GNSS-based receiver autonomous integrity monitoring for aircraft navigation

Imparato, Davide

DOI
10.4233/uuid:a5e5f5be-b1a6-42c2-b285-a4ff01e5bfd2

Publication date
2016

Document Version
Final published version

Citation (APA)
Imparato, D. (2016). *GNSS-based receiver autonomous integrity monitoring for aircraft navigation*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:a5e5f5be-b1a6-42c2-b285-a4ff01e5bfd2

GNSS-BASED RECEIVER AUTONOMOUS INTEGRITY MONITORING FOR AIRCRAFT NAVIGATION

GNSS-BASED RECEIVER AUTONOMOUS INTEGRITY MONITORING FOR AIRCRAFT NAVIGATION

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op maandag 11 april 2016 om 10:00 uur

door

Davide IMPARATO

Master of Science in Aerospace Engineering,
University of Pisa, Pisa, Italy
geboren te Ortona, Chieti, Italy.

Dit proefschrift is goedgekeurd door de

promotor: Prof. dr. ir. P. J. G. Teunissen
copromotor: Dr. ir. C. C. J. M. Tiberius

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. P.J.G. Teunissen,	Technische Universiteit Delft, promotor
Dr. ir. C.C.J.M. Tiberius,	Technische Universiteit Delft, copromotor
Onafhankelijke leden:	
Prof. dr. B. Pervan,	Illinois Institute of Technology (USA)
Prof. dr. ir. M. Mulder,	Technische Universiteit Delft
Prof. dr. ir. B. van Arem,	Technische Universiteit Delft
Dr. ir. B. Ober,	Integricom

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

To my family

SUMMARY

Nowadays, GNSS-based navigation is moving more and more to critical applications. Global Navigation Satellite Systems (GNSS), which in the past used to be represented by the American GPS and the Russian GLONASS are now growing in number and performance. The European system Galileo and the Chinese system Beidou are being deployed, while GPS and GLONASS are being modernized. The availability of a larger number of satellites to provide measurements, together with a new frequency dedicated to civil use, are strongly increasing the application potential of GNSS technology.

To be used in aviation, in particular during critical phases of flight as approach and landing, satellite navigation shall provide a very high level of service. Correctness — within tight bounds — of the position solution, shall be guaranteed to extremely high levels of probability. In operating an aircraft, the risk for so-called Hazardously Misleading Information (HMI) due to the navigation system is typically budgeted at the 10^{-7} to 10^{-9} level. These extremely tight requirements constitute a guarantee of safety, which is called integrity. More formally, integrity is about the trust that a user can have in the navigation service (and more specifically, the indicated position information). The trust is measured by the probability of HMI (or integrity risk), which is the probability that the position error exceeds a certain tolerance, without being detected and an Alert being raised in time.

Commonly, a distinction is made between *system*-level integrity and *user*-level integrity. At system level, integrity is monitored directly by the GNSS control segment and can be monitored by additional external augmentation systems. At user level, integrity is monitored directly by the user via statistical methods. This dissertation focuses on user-level integrity monitoring, also called Receiver Autonomous Integrity Monitoring (RAIM). In a RAIM method, integrity is monitored by exploiting the redundancy of the GNSS signals as collected at the receiver. Calculations are performed within the user equipment itself to check the measurements' consistency. RAIM computations are possible as long as a number of satellites larger than the minimum necessary for a position fix (four in case of single constellation) is visible.

RAIM algorithms have been investigated since the late 1980s, starting with publications by Lee, Brown and Brenner. As main representative and reference of the first generation RAIM algorithms we cite the Weighted RAIM algorithm, also referred to as Least-Squares-Residuals (LS) RAIM, proposed by Walter and Enge. This algorithm is still in use today, typically implemented in aviation grade GPS receivers, to provide low-precision lateral integrity only. As of today no RAIM implementation exists for any application requiring integrity in the vertical plane (i.e. precision approaches), which has more stringent certification requirements. To serve this scope second generation RAIM algorithms are now being developed and tested, as for instance the Advanced RAIM (ARAIM), proposed by the Stanford group.

Different approaches being around show that the community has not reached convergence on the subject of integrity of GNSS for aviation yet, especially on RAIM. The Least Squares residuals RAIM, was the staple of the first generation RAIM: its algorithm was found to be not completely flawless from a theoretical point of view and is not designed to deal

with a multi-constellation system. The ARAIM is currently being tested but has not reached yet a definitive shape and has not fully convinced the community because of its high computational load, its convoluted structure and supposed approximations. New alternative approaches are also being proposed.

