this report, by Stanford University in the US, made the following comment:

{"In contrast to IBPL, and as an improvement on 'traditional' RAIM, the authors support the use of M-RAIM as explained in the paper. While M-RAIM is indeed dependent on assumptions regarding prior probabilities of failures and nominal error distributions, these assumptions are much more realistic and easier to validate than the isotropy assumption required of IBPL."}

1.3.3 Potential of PPP

Stanford University's review of the contents of this report have provided an additional perspective on the potential of precise point positioning (PPP) to address multipath concerns. The review stated:

{"The report does not consider the use of precise point positioning (PPP), which is a class of powerful carrier-based techniques that could very effectively mitigate many large multipath concerns, but it comes with many other issues, including the current lack of an accepted integrity monitoring methodology for PPP. PPP utilizes carrier-phase measurements along with a Kalman filter to estimate the carrier phase ambiguity offsets. It is therefore much less susceptible to the effects of multipath. Sub-meter accuracies are often achieved. However, PPP requires an initialization phase to reduce the uncertainty on the ambiguity estimates. Depending on the quality of the correction information provided, PPP can take more than 15 minutes to achieve meter level accuracies. Further, PPP performs best when it is able to continuously track the phase measurements. Thus, it performs well against multipath, but not as well in obstructed environments. Referenced recent Stanford papers [3, 4] provide useful background on PPP and related integrity methodologies."}

1.4 Discussion of 'Specific' and 'Average' Risk

This section 1.4 is included verbatim from the peer review feedback from Stanford University.

{"The report notes in several places that the protection levels derived from SBAS and RAIM are conservative because they are driven by "worst-case" fault scenarios. As discussed under 'specific' vs. 'average' risk below, some degree of probabilistic averaging over the prior probabilities of faults and (with more difficulty) the distribution of errors generated by these faults may be possible. However, this should be carefully justified, and the resulting reductions in protection levels are likely to be smaller than anticipated.

1.4.1 Definitions of "Specific" and "Average" Risk

In various places in the report, concern is expressed regarding the "conservatism" of the protection levels generated by SBAS and RAIM due to, among other things, inflation of nominal error sigma values *and worst-case assessment of faults and their consequences. Underneath this concern is an issue that relates to the appropriate interpretation of the maritime integrity and continuity requirements. This issue has become known as the difference between "specific risk" and "average risk", where the "risk" involved is the risk of loss of integrity or continuity. Definitions of these terms are given below, and further discussion of how these terms are interpreted in integrity and continuity analysis is provided in Appendix A. Additional background can be found in the reference papers [5, 6].*

***"Specific risk"** is the approach normally used in interpreting aviation integrity requirements, or at least those requirements that protect against events with a severity beyond "Major" (meaning operations where loss of integrity is "Hazardous" or "Catastrophic" as defined in Chapter 3 of the "FAA System Safety Handbook" – see [7].*

The risk of such an event depends on the prior probability of the event (meaning its long-term probability of occurrence without additional real-time information being provided, such as monitor test statistics) and the magnitude of errors that could be generated by the event. Specific risk requires that the worst-case parameters be chosen for both of these values unless substantial, verifiable information exists to support a less-conservative choice.

*Among the possible variations of "average risk," **"fault-averaged risk"** is the closest to "specific risk" and only differs in the degree of probabilistic averaging allowed for rare and unpredictable fault conditions. Where specific risk requires a very strong basis for assigning and making use of prior probabilities of faults, fault-averaged risk allows historical data and engineering judgment to [be used avoid rare faults being assumed to occur with high probability]. Similar sources of information can be used to derive distributions of errors within threat models.*

*"**Geometry-averaged risk**" represents a much more extreme variation of "average risk" in which all parameters that vary with time or location (and thus could be represented by probability distributions) are eligible for probabilistic averaging. This applies not only to rare and unpredictable faults but to parameters that are known and can be calculated. "Geometry" is used to name this variation because, for GNSS applications, a large benefit comes from not using the known GNSS satellite geometry but instead applying a probability distribution of geometries or a distribution of a summary parameter (such as HDOP). Therefore, poor satellite geometries (or poor HDOP) that by themselves cannot provide acceptable protection levels may become acceptable (available for use) if their contribution to the overall probability of geometry or HDOP is sufficiently small.*

As explained further in Appendix A, the authors generally recommend starting with the "specific risk" interpretation used by aviation for integrity and then carefully considering places where "fault-averaged" risk may be justified in view of the nature of the faults and error effects being evaluated and the intent of the underlying integrity requirements. The authors strongly recommend against trying to make use of "geometry-averaged risk" for integrity or continuity analysis, as they deem it improper to average over parameters that are easily known to users."}

2 Technical Analysis and Discussion

2.1 Building on the Foundations of the SEASOLAS Project

Development of EGNOS V3 will enable support for Galileo satellites and the new L5 civilian signals. This development provides the opportunity for other sectors, additional to civil aviation, to influence system development to take account of their own requirements and specificities. The possibility of modifying the service for V3 could enable support for safety-of-life or liability-critical applications extended to the maritime and other sectors.

The GSA's SEASOLAS project was aimed at identifying the user needs and operational requirements for a maritime SBAS service, which could be engineered into the EGNOS V3 design. For various reasons the project's scope did not fully and openly address its maritime objectives, principally because:

1. No modifications to the existing EGNOS (V2) design were considered, excepting multi-constellation dual frequency GNSS capability via additional support messages for Galileo satellites and L5 signals.

2. Alternative user-level RAIM algorithms were not fully considered, given a focus on isotropy-based protection level (IBPL).

3. Key details of IBPL were not disclosed, limiting independent evaluation of the results of the project.

4. Performance analysis was conducted using the PROSBAS software, which is not openly available, and details of this method of analysis were not disclosed.

The MarRINav project offers the following contribution (described within this report) to investigate gaps that remain in a fuller and more open understanding of maritime SBAS and RAIM:

1. The MarRINav team's understanding of IBPL is described in detail, including derivation of the k-factor calculation, which is key to its integrity claims. A method for considering the continuity performance of the algorithm is also described, and a "continuity preserving IBPL" variant is developed and evaluated.

2. An extension of classical RAIM to provide a horizontal protection level (HPL) robust to multiple simultaneous faults is described.

3. The M-RAIM algorithm based on multiple-hypothesis solution-separation (MHSS) is described. This is a modification of ARAIM designed to be used in conjunction with SBAS augmentation to provide a HPL.

If SBAS is to be safely employed by the mariner, then its strengths and weaknesses need to be fully appreciated, and the interaction between SBAS system-level integrity monitoring and RAIM receiver-level monitoring also needs to be understood.

Since M-RAIM is based on ARAIM, and ARAIM (and RAIM in general) does not require SBAS or other augmentation, it is noted that there is potential for application of M-RAIM without augmentation (i.e. supporting "standalone GNSS"). This could be the case for ships outside of SBAS service coverage regions (and, if in mid-ocean, with looser accuracy and HAL requirements than those required for coastal or port & harbour approach phases of the voyage).

2.2 Technical discussion of maritime SBAS and RAIM

In principle, if a system-level problem arises in GPS then it should be very quickly spotted by the network of SBAS receiver-integrity-monitor-stations (RIMS) and an alert can be sent to the user via a number of geostationary satellites. These alerts prevent the navigator from using faulty or inaccurate data from GPS and prevent hazardously misleading information (HMI) which can be a danger to the user's and others' safety.

2.2.1 EGNOS for maritime

SBAS, including EGNOS V2, provides corrections for the GPS satellite's orbital parameters and clock-synchronisation, and corrects for atmospheric delays which affect the GPS signal – all of which improve the accuracy of the user's navigation solution. The SBAS system also broadcasts aviation integrity bounds, which describe the maximum amount of error that could still be present on the GPS signals after correction. These bounds are employed by the user's receiver to estimate a maximum amount of navigational error (a protection level) which surrounds their navigation solution and provides the level of safety the aviation navigator needs.

SBAS is designed primarily to serve the civil aviation sector, and the protection levels it provides are tuned to the particular safety and performance requirements demanded by this industry. Other industries may use the open-access SBAS signal to improve their GPS service, and make use of the integrity information at their own risk without any contractual guarantee of the service performance. So, non-aviation users of current SBAS must be aware that the system is not designed to serve their needs and may not meet their particular expectations.

The maritime sector makes use of GPS on-board vessels for navigation, timing and ancillary purposes. Potentially EGNOS V2 is very valuable to the mariner, providing an alternative augmentation to the IALA Marine Beacon DGPS system, which serves the same purpose of improving accuracy and warning of integrity events affecting the GPS navigation solution. Marine Beacon DGPS has been designed to suit the mariner's needs, providing high accuracy differential corrections for users local to the beacon, and issuing timely warnings if GPS faults are detected.

It is intended that the ‘maritime EGNOS V2 A.1046’ service will augment GPS position fixing by improving accuracy and providing the required level of system integrity for the mariner. The introduction of operational capability for the ‘maritime EGNOS V2 A.1046 service’, expected in early 2022, will provide mariners (of vessels equipped with suitable type approved receivers) with a ‘guarantee’ of the maritime system-level integrity of EGNOS V2 augmented navigation solutions. The performance will be formally declared by a specific System Definition Document (SDD) generated by the European GNSS Authority (GSA) and the ‘guarantee’ will be provided via formal EGNOS Working Arrangements between the EC/GSA and the maritime authorities of EU Member States.

Additionally, Maritime Safety Information (MSI) will be provided in the event of any EGNOS non-availability. This service will provide “integrity at the system level”, warning of GPS faults. However, further changes to an EGNOS maritime service may be necessary in future, potentially as part of EGNOS V3, to contribute to the estimates of GNSS (GPS and Galileo) positional error bounds (protection levels) necessary for the provision of integrity at the user level. Yet such changes to SBAS will not be sufficient by themselves due to the signal effects in the local environment, in particular multipath and non-line-of-sight measurements, which cannot be addressed solely by a change in the parameters sent by EGNOS.

2.2.2 Local GNSS Environment on Vessels

Some environments are potentially quite hostile to reception of the GNSS signals. Noise or interference from other radio systems, or deliberate jamming may obscure the weak GNSS signal. This signal may be received both directly and/or after reflecting from surrounding surfaces such as are common in the built environment. This *multipath* interference is a problem for GNSS receivers, especially if the direct signal is blocked. Urban environments are particularly problematic, especially if the user is surrounded by high-rise buildings, which block the direct line to the satellites and scatter reflections of the GPS signal. Vessels in ports often operate in or near to the built environment and the ship’s structure itself or cargo containers it carries or offloads can be a source of multipath interference, especially if the GNSS receiving antenna is poorly sited.

Neither the Marine Beacon DGPS system nor EGNOS V2 is capable of protecting the navigator from these GPS signal reception hazards. In both cases of DGPS and EGNOS, an integrity risk remains from the local environment which may render GPS navigation potentially unsafe unless the user’s receiver is able to enact some form of receiver autonomous integrity monitoring (RAIM).

2.2.3 Impact of User Level Requirements on Use of SBAS Information

It is important to understand both the challenges of the maritime operational environment and the required level of performance (accuracy, integrity, continuity and availability) of the output position-solution. In particular, the most appropriate maritime requirements are set out in IMO Resolution A.915 [9], which requires:

- Good levels of horizontal accuracy in the solution (the vertical dimension is not generally an issue, neglecting inland waterways);

- Very good availability and continuity performance, potentially even exceeding that demanded from aviation systems;
- Good user-level integrity (although the requirements are not as demanding as aviation);
- Resilience against threats from the maritime environment such as noise and multipath (as well as jamming and spoofing) which may be significant, and resilience against faults of which several may occur simultaneously.

As well as for navigation, GPS data is used by many down-stream services within the typical integrated bridge system on a modern ship. The IMO's e-Navigation initiative to standardise and harmonise the use of electronic data at sea will further the integration of GNSS (GPS, Galileo and other navigation constellations) into automated ship's systems. To ensure the safety and reliability of these systems it is imperative that *all* integrity risks to GNSS are contained or mitigated.

It is unlikely that IALA Marine Beacon DGPS could be easily modified to meet user-level integrity requirements; in fact, SBAS is closer to meeting them. More specifically, beacon DGPS can detect when measurement errors exceed a certain threshold, but it does not place a bound on how large the errors can grow when nothing is detected. This notion of observability is key in SBAS and is the main reason for the observed margin between the actual errors and the broadcast bounds.

The inflated error bounds of aviation SBAS guarantee that the aviation user's position is bounded by the Protection Levels to a risk of 10^{-7} derived from the combined range-error bounds[1]. However, further research is necessary to determine whether the maritime user would be able to derive less stringent position error bounds from the SBAS aviation information, especially if (a) combined with a form of RAIM, and (b) used for operations with integrity-risk requirements on the order of 10^{-5} (such as in IMO Resolution A.915 [9]).

The discrepancy between the needs of the mariner, and the aviation-inspired design specification of SBAS means that the SBAS broadcast data may be unsuitable or require modification before safely being incorporated into the receiver's data processing. The World's SBAS are currently undergoing modification to support new GNSS constellations and new navigation signals. This may represent a time-limited opportunity to update and modernise the service to satisfy mariner's operational needs through appropriate changes to the next generation of SBAS.

[1] When discussing SBAS and EGNOS integrity, it should be noted that SBAS providers do not guarantee the integrity of the pseudoranges and the broadcast error bounds independently. For example, SBAS does not specifically guarantee that its estimates of User Differential Range Error (UDRE) and ionospheric vertical error bounds (GIVE) individually bound clock-ephemeris and ionospheric errors to the 10^{-7} level.

However, the Stanford University authors of the peer review of this report's contents are sceptical that the proposed modification of existing SBAS UDRE and GIVE values to create maritime specific estimates of UDRE and GIVE could be done by SBAS at reasonable cost. The effort needed to create, verify, and monitor such estimates in real-time would create such a burden on existing SBAS that the authors are confident that none of the existing non-European SBAS would consider it. Even if a European mandate were to enforce this responsibility upon EGNOS, the resulting maritime solution would only work within EGNOS coverage. The alternative approach would be for the maritime community to independently determine these estimates, but this would still require detailed insight into each SBAS to be utilized. The implications are discussed in more detail in section 2.7.

2.2.4 RAIM and the Additional Need to Bound Local Errors

It is proposed here that a minimum standard RAIM algorithm is defined to be included in future maritime receivers and to augment SBAS use in the multi-constellation multi-system receiver (MSR). Receiver manufacturers are not limited to the minimum standard and may implement bespoke data processing, but this must be shown to perform at a level not less than the standards dictate. Several options are available and two are introduced below in sections 2.1.5 and 2.1.6.

However, more is needed for any form of RAIM to be useful. In particular, a robust model that bounds the variability of the local environment needs to be established such that RAIM can separate abnormal (and rare) behaviour from troublesome (but unfortunately common) nominal behaviour. The need for GNSS data from ships to help build such models is mentioned in Section 1.2 (e.g. the MARGOT project), but it is important to clarify that RAIM (or any other integrity monitor technique) depends on being able to separate what is relatively frequent and therefore has to be tolerated (and therefore increases protection levels) from what is infrequent and can be targeted for detection and exclusion while still meeting continuity.

There is currently no consistent requirement[2] for all ships' receivers to employ RAIM, but many do and as such the quality of the implementation can and does vary between receivers. The multi-constellation multi-system receiver (MSR) performance standard [8] includes SBAS and RAIM, but its IEC test specification has only recently commenced development.

2.2.5 Maritime RAIM (M-RAIM)

M-RAIM is an adaptation of Advanced RAIM (ARAIM), which itself is a development of TSO-RAIM, the standard RAIM algorithm used for decades by aviation receivers to validate GPS performance. ARAIM[3] aims to detect faulty measurements from the aviator's navigation solution. Both M-RAIM and ARAIM form many solutions from different subsets of GNSS

[2] The IMO GPS Performance Specification requires integrity checks, but does not state how these should be made, with RAIM and DGPS offered as examples. The IEC test specifications for GPS and Galileo include tests for RAIM to confirm functionality, but they do not mandate any specific algorithm.

[3] It should be noted that Advanced RAIM is still in development, so the ARAIM algorithm is not yet widely employed in aviation algorithms. It is the case, however, that algorithms based on the same principles as ARAIM (i.e. solution separation) have been implemented for RAIM and are widely employed.

measurements and can detect and mitigate a wide variety of different faults, including situations where multiple faults or failings happen simultaneously.

The method is critically dependent on a number of assumptions, but this is also true of all other RAIM methods and integrity monitor algorithms in general. Thus, it is not a general weakness of M-RAIM compared to these other methods. The detailed analysis of each of these methods involves assessing the "reasonability" and sensitivity of results to the assumptions required of each method, and this is addressed within the analysis reported here.

M-RAIM is also potentially computationally expensive, demanding the calculation of many subset navigation solutions. However, recent research has found ways to reduce this computational cost greatly. For example, see recent Stanford papers [10, 11].

MRAIM is discussed in detail in section 2.6 and a full mathematical description is provided in Section 3.3.

2.2.6 IBPL

Isotropy-Based Protection Level (IBPL) is an autonomous integrity methodology which attempts to characterise the severity of measurement errors using the degree of consistency of the derived solution. It is computationally much less demanding than M-RAIM, but its capability to deliver the necessary level of integrity in any situation is not proven. The integrity method is also quite conservative and hence the frequency of alarms may be high in some situations. To counter this, the GLA have attempted to develop the concept of 'continuity-preserving IBPL', which includes a calculation of a continuity-preserving bound such that the frequency of interruptions from alarms is kept to a manageable level. However, this version of IBPL remains unsatisfactory for maritime use.

IBPL is discussed in more detail in Section 2.4 and a mathematical analysis of IBPL is provided in Section 3.2.

2.3 Performance Requirements

The performance required from electronic position fixing systems (EPFS) used at sea is described in IMO Resolution A.1046 [2]. This provides minimum performance specifications for any GNSS (or other electronic position fixing system) wishing to be considered as part of the IMO's World Wide Radionavigation System. GNSS identified in this group, along with an appropriate IMO receiver performance requirement, are allowed to be used for safety-of-life applications on a vessel. Therefore, these requirements can be considered as baseline navigation requirements. The requirements listed are limited, particularly as they do not account for integrity threats local to the vessel and only demand a rapid time-to-alarm in case of failures in the GNSS being detected.

Currently, GPS alone is insufficient to meet the IMO specifications; it is necessary to augment GPS with the Marine Beacon DGPS service to meet the time-to-alarm requirement. A proposed future set of requirements A.915 [9] is not in force but exists as a positioning document aimed at guiding future navigation system developments. Resolutions A.1046 and A.915, covered in detail in Work Package 1, are used as a basis for a proposed set of requirements for the SBAS/RAIM analysis of MarRINav, covering the coastal navigation and port & harbour approach phases of the voyage, as explained below.

2.3.1 Requirements for Coastal Navigation and Harbour Approach

The performance requirements used for the MarRINav analysis of SBAS and RAIM application to the coastal and harbour approach phases of the voyage are set and discussed here:

- Horizontal accuracy 10m (95%), which is consistent with the A.1046 requirement for coastal navigation and harbour approach.

  An integrity warning (red light) should be issued to the user if ever their actual navigation solution error exceeds the 25m horizontal alert limit (HAL), which is 2.5 times the 95% accuracy. Exceeding this limit without an alarm is termed hazardously misleading information (HMI)[4]. It is the GLA view that for the coastal and harbour approach phases of the voyage, it could be navigationally unsafe to place the vessel any less than 25m from a hazard using electronic position fixing as sole-means to maintain safety. In these voyage phases, smaller HALs or tighter accuracy requirements are considered not to be operationally sensible.

  Even the best integrity monitors cannot observe the actual position error, so in practice they generate horizontal protection levels (HPLs) that are computed to bound actual errors to the probability of HMI and which issue warnings if the HPL exceeds the HAL. Therefore warnings must occur when the error bound at the required probability (HPL) is too large, which is distinct from the unknown time when the actual error is too large although aimed to coincide. If the HPL is below the HAL then it is determined that adequate integrity is provided.

- If no alarms are present, this situation is expected to continue for the next 15-minutes with probability 99.97%; this is consistent with the current continuity requirements in A.1046.

- The probability of HMI should not exceed the integrity requirement of 10^{-5} per 15-minutes.

  This is a reduction of the operational interval in A.915 which effectively slackens the integrity requirement. The 15-minute interval is here consistent with the continuity requirement and provides a more realistic integrity requirement than the 3-hour interval given in A.915.

- Overall solution availability (plus green light integrity guarantee) of 99.8%, over all time. In practice this is measured per 2 years [12].

[4] The maritime use of the term 'Hazardously Misleading Information' differs from its use for aviation.

Whenever a “red light” integrity warning is issued, the mariner must revert to alternative means of navigation. This potentially causes significant disruption to the navigating officer, and the operation of the vessel. Such an event constitutes a loss of continuity. The loss of availability continues as long as the red light remains on. Ideally these events should be rare. The requirements above are in-line with those of A.1046 and constitute a risk of outage of no more than about once per month.

Reduction of the integrity requirement from 3 hours to 15 minutes represents only about one expected HMI event over every 3 years of any given ship’s operations.

2.3.2 Snapshot Integrity and Continuity Risks

Both continuity and integrity can be thought of in terms of a mean time between events (1 month and 3 years respectively) or as a risk per unit time interval (0.03% and 0.001% per 15-minutes). When considering different time intervals the allowable risk must be adjusted accordingly. As a simple example, the continuity risk allowance 0.03% per 15 minutes could be thought of in terms of three intervals of 5 minutes, each incurring 0.01% risk. The performance requirements may be re-worded over any arbitrary time frame.

Every time the RAIM algorithm is run it incurs a chance that it fails to detect a problem (missed detection, leading to an integrity risk) or that it erroneously raises the alarm (false alarm, potentially leading to discontinuity). The RAIM algorithm must ensure that the cumulative risks to continuity and integrity do not exceed the stated 15-minutes allowances (3×10^{-4} and 10^{-5} respectively).

One might think that the more frequently the RAIM algorithm is called, the faster these risks accumulate, and the less likely the system is to meet the requirements. The issue is correlation – if the time interval is made short enough eventually successive intervals can no longer be considered statistically independent. The results of a RAIM algorithm called at a particular time are highly correlated to the results of calling the algorithm, say, 1 second later. As the time-interval between successive events shortens, the allowable risk gets smaller, until the time intervals approach the length of the correlation-window, at which point the probability “plateaus” to a value, here called the *snapshot risk*.

As an example: superstition says Friday 13th is unlucky. Simple inspection of the calendar shows that the 13th comes around once per month, and each time there is a roughly one in seven chance of this day being a Friday. One might say that the Mean Time Between Friday 13th (MTBF13) is about 7 months, or just over 200 days. Each day picked at random from the calendar therefore has about a one in 200 risk (0.5%) of being Friday 13th. Pick any 2-day interval and the risk of any one of these being F13 approximately doubles to about 1%, pick 3 days and the risk is about 1.5%, and so on.

Now consider shorter times – picking any one-hour interval at random does not cut the risk down by 24 times! A simple calculation: the 200 days (of which one is F13) is 4800 hours (of which 24 are F13). The risk for a one-hour interval is 24 out of 4800, or the same 0.5%. The mathematical reason is that Friday 13th hours are correlated – the day is always 24 hours long.

The *snapshot risk* for a Friday 13th is 0.5%. No matter how short the time-interval (pick any nanosecond at random from history) the risk of it falling on F13 is never less than 0.5%.

So, what is the correlation time for GPS, and what is the snapshot risk for the continuity and integrity requirements? The literature is inconsistent on the correlation time for GPS measurement errors across environments but figures are typically quoted between one and fifteen minutes. In the analysis of this report, an assumption is made that GNSS errors remain correlated over 2.5 minutes (150s) as this agrees with the assumptions made about GPS integrity, when applied to aviation in DO-229 [13].

{The Stanford University peer review authors cautioned to keep in mind that the assumptions in DO-229 reflect airborne velocities with respect to the reflecting surfaces that generate multipath, so it is not guaranteed that correlation times based upon aviation (in flight, at different speeds) fully represent the maritime environment. Therefore analysis of real-world measurements on vessels would be appropriate to determine the correlation window. Correlation times for faster-moving antennae are usually shorter however, so assuming the aviation-specific figure of 150s is probably conservative for a maritime receiver.}

This yields the result that six statistically independent epochs are expected per 15-minute interval, and leads to deriving the per-epoch snapshot risks as 1.667×10^{-6} for integrity and 5×10^{-5} for continuity. These are the risk requirements used to investigate the performance of candidate RAIM algorithms, and for the purpose of this analysis they are considered fixed.