This dissertation offers to the community a critical review of the most popular RAIM algorithms currently available or under development (in particular LS RAIM and ARAIM), and highlights their major strengths and shortcomings. Furthermore it reviews the DIA procedure, a well-established method for gross error detection in geodesy developed by TU Delft, and proposes its application to the RAIM problem. A connection is made from the DIA concept of reliability to integrity risk and a method to evaluate RAIM performance parameters (False Alarm and HMI rates) for a multi-step exclusion/adaptation procedure is proposed (by means of the concept of worst-case bias). The study performed shows the viability of the DIA procedure as an alternative RAIM procedure, and its competitive performance compared to the algorithms currently in use or under development in aviation (LS RAIM and ARAIM). Simulation results show that in several scenarios the DIA method performs significantly better than the others. Points of improvement are nevertheless individuated, also in the DIA, and recommendations are given for the development of the RAIM of the future.

In particular, as a result of the algorithms review and the simulation results, it is concluded that all RAIM algorithms discussed, including the DIA procedure, have room for improvement. Both ARAIM and DIA show safe performance (i.e. risk is never larger than required/announced), but some of the approximations employed in ARAIM seem rather conservative and its exclusion mechanism seems not particularly effective. At the same time also the DIA exclusion mechanism does not appear to perform optimally, from the integrity maximization point of view. From a reliability analysis point of view — i.e. prior computation of the probability of HMI (P_{HMI}) based on the satellite geometry alone — the DIA performs better than ARAIM, i.e. can guarantee higher availability (and higher continuity in a faulty scenario). On the other hand one of the main weaknesses of the RAIM algorithms analyzed is the Exclusion (or Adaptation) mechanism. Both ARAIM and DIA procedures recognize that in many geometries — given a requirement on the continuity — attempting exclusion introduces more risk than just declaring Alert. This suggests that further investigation is required to develop a more robust and reliable exclusion method for integrity.

SAMENVATTING

Tegenwoordig wordt GNSS gebaseerde navigatie meer en meer ingezet voor kritische toepassingen. Global Navigation Satellite Systems (GNSS), in het verleden vertegenwoordigd door het Amerikaanse GPS on het Russische GLONASS, groeien momenteel in aantal en prestatie. Het Europese systeem Galileo en het Chinese system Beidou worden ontwikkelde terwijl GPS en GLONASS gemoderniseerd worden. De beschikbaarheid van een groter aantal satellieten voor metingen, samen met een nieuwe frequentie gewijd aan civiele bebruik, vergroten sterk het toepassings-potentieel van GNSS technologie.

Om gebruikt te kunnen worden in de luchtvaart, en in het bijzonder tijdens kritieke fasen van de vlucht, zoals nadering en landing, dient satelliet-navigatie op zeer hoog niveau dienst te verlenen. De correctheid van de positie-oplossing dient, binnen nauwe grenzen, gegarandeerd te worden met extreem hoge waarschijnlijkheid. In de luchtvaart is de acceptabele waarschijnlijkheid van zogenaamde gevaarlijke misleidende informatie (Hazardously Misleading Information - HMI) ten aanzien van het navigatie-systeem typisch begroot op 10^{-7} tot 10^{-9}. Deze veiligheids-garantie wordt integriteit genoemd. Meer formeel is de integriteit het vertrouwen dat de gebruiker mag hebben in de navigatie-dienst (en meer specifiek, in de gepresenteerde positie-informatie). Het vertrouwen wordt gemeten door de HMI-waarschijnlijkheid (of integriteits-risico), en dit is de waarschijnlijkheid dat de positie-fout een bepaalde grens overschrijd, zonder dat dit door het systeem gedetecteerd wordt en er tijdig een waarschuwing afgegeven wordt.

Gewoonlijk wordt er onderscheid gemaakt tussen integriteit op *systeem* en op *gebruikers* niveau. Op systeem niveau wordt integriteit direct bewaakt door het GNSS controle segment en kan het bewaakt worden door aanvullende externe augmentatie systemen. Op gebruikers-niveau kan integriteit direct bewaakt worden door statistische methoden. Deze dissertatie richt zich op integriteitsbewaking op gebruikers-niveau, ook wel autonome integriteitsbewaking door de ontvanger (Receiver Autonomous Integrity Monitoring - RAIM) genoemd. In een RAIM-methode wordt de integriteit bewaakt door gebruik te maken van de overtalligheid van GNSS metingen, zoals ze door de ontvanger verzameld zijn. Berekeningen worden uitgevoerd op de apparatuur van de gebruiker om de consistentie van de metingen te controleren. RAIM berekeningen zijn mogelijk zolang het aantal zichtbare satellieten groter is dan het minimum van vier dat nodig is voor een positie-oplossing (vier satellieten in geval gebruikt wordt gemaakt van een enkele constellatie).

Onderzoek naar RAIM methoden heeft plaatsgevonden sinds de late jaren '80, beginnend met publicaties door Lee, Brown en Brenner. Als primaire vertegenwoordiging van de eerste generatie RAIM methoden refereren we naar de Weighted RAIM methode door Walter en Enge, ook bekend als Least Squares Residuals (LS) RAIM. Deze methode is vandaag de dag nog steeds in gebruik, en typisch geïmplementeerd in GPS ontvangers voor de luchtvaart, om alleen lage-precisie zijdelingse integriteit te verschaffen. Tot op heden bestaat er geen RAIM implementatie voor toepassingen die integriteit in vertikale zin vereisen, met meer stringente certificerings-eisen. Hiervoor worden tweede generatie RAIM methoden ontwikkeld en getest, zoals bijvoorbeeld het Advanced RAIM (ARAIM), voorgesteld door de Stanford

groep.

De gemeenschap heeft nog geen overeenstemming bereikt op het onderwerp van integriteit van GNSS voor de luchtvaart, in het bijzonder op RAIM. Er circuleren verschilllende methoden voor integriteits-bewaking. De Least Squares residuals RAIM, boegbeeld van de eerste generatie RAIM, is niet volledig waterdicht op theoretisch vlak en is niet ontworpen om met een multi-constellatie systeem om te gaan. ARAIM, zoals momenteel voorgesteld, heeft nog niet haar definitieve vorm bereikt en heeft de gemeenschap niet volledig overtuigd vanwege haar hoge reken-belasting, haar ingewikkelde structuur en veronderstelde benaderingen. Nieuwe alternatieve methoden zijn ook voorgesteld.

Deze dissertatie biedt de gemeenschap een kritische beschouwing van de meest populaire RAIM methoden die momenteel beschikbaar of in ontwikkeling zijn (in het bijzonder LS RAIM en ARAIM), en bespreekt de sterke en zwakke punten. Verder beschouwt deze dissertatie de DIA-procedure, een in de geodesie gevestigde methode (ontwikkeld aan de TU Delft), om grove fouten op te sporen, en wordt de toepassing hiervan op het RAIM probleem behandeld. Het concept van betrouwbaarheid in de DIA procedure wordt gekoppeld aan het integriteits-risico, en een methode wordt gepresenteerd om RAIM prestatie-parameters (onterecht alarm en HMI verhoudingen) te evalueren, door middel van het concept van de meest ongunstige systematische fout, voor het geval van een buitensluiting- en aanpassingsprocedure bestaande uit meerder stappen. De uitgevoerde studie demonstreert de uitvoerbaarheid van de DIA-procedure als een alternatieve aanpak voor RAIM, en haar concurrerende prestaties vergeleken met de methoden die momenteel in gebruik zijn in de luchtvaart (LS RAIM en ARAIM). Simulaties laten zien dat in verschillende scenario's de DIA procedure substantieel beter presteert dan de andere methoden. Mogelijke verbeteringen worden niettemin besproken, ook voor de DIA procedure, en aanbevelingen worden gegeven voor de ontwikkeling van de RAIM methode van de toekomst.

In het bijzonder, als resultaat van de kritische beschouwing en de simulaties, wordt geconcludeerd dat alle besproken RAIM methoden, inclusief de DIA-procedure, verder verbeterd kunnen worden. Zowel ARAIM als de DIA-procedure zijn veilig bevonden (in de zin dat het risico nooit groter is dan vereist/aangekondigd), maar een aantal benaderingen die in ARAIM gebruikt worden lijken behoorlijk conservatief en haar buitensluitingsmechanisme lijkt niet bijzonder effectief. Tegelijkertijd lijkt, vanuit integriteits-oogpunt, ook het buitensluitings-mechanisme van de DIA-procedure niet optimaal te presteren. Vanuit betrouwbaarheids-oogpunt - de a-priori berekening van het integriteits-risico (HMI) op basis van de satelliet-geometrie - presteert de DIA procedure beter dan ARAIM, d.w.z. het kan een hogere beschikbaarheid garanderen (en hogere continuiteit in een scenario met fouten). Het buitensluitings- of aanpassingsmechanisme is één van de belangrijkste zwakke schakels in de geanalyseerde RAIM methoden. Zowel ARAIM als de DIA procedure erkennen dat met veel geometrieën - gegeven een beschikbaarheids-eis - het uitvoeren van een buitensluiting meer risico introduceert dan enkel het afgeven van een waarschuwing. Dit suggereert dat verder onderzoek nodig is om een meer robuuste en betrouwbare buitensluitingsmethode voor integriteit te ontwikkelen.

ACKNOLEDGEMENTS

This thesis would have not been possible without the help and support of numerous people, to whom I am sincerely grateful.

First of all I would like to thank my promotor Peter Teunissen and my copromotor Christian Tiberius. Peter, you taught me a lot with your rigor and scientific integrity. Christian, I will always thank you for your limitless energy and patience in supervising me and providing me your moral support. Your dedication to work and science is an example for me.