It should be noted that this does not imply a form of "average risk" (as explained in section 1.4). Although an "average risk" approach may allow some relief from the conservatism that accompanies "worst-case" assessments, the approach adopted here is to treat each epoch in isolation: so-called "snapshot" operation. The snapshot analysis considers that a string of epochs lasting 15-minutes, with a given correlation-time, should meet the requirements. If six statistically independent epochs occur per 15-minutes, each snapshot application of the RAIM algorithm should not consume more than one sixth of the total budget. This leads to an integrity risk of 1.667×10^{-6} per epoch and 5×10^{-5} continuity risk per epoch.

2.4 IBPL

A number of GPS pseudo-ranges are measured by the receiver, these are used to derive the receiver co-ordinates in three-dimensional space, and the clock offset from GPS time. Usually there are more measurements (satellites in solution) than there are co-ordinates to solve (x, y, z, and t). The co-ordinates are derived for which the measurements *best fit* the solution, the measure of goodness of fit is the sum square of the *residuals* (the difference between the measurement and the expected range to each satellite, from the solution co-ordinates). The receiver obviously doesn't know the accuracy of its estimate of the co-ordinates, but it can calculate the measurement residuals.

2.4.1 Basis of IBPL and its Performance Claims

The basic premise on which Isotropy Based Protection Level (IBPL) is built is error isotropy. This assumes that when an error or measurement fault occurs it has equal probability of affecting any of the measurements, and in any combination. The joint error distribution of all the pseudo-range measurements is required to have an *isotropic* probability distribution function. IBPL involves using the vector of pseudo-range residuals as a way to observe the co-ordinate error directly. The magnitude of the vector of residuals is scaled according to the horizontal dilution of precision (HDOP) and a k-factor chosen to provide a given level of integrity. The method is very easy to compute: the receiver will probably calculate HDOP anyway, and the IBPL's k-factor can be pre-determined and stored in a look-up table.

The method is claimed not to depend on any assumed error models and nor to require any prior knowledge about the kinds of fault that may happen. As such, IBPL is claimed to be very robust to environmental conditions, since it does not rely on any assumptions made about how badly GNSS might be affected. These claims are false, as the basic assumption of isotropy is fundamentally an assumption about error models and, furthermore, one that is demonstrably untrue.

Since IBPL is based on the actual measurement values per epoch, the instantaneous value of the protection level is highly variable and pays little if any regard to the need for continuity. The calculation can be modified to provide a protection level which is not expected to be exceeded more than a given fraction of the time, so offers a "continuity preserving" variant which does not depend on instantaneous measurements but only on solution geometry. However, the basic premise of IBPL (error isotropy) remains demonstrably untrue.

The isotropy assumption can only be valid in very specific circumstances, since generally each satellite pseudo-range measurement would be expected to have a slightly different level of accuracy depending on location in the sky, age, quality and operation of the satellite. When measurement faults or mistakes occur, their distribution is expected to be considerably *anisotropic* and so integrity performance of IBPL may be worse than the isotropy assumption suggests.

2.4.2 Critique of IBPL

The following is an extract from the Stanford University review, providing their input in regard to IBPL:

{"Isotropy is a very strong (and incorrect) assumption on fault behavior and therefore the kinds of faults that may happen (and how badly GNSS might be affected). Here are examples of error characteristics leading to non-isotropic behavior:

- *Satellites with different error distributions (e.g. low-elevation satellites are expected to experience larger errors than high-elevation satellites)*
- *Single satellite faults being more likely to occur than multiple satellite faults*
- *Satellite faults more likely to affect certain satellites than others (e.g. GLONASS vs. GPS, satellites that are near to being occluded resulting in only a reflected path).*

All of the above are consistent with normal and expected behavior. The isotropy assumption may be close to valid in an average sense (i.e. averaged over many different geometries and many user environments), but the report establishes that it is not correct in a specific sense (i.e. with each and every specific geometry and expected error distribution). As such, IBPL is not usable for aviation *integrity applications. If maritime integrity is specified in an average rather than a specific sense, then IBPL could be considered. However, it is not clear that IBPL is valid even in an average sense. Perhaps, at a probability of 10^{-5}, it may be possible to establish this empirically."}*

The IBPL algorithm contains no screening of weak satellite geometries either and so may be vulnerable to HMI under certain conditions. The large k-factors typically used mean that the algorithm appears to be quite pessimistic. It may appear to some observers that IBPL in most conditions achieves reasonable levels of real-world integrity performance despite its limitations. Indeed the magnitude of the residuals vector is an excellent test statistic on which to build a more conventional RAIM algorithm.

However, the nature of the IBPL algorithm means that false-alarm continuity failures are expected to be common unless a large number of very accurate pseudo-range measurements are available for use. There is a concern that if IBPL were to be implemented for maritime safety applications that it may be observed empirically to be effective, not because of its inherent properties, but because GNSS is very reliable. This could be a significant threat to everyone in the future who travels by sea, if IBPL is implemented into future generations of maritime navigation receivers. The very concept of maritime safety could be put at risk if the basis of "empirically verified" were to replace "provably safe".

Hence, serious doubts arise about the maritime suitability of IBPL's integrity and continuity performance on ships and indeed it is concluded that IBPL should not be implemented for safety-of-life maritime operations.

2.4.3 Protection Level Calculation

The residuals vector (ω) is used as a direct measurement for the position error. It assumes that the position error is described (and bounded) by a given multiple of the magnitude of this vector, so that the IBPL protection level (PL) is given by:

$$PL = k_I \; HDOP \; \sqrt{\omega^T \omega}$$

A scaling factor k_I is added to provide sufficient integrity from the protection level. It is a number which ensures that the PL is inflated sufficiently to provide a conservative bound on the position error. An apparent advantage of this technique is that the calculation does not depend on the specifics of an assumed error model and is easy to compute.

Values for k_I which depend on the number of satellites in solution, and the required integrity risk can be pre-computed. The following table illustrates the k_I factor as a function of the number of satellites (N) contributing to the solution and the required integrity risk (α).

The k-factor can be derived by a method explained in Section 3.2.1. It attempts to ensure that the PL of IBPL is larger than the position error with a given level of probability. The principle is that the probability of a measurement error of any particular *magnitude* occurring is not important, since the position error and residuals (hence IBPL) both scale linearly with this, so the integrity risk is the same no matter the size of the errors. As noted elsewhere in this section, the isotrophy assumption is very unrealistic under real-world conditions.

N \α	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
5	14.94	150.0	1500.0	1.5×10^4	1.5×10^5	1.5×10^6	1.5×10^7
6	4.30	14.09	44.70	141.42	447.21	1414.21	4472.13
7	2.67	6.19	13.52	29.22	62.98	135.72	292.40
8	2.03	4.00	7.31	13.11	23.37	41.60	74.00
9	1.68	3.02	4.99	8.03	12.80	20.33	32.25
10	1.46	2.47	3.82	5.74	8.51	12.55	18.46
11	1.30	2.12	3.13	4.49	6.32	8.85	12.35
12	1.18	1.87	2.68	3.71	5.04	6.79	9.10
13	1.09	1.69	2.36	3.18	4.20	5.50	7.16
14	1.02	1.56	2.12	2.80	3.62	4.64	5.90
15	0.96	1.44	1.94	2.52	3.20	4.02	5.02

Table 1 – k-factor for the IBPL algorithm

Table 1 presents the value of the k-factor, indexed by the number of satellites (N) down the left-hand column and required integrity-risk (α) along the top row, as published by InsideGNSS [14].

The claimed performance of IBPL is apparently independent of the actual error distribution. Since the distribution of errors (particularly hazardous integrity faults) are often unknown by the receiver, this purports to make IBPL particularly powerful especially in environments which might severely disrupt the accuracy of GNSS positioning, such as may occur with significant signal obscuration and multipath. This claim is based entirely on the isotropy assumption being true; however, since it is not true, the claim is false.

2.4.4 IBPL and Weak Satellite Geometry

Most integrity algorithms make use of the redundant position-fixing information available to the receiver in the form of the residuals; IBPL is not unique in employing them. When measurement errors are large, these correspond to large position errors in the solution, and usually large measurement residuals. The size (magnitude) of the residuals is a very powerful indicator of the size of the measurement errors and therefore the accuracy of the navigation solution.

There exist certain satellite geometries (combinations of satellite positions in the sky) which may make one satellite critical to positioning. Such a situation can be described by means of a polar sky plot, showing the whole sky as a wheel, with the rim as the horizon, satellite elevation plotted as the closeness to the centre (high elevation) and satellite azimuth plotted as the bearing round from the top (North).

The diagrams in Figure 1 showing the location of 8 satellites with small HDOP (left-hand plot) and large HDOP (right-hand plot) illustrate the HDOP principle, for example the location of the satellites on the right-hand plot in Figure 1 approximately line up across the sky. They provide very good positioning accuracy along this line, but poor accuracy perpendicular to it (in a North-East / South-West direction). In contrast, the left-hand plot shows satellites distributed throughout the sky. For the right-hand plot the user's Horizontal Dilution of Precision (HDOP) would be large and their position accuracy poor. Since IBPL also depends on HDOP, the protection level would also be larger when geometry is poor to account for the fact that the navigation solution is expected to be less accurate.

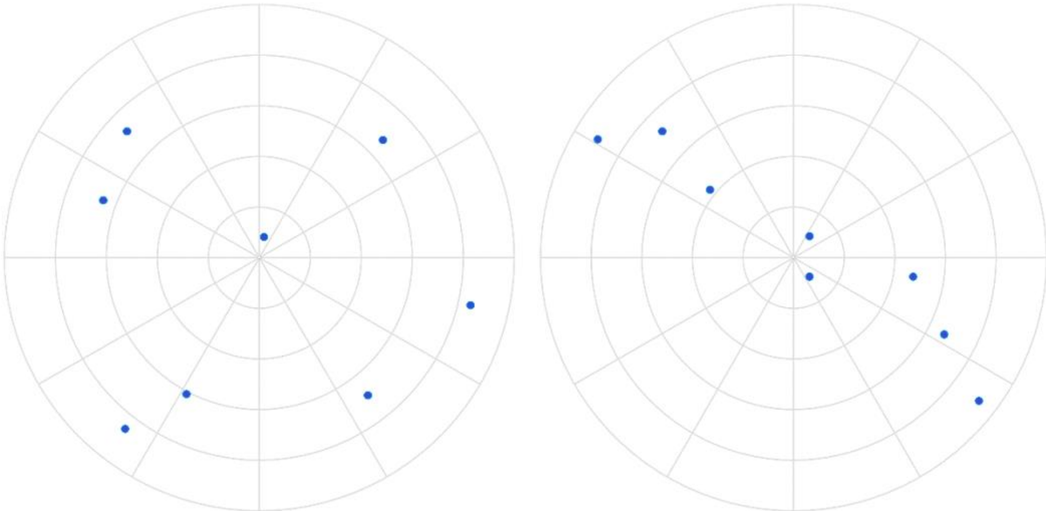

Figure 1 – Example sky-plots of satellite positions

However, if the navigation solution relies on one particular satellite for accuracy in a particular direction then this becomes *critical*. This is not the same as large HDOP or a poor geometry as above, and so this scenario is shown in Figure 2. For example the HDOP of the plot on the right-hand in Figure 1 can be reduced (accuracy improved) if one of the satellites lies away from the line of satellites across the sky, where even though accuracy may be good, a satellite (as circled in Figure 2) is critical and the solution is vulnerable if this one fails.

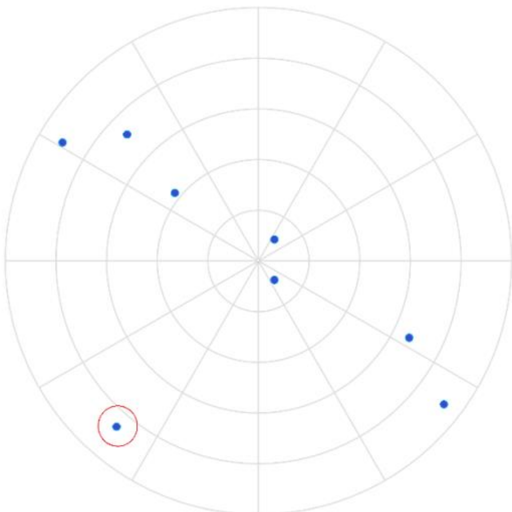

Figure 2 – Example of 'critical satellite' weak geometry integrity-wise

For the scenario of Figure 2, accuracy will be better, and HDOP the protection level will be smaller, but the solution relies heavily on this *critical satellite* (circled in red) for accuracy in a NE / SW direction. This scenario is quite crucially *not the same as a solution with large HDOP*, and should not be thought of as such. HDOP may be quite small for this geometry of satellites.

The problem is that the satellite circled in red is critical for the solution. The residual for this satellite will be small, even if the pseudo-range measurement this satellite provides is greatly in error. This is described as a *weak* geometry since (although accuracy may be good) the solution is reliant on one satellite and its integrity is vulnerable to a failure on this satellite. This is a new definition of a kind of geometry – one with good HDOP, but one which is critically dependent on a single satellite. Hence, the phrase *"weak"* is used for this scenario to differentiate it from a solution with large HDOP for which the word *"poor"* was used. The implication is that the satellite positions shown in **Error! Reference source not found.** describe a good geometry (low HDOP) but one which is easily broken if the critical satellite is lost or faulted, hence *weak.*

We offer an analogy to further explain the differences between our use of the phrases *poor* and *weak*. Suppose there is a police investigation that relies on the testimony of a number of witnesses to establish the facts of a case. If several independent witnesses corroborate each other's story then their information is credible and can be trusted.

If some crucial information is missing (due to a lack of witnesses) then the case may be said to be *poor* as not all of the facts are known or can be proven. Alternatively, when a key detail (such as identifying the suspect) is only provided by a single witness, the whole case can be called *weak*, as it will break down if this one witness is mistaken or lying.

Without this crucial witness, the key information may be missing and there may not be a case at all. It is a judgement call as to which situation is preferable: having no case due to a lack of information (and being aware of this fact), or critical reliance on a single source, which would be catastrophic if this source is false or misleading.

The usual method of determining the integrity of the solution based on the residuals using RAIM is to scan for such weak geometries. Typically, the measures used are the amount of error caused by a failure on one satellite, and the resulting change in residuals. A weak geometry is characterised by a given level of error on the satellite inducing a large position error and a small residual. The ratio of error-to-residual (or error-to-detectability) is used, and is referred to in some RAIM algorithms as the *slope* of the satellite [5].

The protection level can be defined to include this information such that weak geometries require larger protection levels to remain safe. Often a RAIM algorithm will reject weak geometries, as the protection level must be inflated so much that it exceeds the alert-limit.

IBPL makes use of the residuals and the HDOP, but does not consider the effect of critical satellites or weak geometries. This is in part because the assumption is made that errors are isotropic – the isotropy assumption predicts that it is very rare for one satellite to fail in isolation. By contrast, most RAIM processes consider single-satellite failures to be a very significant threat, so take the possibility of weak geometry very seriously. The result is that RAIM and IBPL differ considerably in their treatment of *critical* satellites.

2.4.5 Summary of Problems with IBPL

The IBPL algorithm as it is defined contains certain vulnerabilities:

1. The alarm condition is raised if ever the protection level exceeds the alert limit. The detection threshold for a “fault” is not set based on a false alarm allowance and as such offers no implicit continuity guarantee. Without enough satellites in solution, continuity performance may be poor.

2. The algorithm relies on the isotropy assumption holding true to preserve integrity. Measurement errors generally are not isotopically distributed hence a weighting matrix is usually used in the navigation solution, and faults occur usually on only one or possibly two satellites in isolation so are also considerably *anisotropic*.

3. There is no screening out of weak geometries, for example the case described above of a solution with reasonable HDOP but where a fault on a single satellite could be catastrophic for positioning accuracy. Classical RAIM would reject such a solution, but IBPL does not.

By far the most severe problem with IBPL is not the question of whether it meets the maritime continuity or integrity requirements – but the fact that it does not inform the mariner either way. IBPL may empirically appear to meet the integrity requirements given its terms and assumptions, which neglect the risk of a failure of a single critical satellite due to the isotropy assumption (as bullet 3 above). Lack of geometry screening and lack of protection against large variations in HPL (affecting continuity and availability) are indeed fundamental weaknesses of the algorithm (except as addressed in the following sections 2.4.6 and 2.4.7 and in Section 3.2).

Two possibilities exist:

1. The geometry may be sufficiently weak that a single-satellite fault constitutes a major integrity risk.

2. The measurement errors may be large enough (or the number of satellites in solution small enough) that IBPL is expected to breach the HAL frequently.

In neither case would the algorithm inform the mariner of the potential vulnerability. The biggest problem with the algorithm is not whether it meets the performance requirements, but the fact that when it doesn’t, the mariner is left unaware. Without completely re-designing the algorithm, there may be adjustments that can be made to alleviate the worst of these vulnerabilities, but it remains totally unsatisfactory.

2.4.6 Continuity of IBPL

The IBPL protection level (PL) depends strongly on the magnitude of the residuals. These depend on the individual pseudo-range measurement errors for each satellite. With some simple assumptions, weighting the residuals by the expected accuracy of the measurements, the result is that the sum square of residuals should form a chi-squared consistency test. Chi-squared is often used to analyse raw data such as in medical trials to determine whether the data fits a known assumption (null hypothesis) or whether something unusual has happened (alternate hypothesis). When chi-squared is used in RAIM, the null hypothesis is that the error models that we have assumed are indeed correct, the alternate hypothesis is that a fault has happened and the old models no longer apply. Usually the alternate hypothesis means that the solution cannot be trusted and an integrity-alarm results.

Since the square of the PL is also a multiple of sum square of residuals, we can relate it to a chi-squared variable in a very similar way. The behaviour of a chi-squared variable is a standard result of statistics. It is possible to estimate the probability of it exceeding various levels. We can use this to calculate the expected continuity performance by calculating the probability that the snapshot value of the PL exceeds the HAL. Solutions can be rejected if this probability exceeds the maximum allowable snapshot continuity risk (5×10^{-5}).

Alternatively, starting with the maximum allowable continuity risk and working backwards from the chi-squared distribution, the level can be determined which the PL remains below for 99.995% of all snapshot epochs. This effectively provides a “continuity bound” on the output protection level, and only if *this* value remains below the HAL are both integrity and continuity satisfied. Calculating the chi-squared limit for the appropriate probability (5×10^{-5}) can be done in advance and the results tabulated for easy look-up, just like the k-factors.

A full mathematical description of calculating IBPL continuity and establishing the continuity-preserving bound are given in Section 3.2.2.

2.4.7 Geometry Screening

A similar process to that performed for the classical RAIM algorithm of calculating the slope of each satellite can be performed to protect IBPL against critical satellites and weak geometries. The slope can be used to inflate the protection level to attempt to protect the isotropy assumption from a highly anisotropic fault on a single satellite. This process is described more fully in Section 3.2.3.

2.4.8 IBPL Summary

- The algorithm is computationally very efficient, depending only on a look-up table for the k-factor, and calculation of the HDOP and residuals.

- The simple IBPL process does not depend on any assumptions about the accuracy of the position fixing, or the nature of integrity hazards and so it is very flexible as to which environments it can operate in.

- Integrity depends on the isotropy condition holding true. It does **not**, and this may cause integrity problems, especially for some weak satellite geometries.

- No implicit continuity guarantee is made by the algorithm. The IBPL is expected to fluctuate with the measurement noise and its conservative nature means it may often exceed the HAL.

- A continuity-preserving variant has been investigated which generates a revised protection level that is only exceeded by the IBPL for a small percentage of the time. If this threshold is below the HAL then the algorithm preserves both continuity and integrity (subject to the isotropy condition).

- The continuity-preserving variant is very conservative and re-introduces the need to estimate the accuracy of the measurements, and so loses principal advantages that IBPL offers.

The Stanford University review concluded:

- The large values of the k_I term in the IBPL protection level equation may cover-up weaknesses in the algorithm due to non-isotropic measurement errors, but this would be impossible to prove at the level needed for safety-of-life applications.

- **IBPL should not be used as part of a high-integrity solution for maritime or other GNSS applications. The isotropy assumption that is central to the performance and simplicity of IBPL is simply too far from being valid.**

2.5 RAIM

"Classical RAIM" includes many possible approaches; one of them is the slope-based approach [15] described in this section, while solution separation (as also applied in ARAIM and M-RAIM) is another.

The navigation solution residuals are used as a check of the consistency of the derived solution. The weighted sum of residuals is used, and this forms a chi-squared test as described in section 2.4.6. A maximum allowable limit, or fault detection threshold is set by considering the random variations of a chi-squared statistic. We can use the maximum allowable continuity risk ($5x10^{-5}$) to set the confidence level (in this case 99.995%) and then calculate the chi-squared value (χ_0^2) which will be exceeded exactly this fraction of the time, taking into account the number of degrees of freedom (number of satellites in view minus 4).

Number of satellites	Degrees of Freedom	Limit χ_0^2
5	1	16.4
6	2	19.8
7	3	22.5
8	4	25.0
9	5	27.3
10	6	29.4
11	7	31.5
12	8	33.5
13	9	35.4
14	10	37.3

Table 2 – Chi-squared limits at a confidence level of 99.995%
shows chi-squared limits at a confidence level of 99.995%.

Number of satellites	Degrees of Freedom	Limit χ_0^2
5	1	16.4
6	2	19.8
7	3	22.5
8	4	25.0
9	5	27.3
10	6	29.4
11	7	31.5
12	8	33.5
13	9	35.4
14	10	37.3

Table 2 – Chi-squared limits at a confidence level of 99.995%

When faults are small, they are hard to detect. Only when a fault is large enough to cause the test statistic to exceed the threshold is it considered *detectable*. Faults are expected to occur randomly at a given rate, and a certain level of integrity is demanded – dividing these two yields a minimum tolerable probability of missed-detection: this is the rate at which faults that occur can be missed and the user's integrity requirement will still be met. For a maritime user this missed detection risk will be about one percent; the chi-squared test must be very reliable and have a detection fidelity (positive detection of faults) of at least 99%. The derivation of this is found in Section 3.

The risk of missed detection is given by the statistical variation of the chi-squared variable when a bias is present – the variability of the so-called *non-central* chi-squared distribution is again a standard result of statistics. A value is calculated for the smallest amount of non-centrality which still ensures adequate missed detection – we call this the minimum detectable bias (MDB).

Each satellite is considered in turn, and the magnitude of the error needed for that satellite to cause the MDB is calculated. This gives the smallest satellite error which can be reliably detected (larger errors are easier to spot). It is assumed that an error almost this large can happen at random any time without detection. Depending on the geometry, a similar error magnitude on different satellites yields a different position error – this is the measure of how *critical* each satellite is to the solution.

The satellite that is *most* critical is the one the RAIM algorithm is most interested in – this is the one that is most likely to cause HMI if it fails. A protection level is created which includes a contribution from random errors, and also the MDB for the most critical satellite. This is the worst-case type of failure and it is the condition under which the algorithm must prove its integrity: all other types of satellite failure are less severe and so if integrity is preserved under the worst-case, it is not necessary to check the other cases.

If the calculated HPL exceeds the HAL then this indicates that integrity cannot be provided for the given geometry. An alert needs to be raised to warn the navigator that they cannot guarantee integrity for this navigation solution. In this way classical RAIM screens out weak geometries: solutions which are critically dependent on one satellite's operation are given a very large HPL.

This algorithm suffers from some limitations:

- It is assumed that only one satellite is "faulted" at once. Theoretically the chi-squared test can detect multiple simultaneous failures, but the HPL is calculated based on the assumption that only one fault can happen at a time.

- The algorithm assumes that when a fault happens, it is always the worst-case and the most critical satellite incurs an error of exactly the largest un-detectable bias. Most fault cases will be less severe than this, and so the HPL is conservative.

  Stanford University review authors commented that *{"the worst-case error bounds are generally very robust (subject to the other assumptions that they are based on), but they may be over-conservative compared to the intent of the maritime integrity requirements. The degree to which 'worst-case' analysis is required is a distinction between 'specific' and 'fault-averaged' risk"}* as discussed in section 1.4.

For a maritime receiver employing several GNSS constellations in a potentially complex radio environment the risk of more than one satellite bias (fault) existing simultaneously may be significant. Although RAIM may detect multiple simultaneous failures, the HPL is calculated under the assumption that only one happens at once.

There exist methods for calculating the worst-case combination of two or more simultaneous faults. This process requires more computations but is essentially identical to the method described above for determining a protection-level robust to the occurrence of the fault. Generally the more satellites considered faulty at once, the worse the error can be. This is always true if the satellite faults are all worst-case faults. More simultaneous faults implies the need for a larger HPL, as several individual faults may combine to cause a large position error which nonetheless remains undetectable to the chi-square test. This applies to all forms of GNSS-only consistency checking and not only chi-square based versions. Hence, when considering large number of simultaneous faulty satellite signals the HPL created can be extremely large.