I would like to thank Hans van der Marel, Dennis Odijk and Marc Solé, who collaborated with me during my Ph.D. work and consistently helped me with the simulations set up, with the ARAIM review and with solving several programming issues. I would like to thank Peter Buist for the training on Trimble software, Sandra Verhagen for the training on KALDIA software, and Peter de Bakker for all the precious advices on GNSS topics, Matlab and Latex. I would like to thank Tim Vlemmix, with whom I had a very pleasant experience of work, and who I greatly admire for his professional example. I would also like to acknoledge the chairman and the committee members for their role in the defence cerimony.

I would like to thank Barend Lubbers, project mate for the first three years of Ph.D., and Yanqing Hou, my office mate, who both helped me a lot in my work, especially with Matlab. We worked tightly together during the first year of Ph.D. and our collaboration was very fruitful. They contributed to most part of Section 4.1.

I would like to thank my current and previous colleagues, including Pooja Mahapatra, Joana Martins, Karolina Sarna, Pedro Inácio, Olga Didova, Igor Stepanov, Marcel Kleinherenbrink, Monica Herrero Huerta, Julien Chimot, Thomas Frederikse, Prabu Dheenathayalan, Enayat Hosseini Aria, Mahmut Arikan, Sami Samiei Esfahany, Dimitra Mamali, Freek van Leijen, Gert Mulder, Anneleen Oyen, Jiangjun Ran, Alijafar Mousivand, Ramses Molijn, Ling Chang, Rui Sun, Junchao Shi, Vincent Cabot, Ramon Mendez Gomez, Ramon Hanssen, Jingyu Zhang, Sandra Verhagen, Manuel Delgado Blasco, Miguel Caro Cuenca. I had many intellectually stimulating and fruitful discussions with them, especially at the coffee breaks.

Among the other colleagues I would like to mention in particular Yanqing Hou, Siavash Shakeri, Ricardo Reinoso Rondinel, Lorenzo Iannini, Reenu Todesh, to whom I got more attached throughout the years. Thank you for all the help and support you gave me during my time as Ph.D., and for being the great friends that you are.

I would like to thank Lidwien de Jong, Rebecca Domingo, Marjolein de Niet-de Jager and Danko Roozemond for their kind administrative work during my study.

I would like also to thank all my friends in the Netherlands, in Italy or in elsewhere in the world, for their support and understanding during these years. I would like to thank Dasha, for her fundamental support and love during the last year of my Ph.D. Finally I would like to thank my family for their unconditional love and support throughout all my endeavors.

Davide Imparato
March 21, 2016,
Delft

CONTENTS

ACRONYMS

ADS-B	Automatic Dependent Surveillance-Broadcast
AL	Alert Limit
ANOVA	ANalysis Of VAriance
ARAIM	Advanced RAIM
ARNS	Aeronautical Radio-Navigation Service
ATM	Air Traffic Management
BLUE	Best Linear Unbiased Estimation
CA	Correct Alert
CD	Correct Detection
CNS	Communications Navigation and Surveillance
CONUS	CONtinental US
DGPS	Differential GPS
DIA	Detection Identification Adaptation
EGNOS	European Geostationary Navigation Overlay Service
FA	False Alarm
FA'	False Alert
FAA	Federal Aviation Authority
FDE	Fault Detection and Exclusion
FOC	Fully Operational Constellation
FTE	Flight Technical Error
GAGAN	GPS Aided Geo Augmented Navigation
GBAS	Ground Based Augmentation System
GIC	GNSS Integrity Channel
GLR	Generalized Likelihood Ratio

GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GPWS	Ground Proximity Warning System
HAL	Horizontal Alert Limit
HDT	High Dynamic Threats
HMI	Hazardous Misleading Information
HPL	Horizontal protection Level
ICAO	International Civil Aviation Organization
ICAO ANC	ICAO Air Navigation Conference
ILS	Instrument Landing System
LDT	Low Dynamic Threats
LNAV	Lateral NAVigation
LPV	Localizer Performance with Vertical guidance
LSD	Least Significant Difference
MC	Multiple Comparisons
MCDF	Multi-Constellation Dual Frequency
MDA	Minimum Descent Altitude
MD	Missed Detection
MHSS	Multiple Hypothesis Solution Separation
MOPS	Minimum Operational Performance Standards
MSAS	Multi-functional Satellite Augmentation System
NO	Nominal Operation
PBN	Performance Based Navigation
PDF	Probability Distribution Function
PE	Positioning Error
PF	Positioning Failure
PL	Protection Level
PMF	Probability Mass Function