The RAIM process is based on a null-hypothesis test, either a fault is detected (chi-squared test exceeds threshold and integrity alarm is raised) or it is not. Integrity must be preserved under either eventuality. To ensure that integrity is preserved when no alarms are raised an HPL is defined such that the probability of the error exceeding this level is less than the required integrity risk. This must hold true even if the worst-case type of fault happens. In practice, this makes for a very conservative protection level since it is always critically dependent on the worst-case fault.

Some fault cases might be hazardous in that they cause considerable position error without being detectable, but might also be extremely rare. It may be overly cautious to base the HPL on the worst-case, if this kind of fault is an extremely remote possibility and hence the idea of using a 'fault-averaged risk' approach (as discussed in section 1.4). Since "null-hypothesis" is a choice between faulted or not faulted, it considers the worst-case (specific risk). Some probabilistic model of individual fault likelihoods is needed to make an assessment that is different from one based on the worst-case fault. If such a model is available, other options exist, as suggested by the 'fault-averaged risk' approach. Hence, multiple-hypothesis is an attractive option in considering a whole set of possible faults, any of which might happen and for which modelled probabilities would need to be determined, so that the risk is balanced over all of them.

Ideally a multiple-hypothesis integrity method should be employed, with each potential type of fault considered separately and 'weighted' in the solution based on:

1. How common the fault is;
2. How easy the fault is to detect;
3. How much position error it induces.

There are practical reasons why a multiple-hypothesis RAIM algorithm should use a different fault detection measure other than the chi-squared test. Below, MRAIM is considered as a method based on the Advanced-RAIM (A-RAIM) algorithm proposed for global LPV-200 flight operations [16]. This is an optimised multiple-hypothesis solution-separation (MHSS) test, as further described for aviation vertical integrity [17, 22].

2.6 M-RAIM

M-RAIM is an integrity algorithm based on multiple-hypothesis solution-separation (MHSS) [17, 22]. The logic is this: if a particular set of measurements are faulty and each can incur any amount of error, then the navigation solution can be greatly in error. However, a solution that excludes exactly those faulty measurements will be accurate. By comparing the two (the *all-in-view* solution and the *reduced-subset* solution) then it is possible to see if the excluded set of measurements was faulty or accurate.

For each combination of possible faulty measurements (and there may be lots of different combinations) a reduced subset solution is calculated and compared to the all-in-view solution. If the solutions are significantly different, then this is an indication that a fault has occurred and the alarm should be raised. It is then possible to identify the erroneous measurements, remove them and try again to find a consistent set of solutions.

A detection threshold is set – this is the region around the all-in-view solution which must contain all of the subset solutions to avert an integrity alarm.

A protection level is also set – this is the region around all of the subsets which is expected to also contain the true position of the receiver. If the protection level exceeds the HAL then the solution cannot be used.

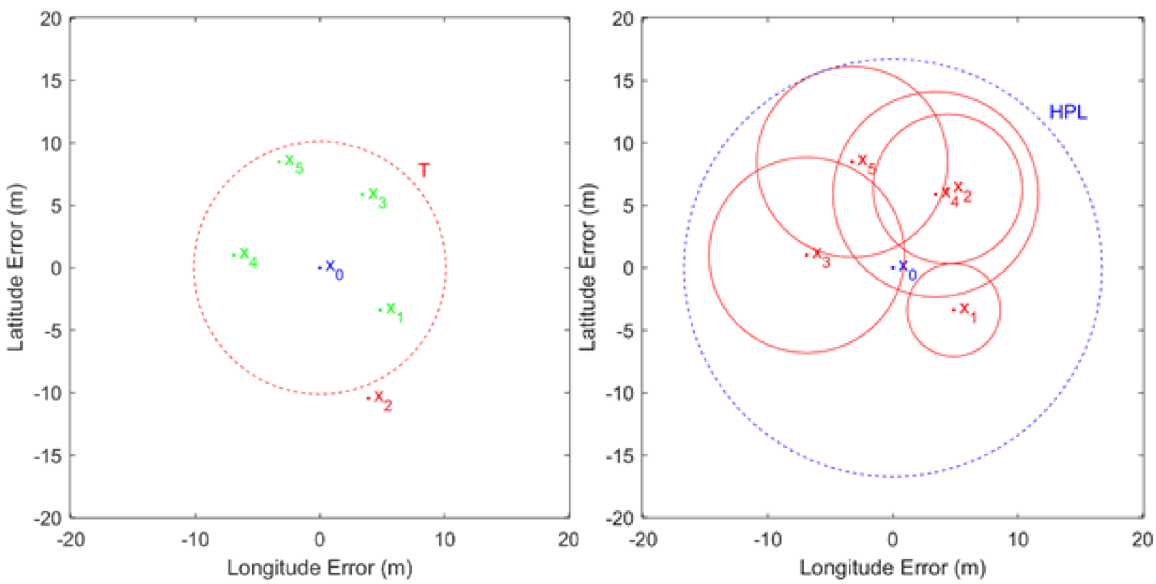

Figure 3 – Example operation of M-RAIM.

In Figure 3, in the left-hand plot, the all in view solution (x_0, blue) is compared to several reduced subset solutions (green). Any that lie outside the threshold (T) are considered erroneous (e.g. x_2, red), and trigger an alarm. In the right-hand plot of Figure 3, the HPL is then the union of all the subset solutions and their respective integrity bounds.

If any of the subset solutions exceed the threshold under nominal measurement conditions or when no underlying faults exist then this is a *false alarm*. For this reason, false alarms are sometimes more precisely described as “fault-free alerts”. Setting the detection threshold too tight causes too many false alarms and this harms the continuity performance of the system.

If the threshold is too lenient then faults might happen and not be detected, such a *missed detection* may cause the true position to lie outside the protection level.

False alarms harm continuity, missed detections harm integrity, so both are carefully budgeted within the receiver [1] when setting the detection threshold and the protection level such that the user's integrity and continuity requirements are met.

This process is mechanically very similar to ARAIM [17, 22]. Indeed just like the aviation equivalent, the process is critically dependent on a number of a-priori assumptions such as the probability of each type of fault occurring, and the assumed "fault free" error models used to estimate pseudo-ranging accuracy at the receiver. M-RAIM is intended to be compatible with existing and future augmentation systems such as DGPS and SBAS (and not as a replacement to them, although it could be used standalone outside the coverage areas of augmentation systems) and so differs from ARAIM in this regard.

The major limitation of this process is the dependency on a-priori assumptions (both nominal fault-free error models and fault-probability values). These assumptions must be conservative for integrity to be guaranteed, and the magnitude of the HPL is critically dependent on them. If these assumptions are too pessimistic, then the HPL may be over-inflated and the availability of the method affected. Also if a large number of potential simultaneous faults are considered then the number of combinations of different kinds of faults can be extremely high resulting in significant computational cost in the receiver.

Stanford University review authors have indicated that recent research has found ways to reduce this computational cost greatly. For example, see the recent Stanford papers [10, 11].

2.6.1 Robust Subset Solutions

A number of reduced-subset solutions are formed, for each subset a particular unique set of pseudo-ranges are removed from solution. The subset solution is said to be *robust* to faults on the set of eliminated measurements since these can take any (large) value and the error on the solution would not change.

A threshold for fault detection is set based on allocation of a certain probability of false alarm. This is to ensure that when there is nothing wrong the occurrence of alarms will be rare enough such that adequate continuity performance is maintained.

Setting the threshold is very important – if it is set too low then solutions may exceed it purely by chance and alarms will be common. If it is set too high then small faults might happen which reduce the accuracy of the solution but do not trigger a fault detection alarm.

2.6.2 Fault Detection and HPL

Missed detection risk can be considered by imagining that a particular type of fault happens. One particular subset of measurements contains an entirely arbitrary measurement error, which incurs an error on the navigation solution but not on the robust subset. If this fault is very large then the navigation solution will be unusable, but such a large error is likely to be easy to detect (the gap between the all-in-view and reduced subset solutions will be large). Likewise, a small error might not be hazardous to the accuracy of the solution, but will be difficult for the fault detection test to spot.

Integrity risk (risk of HMI) is the probability of the position error exceeding the HAL, but the fault not being detected. Therefore, the worst kind of fault lies somewhere in-between large and small errors: too large to be acceptable in the output navigation solution but too small to be easily detectable. The M-RAIM protection level is based upon calculating a precise value for this worst-case size of fault for each reduced subset solution. The integrity risk for any subset solution is never more than the risk of HMI for this corresponding worst-case fault.

A single protection level is defined such that the worst-case risk of the error exceeding this HPL without detection, summed over all of the reduced subsets is equal to exactly the user's integrity requirement. A weighting for this sum is given by the probability of each kind of fault happening (for example it is sensible to assume that single-satellite failures are more common than multiple simultaneous issues affecting many satellites).

Two acts of budget balancing are performed to do this:

1. The fault detection threshold is set by balancing the false alarm risks amongst all of the reduced subsets such that the sum total is equal to the apportioned continuity risk.

2. The HPL is set by balancing the weighted risk of HMI amongst all of the reduced subsets such that the sum total is equal to the apportioned integrity risk.

By balancing these two budgets, the algorithm ensures that the user's continuity and integrity requirements are both met, and the HPL is optimised to be as small as it can be.

This method is critically dependent on several factors:

- The assumed probabilities of each type of fault happening.
- The assumed "fault free" error models used to assess the amount of random noise expected in each of the measurements.
- The method used to apportion false alarm and missed detection risks such that the user's overall continuity and integrity performance requirements are met.

2.6.3 Monitored and Un-Monitored risks

Not all possible faults are included in the subset solutions since the possible number of combinations of large numbers of faulted measurements would be prohibitively high. A large group of faults are designated "un-monitored". These are faults which are thought to occur sufficiently rarely that integrity is not significantly harmed by simply ignoring them.

These faults do not have robust subset solutions defined for them, so it is assumed that they are not detected by the solution separation test. An integrity risk equal to the probability of each of these types of fault occurring is removed "off the top" of the total integrity budget

[1]. The more un-monitored faults there are, the lower the remaining integrity budget left to define a safe protection level.

For the faults that are considered "monitored", the occurrence of any of these will (hopefully!) raise an integrity alarm. A total risk of true alarms has to be taken "off the top" of the continuity budget [1]. The more monitored faults there are, the lower the remaining continuity budget left to allocate to false alarms to set an acceptable detection threshold.

2.6.4 M-RAIM Summary

- The proposed algorithm is a multiple-hypothesis solution separation test akin to aviation Advanced RAIM (ARAIM).
- The test can be made robust to arbitrary faults (or combination of simultaneous faults), a necessary condition for the maritime environment.
- Requires a-priori estimates of *fault-free* errors, and fault probabilities.
- Both integrity and continuity performance requirements are guaranteed by carefully budgeting false alarm and missed detection risks between the different solutions.
- Computationally relatively expensive, especially with large numbers of satellites in solution and when considering simultaneous faults on several measurements at once, although mitigations have been researched [10, 11].

2.7 Consideration of Changes to SBAS for Maritime

The following EGNOS V3 changes are understood by the MarRINav project to be planned from the Version 2 baseline:

- RIMS will track the Galileo satellites, and EGNOS V3 will provide orbit and clock corrections for this constellation in addition to GPS.
- RIMS tracking of L5 signals, and provision of corrections for L5 also.
- Ionosphere data will be retained for legacy L1-only users.

2.7.1 Error Modelling

M-RAIM requires the following:

- A-priori estimates of the pseudo-ranging accuracy which can be expected (fault free).
- Probability that the fault free Gaussian model is exceeded (fault probability).
- Lists of which faults need to be considered (monitored and un-monitored faults), or simple decision-making rules to enumerate these.
- Augmentation in the form of local differential corrections will improve accuracy and can help with the error-bounding process.

Critically, the error models needed by M-RAIM, and those provided by SBAS are not equivalent. SBAS provides the inflated standard deviation parameters of a Gaussian over-bound model conservative such that a VPL can be defined, and is safe down to 10^{-7} risk. Applying these bounds, M-RAIM in conjunction with SBAS would not have to consider satellite-based faults as significant, since their probability of occurring is kept below the maritime user's integrity requirement.

M-RAIM does still have to consider local hazards such as multipath, and SBAS neither provides a fault-free model, or a likely fault probability for this problem. Aviation Minimum Operational Performance Standards (MOPS) for SBAS [7] do include a multipath bounding model, but this is appropriate to a type-approved antenna installation on an aircraft in flight, not a typical GNSS antenna on a ship operating in a potentially complex radio environment.

It is necessary to determine a nominal vessel multi-path model, and the associated probability that instantaneous measurements exceed this model (fault probability). The Stanford University review team noted the following: *{" this as a very important point, and one that is perhaps unsolvable without more information on the environment (and how the environment varies under different operational conditions) as real-world measurements."}*

M-RAIM requires to be equipped with estimates of the fault-free error distribution on each parameter, and the probability of this distribution being exceeded. The M-RAIM process then determines, given these starting points, an estimate for the integrity over-bound appropriate to the user's required integrity-risk. M-RAIM performance would be sub-optimal if the error estimates were excessively inflated, as would be the case if existing SBAS (EGNOS) information were used directly as provided, representing massively inflated (x300%) integrity bounds for aviation integrity (10^{-7}). Hence, M-RAIM must be provided with or be able to derive the fault-free estimates of error distributions.

Applying the SBAS Gaussian over-bounds as estimates of the fault-free error distribution will greatly over-estimate the amount of error expected on each parameter. As such the receiver will believe that the quality of all of its measurements are considerably worse than they really are. This will lead to very lenient detection thresholds, low alarm rates, poor fault detection, worse integrity performance and a very inflated HPL. It is also possible that these "very inflated" HPLs would still be acceptable relative to HAL, at least for some maritime applications.

Ideally M-RAIM should be provided with best-estimate "fault free" error models (and associated fault probabilities) not inflated integrity over bounds.

The Stanford University review team commented as follows:

{"This ideal is not completely practical, and some level of conservatism (inflation for "over-bounding") will still be needed in the M-RAIM fault free error models. It is correct that SBAS overbounds may be too conservative (in the context of M-RAIM and maritime users in general) because they are constructed to mitigate integrity risks specific to SBAS and assessed very conservatively for aviation applications. But even without this, constructing maritime fault-free error models that bound to relatively low probabilities (10^{-3} to 10^{-5}) requires that many imperfect or "off-nominal" conditions be covered, and the combination of these conditions requires at least some inflation of the "fault-free" standard deviation. This is true even if a separate bound on un-modeled bias error is also provided, as suggested later in Section 2.7.2. While providing a separate bias error bound helps remove

conservatism from the standard deviation, some inflation of the standard deviation will still be needed (in most cases)."}

2.7.2 Stanford Proposals for DFMC SBAS

The Stanford research team have previously suggested that SBAS should be modified to include a biased-Gaussian model to over bound errors [19]. This would see the integrity bound broadcast as pairs of mean and standard deviation parameters for each satellite. The team had also suggested that fault-free error models, where known or suspected biases may play a significant factor (for example signal deformations), should also be modelled using separate nominal standard deviation and bias terms.

Here we shall consider the effect of broadcasting error bounds as pairs of bias and standard deviation parameters. The user would then re-create their bound by multiplying the standard deviation by an appropriate k-factor, and adding it to the bias. Therefore, the broadcast standard deviation would receive *no* inflation over the internal SBAS assumed error, and the integrity bound due to the effect of undetectable faults would be broadcast as the bias term. The user's receiver may then choose to ignore this term if it performs its own fault-detection by implementing RAIM.

The Stanford University review team provided the following comments:

1. *{"The broadcast of a tighter sigma to represent the "core" of the distribution is not a bad proposal and is worth further study."}*

2. *{"This description is also akin to the original Galileo safety of life concept (which uses the SISA and SISMA)."}*

3. *{"The 2010 paper (reference [19] of the GLA report) that includes [Stanford's proposal] floated an idea that failed to gain any traction. Much of the resistance was that this would represent a very significant change to the ground monitoring, broadcast information, and integrity analysis. The paper speculated on values for fault-free sigmas and bias bounds that would need to be validated. Under strict integrity analysis, values such as these are almost always increased from before to meet the same documented level of integrity. Thus, the benefit would likely be somewhat reduced. Although this idea might be worth revisiting for maritime depending on the integrity requirements, it is likely to be a very costly approach.*

   If this idea were at least partially implemented, is unlikely that SBAS would broadcast these fault free-sigmas, nominal biases, and faulted biases with any great frequency. Much more likely would be to develop a table that allows their determination from already broadcast UDREIs (or DFREIs)."}

It is noted that the work of RTCM SC-134 working group looking at standards definitions for future high-accuracy integrity services has also proposed the same mean-and-variance method to more concisely communicate integrity bounds to users in a variety of different industry sectors including road and rail.

Although it is proposed here that the capability is investigated for future SBAS integrity bounds to be broadcast as separate mean and standard deviation figures, allowing the

broadcast error to more closely match the expected fault-free error without excessive inflation, it is accepted that it may not be possible to make changes to the SBAS/EGNOS V3 broadcast or operation.

It is recommended that the Stanford idea to develop a table that allows the determination of fault free-sigmas, nominal biases, and faulted biases from already broadcast UDREIs (or DFREIs) is investigated further.

2.7.3 Initial Proposals for DFMC SBAS Maritime Service supporting User-Level Integrity

2.7.3.1 Primary Concerns

The following is offered as an outline description of how a future SBAS maritime service might support a user-level integrity capability for maritime. This is driven by several primary concerns:

1. The receiver must implement some form of RAIM (M-RAIM is recommended) to detect and mitigate faults or failings, which may occur due to the local environment of the vessel. This is entirely because maritime receivers operate in potentially a far less clean radio environment compared to aviation receivers. Local noise, interference, jamming, spoofing, multi-path reflections and possibly signal obscuration may all pose a threat to a marine GNSS receiver. All of these are more easily controlled in an aviation setting and are less likely to affect an aircraft in flight. Note: RAIM will not protect the user from interference and jamming that denies use of GNSS signals and will only protect partially against spoofing (because advanced spoofing might be designed to defeat RAIM monitoring). Mitigation of these threats is discussed in Deliverable Report D6 [1], which includes jamming and spoofing detection in the MSR together with backup systems to provide resilient positioning.

2. If the receiver employs inflated error estimates as inputs to the RAIM algorithm then this may result in an overly inflated fault-detection threshold, poor fault detection, and an excessively large HPL. The availability of the maritime solution relies on keeping error estimates as tight as possible to the true expected distribution of errors and using "fault free" estimates where possible in place of inflated models.

3. The Stanford University review provided the following further clarification:

   {"There are (at least) two factors that make the SBAS error inputs poor choices (if not suitably modified): (1) they are inflated to bound rare-event errors, and (2) the error variances suffer extra, unpredictable inflation due to quantization (to one of 14 levels) within the SBAS message transmission."}

4. The SBAS broadcast integrity bounds are likely to be considerably inflated estimates of the various error components, such as the residual error in the orbit and clock corrections (UDRE) and grid ionosphere vertical error (GIVE). As described by point #2 above, using the integrity bounds designed for the aviation user as input to the RAIM

algorithm may inflate the maritime HPL excessively and harm the availability of solution.

5. The provision of a conservative model to estimate local code noise and multi-path (CNMP) that is universal and usable by any vessel, with any GNSS antenna, and during any stage of the voyage will be a challenge and may result in overly conservative model having to be assumed. The accuracy performance of EGNOS V3 will be very good; the largest component of pseudo-range measurement error assumed by the receiver is likely to be due to the antenna multi-path model. Again, this has the potential to harm solution availability as described by point #2, above.

   The Stanford University review team offered a very helpful comment:

   {"This is especially true in ports or harbors with many nearby reflecting surfaces off-ship, but even shipboard environments not considering outside influences ("own-vessel multipath") tend to be worse than aircraft. If possible, separate antenna multipath models should be established for oceanic operations (and others where no other surfaces outside the ship in question would apply) vs. port/harbor conditions including external reflections. These models will need to be conservative to represent a wide variety of shipboard and external environments within each category. Most importantly, to the degree that the shipboard antenna multipath model dominates the other GNSS error sources, less emphasis can be placed on reducing the conservatism in the models representing these other errors (e.g. UDRE and GIVE).

   Aircraft type acceptance in which the installation is accepted for a specific aircraft type given the antenna location and aircraft environment is crucial to maintaining the airborne error bounds. This section is correct that a single model with small error bounds cannot be universally applied to all potential installations. A user error bounding model will be critically dependent on the quality of the antenna location and its surrounding environment. It is impossible to achieve high levels of availability and continuity if the antenna is poorly installed or can only be located in a poor environment."}

6. When the receiver employs the dual-frequency (L1 + L5) ionosphere-free combination to provide a pseudo-range measurement free of ionosphere bias, this necessarily means inflating its assumptions about local CNMP by some 260%. Given that this model might have to be an excessively conservative one to begin with, the use of dual-frequency combination of pseudo-ranges may be incompatible with the need to keep the HPL small.

2.7.3.2 Initial Proposed Maritime Method

Each of these points above leads to a proposed maritime method for employing SBAS (potentially EGNOS V3), for which the primary design requirement is to minimise the pseudo-range error model that must be assumed by the maritime receiver. It is proposed initially that:

1. The user's receiver tracks only the L5/E5a signals from the GPS/Galileo constellation. It may also measure L1/E1 but these measurements are not relevant to the SoL service.

2. The receiver will not employ the dual-frequency ionosphere-free combination of pseudo-ranges to determine the ionosphere delay, as this excessively inflates local code noise and multi-path (CNMP) errors. Only L5/E5a measurements are to be used. It is believed that the L5 signal design will offer improved multi-path mitigation performance also.

   The recommendation is made here that a trade study should be conducted to consider the use of L1/E1, L5/E5a and combinations, their benefits and drawbacks in the maritime environment.

3. The receiver employs EGNOS broadcast satellite orbit and clock corrections for all of the GPS and Galileo satellites in view.

4. The receiver applies the existing SBAS MOPS model to estimate the troposphere delay.

5. The receiver applies a suitable scale factor (x1.8) to convert the EGNOS broadcast ionosphere delay map and GIVE values to corrections to L5 (from L1). This data is then applied as per the current MOPS to correct the ionosphere delay bias on the L5 / E5 measurements.

6. The error on each pseudo-range is estimated as a root-sum-square (RSS) addition of errors from: Orbit and clock (σ_{UDRE}); ionosphere (σ_{GIVE}); residual troposphere error (σ_{tropo}) and code noise and multipath (σ_{CNMP}). Code noise and multipath error are pre-determined maritime-specific functions of satellite elevation and are hard-coded into the receiver. UDRE and GIVE are broadcast in real time by EGNOS, three options exist:

   - A bespoke SBAS maritime service message broadcasts *fault-free* estimates of UDRE and GIVE, recognising this may not be a cost-effective option.

   - A single bespoke data-field within another SBAS message defines a maritime scale-factor (e.g. x2.5) by which to reduce the aviation UDRE and GIVE bounds to yield effective maritime fault-free estimates of these parameters.

   - A bespoke table to convert broadcast UDRE and GIVE index parameters to appropriate fault-free error estimates, employed exclusively by maritime receivers. This removes the need to modify the SBAS broadcast message.

   The Stanford team commented that:

   {"As mentioned earlier, using a description of the core of the distribution (meaning an uninflated version of the distribution) is a good idea, but it might be difficult and expensive to implement. In any case, existing UDREs might actually be low enough to result in acceptable HPLs. As acknowledged below, it would be more difficult to produce and justify smaller GIVEs [which may support the use of dual-frequency when considered in a trade-off analysis]."}

7. The receiver applies a suitably modified version of the ARAIM algorithm, such as M-RAIM, to detect and exclude autonomously faulty measurements made by the receiver, which

are induced by the local environment (and are hence un-observable to SBAS). This algorithm determines a HPL appropriate to the navigation solution and the maritime navigation performance requirements (continuity and integrity).

The Stanford team provided important comments on points 1 and 2 above, suggesting that these areas need further investigation:

{"[Points 1 and 2] propose the use of L5/E5a without L1/E1 and avoiding dual-frequency ionosphere calculations. If one wants to avoid the multiplication factor on multipath imposed by the iono-free measurement combination, there are other combinations that can be applied (for example, averaging L1 and L5). However, these leave the user vulnerable to uncertainty on the ionospheric corrections and to potentially un-monitored inter-frequency signal biases. Dual-frequency GBAS considered a divergence-free version of L1. PPP uses iono-free measurements but is carrier based, so the multipath is not the dominant contributor. Other frequency combinations are possible. However, none of these other choices are truly supported by SBAS. L5 is potentially a poor choice because the ionospheric delay is 1.8 times larger, but more importantly so is the uncertainty. So it is not just the correction to delay that needs to be multiplied by 1.8, but the GIVE (and corresponding UIRE) also needs to be multiplied by this number.