RAIM	Receiver Autonomous Integrity Monitoring
RF	Radio-Frequency
RRAIM	Relative RAIM
RTCA	Radio Technical Committee for Aeronautics
SBAS	Space Based Augmentation System
SDCM	System for Differential Corrections and Monitoring
SIS	Signal in Space
SoL	Safety of Life
SS	Solution Separation
SV	Space Vehicle
TTA	Time-to-Alert
UMPI	Uniformly Most Powerful Invariant
URA	User Range Accuracy
UTC	Coordinated Universal Time
VAL	Vertical Alert Limit
VPL	Vertical Protection Level
WAAS	Wide Area Augmentation System
WD	Wrong Detection
WE	Wrong Exclusion

SYMBOLS

Symbol	Units	Description
a	–	Availability of the navigation system, Equations (2.7), (2.8)
A	–	Geometry matrix, defining the relationship between observations and unknown parameters, Equation (2.2)
c	–	Continuity, Equation (2.6)
c_0	–	Requirement on continuity, Equation (2.14)
C_y	–	Matrix defining the anomaly characteristic/signature, Equation (2.34)
$\underline{e}$	m	Error vector ($\underline{e} = \underline{y} - E(\underline{y})$), Equation (2.2)
H_0	–	Null hypothesis (statistical hypothesis testing), Equations (3.15), (3.21)
H_a	–	(General) alternative hypothesis (statistical hypothesis testing), Equations (3.15), (3.21)
H_i	–	i^{th} alternative hypothesis (statistical hypothesis testing), Equations (2.37), (3.15)
I	–	Identity matrix
k	–	Test threshold (general), Equation (3.25)
K	–	Test critical region (general), Equation (3.17)
$N(\mu, Q)$	–	(Multivariate) Normal distribution with mean μ and variance matrix Q
p_0, p_i	–	Prior probabilities of occurrence of hypotheses H_0, H_i, Equation (2.37)
P_A	–	Projector (matrix) onto the vector space $R(A)$ (defined by the suffix), $P_A = A(A^T W A)^{-1} A^T W$, Equation (4.3)
$P_A^{\perp}$	–	Projector (matrix) onto the vector space perpendicular to $R(A)$ (defined by the suffix), $P_A^{\perp} = I - A(A^T W A)^{-1} A^T W$, Equation (5.1)
P_{FA}	–	Probability of False Alarm (statistical hypothesis testing), Equation (3.18)
$P_{FA_{tot}}$	–	Total probability of False Alarm (statistical hypothesis testing), Equation (2.29)
$P_{FA'}$	–	Probability of False Alert, Equation (2.28)
$P_{FA'}^{req}$	–	Requirement on the Probability of False Alert, Table 7.3

P_{HMI}	–	Probability of Hazardous Misleading Information, Equation (2.26) (general definition), Equation (2.27) (RAIM definition)
P_{HMI}^{req}	–	Requirement on the P_{HMI}, Equation (2.12)
P_{MD}	–	Probability of Missed Detection (statistical hypothesis testing), Equation (3.19)
P_{PF}	–	Probability of Positioning Failure, Equation (2.24)
P_{WD}	–	Probability of Wrong Detection (statistical hypothesis testing), Equation (5.35)
Q	m^2	Covariance matrix (variable in subscript)
S	–	BLUE matrix $S = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$, Equation (4.42)
$\underline{t}$	m	Misclosures vector, Section 6.6.3
$\underline{T}$	–	Test statistic (general), Equation (3.17)
$\underline{T}_q$	–	UMPI test statistic, Equation (3.22)
$\underline{w}$	–	w-test statistic, Equation (5.5)
x	m	Vector of the true unknown parameters of interest (e.g. in the GNSS case user position and clock offset), Equations (2.1) and (2.2)
$\underline{y}$	m	Observable vector (e.g. in the GNSS case measurements from GNSS satellites), Equations (2.1) and (2.2)
α	–	Probability of False Alarm, $\alpha \equiv P_{FA}$ (statistical hypothesis testing), Equation (3.18)
β	–	Probability of Missed Detection $\beta \equiv P_{MD}$ (statistical hypothesis testing), Equation (3.19)
$\chi^2(m,\lambda)$	–	χ^2 distribution with m degrees of freedom and non-centrality parameter λ
φ	rad	Angle between bias vector and the detection space, $R(A)^{\perp}$, Equation (6.10)
γ	–	Detection power of a test (statistical hypothesis testing), Equation (3.20)
Γ	–	Wrong Detection region (DIA procedure), Equation (5.36)
λ	–	Non-centrality parameter of the chi^2 UMPI test statistic distribution under an alternative hypothesis, measure of the internal reliability (TU Delft testing theory), Equation (4.7)
λ_x	–	Normalized squared bias in the position domain, measure of the external reliability (TU Delft testing theory), Equation (F.1)
λ_y	–	Normalized squared measurement bias (TU Delft testing theory), Equation (F.2)
Λ	–	Superset of Γ, Equation (5.40)
κ	–	Probability of Wrong Detection $\kappa \equiv P_{WD}$ (statistical hypothesis testing), Equation (5.35)