While it is true that the dual-frequency combination increases the effect of multipath, this increase is at least partially offset by removing the uncertainty on ionospheric delay, which is currently the main factor limiting availability in SBAS. The relative benefit of this step depends on the degree, if any, to which SBAS GIVEs can be reduced for maritime users, as proposed in [bullet 6 above]. Even if single-frequency L5/E5a were the approach taken, the parallel use of L1/E1 measurements should still be considered (if not mandated) as a potential backup (if L5/E5a is unusable)."}

The Stanford team further commented:

{"The SBAS corrections on L1 are for the L1 signal only. The SBAS corrections on L5 are for the L1/L5 iono-free combination only. Any application to other frequency combinations runs the risk that unmonitored inter-frequency bias offsets will affect that pseudorange combination. These inter-signal biases are often small and usually do not change rapidly (unless there is a signal path reconfiguration on the satellite). However, there would need to be a conscious decision to accept such a risk (i.e. it is not one that is monitored or protected against by SBAS, and it may be expensive to add such monitoring and certainly to add a broadcast alerting function for a threat that does not affect currently targeted SBAS users).

The [Stanford] authors strongly suggest… retaining both single and multiple-frequency candidates and making a decision later [as a result of a trade-off analysis], when a better understanding of the bounding multipath errors and GIVE values for maritime users is available."}

EGNOS V3 will then only guarantee a minimum performance for a maritime safety-of-life service to vessels equipped with a suitable type-approved receiver. The relevant IEC receiver specification will need developing, including receiver testing procedures to ensure that the EGNOS data is employed as described above, and that an adequate RAIM algorithm is applied to the data. M-RAIM will be offered freely as a generic default, but any processing which meets the same performance specifications will be allowed.

The antenna type and installation should also be certified to ensure that installation is such that own-vessel multipath and signal-obscuration is controlled. This process should also ensure that measurement errors made by the receiver are adequately described by the mandated code noise and multipath model appropriate to the maritime environment.

There may also have to be allowances made for different phases of the voyage. For example, operation in ports, underneath cranes or bridges, and seaways through city centres may pose a more challenging multi-path environment. It may be appropriate to mandate that non-GNSS navigation services be employed in some locations.

It may not be appropriate to scale-down the GIVE values for a maritime receiver. This is due to the fact that the ionosphere threat-model used to inflate GIVE considers solar storms and excited ionospheric conditions as potentially hazardous. This threat has the potential to induce errors on all satellites in view. The RAIM process proposed for maritime receivers is not designed to mitigate such a threat so may need to employ the SBAS/EGNOS GIVE values un-corrected.

The Stanford review team commented:

{"[Scaling down of GIVE values] is a critical point to consider if dual-frequency estimation of ionospheric delays will not be conducted. SBAS protection against ionospheric threats is very conservative and defends a more demanding integrity risk requirement than what is proposed for maritime users in Section 2.3. However, since maritime users will rely on SBAS to some degree for protection against ionospheric anomalies, the 'scaling down of GIVE' cannot completely remove GIVE inflation provided for this protection. This also raises the issue of maritime users at the boundaries of SBAS coverage, where SBAS monitoring of the ionosphere is limited and some satellites may be 'not monitored' or have very high GIVE values.

In order to scale down GIVE (or really to implement any of these SBAS changes), a new set of integrity requirements need to be specified (e.g. 10^{-5} and interpreted in an average risk sense rather than a specific risk sense). Also, methods to broadcast the required information to the user need to be specified. Then a full analysis of the impact of such changes would need to be made."}

It is clear that further research is needed to establish the most cost-effective way to employ SBAS broadcast data in conjunction with a RAIM process to ensure adequate user-level integrity at sea.

3 Mathematical Description

3.1 Navigation Solution

We start with a mathematical description of GNSS position fixing process. A position-solution is derived by a process of iterated weighted least squares of linearized pseudo-range equations:

Measurements of the time-of-arrival of the GNSS navigation signals provide pseudo-range measurements (time-of-flight range measurements made against a common, inaccurate receiver clock). It is difficult to directly convert pseudo-ranges to a solution and so the process is linearized. The Geometry matrix G defines the amount by which each pseudo-range measurement (y) is expected to change for a small change in position-co-ordinates (x):

$$G = \frac{dy}{dx}$$
$$y_{obs} - y_{AP} = G(x_{new} - x_{AP})$$

Here x_{AP} is an assumed position (AP) about which the linearization is performed. The estimated range to each satellite is calculated from the AP and is given as y_{AP}. The co-ordinates x_{new} represent some location near to the AP and y_{obs} are the corresponding pseudo-range observations made by the GNSS receiver.

For a given set of pseudo-range measurements, the best linear un-biased estimate of an update to the position-co-ordinates is given by the pseudo-inverse of G:

$$(x_{new} - x_{AP}) = (G^T G)^{-1} G^T (y_{obs} - y_{AP})$$

Since not all measurements are equally accurate, ideally a weighting should be applied which is a best estimate of the inverse of the covariance of the measurements:

$$(x_{new} - x_{AP}) = (G^T W G)^{-1} G^T W (y_{obs} - y_{AP})$$
$$W^{-1} = \langle x | x^T \rangle$$

The estimated solution co-ordinates are determined by iterating this equation and updating the assumed-position until changes are small and $x_{new} \approx x_{AP}$. We define the projection matrix (K) which projects a change in pseudo-ranges to the co-ordinate space:

$$K = (G^T W G)^{-1} G^T W$$

Often, we neglect to explicitly define the linearization and iteration process and make the substitutions ($\hat{x} = x_{new} - x_{AP}$) and ($y = y_{obs} - y_{AP}$) which simplifies the writing of the solution:

$$\hat{x} = (G^T W G)^{-1} G^T W y$$
$$\hat{x} = K y$$

If there are at least the same number of pseudo-range measurements as co-ordinates, then a unique solution can be found. If there are more pseudo-range measurements then the solution is over-determined and the *residuals* can be found. These are the difference between the measured pseudo-ranges and those we would expect if the receiver were actually located at the solution-location $\hat{x}$.

$$\hat{y} = G\hat{x}$$

The estimated pseudo-range residuals are given as:

$$\omega = y - \hat{y}$$
$$\omega = y - GKy = (I_n - GK)y$$

We define:

$$A = I_n - GK$$

As the matrix which projects the pseudorange measurements to the residuals:

$$\omega = Ay$$

The co-variance of the resulting navigation solution is given by:

$$C_x = (G^T W G)^{-1}$$

The dilution of precision (DOP) of the solution is given by the “unweighted” co-variance of the solution

$$C_{DOP} = (G^T G)^{-1}$$

3.2 IBPL

Isotropy Based Protection Level (IBPL) involves using the vector of pseudo-range residuals as way to observe the instantaneous position error directly. The residuals vector magnitude is scaled according to HDOP and a pre-computed k-factor chosen to provide a given level of integrity. The method is very easy to compute: the k-factor can be pre-determined and stored for easy look up. The method does not depend on a-priori error models or any assumptions about the kinds of fault which may happen.

IBPL is probably very useful for land mobile and hand held GNSS users who need to be able to determine if their fix is acceptable, but do not have any expectations regarding the availability or continuity of the solution. The algorithm is very computationally efficient demanding little processing power – again an advantage for hand held, miniaturised or power critical applications.

IBPL is based on the actual measurement values per epoch, as such the instantaneous value of the protection level is highly variable. The calculation can be modified to provide a protection level which is not expected to be exceeded more than a given fraction of the time, so offers a “continuity preserving” variant which does not depend on instantaneous measurements but only on solution geometry.

The basic premise on which IBPL is built (error isotropy) is fundamentally not proven. Under the most common fault conditions, the error is expected to be considerably *anisotropic* and so integrity performance of IBPL may be worse than the isotropy assumption suggests. The algorithm contains no screening of weak geometries and so may be vulnerable to HMI under certain conditions. The algorithm is, however, very conservative and is likely to achieve reasonable levels of real-world integrity performance despite its limitations. The conservative nature of this algorithm means that a large number of very precise pseudo-range measurements are needed for the protection level to be reliably below the 25m HAL.

The isotropy based protection level (IBPL) is determined by multiplying the magnitude of the residuals vector by the HDOP of the navigation solution and an appropriate choice of k-factor:

$$IBPL = HDOP.k_I.\sqrt{\omega^T \omega}$$

We shall now describe how to calculate the k-factor, and the logic behind the claim that integrity is preserved whenever IBPL<HAL.

The Stanford University review team has provided the following input, together with Appendix B of mathematical corrections and simplifications to the GLA derivation:

{“In general, this section describes very well the IBPL approach and its shortcomings. Appendix B provides a detailed mathematical discussion of this section in which we:

- *provide some corrections to the derivations;*
- *re-derive the inflation factors (in a similar way but without geometric considerations);*

- *provide examples where IBPL might not provide adequate protection (to reinforce the points made in the report."}*

We need to think of a set of pseudo-range measurements as a vector in n-dimensional space. The origin at zero corresponds to a set of observations which are all exactly perfect (all measurements contain zero error). Each of the base vectors of our n-dimensional "measurement space" corresponds to each one of the n satellites in solution. For example, we can write the error-vector as:

$$y = y_{obs} - y_{true} = [y_1 \quad y_2 \quad y_3 \quad \cdots \quad y_{n-1} \quad y_n]^T$$

The amount of error on any one pseudo-range measurement describes the extent of the vector in that dimension of measurement space. If we consider now the space of all measurements, for which the error vector is a constant magnitude:

$$|y| = R$$

This is described in measurement space as the surface of a sphere, radius R and centred on the origin.

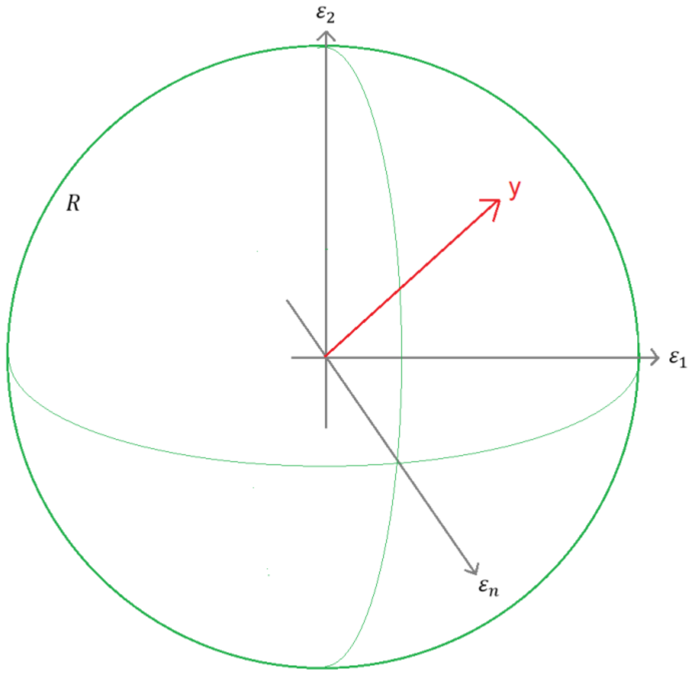

Figure 4 – Error-vector and sphere radius R in measurement-space

Figure 4 depicts an n-dimensional vector space spanned by some orthonormal bases {ε_i} which indicate the direction of errors on the individual satellites in solution. This error vector is divided into two component parts: the position error and the residuals. We have used the term 'projection matrix' for the matrix K that maps the observed pseudo-ranges to position co-ordinates. We can see that K maps the error vector to a position co-ordinate error.

$$\hat{x} = Ky$$

The operation is termed 'projection' as it takes an n-dimensional vector and finds its image on a 4-dimensional space defined by the parameters (lat, lon, ht, t) much like projecting the 2D shadow of a 3D object onto a screen. An image of this error is found in a 4D subspace within the original measurement space. We call this the co-ordinate *image.*

$$\hat{y} = GKy$$

The residuals are given as what is 'left over'.

$$\omega = y - \hat{y} = (I_n - GK)y$$

This is effectively a decomposition of the measurement vector (y) into two orthogonal components. A co-ordinate image ($\hat{y}$) which exists within a 4D subspace of the n-dimensional measurement space, and the residuals (ω) which exist within the orthogonal complement which is an (n-4) dimensional space.

For purposes we shall employ later, we define the angle θ which is the angle between a given error vector (y) and its co-ordinate image ($\hat{y}$). Using the definition of the scalar product:

$$y.\hat{y} = |y||\hat{y}| \cos(\theta)$$

We might also define the angle θ as:

$$\tan(\theta) = \frac{|\omega|}{|\hat{y}|}$$

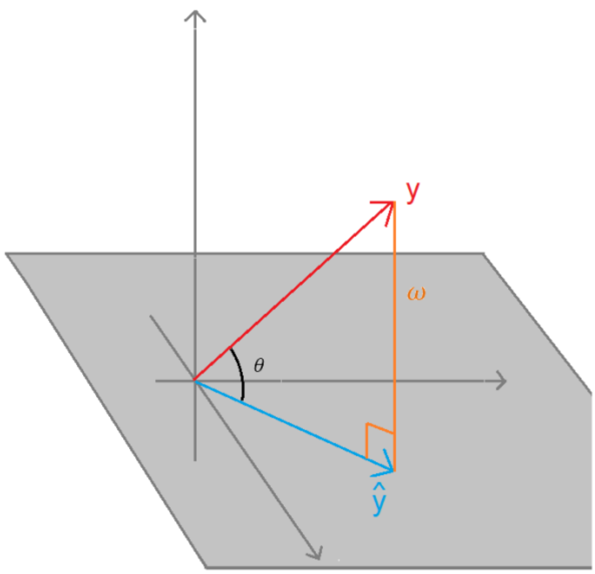

Figure 5 – Simple 3D image of the perpendicular projection of a vector y

Figure 5 depicts the perpendicular projection of a vector y onto its flat (2D) image $\hat{y}$, the remaining vertical component ω and the angle θ. We argue that for a given multiple of the magnitude of the residuals ($k|\omega|$) to be a conservative estimate over the magnitude of the co-ordinate image ($|\hat{y}|$) this requires that the angle θ be at least a given magnitude:

$$k|\omega| \geq |\hat{y}|$$
$$\tan(\theta) = \frac{|\omega|}{|\hat{y}|} \geq \frac{1}{k}$$
$$\theta \geq \theta_{min} = \tan^{-1}\left(\frac{1}{k}\right)$$

Looking at the sphere of radius R in measurement space this condition ($\theta \geq \theta_{min}$) is satisfied for a particular region of its surface area. Assuming that the orientation of the measurement error vector in this space is completely random, the probability of the angle θ exceeding this minimum (θ_{min}) is given by the ratio this region of the sphere to its total surface area. This is the way in which the algorithm makes use of rotational symmetry and isotropy to argue its case. Crucially it can be seen that this calculation only holds true if the orientation is truly random and isotropic.

We define integrity risk as the probability that the condition ($k|\omega| \geq |\hat{y}|$) does not hold true – i.e. co-ordinate error exceeds our estimate based on the residuals. We can see that the calculation of integrity risk only depends on the shape of the sphere and is independent of the radius R, so it holds for all magnitudes of error.

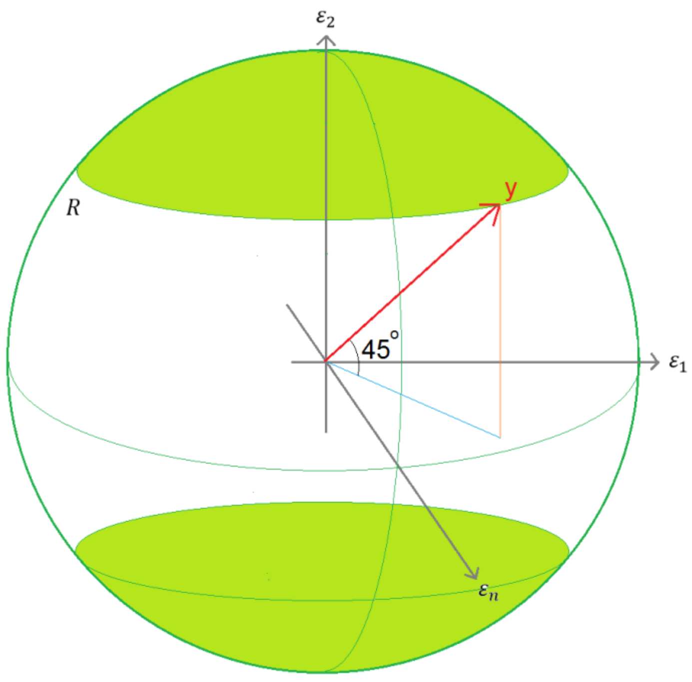

Figure 6 – Simple example of the use of the area of the sphere to determine integrity

Figure 6 depicts a simple example of the use of the area of the sphere to determine integrity. In this example a 3D sphere is shown, $\theta_{min} = 45^o$ and the same horizontal and vertical de-composition of the vector y is shown. The integrity condition ($|\omega| \geq |\hat{y}|$) is satisfied for the highlighted spherical caps (only about 30% of the total area). A larger k-factor means a smaller angle (θ_{min}) and larger spherical caps, so better integrity.

The probability that $k|\omega| \geq |\hat{y}|$ is given by a surface area calculation defined by the angle θ_{min}. We can relate the co-ordinate image to the co-ordinates themselves using the Geometry matrix G:

$$\hat{y} = G\hat{x}$$
$$|\hat{y}|^2 = \hat{y}^T\hat{y} = \hat{x}^T G^T G\hat{x}$$

Because $G^T G$ is a positive-definite symmetric matrix we can make use of its properties (but note the correction to this in Appendix B):

$$|\hat{y}|^2 \geq |\hat{x}|^2 \, Tr(G^T G)$$

This relates the co-ordinates to their image using the dilution of precision (DOP) of the solution. DOP is given by the trace of the matrix $(G^T G)^{-1}$. HDOP is the horizontal DOP and is given by the sum of only the first two terms. We can see:

$$DOP = \sqrt{Tr[(G^T G)^{-1}]}$$
$$HDOP = \sqrt{[(G^T G)^{-1}]^{(1,1)} + [(G^T G)^{-1}]^{(2,2)}}$$
$$HDOP \leq DOP = \frac{1}{Tr[(G^T G)]}$$

We derive:

$$|\hat{y}|^2 \geq \frac{|\hat{x}|^2}{HDOP^2}$$
$$k|\omega| \geq |\hat{y}| \geq \frac{|\hat{x}|}{HDOP}$$

This yields the IBPL definition:

$$IBPL = k \; HDOP \; |\omega|$$

Integrity requires that IBPL is a conservative estimate of the actual co-ordinate errors:

$$IBPL \geq |\hat{x}|$$

This holds for a given percentage of measurement errors defined by the angle θ_{min}:

$$\theta_{min} = \tan^{-1}\left(\frac{1}{k}\right)$$

If we require a particular level of integrity risk, we can determine this by performing the area integral on the sphere as a ratio of the total surface. By setting this ratio equal to our particular integrity risk requirement the necessary value of θ can be found. This then yields the required k-factor for the IBPL calculation.

3.2.1 Spherical Integral

The level of integrity risk associated with a given k-factor is equal to the area of the surface of the n-sphere within an angle θ of the 4D co-ordinate image ($\hat{y}$) as a ratio of the total surface area of the sphere. The integrity problem has now been reduced to one of spherical geometry.

Note here the slightly different approach to spherical integrals contained in the Stanford University review – see Appendix B.

To find the answer we must address some other questions first:

1. What shape is the intersection between the n-sphere and the 4D co-ordinate image space?
2. What shape is the area on the sphere's surface within an angle θ of this intersection?
3. How do we calculate its size?
4. Is there a general form for the ratio of this area to the total surface area of the n-sphere?

To answer the first question, we can think of the 4D subspace in $\mathbb{R}^n$ as a set of constraints – it is the set of n-dimensional points (for a given set of bases), such that n-4 of these must be zero. Considering the intersection of this 4D space with an n-dimensional sphere, we see that because of the rotational symmetry of the sphere we are free to pick our base-vectors. It is always possible to define an orthonormal set of bases such that n-4 co-ordinates are zero at the intersection of the sphere with the 4D subspace.

The intersection is a sphere constrained such that only four of its co-ordinates are non-zero. This is a 4D sphere.

To answer the question about the shape of the region within an angle θ of this intersection, we can look at simpler examples first. We imagine the intersection of a 3D sphere with a 2D plane as the sphere's equator. The regions on the surface within an angle θ of the equator are given as a 'belt' around the middle of the sphere, as shown in Figure 7:

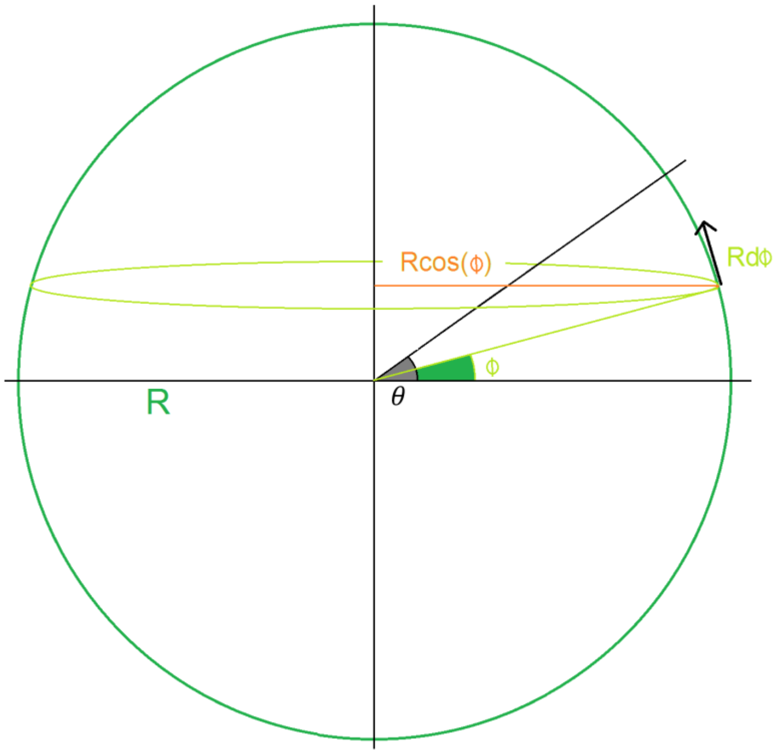

Figure 7 – Method to integrate the area of a belt around the equator of a sphere

We can look at the answer to the third question by considering finding the area of this belt by a surface integral. We sum up the circumference of a set of circles radius $R\cos\phi$ with the angle ϕ running from zero to θ. Actually the area is twice this as we have to consider the area south of the equator also. The integration variable is the arc distance $Rd\phi$ as shown in Figure 7.

$$A_{belt}(R,\theta) = 2\int_0^\theta A_2(R\cos\phi)Rd\phi = 2\int_0^\theta 2\pi R\cos\phi\, Rd\phi$$
$$A_{belt}(R,\theta) = 4\pi R^2[\sin\phi]_0^\theta = 4\pi R^2\sin\theta$$

We have used the notation $A_2(R)$ for the surface of the 2-sphere (circumference of a circle), the general result for the area of an n-sphere can be derived, but it is given as a standard result:

$$A_n(R) = \frac{2\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}R^{n-1}$$

In this example, the "integrity risk" is the ratio of this area to the surface of the 3-sphere:

$$I_3(\theta) = \frac{A_{belt}^3(R,\theta)}{A_3(R)} = \frac{4\pi R^2\sin\theta}{4\pi R^2} = \sin\theta$$

Looking back to Figure 6, for $\theta = 45^o$ as shown on the diagram, this is about 70%.

In fact, Figure 7 is a good conceptual model of how we will perform the integral in 5 dimensions. We don't show every dimension, but the green circle suffices to show the outline of a 5D sphere. The horizontal line through the origin intersecting this sphere is the 4D co-ordinate image space. One dimension remains (the vertical) along which we integrate. Using the surface area-of the 4-sphere:

$$A_4(R) = 2\pi^2 R^3$$

The area of the 5D belt is given by exactly the same integral:

$$A^5_{Belt}(R,\theta) = 2\int_0^\theta A_4(R\cos\phi)Rd\phi = 2\int_0^\theta 2\pi^2(R\cos\phi)^3 Rd\phi$$

Which we can perform with some substitution:

$$A^5_{Belt}(R,\theta) = 4\pi^2R^4\int_0^\theta \cos^3\phi\, d\phi = 4\pi^2R^4\int_0^\theta (\cos 3\phi + 3\sin^2\phi\cos\phi)\, d\phi$$

$$A^5_{Belt}(R,\theta) = 4\pi^2R^4\left[\frac{1}{3}\sin 3\phi + sin^3\phi\right]_0^\theta = 4\pi^2R^4\left[\frac{1}{3}(3\sin\phi\,(1-\sin^2\phi) - \sin^3\phi) + sin^3\phi\right]_0^\theta$$

$$A^5_{Belt}(R,\theta) = 4\pi^2R^4\left(sin\theta - \frac{1}{3}\sin^3\theta\right)$$

We can determine the integrity risk by dividing by the total surface area of the 5-sphere:

$$I_5(\theta) = \frac{A^5_{Belt}(R,\theta)}{A_5(R)}$$

$$I_5(\theta) = \frac{1}{2}[3\sin\theta - \sin^3\theta]$$

By setting $\theta = \tan^{-1}\left(\frac{1}{k}\right)$ we can also now determine the k-factor for given level of integrity (but only with 5 satellites in solution).

Integrity	k-factor
10^{-1}	14.944
10^{-2}	149.994
10^{-3}	$1.5\text{x}10^{3}$
10^{-4}	$1.5\text{x}10^{4}$
10^{-5}	$1.5\text{x}10^{5}$
10^{-6}	$1.5\text{x}10^{6}$
10^{-7}	$1.5\text{x}10^{7}$
10^{-8}	$1.5\text{x}10^{8}$

Table 3 – Derivation of k-factor for n=5

The values of k-factor in Table 3 agree with the results given in Table 1, in section 2.4.3.

We now need to extend this approach to consider more satellites in solution. As we add more satellites and increase the number of dimensions, the intersection is still a 4-sphere, but there are more dimensions through which the integration extends, we now need to consider it as a solid angle through several dimensions (Ω). The arc distance for the integration is also over the solid angle ($Rd\Omega$).

This is harder to conceptualise, but if we show this as a region on a regular sphere and reduce the 4-sphere to just a line showing its radius we get an idea, as shown in Figure 8:

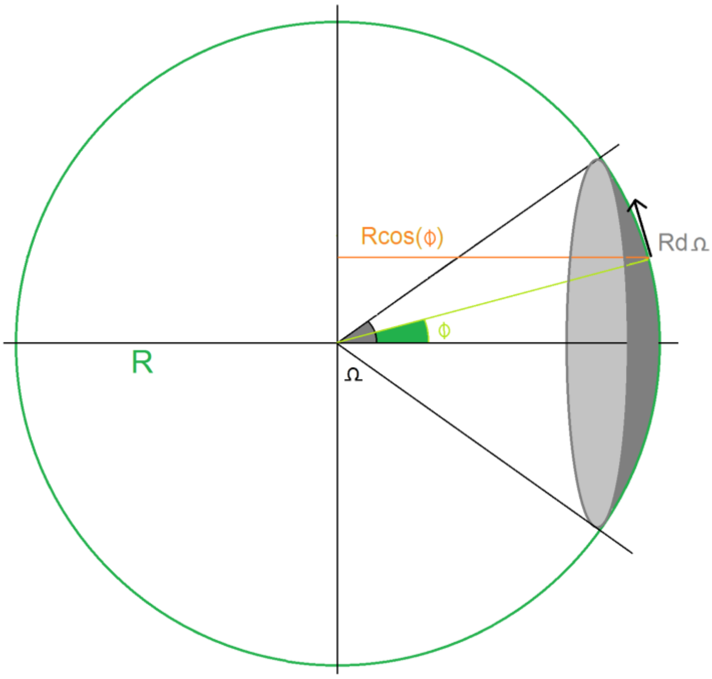

Figure 8 – Integration over a 2D solid-angle Ω to evaluate 6-satellite integrity.

In Figure 8, the green circle represents the surface of a sphere in 6-dimensions. The orange line extends horizontally from the vertical axis until it touches the grey cap and represents the radius of a smaller 4-dimensional sphere (these dimensions are not shown!). The grey cap is the region over which we have to integrate the area of the 4-spheres of radius $Rcos(\phi)$.

The cap of the solid angle can be considered as a collection of concentric circles. We introduce another integration variable: the angle ψ which goes from 0 to 2π round the centre of the cap. This defines a circle of radius $R\sin\phi$, as shown in Figure 9. We integrate over the whole solid angle by integrating over both ψ and ϕ:

$$Rd\Omega = \mathrm{R}\sin\phi\, d\psi\, R\, d\phi$$

In Figure 9, the diagram shows the inside surface of the cap as seen from a view-point within the sphere near to the origin (O). The new angle ψ describes a rotation around the centre of the cap, the integration variable is $\mathrm{R}\sin\phi\, d\psi$. The angle ϕ goes from zero to θ, with integration variable $\mathrm{R}d\phi$.

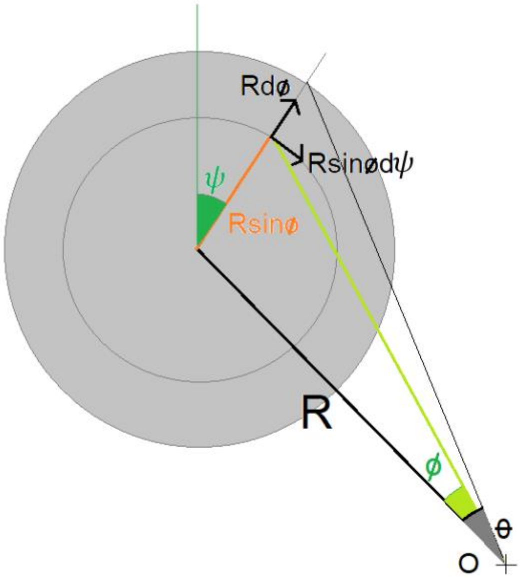

Figure 9 – Integration around the cap of the solid angle Ω.

Substitution into the belt integral gives:

$$A_{belt}^{6}(R,\theta) = \int_0^{\theta}\int_0^{2\pi} A_4(R\cos\phi)\, R\sin\phi\, d\psi\, Rd\phi$$
$$A_{belt}^{6}(R,\theta) = 2\pi.2\pi^2 R^5 \int_0^{\theta} \cos^3\phi \sin\phi\, d\phi = 4\pi^3 R^5 \left[-\frac{1}{4}\cos^4\phi\right]_0^{\theta}$$
$$A_{belt}^{6}(R,\theta) = \pi^3 R^5 (1-\cos^4\theta) = \pi^3 R^5 (2\sin^2\theta - \sin^4\theta)$$

This yields the integrity-risk calculation for 6 satellites:

$$I_6(\theta) = \frac{A_{Belt}^{6}(R,\theta)}{A_6(R)}$$
$$I_6(\theta) = \frac{\pi^3 R^5 (2\sin^2\theta - \sin^4\theta)}{\pi^3 R^5} = 2\sin^2\theta - \sin^4\theta$$

This process is endlessly extensible to more dimensions, each time only requiring the integration of 4-spheres round a solid angle. The number of dimensions of this solid angle is n-4 and so each time we add another satellite to solution we get another integration-variable to go around this solid angle:

$$A_{belt}^{7}(R,\theta) = \int_0^{\theta}\int_0^{\pi}\int_0^{2\pi} A_4(R\cos\phi)\, R\sin\phi\sin\varphi\, d\psi\, R\sin\phi\, d\varphi\, Rd\phi$$
$$A_{belt}^{8}(R,\theta) = \int_0^{\theta}\int_0^{\pi}\int_0^{\pi}\int_0^{2\pi} A_4(R\cos\phi)\, R\sin\phi\sin\varphi\ \sin\psi\, d\gamma\, R\sin\phi\sin\varphi\, d\psi\, R\sin\phi\, d\varphi\, Rd\phi$$

Etc.…

A similar approach is described in [18] which first flattens the spherical cap through perpendicular projection onto the co-ordinate image space. This introduces an additional term into the integral, but yields an identical result.

After some work, we arrive at a generic equation for the integral of a 4-dimensional belt of thickness θ around the equator of an n-dimensional sphere:

$$A_{belt}^{n}(R,\theta) = 4\pi^2 R^{n-1} \frac{\pi^{\frac{n-4}{2}}}{\Gamma\left(\frac{n-4}{2}\right)} \left(\frac{1}{n-4}\sin^{\mathrm{n-4}}\theta - \frac{1}{n-2}\sin^{\mathrm{n-2}}\theta\right)$$

The total area of the n-sphere is given above as:

$$A_n(R) = \frac{2\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} R^{n-1}$$

We define the integrity as the ratio of the area of the belt to the area of the sphere:

$$I_n(\theta) = \frac{A_{belt}^{n}(R,\theta)}{A_n(R)}$$

$$I_n(\theta) = 4\pi^2 R^{n-1} \frac{\pi^{\frac{n-4}{2}}}{\Gamma\left(\frac{n-4}{2}\right)} \left(\frac{1}{n-4}\sin^{\mathrm{n-4}}\theta - \frac{1}{n-2}\sin^{\mathrm{n-2}}\theta\right) \left[\frac{2\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} R^{n-1}\right]^{-1}$$

With some cancellation of terms, this becomes…

$$\begin{aligned} I_n(\theta) &= 2\frac{R^{n-1}}{R^{n-1}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}-2\right)} \frac{\pi^{\frac{n}{2}}}{\pi^{\frac{n}{2}}} \left(\frac{1}{n-4}\sin^{\mathrm{n-4}}\theta - \frac{1}{n-2}\sin^{\mathrm{n-2}}\theta\right) \\ &= 2\left(\frac{n}{2}-1\right)\left(\frac{n}{2}-2\right)\left(\frac{1}{n-4}\sin^{\mathrm{n-4}}\theta - \frac{1}{n-2}\sin^{\mathrm{n-2}}\theta\right) \end{aligned}$$

$$\boldsymbol{I_n(\theta) = \frac{1}{2}\left[(n-2)\sin^{n-4}\theta - (n-4)\sin^{n-2}\theta\right]}$$

Further simplification may be possible, but this form of the equation has a nice aesthetic and the symmetry of the n-4 and n-2 terms is pleasing.

Helpfully this is a nice and simple equation that can be evaluated for any value of n without much difficulty. It is then a simple process of determining the angle θ that provides the necessary integrity-risk for a given number of satellites, and converting this to a k-value.

$$k_I = \frac{1}{\tan\theta}$$

The k-values are enumerated in Table 4, with the number of satellites down the left-hand column and integrity-risk across the top row. This method evaluates exactly the k-factor calculation used for IBPL. Sensible, independent and unbiased evaluation of the performance of IBPL is not possible without the ability to determine the k-factor.

N / Risk	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}
5	1499.999	$1.5x10^4$	$1.5x10^5$	$1.5x10^6$	$1.5x10^7$	$1.5x10^8$
6	44.705	141.416	447.212	1414.213	4472.136	$1.4142x10^4$
7	13.520	29.216	62.985	135.716	292.399	629.959
8	7.310	13.110	23.375	41.602	73.999	131.602
9	4.987	8.026	12.797	20.330	32.251	51.134
10	3.823	5.740	8.511	12.549	18.459	27.121
11	3.135	4.486	6.325	8.853	12.348	17.190
12	2.683	3.708	5.039	6.789	9.104	12.179
13	2.363	3.181	4.204	5.503	7.163	9.294
14	2.124	2.802	3.625	4.638	5.897	7.470
15	1.939	2.517	3.200	4.021	5.018	6.235
16	1.792	2.294	2.876	3.561	4.376	5.353

Table 4 – k-factor calculations according to the above method.

3.2.2 Continuity Calculation

The Stanford University review team commented here:

{"The continuity preserving version of IBPL proposed here could solve one of the issues with it. It does assume that the nominal noise is the same for all satellites, which is not usually the case. Because of the mismatch between the (actual) nominal noise error and the isotropy assumption, the actual statistics of the rms of ω is a sum of weighted chi-squares, which has a more complex (and non-standard) distribution. It would be of course possible to assume that the errors are isotropic after normalization with the standard deviations (which highlights how arbitrary the isotropy assumption is)."}

{"If it is assumed that all pseudorange measurements have the same error sigma under nominal ('fault-free') conditions, the isotropy assumption is not as much of a stretch, although (as noted in other comments) nominal errors on different satellites (at varying elevation angles) typically do not have the same standard deviation. As noted elsewhere, the isotropy assumption is very doubtful under faulted conditions."}

We can consider the continuity performance of IBPL by relating it to a chi-squared variable. The weighted sum square of residuals is often used as a fault detection measure. If the measurement errors are well behaved this is distributed as a chi-squared variable with n-4 degrees of freedom.

$$t^2 = \omega^T W \omega \sim \chi^2(n-4)$$

If we assume that the isotropy condition holds (even though we know it does not!) and the error on each pseudo-range measurement is approximately the same, described by some nominal standard deviation (σ). We can re-write the weighting matrix and the chi-squared statistic:

$$W = \frac{1}{\sigma^2} I_n$$
$$t^2 = \frac{1}{\sigma^2} \omega^T \omega$$

The IBPL parameter can be related to chi-squared by squaring the value:

$$IBPL^2 = (k_I . HDOP)^2 \omega^T \omega$$

We can now relate a multiple of the square of the IBPL to chi-squared statistic:

$$\left(\frac{IBPL}{k_I . HDOP . \sigma}\right)^2 = \frac{1}{\sigma^2} \omega^T \omega = t^2 \sim \chi^2(n-4)$$

If we have a reasonable estimate of σ, the (fault free) measurement errors, we can calculate the probability of IBPL exceeding some level by considering the chi-squared distribution:

$$p(IBPL^2 > x^2) = p\left(\chi^2 > \left(\frac{x}{k_I . HDOP . \sigma}\right)^2\right)$$

By setting (x) equal to the HAL, it is possible to calculate the continuity risk of IBPL, and hence the level of continuity which will be achieved over 15-minutes. We define the chi-squared "alarm limit" (χ^2_{alarm}) as the value which corresponds to IBPL exceeding the HAL:

$$C_{risk} = p(IBPL^2 > HAL^2) = p\left(\chi^2 > \left(\frac{HAL}{k_I . HDOP . \sigma}\right)^2\right) = p(\chi^2 > \chi^2_{alarm})$$

The probability of chi-squared exceeding this value (the continuity risk C_{risk}) can be determined. A simple example follows for nominal geometries with a different numbers of satellites. We assume the geometry is good (HDOP = 0.75) and pseudo-range accuracy (σ) is 1m.

Stanford University review team commented here:

{"This [assumption of HDOP and σ] is adequate for the purpose of the example that follows, but the limitations of these assumptions should be kept in mind and stated here or after the results that follow. First, this low a value of HDOP is realistic for 12 or more satellites in view (where continuity is or is nearly met), but not so much for fewer than 10 satellites (where continuity is not met in any case). Second, the 1-meter pseudorange accuracy assumption is conservative for satellites corrected by SBAS except for those at low elevation angles. In future work, we suggest using specific example geometries with few and many satellites along with non-identical error sigma values (generated by

standard SBAS error models) to give a clearer picture of which geometries will and will not meet the continuity test proposed here."}

Continuity risk over 15-minutes can then be found as the probability that six successive epochs (of 150 seconds correlation time) are all free of alarms:

$$C_{15\,min} = (1 - C_{risk})^6$$

Stanford University review team commented here:

{"The equation [for continuity risk over 15 minutes] that translates continuity risk per 150-second period into 15-minute continuity risk shows the extreme dependence on the assumption of 150 seconds representing the time between independent samples. If, for example, the time between independent samples were 300 seconds under certain conditions (where a ship is moving more slowly and multipath error is changing more slowly as a result), the right hand side of this equation would be raised to the power of 3 instead of 6, likely giving very different results."}

{"Furthermore, this equation underestimates the effect of temporal decorrelation in a more fundamental way. This criterion to translate the integrity risk over time to per-exposure integrity risk has been used extensively in the past, but it has been shown in recent work to underestimate the effect of temporal exposure. See the following papers for more details [20, 21]."}

Number of satellites	Degrees of Freedom	IBPL k-factor (k_I)	Chi-squared AL (χ^2_{alarm})	Continuity risk (C_{risk})	Performance over 15-mins
5	1	9×10^5	1.4×10^{-9}	>0.9999	<0.01%
6	2	1.10×10^3	9.26×10^{-4}	0.9995	<0.01%
7	3	114.5	0.08	0.994	<0.01%
8	4	36.61	0.83	0.935	<0.01%
9	5	18.35	3.30	0.654	0.17%
10	6	11.52	8.38	0.212	23.98%
11	7	8.22	16.44	0.021	87.86%
12	8	6.36	27.50	5.81×10^{-4}	99.65%
13	9	5.19	41.30	4.42×10^{-6}	**99.99%**
14	10	4.39	57.55	1.05×10^{-8}	**>99.99%**
15	11	3.83	75.94	8.9×10^{-12}	**>99.99%**
16	12	3.40	96.17	3.1×10^{-15}	**>99.99%**

Table 5 – k-factor calculations

Table 5 shows the k-factor calculations for 1.667x10^{-6} integrity risk and chi-squared limits corresponding to integrity alarms (χ^2_{alarm}) and corresponding continuity risk performance.

Alternatively, if we have a fixed continuity requirement, we can determine the chi-squared value only exceeded for this percentage of the time:

$$C_{req} = 5 \times 10^{-5} = p(\chi^2 > \chi_0^2)$$

This can be determined in advance and tabulated for easy look-up. Below we calculate the chi-squared limits that preserve adequate continuity performance. A value for the IBPL that is expected to be exceeded only a small amount of the time can be calculated by working backwards from this value:

$$\left(\frac{IBPL_C}{k_I.HDOP.\sigma}\right)^2 = \chi_0^2$$
$$IBPL_C = k_I.HDOP.\sigma.\chi_0$$

The subscript c is used to indicate that this is a continuity-preserving bound, and hence we term this *Continuity-Preserving IBPL*.

Stanford University review team commented here:

{"The entire continuity risk is assigned to the probability that the chi-square test statistic exceeds a limiting value. This makes sense if the isotropy assumption is assumed to hold over all conditions, but as noted above, it is more realistic to assume that the chi-square test statistic holds under nominal conditions only. In that case, the chi-square test statistic relationship only covers nominal conditions, and a separate allocation of continuity risk must be made for faulted conditions, when it is conventionally (and conservatively) assumed that continuity is always lost (as is done for M-RAIM under 'Risk Budgeting' in Section 3.3). For example, if the total probability of faults were 10^{-5} per epoch (snapshot test), the remaining continuity risk that would be assigned to the chi-square test statistic would be 4 × 10^{-5} per epoch."}

Table ***6*** shows k-factor calculations for 1.667x10^{-6} integrity risk, chi-squared limits corresponding to a continuity risk of 5x10^{-5} (χ_0^2), and the continuity preserving bound.

The two tabulated calculations in Table 5 and
Table ***6*** above can be seen to be equivalent. With only 12 satellites in-view, the achievable continuity performance (at 25m HAL) will be about 99.65%. This only just misses the requirement 99.97%. With 13 satellites in view (assuming the same levels of HDOP and accuracy), the requirement is met. Conversely, the continuity-preserving bound (the bound only exceeded 0.03% in 15-minutes) for the 12-satellite solution just exceeds the 25m HAL, at 27.6m. For 13 satellites, this bound would be 23.2m, which lies just within the 25m HAL and shows that continuity is preserved for with 13 satellites.

Either calculation (continuity of IBPL or calculation of a continuity-preserving bound) can be seen to be equivalent and to show exactly the same information. They both indicate whether or not a given level of continuity performance can be achieved for a given position epoch.

Number of satellites	Degrees of Freedom	IBPL k-factor (k_I)	Chi-squared Limit (χ_0^2)	Continuity bound ($IBPL_C$)
5	1	$9\text{x}10^5$	16.4	2.7Mm
6	2	$1.10\text{x}10^3$	19.8	3.6km
7	3	114.5	22.5	407m
8	4	36.61	25.0	137m
9	5	18.35	27.3	71.9m
10	6	11.52	29.4	46.9m
11	7	8.22	31.5	34.6m
12	8	6.36	33.5	27.6m
13	9	5.19	35.4	**23.2m**
14	10	4.39	37.3	**20.1m**
15	11	3.83	39.1	**17.9m**
16	12	3.40	40.9	**16.3m**

Table 6 – k-factor calculations and the continuity preserving bound

A critique of the IBPL algorithm is not that it fails to meet the required levels of continuity *per se*, but that it fails to inform the navigator of whether continuity is met. Additional calculations are required to inform the user whether or not continuity is met for each specific epoch in time. Without this information, it is not possible for the user to assess whether frequent continuity alarms are expected in the next few minutes. Therefore, they cannot make an informed decision about whether to embark on a manoeuvre for which good continuity is required.

3.2.3 Geometry Screening

We can consider the effect on the navigation solution of an arbitrary bias on one satellite. The classical RAIM process [15] defines the slope of the satellite as the ratio of increase in position-error to increase in the residuals statistic for such a bias:

$$S_{v,i} = \frac{|h_i|}{\lambda_i}$$

Where $|h_i|$ is the vertical error induced by a bias (μ) on the i^{th} satellite and is derived from the third row and k^{th} column of the projection matrix:

$$|h_i| = |K_{(3,i)}\mu|$$

And the non-centrality parameter λ_i is given by the *observability* matrix (M) which determines the effect the bias (μ) has on the residuals:

$$\lambda_i = \mu\sqrt{M_{(i,i)}}$$
$$M = (I_n - GK)^T W(I_n - GK)$$

We can consider the effect a general vector of biases has on the residuals:

$$\Delta\omega = \mu(I_n - GK)$$

Again assuming the isotropy condition holds for the un-faulted measurements:

$$W = \frac{1}{\sigma^2} I_n$$
$$\sigma^2 M = (I_n - GK)^T (I_n - GK)$$

We can consider the effect the bias (μ) on the i^{th} satellite has on the IBPL:

$$\Delta IBPL = k_I.HDOP.\sqrt{\Delta\omega^T\Delta\omega} = k_I.HDOP\sqrt{\mu^2(I_n - GK)^T(I_n - GK)}$$
$$\Delta IBPL = k_I.HDOP.\sigma.\mu\sqrt{M_{(i,i)}}$$

Substituting the non-centrality parameter:

$$\Delta IBPL = k_I.HDOP.\sigma.\lambda_i$$

The integrity condition demands that IBPL be larger than the co-ordinate error:

$$IBPL \geq |\hat{x}|$$

The computation of the k-factor ensures this is true (up to a given level of risk) when errors are isotropic. To ensure this remains true when a bias is incurred, the change in IBPL must exceed the change in co-ordinate error:

$$\Delta IBPL > |\Delta\hat{x}|$$

This co-ordinate error is given by the root sum square of the i^{th} column of the projection matrix:

$$|\Delta\hat{x}| = \mu \sqrt{\sum_{j=1}^{4} \left(K_{(j,i)}\right)^2}$$

If we slightly re-define the slope of the satellite to be:

$$S_{x,i} = \frac{|\Delta\hat{x}|}{\lambda_i}$$

The integrity condition holds if:

$$\Delta IBPL = k_I.HDOP.\sigma.\lambda_i > S_{x,i}\lambda_i$$

Or if:

$$k_I.HDOP.\sigma > S_{x,i}$$

It is a simple enough task to calculate the slope of each satellite, and consider whether the factor ($k_I.HDOP.\sigma$) exceeds this. If it does not, then this means that IBPL is potentially vulnerable to a fault on that satellite.

Two options are available:

1. Reject solutions which have a critical satellite for which the slope exceeds ($k_I.HDOP.\sigma$)
2. Inflate the k-factor above the isotropic value such that:

$$k_{new} = \max\left\{k_I\,,\frac{S_{x,i}}{HDOP.\sigma}\right\}$$

The second option is probably better since it does not necessarily result in an integrity alarm being sounded.

The anisotropy-tolerant, continuity-preserving IBPL can then be defined as:

$$IBPL_{C,I} = \chi_0 \max\left\{k_I HDOP.\sigma\,,\, S_{x,k}\right\}$$

Stanford review team commented here:

{"Regarding $IBPL_{C,I}$ equation, it should be pointed out that this approach "tolerates" anisotropy to a limited extent but not fully, as the text below illustrates (i.e. multiple simultaneous faults on two or more satellites are not yet considered, and doing so would be cumbersome). Thus, the assumption of significant isotropy is still present. How much remaining "isotropy" is too much, especially when the remaining reliance on isotropy is hard to quantify?"}

We have not considered the possibility that a simultaneous fault on two (or more) satellites potentially violates the IBPL integrity condition. Including calculations for the multi-fault slope values (as described above) is potentially quite a computationally intensive process, and removes the main selling point of IBPL (its simplicity) as an integrity algorithm. Investigation into this aspect is left as a potential direction for future research.