ρ	m	(Phase-smoothed) pseudorange, Equation (2.39)
ζ	–	Probability of simultaneous rejection of OMT and acceptance of all w-tests (in the DIA procedure), Figure 5.14
Ψ	–	Superset of Γ, Equation (5.40)
Ω	m	Acceptance region for the measurements y of a RAIM algorithm, Equation (2.11)
Ω_{AL}	m	Acceptance region for the Positioning Error $\hat{x} - x$, Equation (2.12)
$\Omega_T, \Omega_k, \Omega_w$	m	Acceptance region for the test statistics (in subscript), e.g. Equation (3.30)
Ω_x	m	Acceptance region for the position estimator $\hat{x}$, Section 5.1.6
∇	m	Vector quantifying the size of the biases affecting the measurements in case of anomaly, Equation (2.34)
$\|v\|_M$	–	Norm of a vector in the metric defined by the positive definite matrix M, defined as $\|v\|_M = \sqrt{v^T M v}$.

List of Figures

LIST OF TABLES

1

INTRODUCTION

The topic of this dissertation is GNSS-based Receiver Autonomous Integrity Monitoring (RAIM), with specific focus to its aircraft navigation application. But what is RAIM? What is integrity of navigation? How does Global Navigation Satellite System (GNSS) based navigation apply to aircraft navigation? In this Chapter we present the background of this work, its reasons and objectives, and the contribution that through this work is given to the community.

We first describe the main GNSS applications in aviation, and we provide an overview of the recent progress in GNSS. Following we introduce the concept of integrity of navigation (that will be further developed in Chapter 2) and the concepts of integrity monitoring systems, present and under design. Finally we discuss the objectives, and the original contribution of this research to the scientific community.

1.1. THE ROLE OF GNSS IN AVIATION

The use and acceptance of Global Navigation Satellite Systems (GNSS) in aviation has — so far — not been a smooth and straight ride. During the early years of GPS, with the so-called block I satellites, launched from 1978 to 1985, the capability of satellite navigation was successfully demonstrated. In the eighties and nineties of last century, there was a lot of activity on the subject of integrity (reliability/soundness), most of it, connected to the use of satellite navigation in aviation. Then, GPS became fully operational (FOC with 24 satellites) in 1995. There was an enormous, sky-reaching believe in the potential of satellite navigation. The USA FAA stated, in 1996, that all air navigations' aids, current at that time, would be phased out by 2010. Only GPS and augmented GPS systems were to be provided for civil aviation by the government after 2010. As we know today, this has not become reality.

During the nineties there was mentioning and demonstration of possibilities of interference to GPS signals. This materialized eventually in the final Volpe report on vulnerability, in 2001 [106]. A major finding was that potential interference to the GPS L1 civil signal was a serious problem. This was a 'big-blow' to ambitious plans, the phase of disillusionment. At the same time, however, technological progress was made. In 2000, Europe announced the plan for Galileo, a European civilian multi-frequency satellite navigation system. A wide-band GPS signal on the L5 frequency was specified (to be the third GPS frequency), and the design of the second frequency signal for civilian GPS users was concluded. The status of today is that GPS has a more than nominal constellation, with modernization and additional signals

underway, Galileo is under deployment with 8 satellites already flying, the Russian Glonass is again fully operational, the Chinese Compass-Beidou system has already 17 satellites in orbit, and the Space Based Augmentation Systems (SBAS) in the USA and in Europe (WAAS and EGNOS respectively) are operational.

These recent developments provide *diversity* in system control, with many satellites (over 100 expected in several years from now), and *diversity* in radio frequency spectrum, with signals in the range from 1.1 to 1.6 GHz (some of them wide-band), occupying several tens of MHz and mitigating the earlier concerns on risks and interference. We are now making steady progress towards the plateau of productivity.

GNSS is nowadays a key technology for the Communications, Navigation and Surveillance (CNS) infrastructure, necessary for the introduction of Performance Based Navigation (PBN) and Automatic Dependent Surveillance-Broadcast (ADS-B). It is employed in safety related systems such as GPWS (Ground Proximity Warning Systems), and provides the time reference that is used to synchronise many systems and operations in Air Traffic Management (ATM). About 90% of the fleet operating in Europe is already equipped with GPS receivers and it is expected to reach 100% before 2020.