The Stanford University review team commented:

{"The modification of IBPL described [here] is essentially a re-derivation of chi-square RAIM assuming that there is no nominal noise (e.g. the formulation of $S_{x,k}$ *does not include noise, and the error bound ends up being the slope times the root square of the chi-square statistic). Because noise is neglected in the derivation of the added term, this approach is therefore not at the same standard of integrity as ARAIM or its variant M-RAIM. Still, the idea of using both IBPL and M-RAIM might be worth exploring (e.g. taking the maximum of IBPL and M-RAIM protection levels assuming single-measurement faults only in M-RAIM)."}*

However, given the fundamental false premise of isotropy in IBPL and supported by the overwhelming strength of Stanford review comments regarding IBPL for maritime, it is concluded that IBPL (with or without a continuity preserving bound or with anisotrophic tolerance) is totally unsuitable for general maritime navigation and will not be considered further.

3.3 M-RAIM

An integrity algorithm based on multiple-hypothesis solution-separation (MHSS) is proposed. An all-in-view position solution is derived, as are several reduced subset solutions each excluding a unique set of pseudo-range measurements. The distance between the all in view solution and each subset solution is used as a fault detection measure. When the detection measure exceeds the threshold for any subset, an alarm is raised indicating a fault has occurred. Either the user can be warned by means of an integrity alert, or fault exclusion can be attempted.

Each type of fault is expected to occur with a nominal *a-priori* fault probability and each of the corresponding solution separation tests will have some allocation of false alarm and missed detection risk. False alarms harm continuity, missed detections harm integrity, so both are budgeted within the receiver such that the user's operational integrity and continuity requirements are met.

Individual detection thresholds are assigned to each subset solution dependent on the budgeted false alarm risk. In addition, each subset solution will have a given horizontal protection level (HPL) for which the risk of the fault causing the error to exceed the HPL without detection is given by the missed detection risk allowance for that subset. The HPL is defined as the smallest region over which this risk is met, assuming that each of the faulted measurements can incur arbitrary and unknown levels of position error (bias).

If no alarms are raised, the all-in-view solution is passed to the navigator. The HPL associated with this solution is the maximum of all the subset HPLs calculated. HPL can be minimised by optimising the allocation of false alarm and missed detection risk budgets between the various subsets such that all detection thresholds and all HPLs are equal in size. If the all in view HPL exceeds the horizontal alert limit (HAL) for the current operation then the integrity alert must be raised and the solution cannot be used.

This process is mechanically very similar to ARAIM [16]. Indeed just like the aviation equivalent, the process is critically dependent on a number of a-priori assumptions such as the probability of each type of fault occurring, and the assumed "fault free" error models used to estimate pseudo-ranging accuracy at the receiver. M-RAIM is intended to be compatible with existing and future augmentation systems such as DGPS and SBAS (and not as a replacement to them) and so differs from ARAIM in this regard. The HPL calculation is also derived by a slightly different method.

The biggest limitations of this process are the dependency on a-priori assumptions (both nominal fault-free error models and fault-probability values). These assumptions must be conservative for integrity to be guaranteed, and the magnitude of the HPL is critically dependent on them. If these assumptions are too conservative then the HPL may be inflated too far and the availability of the method may be damaged. If a large number of potential simultaneous faults are considered then the number of combinations of different kinds of faults can be extremely high resulting in significant computational cost in the receiver.

3.3.1 Rayleigh and the Semi-Major Axis

If we look closely at the defined co-variance of the navigation solution, we can describe the properties of the distribution of position-errors. We use the subscript zero to differentiate the all in view solution from any others we will define later.

$$C_0 = (G^T W G)^{-1} = \begin{bmatrix} \sigma_{x1}^2 & \sigma_{x1x2} & & \cdots \\ \sigma_{x1x2} & \sigma_{x2}^2 & & \\ & \vdots & \sigma_{x3}^2 & \ddots \\ & & \ddots & \sigma_{x4}^2 \end{bmatrix}$$

The horizontal position-error is described by the four elements in the top-left 2x2 cell of the matrix. This describes a 2D Normal distribution with unequal variances ($\sigma_{x1}^2, \sigma_{x2}^2$), and correlation ($\sigma_{x1x2}$). The term "error ellipse" is often used since the lines of constant probability density (*isoprobability*) are similar concentric curves, elliptical in shape. The standard deviations of the distribution along the semi-major and semi-minor axes (a, b) of the ellipse are given such that the distribution is un-correlated along these directions:

$$\sigma_{a,0} = \sqrt{\frac{1}{2}\left(\sigma_{x1}^2 + \sigma_{x2}^2 + \sqrt{(\sigma_{x1}^2 - \sigma_{x2}^2)^2 + 4\sigma_{x1x2}^2}\right)}$$

$$\sigma_{b,0} = \sqrt{\frac{1}{2}\left(\sigma_{x1}^2 + \sigma_{x2}^2 - \sqrt{(\sigma_{x1}^2 - \sigma_{x2}^2)^2 + 4\sigma_{x1x2}^2}\right)}$$

Again, the subscript zero differentiates the errors on the all in view solution from any other solutions we may derive later.

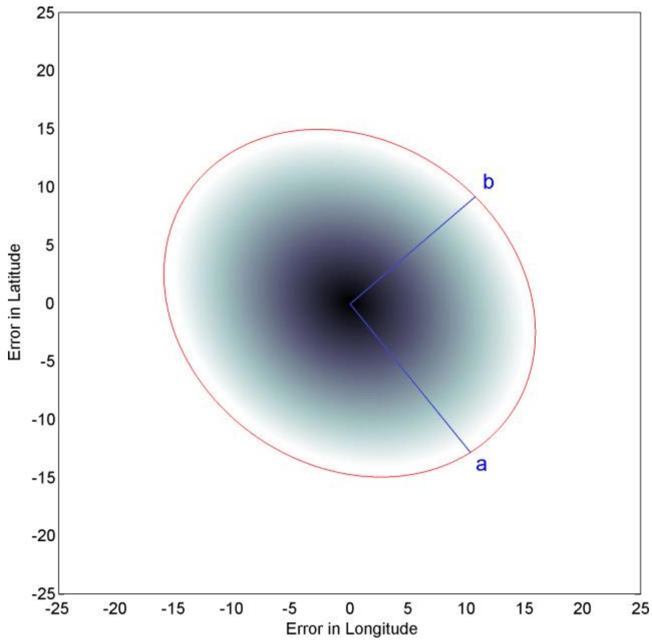

Figure 10 – Horizontal error distribution as 2D Normal

We make an approximation that the error is conservatively described by a circular 2D Normal distribution, as shown in Figure 10, with an iso-probability error ellipse (red) with semi-major and semi-minor ellipse axes (blue) and standard deviation given by the larger of these ($\sigma_{a,0}$). We choose to apply a Rayleigh probability distribution to describe the horizontal error. The cumulative distribution function (CDF) and its derivative, the probability distribution function (PDF) of the Rayleigh distribution are given as:

$$F(r) = 1 - e^{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2}$$
$$F'(r) = \frac{r}{\sigma^2} e^{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2}$$

In our case, the CDF can be interpreted as a conservative bound on one minus the probability of the radial error (r) on the solution exceeding a given radial distance (R) from the true position as:

$$p\{r > R\} \leq 1 - F(R) = e^{-\frac{1}{2}\left(\frac{R}{\sigma_{a,0}}\right)^2}$$

We assume that the pseudo-range measurements and the solution errors are approximately un-biased Normal variables. A-RAIM makes use of nominal biases to minimise the standard deviation parameters needed to conservatively bound ranging errors [16, 17] but within the timescale available for the MarRINav project we shall not.

3.3.2 Probability of Misleading Information ($\boldsymbol{p_{MI}}$)

We can consider the probability that the error on the navigation solution exceeds a particular level. For example we might define a HPL as the radial error expected to be exceeded only a small percentage (p) of the time, we can derive this by inverting the CDF equation:

$$p = e^{-\frac{1}{2}\left(\frac{HPL}{\sigma_{a,0}}\right)^2}$$
$$HPL = \sigma_{a,0}\sqrt{-2\ln(p)}$$

This is what we call the fault-free model – we assume that all of the measurement errors are conservatively described by Normal distributions with variances given by the modelled errors (σ_i^2) used in the weighting matrix. We know that these nominal errors will only prove to be conservative when the measurement errors are well behaved. The models are not expected to hold when integrity threats occur such as a satellite failure, jamming, or non-line-of-sight (NLOS) reception of a GPS signal.

We might sub-divide the total allowable integrity risk into two portions: a fault-free and faulted case. For the fault-free case integrity can be preserved by ensuring the HPL is a conservative bound up to the apportioned level of risk.

$$I_{req} = I_{fault} + I_{free}$$
$$HPL_0 = \sigma_{a,0}\sqrt{-2\ln\left(I_{free}\right)}$$

Again we use the subscript zero to indicate that this is the fault free HPL, and to differentiate it from other HPLs we may define later. It is then also necessary to ensure that when faults occur, they can be detected with adequate fidelity that integrity threats do not induce more risk of HMI than apportioned (I_{fault}).

We can now look at the probability of the error exceeding a particular bound when faults *do* occur. Assuming that a fault may occur which causes one or more of the measurement errors to become biased, this will incur a bias in the navigation solution and move the centre of the distribution a distance (μ) from the true position. We can conservatively apply the Rayleigh distribution in this case, provided that the bias (μ) is less than the value R:

$$p\{(r > R)|(\mu \neq 0)\} \leq 1 - F(R - \mu) = e^{-\frac{1}{2}\left(\frac{R-\mu}{\sigma_{a,0}}\right)^2}$$

Our approximation states that when the bias approaches the value of R we assume 100% probability of the error exceeding this value:

$$p\{(r > R)|(\mu = R)\} \leq 1 - F(R - R) = e^{-\frac{1}{2}\left(\frac{0}{\sigma_{a,0}}\right)^2} = 100\%$$

This calculation is useful, for example when considering the probability that the biased error distribution exceeds the estimated HPL. We call this the probability of misleading information (p_{MI}):

$$p_{MI}(\mu) = e^{-\frac{1}{2}\left(\frac{HPL-\mu}{\sigma_{a,0}}\right)^2} for\ \mu \leq HPL$$

$$p_{MI}(\mu) = 1\ for\ \mu \geq HPL$$

Generally, small biases are not a problem and larger biases are more hazardous. The risk of the error exceeding the HPL is assumed to approach certainty as the bias reaches the HPL in magnitude.

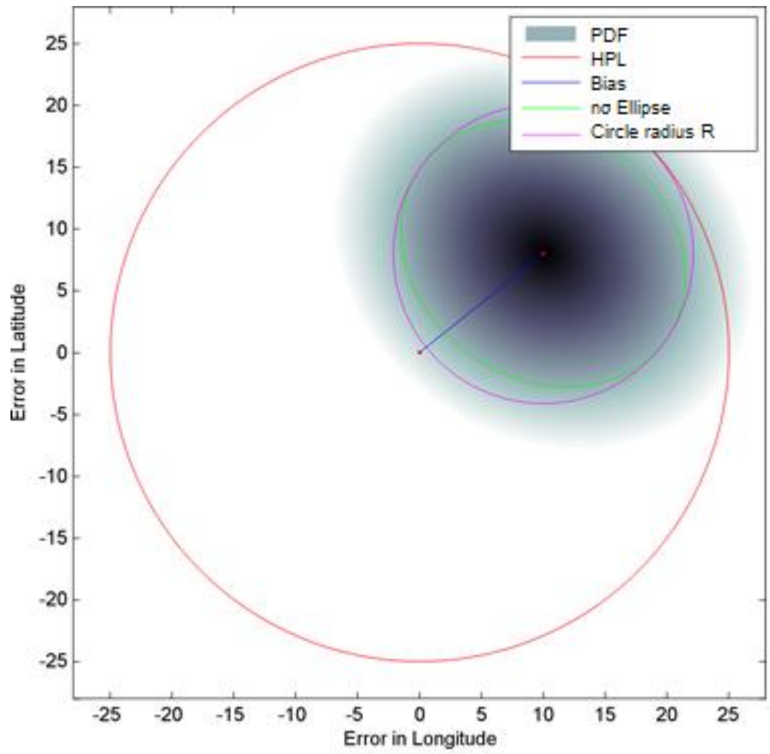

Figure 11 – Graphical representation of using the Rayleigh (circular)

Figure 11 depicts a graphical representation of using the Rayleigh (circular) distribution to over-estimate the probability of the error exceeding some HPL. More of the distribution exceeds the green ellipse than exceeds the circular bound given by the semi-major axis (the basis of the Rayleigh model). More of the error exceeds the Rayleigh bound (magenta circle) than lies outside the HPL (red circle). The Rayleigh model with $R = HPL - \mu$ is a conservative way of estimating the risk of the error exceeding the HPL.

Since the nature of the fault is unknown, the bias (μ) is essentially arbitrary, and if it grows too large there will always be 100% probability of the error exceeding *any* HPL we define. A fault detection process is needed to actively seek out these errors and prevent hazardous position biases from corrupting the receiver's output navigation solution.

3.3.3 Reduced-Subset Solutions

A number of reduced-subset solutions are formed, indexed by the letter (j). For each subset, a particular unique set of pseudo-ranges $\{y_{j1}, y_{j2}, \dots\}$ are removed from solution. The weighting matrix (W_j) is defined which sets the on-diagonal elements corresponding to $\{y_{j1}, y_{j2}, \dots\}$ to zero.

$$W_j = \begin{bmatrix} \sigma_1^2 & 0 & & \cdots \\ 0 & 0 & & \\ \vdots & & \sigma_3^2 & \cdots \\ & & \vdots & \ddots \end{bmatrix}^{-1}$$

A new projection-matrix is defined:

$$K_j = \left(G^T W_j G\right)^{-1} G^T W_j$$

And a new solution is formed:

$$x_j = K_j y$$

With co-variance given by:

$$C_j = \left(G^T W_j G\right)^{-1}$$

The difference between the all-in-view solution and the reduced subset is used as a fault detection parameter:

$$d_j = x_j - x_0 = \left(K_j - K_0\right) y$$

This has co-variance defined by:

$$C_{d,j} = C_j - C_0$$

We note that this is a non-trivial result derived from the fact that the error on the reduced subset solution is considerably correlated with the error on the all in view solution. A proof of this can be derived, and is also given in the Appendix to [7]. The parameter d_j has four components – the 3D geographical separation between the solutions and the difference in the clock-solution.

$$d_j = [d_{x1} \quad d_{x2} \quad d_{x3} \quad d_t]_j{}^T$$

We are only interested in the horizontal component; this is the root sum square of the first two:

$$d_{H,j} = \sqrt{d_{x1}^2 + d_{x2}^2}$$

This parameter has a 2D Normal (elliptical) distribution for which the semi-major axis ($\sigma_{d,j}$) is calculated from the co-variance $C_{d,j}$, and once again a Rayleigh approximation is applied. We note (but do not prove for now) that the co-variance matrix $C_{d,j}$ is degenerate if the subset solution excludes three or fewer measurements, and that the error ellipse for $d_{H,j}$ has eccentricity equal to one for single-satellite faults. A regular 1D Normal distribution would be more appropriate to describe a single-satellite fault, but for reasons we will derive later, the Rayleigh distribution suffices as a conservative bound.

A fault detection threshold is set for the j^{th} solution employing the inverse Rayleigh CDF such that the threshold is expected to be exceeded only a small fraction of the time:

$$t_j = \sigma_{d,j} \sqrt{-2 \ln\left(p_{FA,j}\right)}$$

Here $p_{FA,j}$ is the amount of false alarm risk allocated to this subset, subject to the sum total over all subset solutions equalling a maximum allowance:

$$\sum_{all\ j} p_{FAj} = p_{FA}$$

We shall deal with how to calculate this allowance and how to allocate this risk budget between the individual solutions later. For now it suffices to say that the risk of false alarm is budgeted and controlled to preserve Continuity.

Stanford review team commented here: {“This section proposes to preserve continuity by sub-allocating false-alarm risk. However, depending on the operation, exclusion capability might also be needed to meet the continuity requirements.”} This would require further research, noting that determining the impact of Fault Detection and Exclusion (FDE) on continuity and availability is non-trivial.

3.3.4 Fault Detection and Missed Detection

Only if all separation distances ($d_{H,j}$) are below their respective thresholds (t_j) then the all-in-view solution is accepted, the HPL is calculated and the green light integrity guarantee can be shown.

If one of the separation distances ($d_{H,j}$) exceeds its respective threshold (t_j) then the corresponding subset $\{y_{j1}, y_{j2}, \dots\}$ are deemed to contain a fault and may be considered for removal. Successful fault detection and exclusion (FDE) means that the solution can be given the green light integrity guarantee after removal the faulty measurements – potentially this improves user level continuity significantly. If successful exclusion is *not* achieved then the red light integrity alarm is raised to warn the navigator of a problem with the GNSS solution. FDE may not always be available and is considered a ‘nice to have’ feature. We do not deal specifically with FDE in this report.

To calculate the probability of misleading information (p_{MI}) we considered the case when a particular subset was assumed to be faulted, and the effect this had on the navigation solution. We now also consider the effect of this same bias, and the effect it has on the fault detection process. The horizontal position error bias (μ) incurred on the all in view solution relates to faults on a particular subset of measurements $\{y_{j1}, y_{j2}, \dots\}$, since these measurements are absent from the j^{th} subset solution this solution incurs no bias. The j^{th} separation distance is conservatively described by a biased Rayleigh distribution and the probability of the value lying below the threshold can be found:

$$p\left(d_{H,j} < t_j\right)|(\mu \neq 0) \leq 1 - F\left(\mu - t_j\right) = e^{-\frac{1}{2}\left(\frac{\mu - t_j}{\sigma_{d,j}}\right)^2}$$

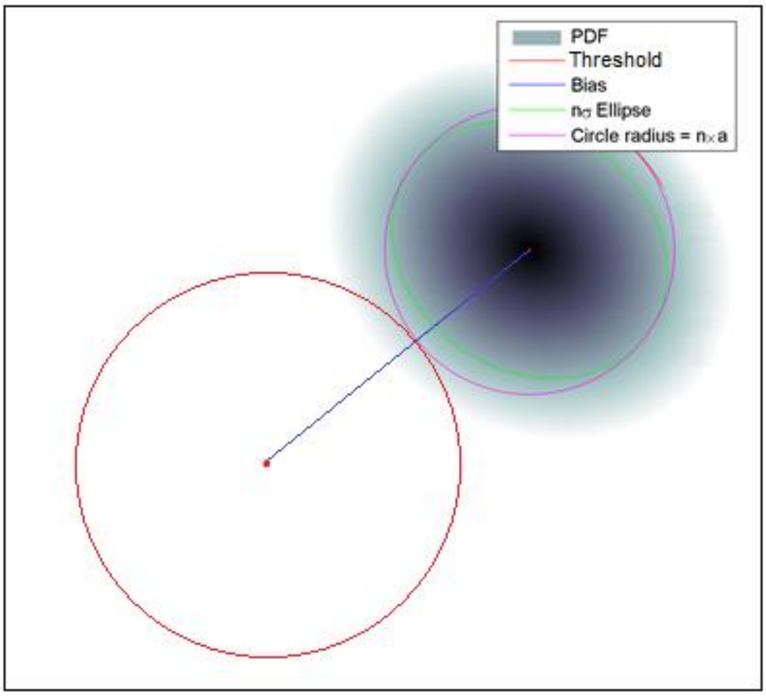

Figure 12 – Example of using the Rayleigh distribution to over-bound the risk of missed-detection.

Figure 12 shows an example of using the Rayleigh distribution to over-bound the risk of missed-detection. Any solution lying within the threshold (red circle) is missed – the same over-bounding argument using a circular approximation is applied as before, this time $R = \mu - t_j$.

Once again, we have made an approximation: the bias must be sufficient to place the majority of the distribution outside the circle defined by the detection threshold:

$$\mu \geq t_j$$

We have to consider the alternative, and assume a 100% probability of missed detection for small biases:

$$p(d_{H,j} < t_j)|(\mu = t_j) \leq 1 - F(t_j - t_j) = e^{-\frac{1}{2}\left(\frac{0}{\sigma_{d,j}}\right)^2} = 100\%$$

We now have a function for the probability of missed detection (p_{MD}):

$$p_{MD}(\mu) = e^{-\frac{1}{2}\left(\frac{\mu - t_j}{\sigma_{d,j}}\right)^2} \; for \; \mu \geq t_j$$
$$p_{MD}(\mu) = 1 \; for \; \mu \leq t_j$$

Generally, very large biases are less of a problem, since they are easy to detect. Likewise very small biases do not deflect the navigation solution far from the truth and are not a hazard.

Mid-range biases are more insidious, causing significant navigation error, without being easy to detect. It is these that the integrity algorithm must be most wary of, and these that drive the calculation of the protection level.

3.3.5 Integrity Risk and Protection Level

We now define the integrity risk for the process. For each subset, a risk exists that the fault occurs (with some nominal *a-priori* fault probability $p_{f,j}$). If it does, and the fault causes the error to exceed the HPL then this is *misleading information*. If this goes un-detected then an incident of *hazardously misleading information* (HMI) happens.

The Stanford review team commented here:

{"The definitions of "misleading information" and "hazardously misleading information" do not match the meanings they have in aviation. Both terms normally represent the case where no alert or annunciation occurs within the time to alert. Usually "Hazardous" is added when the HPL is also above the Horizontal Alert Limit (without annunciation). "Hazardous" also implies conditions which have a severity level of "Hazardous" as classified by aviation, which does not appear to be the case here (since a 10^{-5} -level integrity requirement implies at most "Major" severity in aviation terms)."}

The GLA authors note this difference between maritime and aviation definitions.

The risk of HMI for each subset is given as a function of the magnitude of the fault:

$$p_{HMI,j}(\mu) = p_{f,j} p_{MD,j}(\mu) p_{MI,j}(\mu)$$

This is given as:

$$p_{HMI,j}(\mu) = p_{f,j} e^{-\frac{1}{2}\left(\frac{\mu - t_j}{\sigma_{d,j}}\right)^2} e^{-\frac{1}{2}\left(\frac{HPL-\mu}{\sigma_{a,0}}\right)^2} \; for\ t_j \leq \mu \leq HPL$$

$$p_{HMI,j}(\mu) = p_{f,j} e^{-\frac{1}{2}\left(\frac{HPL-\mu}{\sigma_{a,0}}\right)^2} \; for\ \mu \leq t_j$$

$$p_{HMI,j}(\mu) = p_{f,j} e^{-\frac{1}{2}\left(\frac{\mu - t_j}{\sigma_{d,j}}\right)^2} for\ \mu \geq HPL$$

For the small-bias region ($\mu \leq t_j$) faults are assumed to be undetectable and p_{MI} is a monotonically increasing function of bias, so has positive derivative attaining a maximum at ($\mu = t_j$). Likewise, for the large-bias region ($\mu \geq HPL$) the solution is assumed to always exceed the HPL and p_{MD} is a monotonically decreasing function of bias, with negative derivative and a maximum at ($\mu = HPL$). It can be seen that risk of HMI as a function of bias is continuous, differentiable and has continuous derivative throughout.

A single point of maximum risk exists, and it is always within the region ($t_j \leq \mu \leq HPL$). We find what this maximum point is by differentiating the risk with respect to μ and setting this equal to zero to find the turning point, which then constitutes the maximum risk over any magnitude of fault.

$$\frac{dp_{HMI,j}(\mu)}{d\mu} = \left(\frac{HPL - \mu}{\sigma_{a,0}^2} - \frac{\mu - t_j}{\sigma_{d,j}^2}\right) p_{HMI,j}(\mu)$$

This has a zero at a bias value that we shall call the *worst-case fault*:

$$\mu_0 = \frac{\sigma_{d,j}^2 HPL - \sigma_a^2 t_j}{\sigma_{d,j}^2 + \sigma_a^2}$$

This is the worst case since the risk of HMI for any biases other than this exact value will be lower. Since we do not know what magnitude the bias can take, it is conservative to assume that when a fault occurs it always takes exactly this value. We are now able to provide an over-bound for the integrity risk on the j^{th} subset *independent* of the magnitude of the fault:

$$p_{HMI,j}(\mu_0) \geq p_{HMI,j}(\mu) \; \forall \; \mu$$

And so by pre-allocating a portion of the total integrity risk budget to each subset (I_j) we can ensure that the sum total over all of the subsets never exceeds some maximum total allowance:

$$p_{HMI,j}(\mu_0) = I_j$$
$$\sum_{all\ j} I_j = I_{total}$$

This is how we define the HPL for each subset. It is set such that the maximum possible risk never exceeds the allocation:

$$p_{HMI,j}(\mu) \leq p_{HMI,j}(\mu_0) = p_{f,j} e^{-\frac{1}{2}\left(\frac{\mu_0 - t_j}{\sigma_{d,j}}\right)^2} e^{-\frac{1}{2}\left(\frac{HPL_j - \mu_0}{\sigma_{a,0}}\right)^2} = I_j$$

Substituting the form of μ_0 from above, and re-arranging to make HPL_j the subject of the formula (after some work) this yields:

$$HPL_j = t_j + \sqrt{-2\left(\sigma_{a,0}^2 + \sigma_{d,j}^2\right)\ln\left(\frac{I_j}{p_{f,j}}\right)}$$

We have already seen that the co-variance of the all in view and reduced subset solutions are related:

$$C_{d,j} = C_j - C_0$$

We determine that the error on the reduced subset solution is effectively the sum of two random, independent variables: the error on the all in view solution; and the separation distance:

$$C_j = C_0 + C_{d,j}$$

This allow us to use the semi-major axis of the horizontal error ellipse of the reduced subset solution as an upper bound:

$$\sigma_{a,j}^2 \geq \sigma_{a,0}^2 + \sigma_{d,j}^2$$

The Stanford review team noted that:

{"The semi-major axis is an upper bound. This is demonstrated by noting that the maximum eigenvalue of a sum of matrices is always smaller than the sum of the maximum eigenvalue of each matrix."}

Substituting this, and the definition of the detection threshold into the HPL equation, and by defining an allocated missed detection probability for each subset:

$$p_{MD,j} = \frac{I_j}{p_{f,j}}$$

We arrive at a pleasing and symmetrical closed-form solution for the subset HPL:

$$HPL_j = \sigma_{d,j}\sqrt{-2\ln\left(p_{FA,j}\right)} + \sigma_{a,j}\sqrt{-2\ln\left(p_{MD,j}\right)}$$

We also now remind ourselves of the fault free HPL:

$$HPL_0 = \sigma_{a,0}\sqrt{-2\ln\left(I_{free}\right)}$$

The HPL which is passed to the navigator, and which must lie below the HAL of 25m for the solution to be accepted, is given as the maximum over all HPLs defined:

$$HPL = \max\{HPL_j\}\ all\ j \geq 0$$

3.3.6 Risk-Budgeting

Not all possible faults are included in the subset solutions since the number of permutations for large numbers of faulted measurements would be prohibitively high. This is as defined for the A-RAIM process: a large group of faults are designated "un-monitored", these are faults which are thought to occur sufficiently rarely that integrity is not significantly harmed by ignoring them. We sum up the a-priori fault probability for all these:

$$p_{un\ monitor} = \sum_{all\ k \neq j} p_{f,k}$$

We assume that HMI can occur whenever any of these faults are present (since we conservatively assume they cannot be detected). A total integrity risk budget can now be drawn up, this is the sum of risks from all un-monitored faults; plus the fault free risk allocation; plus the sum of all integrity risks (I_j) apportioned to each subset. This has to be equal to the maximum allowable integrity risk requirement ($I_{req} = 1.667 \times 10^{-6}$):

$$I_{req} = \sum_{all\ j} I_j + I_{free} + p_{un\ monitor}$$

For all of the faults which we do monitor, it is assumed that any one of them can cause an integrity alarm (hence loss of continuity) when they happen. The sum of the a-priori fault probabilities for the monitored threats gives the contribution to continuity risk:

$$p_{monitor} = \sum_{all\ j} p_{f,j}$$

A continuity budget can now also be drawn up, by adding all the false alarm allowances from each subset to the "true alarm" probabilities given by $p_{monitor}$. Again, we demand that the total must equal the maximum allowable continuity risk allowance ($C_{req} = 5 \times 10^{-5}$).

$$C_{req} = \sum_{all\ j} p_{FA,j} + p_{monitor}$$

We note here that fault probabilities in the maritime environment may prove to be prohibitively large. To preserve integrity, $p_{un\ monitor}$ must be small. We may have to 'monitor' a very large number of possible permutations of different kinds of faults to ensure this – potentially hundreds if not thousands (!) of subset solutions might have to be defined.

Since each of these monitored faults is a potential integrity alarm, $p_{monitor}$ may also be large. Indeed, it may actually exceed the total continuity risk. In this case, continuity of electronic positioning cannot be guaranteed unless the vessel has some resilient fall back system to continue operations when the GNSS receiver raises an alarm. Alternatively FDE will have to become mandatory as opposed to being a 'nice to have' functionality.

Both of these possibilities (large processor burden, or the need for a backup PNT system) depend on the severity of the maritime environment and the nature of the threats against GNSS reception.

A long term study evaluating suitable fault free receiver error models, and the risk of these models being violated, is a necessary pre-requisite to providing user-level integrity at sea.

Looking again at the HPL equation:

$$HPL_j = \sigma_{d,j}\sqrt{-2\ln\left(p_{FA,j}\right)} + \sigma_{a,j}\sqrt{-2\ln\left(p_{MD,j}\right)}$$

We can 'spend' each of our budgets allocating a certain amount of risk to each of the subset solutions, in doing so we reduce the corresponding HPL. Continuity risk (false alarm) spent on a subset lowers the detection threshold and reduces the first term; integrity risk (missed detection) spent on a subset lowers the integrity bound and reduces the second term.

Since the solution HPL is the maximum of all of these:

$$HPL = \max\left\{HPL_j\right\}\ all\ j \geq 0$$

It is optimal to allocate the budgets such that all HPLs are created equal. There exists an optimal allocation of both budgets which produces a globally smallest HPL, but we do not derive this method. Instead we use a simpler method to balance each budget independently.

3.3.7 Budget Allocation and Optimised HPL

The question is how to allocate probabilities of false alarm ($p_{FA,j}$) and missed detection ($p_{MD,j}$) to create a number of protection levels:

$$HPL_j = \sigma_{d,j}\sqrt{-2\ln\left(p_{FA,j}\right)} + \sigma_{a,j}\sqrt{-2\ln\left(p_{MD,j}\right)}$$

Such that the total continuity and integrity budgets are balanced.

$$\sum_{all\ j} p_{FA,j} = C_{req} - p_{monitor}$$
$$\sum_{all\ j} p_{f,j}p_{MD,j} = I_{req} - I_{free} - p_{un\,monitor}$$

Ideally this is done such that all HPLs take the same value. This can be achieved by balancing each budget separately such that both the false alarm and missed detection terms in the HPL are the same for all subsets, defining:

$$HPL_j = t_j + l_j$$
$$t_j = \sigma_{d,j}\sqrt{-2\ln\left(p_{FA,j}\right)}$$
$$l_j = \sigma_{a,j}\sqrt{-2\ln\left(p_{MD,j}\right)}$$

Such that every t_j is the same, and every l_j is also:

$$t_j = t_k = t\ \forall\ j,k$$
$$l_j = l_k = l\ \forall\ j,k$$

We can also balance the allocation of fault free and faulted risk budgets such that the fault free HPL also takes the same value:

$$HPL_0 = HPL_j = t + l$$

If we define:

$$T_j = e^{-\frac{1}{2}\left(\frac{1}{\sigma_{d,j}}\right)^2}$$
$$p_{FA,j} = e^{-\frac{1}{2}\left(\frac{t}{\sigma_{d,j}}\right)^2} = \left[T_j\right]^{t^2}$$

Likewise:

$$L_j = e^{-\frac{1}{2}\left(\frac{1}{\sigma_{a,j}}\right)^2}$$
$$p_{MD,j} = e^{-\frac{1}{2}\left(\frac{l}{\sigma_{a,j}}\right)^2} = \left[L_j\right]^{l^2}$$
$$I_{free} = e^{-\frac{1}{2}\left(\frac{t+l}{\sigma_{a,0}}\right)^2} = [L_0]^{(t+l)^2}$$

The factors T_j and L_j depend only on the solution geometry and can be calculated before we allocate the risk budgets or define the detection thresholds or HPL. Our budget-balancing problem is now:

$$\sum_{all\ j} \left[T_j\right]^{t^2} = C_{req} - p_{monitor}$$
$$[L_0]^{(t+l)^2} + \sum_{all\ j} \left[L_j\right]^{l^2} = I_{req} - p_{un\,monitor}$$

Since all T_j are less than 1, and t^2 is positive, the function $\left[T_j\right]^{t^2}$ is a monotonically decreasing function of t. We can define a search interval by calculating maximum and minimum possible values for t; a simple way to do this is by allocating the continuity budget equally between solutions (where there are N monitored subset solutions):

$$p_{FA} = \frac{1}{N}\left(C_{req} - p_{monitor}\right)$$
$$t_j = \sigma_{d,j}\sqrt{-2\ln(p_{FA})}$$
$$t_{min} = \min\{t_j\}$$
$$t_{max} = \max\{t_j\}$$

The search begins by determining the sum for the half way point in the search interval:

$$t_0 = \frac{1}{2}(t_{min} + t_{max})$$
$$S = \sum_{all\ j} \left[T_j\right]^{t_0^2}$$

If $(S > C_{req} - p_{monitor})$ then a new search begins at the centre of the upper half-interval $[t_0, t_{max}]$, otherwise the search proceeds at the mid-point of the lower half-interval $[t_{min}, t_0]$. This continues until the search interval is reasonably small, say 1cm. Since the value t_{max} will rarely exceed a few meters, this should converge in less than 10 iterations.

Once the value for t is found, this can be substituted into the integrity budget, and then the value for l found in a similar way. A single HPL is then found:

$$HPL = t + l$$

Only if all of the separation distances (d_j) lie below the threshold (t), and the HPL lies below the HAL is the navigation solution considered acceptable for use.

4 Conclusions and Recommendations

1. It would be more difficult and costly to utilise IALA Beacon DGPS than EGNOS V3 (or alternative SBAS) for the future provision of maritime navigation integrity at user-level.

   Recommendation: The dual frequency multi constellation (DFMC) capability of EGNOS V3 (or alternative SBAS), supported by the ship's Multi System Receiver (MSR), should be used in the development of position integrity for vessels rather than modifying the beacon system.

2. SBAS (EGNOS V3) alone will be insufficient to address user-level integrity for general maritime navigation due to the local GNSS signal reception environment (noise, interference, multipath and non-line-of-sight reception) on vessels.

   Recommendation: Receiver algorithms for receiver autonomous integrity monitoring (RAIM) should be designed, and an appropriate IEC test specification produced to ensure future type approved receivers adequately protect the user from potentially misleading GNSS errors caused by effects local to the vessel.

3. Maritime RAIM (M-RAIM) is a method that shows considerable promise as a candidate form of RAIM for inclusion in the maritime user-level integrity solution.

   Recommendation: M-RAIM should be researched further and evaluated for implementation in future maritime receivers when used either in combination with SBAS (e.g. EGNOS V3) or standalone (for locations outside SBAS coverage).

4. The fundamental assumption of IBPL autonomous receiver monitoring is not generally valid in maritime operations and IBPL cannot be relied upon to provide user-level integrity and continuity on vessels.

   Recommendation: IBPL should not be implemented in receivers for general maritime navigation.

5. Existing SBAS (EGNOS V3) information planned to be provided for aviation is not ideal for determining maritime user-level integrity and hence provision of the underlying SBAS error statistics[5] would assist solutions for user-level integrity.

[5] Ideally a bespoke SBAS maritime service message would broadcast fault-free estimates of UDRE and GIVE, recognising this may not be a cost-effective option. Other options that could be considered are (i) a single bespoke data-field within another SBAS message defines a maritime scale-factor (e.g. x2.5) by which to reduce the aviation UDRE and GIVE bounds to yield effective maritime fault-free estimates of these parameters, and (ii) a bespoke table is implemented to convert broadcast UDRE and GIVE index parameters to appropriate fault-free error estimates, employed exclusively by maritime receivers

Recommendation: Changes to EGNOS parameters or transmission of an additional maritime message should be investigated to evaluate whether the provision of maritime specific information would be cost-effective.

6. It is recognised that the capability for future SBAS integrity bounds to be broadcast as separate mean and standard deviation figures, allowing the broadcast error to more closely match the expected fault-free error without excessive inflation, may not be cost-effective.

   Recommendation: The feasibility of making changes to the broadcast SBAS information should be investigated further; also the idea suggested by Stanford to develop a table that allows the determination of fault free-sigmas, nominal biases, and faulted biases from already broadcast UDREIs (or DFREIs).

7. Protection levels derived from SBAS and RAIM may be overly conservative if they are driven by “worst-case” fault scenarios and a “specific risk” integrity design.

   Recommendation: Consideration should be given to “specific” vs. “average” risk; a “fault-averaged risk” approach would provide some degree of probabilistic averaging over the prior probabilities of faults. It is noted that M-RAIM adopts a “fault-averaged risk” approach based on a-priori fault probabilities.

8. The use of dual frequency combinations in maritime may lead to an inflation of error bounds due to a multiplication factor on multipath imposed by the iono-free measurement combination. However, single frequency L5/E5a is potentially a poor choice because the ionospheric delay is 1.8 times larger than L1/E1, and more importantly so is the uncertainty (GIVE and corresponding UIRE would need to be multiplied by this number).

   Recommendation: The advantages of dual-frequency L1/L5 (E1/E5a) use against the use of single-frequency L1/E1 or L5/E5a for maritime positioning should be investigated further by trade-off analysis.

9. For M-RAIM, it is necessary to determine a nominal vessel multi-path model, and the associated probability with which instantaneous measurements exceed this model (fault probability).

   Recommendation: More information should be gathered from real-world measurements in the maritime environment (including how the environment varies under different operational conditions) to establish a multipath model, and the associated probability with which instantaneous measurements exceed this model (fault probability).

10. PPP is a powerful technique to combat multipath, but user-level integrity for PPP has not yet been developed.

Recommendation: PPP for maritime applications should be researched further but PPP should not be used for general maritime navigation until and unless a user-level integrity solution has been developed.

References

[1] MarRINav Deliverable D6, "MSR Integrity Report", Version 1.1, 18th October 2019.

[2] IMO Resolution A.1046 (27) – 'World-Wide Radio-Navigation System'. IMO 30th November 2011.

[3] Gunning, Kazuma, Blanch, Juan, Walter, Todd, de Groot, Lance, and Norman, Laura, "Design and Evaluation of Integrity Algorithms for PPP in Kinematic Applications." *Proceedings of ION GNSS+ 2018*, Miami, FL, September 2018, pp. 1910-1939. http://web.stanford.edu/group/scpnt/gpslab/pubs/papers/GunningIONGNSS2018PPPintegrity.pdf

[4] Gunning, Kazuma, Blanch, Juan, Walter, Todd, de Groot, Lance, and Norman, Laura, "Integrity for Tightly Coupled PPP and IMU." *Proceedings of ION GNSS+ 2019*, Miami, FL, September 2019, pp. 3066-3078. http://web.stanford.edu/group/scpnt/gpslab/pubs/papers/Gunning_Integrity_for_PPP_and_IMU_IONGNSS2019.pdf

[5] Walter, Todd, Enge, Per K., and DeCleene, Bruce, "Integrity Lessons from the WAAS Integrity Performance Panel (WIPP)," *Proceedings of the ION 2003 National Technical Meeting*, Anaheim, CA, January 2003, pp. 183-194. http://web.stanford.edu/group/scpnt/gpslab/pubs/papers/Walter_IONNTM_2003.pdf

[6] Pullen, Samuel P., Walter, Todd, and Enge, Per K., "SBAS and GBAS Integrity for Non-Aviation Users: Moving Away from 'Specific Risk'," *Proceedings of the ION 2011 National Technical Meeting*, San Diego, CA, January 2011, pp. 533-545. http://web.stanford.edu/group/scpnt/gpslab/pubs/papers/Pullen_IONITM_2011_Integrity_for_Non-Aviation_Users.pdf

[7] https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/risk_management/ss_handbook/media/chap3_1200.pdf

[8] IMO Resolution MSC.401(95) – 'Performance Standards for Multi-System Shipborne Radionavigation Receivers', Adopted 8 June 2015.

[9] International Maritime Organization (2001), Resolution A.915(22): *Revised Maritime Policy and Requirements for a Future Global Navigation Satellite System (GNSS)*, adopted 29 November 2001.

[10] Blanch, Juan, Walter, Todd, Enge, Per, "A Formula for Solution Separation without Subset Solutions for Advanced RAIM," *Proceedings ION/IEEE PLANS*, Monterey, California, April 2018. http://web.stanford.edu/group/scpnt/gpslab/pubs/papers/Blanch_IEEEIONPLANS_2018_ARAIM.pdf

[11] Blanch, Juan, Walter, Todd, Enge, Per, "Fixed Subset Selection to Reduce Advanced RAIM Complexity," *Proceedings of the 2018 International Technical Meeting of The Institute of Navigation*, Reston, Virginia, January 2018, pp. 88-98. http://web.stanford.edu/group/scpnt/gpslab/pubs/papers/Blanch_IONITM_2018_FixedSubsetARAIM.pdf

[12] IALA Recommendation R-121 & Guideline 1112.

[13] RTCA DO-229/D – 'MOPS for GPS / SBAS airborne equipment', RTCA SC 159, 2006.

[14] Inside GNSS article "Autonomous Integrity – An Error Isotropy-Based Approach for Multiple Fault Conditions" https://www.insidegnss.com/auto/janfeb09-azaoli.pdf

[15] Walter, Enge, "Weighted RAIM for Precision Approach", ION GPS 1995.

[16] EU US Cooperation on Satellite Navigation. Working Group C, ARAIM Technical Subgroup, "Milestone Report", (various)

[17] Blanch, J., Walter, T., Enge, P., "RAIM with Optimal Integrity and Continuity Allocations Under Multiple Failures," IEEE Transactions on Aerospace and Electronic Systems Vol. 46, No. 3, July 2010. https://doi.org/10.1109/TAES.2010.5545186

[18] Welte, A. "Protection Levels for High Integrity Localization for Autonomous Driving", MSC Thesis, Ensta Bretagne, Aug 20th 2017

[19] Walter, Blanch, Enge, "VPL Equations for Dual Frequency SBAS", ION GNSS 2010

[20] Bang, Eugene, Milner, Carl, Macabiau, Christophe, and Estival, Philippe, "Sample Temporal Correlation Effect on PHMI," Proceedings of the 2019 International Technical Meeting of The Institute of Navigation, Reston, VA, January 2019, pp. 85-99. https://doi.org/10.33012/2019.16684

[21] Pervan, Boris, Khanafseh, Samer, and Patel, Jaymin, "Test Statistic Auto-and Cross-correlation Effects on Monitor False Alert and Missed Detection Probabilities," Proceedings of the 2017 International Technical Meeting of The Institute of Navigation, Monterey, CA, January 2017, pp. 562-590. https://doi.org/10.33012/2017.14874

[22] Blanch, Juan A., Walter, Todd, Enge, Per K., Lee, Y., Pervan, Boris S., Rippl, M., Spletter, A., Kropp, V, "Baseline Advanced RAIM User Algorithm and Possible Improvements," IEEE Transactions on Aerospace and Electronic Systems, Volume 51, No. 1, January 2015. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=7073524

Appendix A: Additional Discussion of "Specific" and "Average" Risk

A.1 Appendix provided by the Stanford review team

In section 1.4 of the main body of this report, the section named "Definitions of 'Specific' and 'Average' Risk" provided definitions of these concepts and references to papers discussing them. This appendix clarifies the distinction between these different interpretations of risk and how they have affected integrity assessments for GNSS aviation systems in the past.

"Specific" in "specific risk" refers to the method by which the effects of faults or anomalies that could lead to these events with severe (or worse) consequences are calculated. The focus on this method is on the worst specific circumstances that could be encountered rather than a probabilistic average over all such circumstances of a given class (for example, GNSS satellite clock failures). As explained above, "specific risk" requires that "worst-case" parameters be chosen for both prior probabilities of faults and magnitudes of errors resulting from these faults unless substantial, verifiable information exists to support a less-conservative choice. For prior probabilities, commitments by GNSS service providers often form the basis for using values lower than 1.0 (1.0 being the worst case, as it assumes that the event in question always occurs). For error magnitudes, threat models corresponding to each fault place bounds on the size of errors, but it is typically difficult to justify a specific choice of probability distribution for errors within the threat model that would allow errors other than the worst-case (the "threatening" value that is the most difficult to detect) to be considered in a probabilistic calculation.

The key difference between "fault-averaged risk" and "specific risk" is that probabilistic averaging is allowed in the derivation of both prior fault probabilities and distributions of error magnitudes as long as sufficient justification exists to do this and the faults in question are both rare and unpredictable. Both of these assessments still require conservatism, particularly in the case of error distributions, where the assumption of Uniform distributions over a very large range of magnitudes may underweight the errors that are most threatening. Averaging is not forbidden in "specific risk," but it is typically limited to prior-probability assessments, as faults occurring rarely tend to provide few samples of the resulting errors that would clearly support non-worst-case modeling of these errors.

An example of ionospheric spatial gradient threat modeling for CAT I precision approaches supported by the Ground-based Augmentation System (GBAS) may help differentiate these two approaches. A threat model giving minimum and maximum magnitudes of spatial gradients (along with other relevant parameters) was derived from a limited amount of historical data, since large spatial gradients are rare in mid-latitude regions. In addition, a probabilistic analysis based on historical data suggested that the prior probability of large gradients within the "threatening" part of the threat model was on the order of 10^{-5} per operation.

After extensive review and discussion, it was decided by the FAA that no explicit use of this probability (or any prior probability below 1.0) could be used in calculations to determine integrity risk under this scenario or its effect on protection levels. In addition, the worst-case gradient (usually the largest one in the threat model) had to be assumed at all times. This is an extreme example of the consequences of specific risk on aviation integrity. It had a major effect on GBAS ground system design and caused a significant loss in availability. This very conservative approach was partially due to uncertainty regarding the nature and scope of the ionospheric threat at the time the initial decision was taken along with the knowledge that distributed networks of GPS receivers (such as those used in SBAS) could provide advance warning of severe ionospheric gradients. As a result, these gradients were unpredictable to GBAS because GBAS did not have guaranteed access to such networks and only made observations from nearby locations at a single airport.

After years had passed and the analysis was repeated to support a more-demanding operation (CAT II/III precision approach to a lower decision height), a revised and more conservative assessment of the prior probability of anomalous ionospheric gradients (10^{-3} per approach) was developed and accepted for use, even under “specific risk.” What had changed over time was years of additional experience reinforcing the fact that large gradients were indeed very rare. In addition, it had become understood that, when anomalous gradients did occur, smaller gradients within the threat model were much more likely than the worst-case ones at the extremes which drove the integrity assessment. Despite this, the revised analysis maintained the use of the worst-case gradient at all times.

The approach that would have been taken under “fault-averaged risk” interpretation would likely have been different. For the CAT I risk assessment, a prior probability lower than 1.0 but not as low as 10^{-5} (probably 10^{-3}) would likely have been accepted, and the historical data of gradient magnitudes already available could have been used to fit a bounding (conservative) distribution of magnitudes within the threat model that would have removed the entire analysis being based on the very largest gradient observed. The impact on GBAS would have been much less severe while still meeting the same integrity requirement (under this interpretation of risk).

Regarding the difference between “specific risk” and “fault averaged risk,” the authors’ experience with aviation integrity leads them to prefer “specific risk” overall but with a willingness to consider “fault averaged risk” where (a) a given technique of probabilistic averaging is supported by a consensus of experts, and (b) probabilistic averaging has a significant performance benefit that is worth exploiting. The example given above regarding the GBAS ionospheric gradient threat is an extreme example that clarifies the differences between the two approaches and illustrates the potential impact. However, for the majority of rare events, use of a prior probability below 1.0 is justified under “specific risk,” while the basis for averaging error magnitudes is not as clear-cut. In practice, averaging the probability of which satellite measurement is faulted among all measurements (instead of assuming that the measurement having the worst effect is always faulted with the probability applying to all of them) produces a risk reduction of roughly one order of magnitude (given an average of approximately 10 satellites in view). However, since averaging over error magnitudes is harder to justify, the benefit that can be claimed with sufficient conservatism varies greatly (when it can be claimed) and may be less than another factor of 10.

The authors' perspective above applies mainly to aviation applications with integrity risk requirements on the order of 10^{-7} or lower. The interpretation of integrity risk for maritime applications may be different. Fault-averaging is more common in aviation when continuity risk at the level of 10^{-5} per operation or higher is evaluated (loss of continuity typically has "Minor" severity). Therefore, given that the maritime integrity requirement expressed in Section 2.3 of the GLA paper is 10^{-5} per 15 seconds, an integrity risk evaluation approach more tolerant of averaging than the authors' experience with aviation may be justified.