The paper [87], of 2008, presents the EUROCONTROL policy on GNSS for navigation applications in the civil aviation domain, with implications to the SESAR implementation phase. The policy is based on a gradually increasing reliance on satellite navigation, based on a Multi-Constellation, and Multi-Frequency GNSS. A similar policy has been put in place already by the FAA in the USA, with the Next Generation Air Transportation System (NextGen) program [30] “see where we are going!”.

1.1.1. AIR TRAFFIC MANAGEMENT

Presentation [74] gives the view of Boeing on future CNS/ATM (Communications, Navigation and Surveillance/Air Traffic Management) and evolution towards 4D trajectory management, with a substantial role for GNSS positioning. Determination of position and time — 4D positioning — is a cornerstone for future ATM developments. ATM moves from what is essentially a space-based system (primarily based on centralized radar-surveillance), to a time-based system. Aircraft would be required to ‘turn up’ at a very precise departure and arrival time for use of the runway. The overall goal is to improve airspace (and runaway) capacity, safety and efficiency, and reduce environmental impact.

Automated decision support tools are envisioned to be used in future ATM systems, with the purpose of providing safe (and efficient) services in increasingly congested skies [83]. Decision support tools should provide conflict-free flight trajectories in real time, which must be accurately adhered to by each aircraft to maintain sufficient separation with surrounding traffic.

Airborne autonomous operations will rely — in terms of technical resources — on satellite-based 4D navigation, and on data-link communication. Next, an important element consists of trajectory prediction, which is used in conflict resolution [83]. Crucial input to trajectiory prediction and conflict resolution, is the current aircraft position, or more general, the current aircraft *state*, with the highest possible quality (best possible state estimate, together with an adequate description, or statement of the quality).

1.1.2. APPROACH AND LANDING

Integrity is particularly important during the approach phase, which is the most demanding phase of flight. The associated requirements on the navigation system are extremely strict during this phase. Navigation must be available more than 99% of the time regardless of the weather, and the navigation system must be reliable especially after an aircraft approach has commenced. A break in the continuity of service must affect less than one approach in 100 000. At the same time, integrity has to be guaranteed, i.e. the system has to guard against the occurrence of Hazardous Misleading Information (HMI). This means that possible failures which may lead to dangerous navigation errors must be detected within few seconds (referred to as Time-to-Alert, TTA). The risk of a HMI — an integrity failure — must be less than 10^{-7} to 10^{-9} per approach (depending on the minimum decision height). Figure 1.1 shows the different types of approach procedures with their respective Minimum Descent Altitudes (MDA) and Alert Limits (AL). The MDA is a specific altitude below which descent must not be made without the required visual reference, whereas the ALs (Vertical or Horizontal, VAL and HAL) define the maximum position errors that are not classified as dangerous for navigation. Currently, as from the Figure, the simple GPS is able to support only Lateral Navigation (LNAV) approaches, whereas more demanding approaches require aid from additional systems (e.g. SBAS, see Section 1.5).

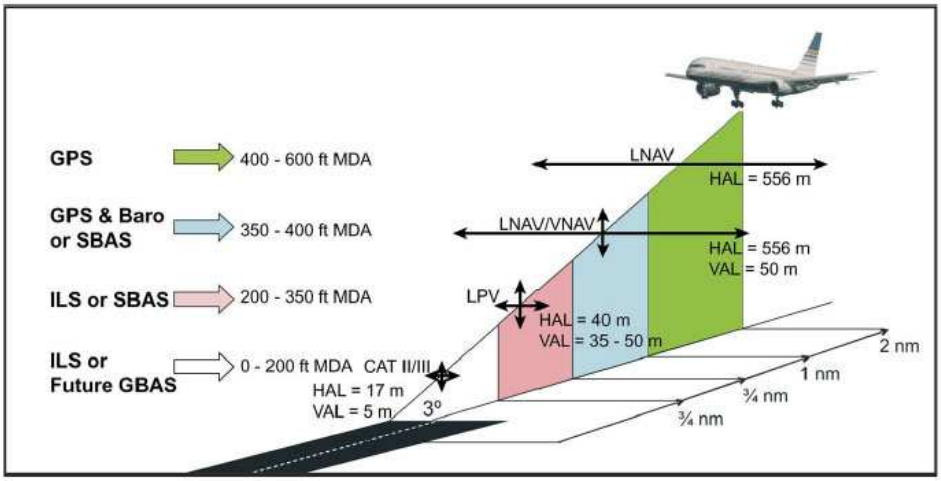


Figure 1.1: Approach procedures. On the left are the navigation systems employed ordered by decreasing MDA. On the right the ALs are reported for the different type of approaches, also decreasing with decreasing MDA. Figure taken from [110].

LNAV approach refers to approach procedures where the radio equipment provides lateral guidance only and the vertical information comes from barometric altimetry. The LNAV approaches are characterized by 'drive' and 'dive' procedures, during which the aircraft reduces its altitude step by step as it approaches the airport. These procedures are not favoured by the pilots because of the quite high workload required. Precise vertical guidance presents a much more manageable workload to the pilot and thus is significantly safer.