"Geometry-averaged risk" is seen quite differently from the above two interpretations of risk because it averages over parameters, such as satellite geometry, that are observable to all users. Thus, probabilistic averaging in this case is extended to parameters that are (or should be) known. The only advantages of this approach are simplicity and lower protection levels. Note that, if this interpretation were used, the assumptions underlying IBPL become less objectionable because errors can be averaged over the entire distribution of satellite geometries and weak geometries critically dependent on specific satellites (as illustrated in Section 2.4.1 of the GLA paper) contribute much less to the overall integrity risk.

To the authors' knowledge, geometry-averaged risk is not implemented in any GNSS applications with demanding integrity requirements, and the authors strongly recommend against it. In the authors' view, it is inappropriate to probabilistically average over parameters that are actually known (or can be easily computed) with information available in real time. Integrity risk (and protection levels generated to satisfy a given integrity risk requirement) should always account for all such known parameters.

Appendix B: Mathematical Discussion of Section 3.2 ("IBPL")

B.1 Appendix provided by the Stanford review team

There are some issues with the GLA derivation in Section 3.2.

The first inequality below Figure 6 in Section 3.2 states

$$|\hat{y}|^2 \geq |\hat{x}|^2 \, Tr(G^T G)$$

But this is not true in general. All we can say is that:

$$\hat{x}^T G^T G \hat{x} \geq \hat{x}^T \hat{x} \min \lambda\left(G^T G\right) \tag{1}$$

Where lambda designates the vector of eigenvalues of the matrix within the parenthesis above. The trace is equal to the sum of the eigenvalues. Here is a counter example: Take

$$G = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{and } x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The first term above is 0 and the second one is 1, therefore the inequality does not hold.

The following equation for the DOP (on page 30; right-hand part of last equation on the page):

$$DOP = \frac{1}{tr\left(G^T G\right)} \tag{2}$$

is incorrect. This is not equal to $DOP = \sqrt{tr\left(\left(G^T G\right)^{-1}\right)}$.

Here is a counterexample: take G = I_2 (the unit matrix of size 2). We have:

$$\frac{1}{tr\left(G^T G\right)} = \frac{1}{2} \tag{3}$$

whereas:

$$DOP = \sqrt{tr\left(\left(G^T G\right)^{-1}\right)} = \sqrt{2} \tag{4}$$

Proof for relation between HDOP and IBPL

Although these two equations are incorrect, the conclusion that

$$|\hat{y}|^2 \geq \frac{|\hat{x}|^2}{HDOP^2}$$
$$k|\omega| \geq |\hat{y}| \geq \frac{|\hat{x}|}{HDOP}$$

is still correct. Here is a proof:

$$\hat{x} = \left(G^T G\right)^{-1} G\hat{y} \tag{5}$$

Let us look at each component (e_k is the vector with all zeros except a one in the kth position):

$$e_k^T \hat{x} = e_k^T \left(G^T G\right)^{-1} G\hat{y} \tag{6}$$

Applying the Cauchy-Schwarz inequality to the vectors, we get:

$$\left(e_k^T \hat{x}\right)^2 \leq \left|e_k^T \left(G^T G\right)^{-1} G\right|^2 |\hat{y}|^2 \tag{7}$$

But we have:

$$\left|e_k^T \left(G^T G\right)^{-1} G^T\right| = e_k^T \left(G^T G\right)^{-1} G^T G \left(G^T G\right)^{-1} e_k = e_k^T \left(G^T G\right)^{-1} e_k \tag{8}$$

Therefore:

$$\left(e_1^T \hat{x}\right)^2 + \left(e_2^T \hat{x}\right)^2 \leq \left(e_1^T \left(G^T G\right)^{-1} e_1 + e_2^T \left(G^T G\right)^{-1} e_2\right) |\hat{y}|^2 = HDOP^2 |\hat{y}|^2 \tag{9}$$

The left side of the inequality is the horizontal error, so we have (as shown on page 31):

$$\left|\hat{x}_{hor}\right| \leq HDOP|\hat{y}| \tag{10}$$

Spherical integrals

As a check, we propose a slightly different approach to derive the IBPL inflation factors. The derivation included in the report makes a lot of sense, but it does require a very good grasp on n-dimensional geometry and makes at least one big leap (on page 36, where at “etc…”). The conclusion is the same.

We perform the change of variables (which is a modification to the usual n-dimensional spherical coordinates):

$$\begin{aligned}
y_1 &= r\cos(\theta)\cos(\alpha_1) \\
y_2 &= r\cos(\theta)\sin(\alpha_1)\cos(\alpha_2) \\
y_3 &= r\cos(\theta)\sin(\alpha_1)\sin(\alpha_2)\cos(\alpha_3) \\
y_4 &= r\cos(\theta)\sin(\alpha_1)\sin(\alpha_2)\sin(\alpha_3) \\
y_5 &= r\sin(\theta)\cos(\beta_1) \\
y_6 &= r\sin(\theta)\sin(\beta_1)\cos(\beta_2) \\
&\vdots \\
y_{n-1} &= r\sin(\theta)\sin(\beta_1)\cdots\sin(\beta_{n-6})\cos(\beta_{n-5}) \\
y_n &= r\sin(\theta)\sin(\beta_1)\cdots\sin(\beta_{n-6})\sin(\beta_{n-5})
\end{aligned} \tag{11}$$

All angular variables range from 0 to π except for α_3 and β_{n-5} which range from 0 to 2π.

With the notations from the report:

$$\hat{y} = [x_1 \quad x_2 \quad x_3 \quad x_4 \quad 0 \quad \cdots \quad 0] \tag{12}$$

We want to compute the probability that:

$$\frac{|\hat{y}|}{|\omega|} > K \tag{13}$$

This is equivalent to:

$$\begin{aligned}
&\frac{|\hat{y}|}{|\omega|} > K \\
&|\hat{y}|^2 > K^2|\omega|^2 \\
&|\hat{y}|^2 + K^2|\hat{y}|^2 > K^2|\omega|^2 + K^2|\hat{y}|^2 = K^2|y|^2 \\
&\frac{|\hat{y}|^2}{|y|^2} > \frac{K^2}{1+K^2}
\end{aligned} \tag{14}$$

We have:

$$|\hat{y}|^2 = \sum_{k=1}^{4} x_k^2 = r^2\cos^2(\theta) \tag{15}$$

(14) is equivalent to:

$$\cos^2(\theta) > \frac{K^2}{1+K^2} \tag{16}$$

and:

$$\frac{1}{1+K^2} > \sin^2(\theta) \tag{17}$$

The above inequality defines the range of θ for which there would be a loss of integrity event.

It can be shown that the volume element in spherical coordinates is of the form (see below):

$$d^nV = \left|\cos^3(\theta)\right|\sin^{n-5}(\theta)\, f(\alpha,\beta,r)\, d\varphi_1 d\varphi_2 \cdots d\varphi_{n-1} dr \tag{18}$$

The arguments in the integrand are separable, so the total volume is given by:

$$\begin{aligned} V_n &= \int_{\alpha,\beta,r} \int_{\theta=0}^{\pi} d^nV = \int_{\alpha,\beta,r} \int_{\theta=0}^{\pi} \left|\cos^3(\theta)\right|\sin^{n-5}(\theta)\, f(\alpha,\beta,r)\, d\varphi_1 d\varphi_2 \cdots d\varphi_{n-1} dr d\theta \\ &= \left(\int_{\alpha,\beta,r} f(\alpha,\beta,r)\, d\varphi_1 d\varphi_2 \cdots d\varphi_{n-1} dr\right)\left(\int_{\theta=0}^{\pi} \left|\cos^3(\theta)\right|\sin^{n-5}(\theta)\, d\theta\right) \end{aligned} \tag{19}$$

Similarly, the volume corresponding to the loss of integrity is given by:

$$\begin{aligned} V_n(K) &= \int_{\alpha,\beta} \int_{\frac{1}{1+K^2} > \sin^2(\theta)} d^nV \\ &= \left(\int_{\alpha,\beta,r} f(\alpha,\beta,r)\, d\varphi_1 d\varphi_2 \cdots d\varphi_{n-1} dr\right)\left(\int_{\frac{1}{1+K^2} > \sin^2(\theta)} \left|\cos^3(\theta)\right|\sin^{n-5}(\theta)\, d\theta\right) \end{aligned} \tag{20}$$

The probability of loss of integrity is given by:

$$\frac{V_n(K)}{V_n} = \frac{\int_{\frac{1}{1+K^2} > \sin^2(\theta)} \left|\cos^3(\theta)\right|\sin^{n-5}(\theta)\, d\theta}{\int_{\theta=0}^{\pi} \left|\cos^3(\theta)\right|\sin^{n-5}(\theta)\, d\theta} = \frac{\int_{\theta=0}^{\arcsin\left(\sqrt{\frac{1}{1+K^2}}\right)} \cos^3(\theta)\sin^{n-5}(\theta)\, d\theta}{\int_{\theta=0}^{\frac{\pi}{2}} \cos^3(\theta)\sin^{n-5}(\theta)\, d\theta} \tag{21}$$

We have:

$$\begin{aligned} &\int_{\theta=0}^{\theta_{\min}} \cos^3(\theta)\sin^{n-5}(\theta)\, d\theta = \int_{\theta=0}^{\theta_{\min}} \left(1-\sin^2(\theta)\right)\sin^{n-5}(\theta)\cos(\theta)\, d\theta = \\ &\int_{\theta=0}^{\theta_{\min}} \left(\sin^{n-5}(\theta) - \sin^{n-3}(\theta)\right)\cos(\theta)\, d\theta = \left[\frac{\sin^{n-4}(\theta)}{n-4} - \frac{\sin^{n-2}(\theta)}{n-2}\right]_0^{\theta_{\min}} = \frac{\sin^{n-4}(\theta_{\min})}{n-4} - \frac{\sin^{n-2}(\theta_{\min})}{n-2} \end{aligned} \tag{22}$$

Therefore, as in the report:

$$\frac{V_n(K)}{V_n}=\frac{1}{2}\left((n-2)\sin^{n-4}(\theta_{\min})-(n-4)\sin^{n-2}(\theta_{\min})\right)=\frac{1}{2}\left((n-2)\left(1+K^2\right)^{-\frac{n-4}{2}}-(n-4)\left(1+K^2\right)^{-\frac{n-2}{2}}\right)$$
$$=\frac{1}{2}\left((n-2)\left(1+K^2\right)^{-\frac{n-4}{2}}-(n-4)\left(1+K^2\right)^{-\frac{n-2}{2}}\right)=\frac{\left(1+K^2\right)^{1-\frac{n}{2}}}{2}\left(K^2(n-2)+2\right) \tag{23}$$

Volume element of change of coordinates in Equation (18)

We can re-write the change of variable as follows:

$$\begin{aligned} y_{1-4}&=r\cos(\theta)u(\alpha)\\ y_{5-n}&=r\sin(\theta)v(\beta)\end{aligned} \tag{24}$$

The Jacobian of this transformation can be written:

$$J=\begin{bmatrix}\cos(\theta)u(\alpha) & -\sin(\theta)u(\alpha) & r\cos(\theta)\frac{\partial u(\alpha)}{\partial\alpha} & 0\\ \sin(\theta)v(\beta) & \cos(\theta)v(\beta) & 0 & r\sin(\theta)\frac{\partial v(\alpha)}{\partial\alpha}\end{bmatrix}$$
$$=\begin{bmatrix}u(\alpha) & 0 & r\cos(\theta)\frac{\partial u(\alpha)}{\partial\alpha} & 0\\ 0 & v(\beta) & 0 & r\sin(\theta)\frac{\partial v(\alpha)}{\partial\alpha}\end{bmatrix}\begin{bmatrix}\cos(\theta) & -\sin(\theta) & 0\\ \sin(\theta) & \cos(\theta) & 0\\ 0 & 0 & I_{n-2}\end{bmatrix} \tag{25}$$

As a consequence:

$$|J|=$$
$$=\left|\det\begin{bmatrix}u(\alpha) & 0 & r\cos(\theta)\frac{\partial u(\alpha)}{\partial\alpha} & 0\\ 0 & v(\beta) & 0 & r\sin(\theta)\frac{\partial v(\alpha)}{\partial\alpha}\end{bmatrix}\right|\left[\det\begin{bmatrix}\cos(\theta) & -\sin(\theta) & 0\\ \sin(\theta) & \cos(\theta) & 0\\ 0 & 0 & I_{n-2}\end{bmatrix}\right]= \tag{26}$$
$$=\left|\cos^3(\theta)\right|\sin^{n-5}(\theta)\det\begin{bmatrix}u(\alpha) & 0 & r\frac{\partial u(\alpha)}{\partial\alpha} & 0\\ 0 & v(\beta) & 0 & r\frac{\partial v(\alpha)}{\partial\alpha}\end{bmatrix}$$

Example where IBPL breaks down

We build an example that illustrates the weaknesses of IBPL. This example builds on the analysis given in Section 2.4 of the report.

IBPL assumes that faults from line of sights that are close by are completely uncorrelated. This does not seem to be a good assumption for multipath.

Let us consider a geometry with 4 lines of sight. Let us now consider a geometry where three of those are repeated *n* times.

According to IBPL, the PL will correspond to the DOP times a factor, which would go down with increasing *n*, it would therefore provide a finite PL. However, this is a geometry in which RAIM would not be able to provide a finite PL, because a fault in the measurement corresponding to the non -repeated line of sight would be undetectable (the subset corresponding to this fault does not have enough geometry diversity to provide a solution, or equivalently, the slope is infinite).

This example although artificial, is not very far from many situations where many of the lines of sight are clustered together in sets. It also shows that there is no reason IBPL is expected to work well in multipath prone environments, since multipath is somewhat spatially correlated.

Appendix C: Mathematical Discussion of Section 3.3 ("M-RAIM")

C.1 Appendix provided by the Stanford review team

There is a somewhat easier way of demonstrating the HPL defined in section 3.3 by the symmetrical closed-form solution for the subset HPL:

$$HPL_j = \sigma_{d,j}\sqrt{-2\ln(p_{FA,j})} + \sigma_{a,j}\sqrt{-2\ln(p_{MD,j})}$$

In addition, this alternate method shows how to integrate the effect of the nominal biases. (Note that the equation numbers in this appendix continue onward from those in Appendix B.)

We have:

$$\left(\left|\hat{x} - x\right| > L, \left|\hat{x} - \hat{x}^{(j)}\right| \leq t_j\right) \Rightarrow \left|x - \hat{x}^{(j)}\right| \geq L - t_j \qquad (27)$$

where x designates the horizontal position. This is a direct consequence of the triangle inequality:

$$\left|x - \hat{x}^{(i)}\right| + \left|\hat{x} - \hat{x}^{(i)}\right| \geq \left|\hat{x} - x\right| \qquad (28)$$

From this inequality we deduce that:

$$\left(\left|\hat{x} - x\right| > L, \left|\hat{x} - \hat{x}^{(j)}\right| \leq t_j\right) implies \left(\left|x - \hat{x}^{(j)}\right| \geq L - t_j\right) \qquad (29)$$

As a consequence:

$$P\left(\left|\hat{x} - x\right| > L, \left|\hat{x} - \hat{x}^{(j)}\right| \leq t_j\right) \leq P\left(\left|x - \hat{x}^{(j)}\right| \geq L - t_j\right) \qquad (30)$$

Based on the result from p44-45 applied to the subset *j*, we have:

$$P\left(\left|x - \hat{x}^{(j)}\right| \geq L - t_j\right) \leq e^{-\frac{1}{2}\left(\frac{L - t_j}{\sigma_{a,j}}\right)^2} \qquad (31)$$

If we want:

$$P\left(\left|\hat{x}-x\right|>L,\left|\hat{x}-\hat{x}^{(i)}\right|\leq t_j\right)\leq p_{md,j} \tag{32}$$

It is sufficient to ensure that:

$$e^{-\frac{1}{2}\left(\frac{L-t_j}{\sigma_{a,j}}\right)^2}\leq p_{md,j} \text{ and } L-t_j\geq 0 \tag{33}$$

which is equivalent to:

$$L-t_j\geq\sigma_{a,j}\sqrt{-2\ln p_{md,j}} \tag{34}$$

And this inequality shows that the HPL as defined at the bottom of p.52 meets the integrity allocation.

Case with nominal biases

This approach has the advantage of simplifying the case with biases.

We have:

$$\left|x-\hat{x}^{(j)}\right|=\left|\hat{x}^{(j)}-x-b^{(j)}+b^{(j)}\right|\leq\left|\hat{x}^{(j)}-x-b^{(j)}\right|+\left|b^{(j)}\right| \tag{35}$$

where $b^{(j)}$ is the effect of the nominal bias on the horizontal position solution. Following the same approach, we have:

$$\left(\left|x-\hat{x}^{(j)}\right|\geq L-t_j\right)\Rightarrow\left|\hat{x}^{(j)}-x-b^{(j)}\right|+\left|b^{(j)}\right|\geq L-t_j \tag{36}$$

which implies:

$$\left(\left|x-\hat{x}^{(j)}\right|\geq L-t_j\right)\Rightarrow\left|\hat{x}^{(j)}-x-b^{(j)}\right|\geq L-t_j-\overline{b}^{(j)} \tag{37}$$

where $\overline{b}^{(j)}$ is an upper bound on the maximum bias of the position solution caused by the nominal bias (we compute it later).

We therefore have:

$$P\left(\left|x-\hat{x}^{(j)}\right|\geq L-t_j\right)\leq P\left(\left|\hat{x}^{(j)}-x-b^{(j)}\right|\geq L-t_j-\overline{b}^{(j)}\right) \tag{38}$$

The random variable $\hat{x}^{(j)}-x-b^{(j)}$ is a zero mean 2D Gaussian, and again, the previous result applies, so we have:

$$P\left(\left|x-\hat{x}^{(j)}\right| \geq L-t_j\right) \leq e^{-\frac{1}{2}\left(\frac{L-t_j-\left|\bar{b}^{(j)}\right|}{\sigma_{a,j}}\right)^2} \tag{39}$$

And therefore, we can define:

$$HPL_j = t_j + \bar{b}^{(j)} + \sigma_{a,j}\sqrt{-2\ln p_{MD,j}} \tag{40}$$

To ensure that the probability of false alert is met, we need to take into account the biases in the solution separation statistic. Following the same reasoning, we can show that it is sufficient to define t_j as follows:

$$t_j = \sigma_{d,j}\sqrt{-2\ln p_{FA,j}} + \bar{b}_{ss}^{(j)} \tag{41}$$

where $\bar{b}_{ss}^{(j)}$ is an upper bound of the effect of the nominal biases on the solution separation statistic.

We now develop formulas for both $\bar{b}_{ss}^{(j)}$ and $\left|\bar{b}^{(j)}\right|$. We have:

$$\left|b^{(j)}\right| = \sqrt{\left(\sum_{k=1}^{n} K_{1,k}^{(j)} b_k\right)^2 + \left(\sum_{k=1}^{n} K_{2,k}^{(j)} b_k\right)^2} \leq \sqrt{\left(\sum_{k=1}^{n} \left|K_{1,k}^{(j)}\right| \bar{b}_k\right)^2 + \left(\sum_{k=1}^{n} \left|K_{2,k}^{(j)}\right| \bar{b}_k\right)^2} \tag{42}$$

where $\bar{b}_k$ is a known upper bound on the absolute value of the nominal bias affecting measurement k. This is not the smallest upper bound that can be computed, but it is very easy to obtain and probably not too far from the smallest. We can define:

$$\bar{b}^{(j)} = \sqrt{\left(\sum_{k=1}^{n} \left|K_{1,k}^{(j)}\right| \bar{b}_k\right)^2 + \left(\sum_{k=1}^{n} \left|K_{2,k}^{(j)}\right| \bar{b}_k\right)^2} \tag{43}$$

The formula for the bias in the solution separation is very similar:

$$\bar{b}_{ss}^{(j)} = \sqrt{\left(\sum_{k=1}^{n} \left|K_{1,k}^{(j)} - K_{1,k}^{(0)}\right| \bar{b}_k\right)^2 + \left(\sum_{k=1}^{n} \left|K_{2,k}^{(j)} - K_{2,k}^{(0)}\right| \bar{b}_k\right)^2} \tag{44}$$

Case with nominal biases

This approach has the advantage of simplifying the case with biases.
We have:

$$\left|x-\hat{x}^{(j)}\right| = \left|\hat{x}^{(j)} - x - b^{(j)} + b^{(j)}\right| \leq \left|\hat{x}^{(j)} - x - b^{(j)}\right| + \left|b^{(j)}\right| \tag{45}$$

where $b^{(j)}$ is the effect of the nominal bias on the horizontal position solution. Following the same approach, we have:

$$\left(\left|x-\hat{x}^{(j)}\right| \geq L-t_j\right) \Rightarrow \left|\hat{x}^{(j)}-x-b^{(j)}\right|+\left|b^{(j)}\right| \geq L-t_j \tag{46}$$

which implies:

$$\left(\left|x-\hat{x}^{(j)}\right| \geq L-t_j\right) \Rightarrow \left|\hat{x}^{(j)}-x-b^{(j)}\right| \geq L-t_j-\overline{b}^{(j)} \tag{47}$$

where $\overline{b}^{(j)}$ is an upper bound on the maximum bias of the position solution caused by the nominal bias (we compute it later).

We therefore have:

$$P\left(\left|x-\hat{x}^{(j)}\right| \geq L-t_j\right) \leq P\left(\left|\hat{x}^{(j)}-x-b^{(j)}\right| \geq L-t_j-\overline{b}^{(j)}\right) \tag{48}$$

The random variable $\hat{x}^{(j)}-x-b^{(j)}$ is a zero mean 2D Gaussian, and again, the previous result applies, so we have:

$$P\left(\left|x-\hat{x}^{(j)}\right| \geq L-t_j\right) \leq e^{-\frac{1}{2}\left(\frac{L-t_j-\left|\overline{b}^{(j)}\right|}{\sigma_{a,j}}\right)^2} \tag{49}$$

And therefore, we can define:

$$HPL_j = t_j + \overline{b}^{(j)} + \sigma_{a,j}\sqrt{-2\ln p_{MD,j}} \tag{50}$$

To ensure that the probability of false alert is met, we need to take into account the biases in the solution separation statistic. Following the same reasoning, we can show that it is sufficient to define t_j as follows:

$$t_j = \sigma_{d,j}\sqrt{-2\ln p_{FA,j}} + \overline{b}_{ss}^{(j)} \tag{51}$$

where $\overline{b}_{ss}^{(j)}$ is an upper bound of the effect of the nominal biases on the solution separation statistic.

We now develop formulas for both $\overline{b}_{ss}^{(j)}$ and $\left|\overline{b}^{(j)}\right|$. We have:

$$\left|b^{(j)}\right| = \sqrt{\left(\sum_{k=1}^{n} K_{1,k}^{(j)} b_k\right)^2 + \left(\sum_{k=1}^{n} K_{2,k}^{(j)} b_k\right)^2} \leq \sqrt{\left(\sum_{k=1}^{n} \left|K_{1,k}^{(j)}\right| \overline{b}_k\right)^2 + \left(\sum_{k=1}^{n} \left|K_{2,k}^{(j)}\right| \overline{b}_k\right)^2} \tag{52}$$

where $\overline{b}_k$ is a known upper bound on the absolute value of the nominal bias affecting measurement k. This is not the smallest upper bound that can be computed, but it is very easy to obtain and probably not too far from the smallest. We can define:

$$\bar{b}^{(j)} = \sqrt{\left(\sum_{k=1}^{n}\left|K_{1,k}^{(j)}\right|\bar{b}_k\right)^2 + \left(\sum_{k=1}^{n}\left|K_{2,k}^{(j)}\right|\bar{b}_k\right)^2} \tag{53}$$

The formula for the bias in the solution separation is very similar:

$$\bar{b}_{ss}^{(j)} = \sqrt{\left(\sum_{k=1}^{n}\left|K_{1,k}^{(j)} - K_{1,k}^{(0)}\right|\bar{b}_k\right)^2 + \left(\sum_{k=1}^{n}\left|K_{2,k}^{(j)} - K_{2,k}^{(0)}\right|\bar{b}_k\right)^2} \tag{54}$$

It is not clear whether the algorithm presented here offers a performance benefit over the formulation given in the ARAIM reference algorithm. It might be worthwhile performing a comparison of both methods. This method has the benefit of being invariant with respect to the horizontal coordinate system, but it does so at the expense of a potentially conservative bound.

The report says that "if these are too conservative, then the HPL may be inflated too far and the availability of the method may be damaged". From an integrity point of view, the fault probabilities and nominal error models should reflect the amount of knowledge, if these models lead to unavailability, there is little one can do to improve the situation using the same measurements. If that happens, then we must work harder to guarantee smaller probabilities or a smaller nominal error model.

Project Manager
Richard Greaves
Richard.greaves@nlaltd.co.uk

Registered office: Mill House Farm, Blackthorn Hill, Blackthorn, Oxfordshire OX25 1TJ
Registered in England: 10801372