Worldwide vertically guided approach is an abiding goal of the aviation community. We would like to fly an aircraft down to an altitude of 70m anywhere in the world regardless of weather and time of day, and without the use of any dedicated navigation equipment at the receiving airfield. At TU Delft we believe that satellite navigation, with the modernized

1

GNSSs, will achieve this goal.

1.2. GNSS CONTEXT

GNSS capability has been growing very rapidly in recent times. In particular, two main factors are leading the progress of GNSS: 1. the deployment of new constellations and 2. the introduction of new frequency signals. The European system, Galileo, has now 8 satellites in orbit available and functioning, sufficient to generate a position fix in the European area, whereas the Chinese system, Beidou, deployed 17 satellites already and is expected to be fully operational in few years time. At the same time the GPS is being upgraded and the Russian Glonass constellation has reached fully operational status again.

The resulting large number of usable GNSS signals can strongly increase the availability of the GNSS service and consequently, as we will see, its integrity. In fact the integrity of the positioning service is strongly related to the redundancy of observations available.

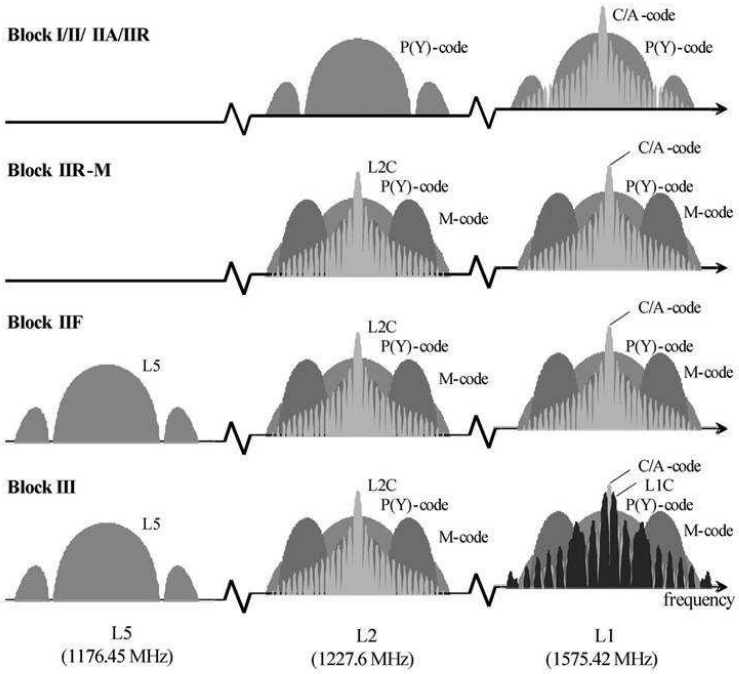


Figure 1.2: GPS signals: present and future. For each generation of satellite GPS satellites (on each row/trace) the broadcast signals are represented, for each of the frequencies L5, L2 and L1. The Block IIR-M satellites started to broadcast in 2005, the Block IIF in 2009, also in the new frequency L5. The first Block III satellite is planned to be launched in 2017. Figure taken from [110].

Up to 2005, GPS satellites broadcast the navigation signals in the two frequency bands shown in the top trace of Figure 1.2. L1 denotes the broadcast at 1575.42 MHz and L2 denotes the broadcast at 1227.60 MHz. As shown, two signals are broadcast on frequency L1, a narrow-band signal and a wide-band signal. The narrow-band signal is modulated by a spread-spectrum code called C/A code. This code has a modest chipping rate of 1.023 Mcps, and consequently a null-to-null bandwidth of 2.046 MHz. The wide-band signal is modulated by the P(Y) code, which has a chipping rate of 10.23 Mcps and a null-to-null bandwith

of 20.46 MHz. The C/A code is available to everyone and it is at the basis of the majority of today's civil applications [40], [93]. The P(Y) code is instead primarily for US military use.

Starting from 2005, new GPS satellites began to broadcast the signals shown on the second trace of Figure 1.2 [31]. They continued to send the 'old' signals, but introduced as well new military sugnals at L1 and L2 (M-code and P(Y) code) and a civil signal at L2. This new signal however has not much importance for aviation since it does not lie in the Aeronautical Radio-Navigation Service (ARNS) portion of the spectrum. All civil aviation organizations demand that the signals employed be in ARNS bands so that they have institutional control over this spectrum and maintain legal protection from interference [40].

In 2009 new GPS satellites were deployed able to broadcast signals in a third band, called L5 [40], [104] (L3 and L4 carry non-navigation information for the military). L5 is located in an ARNS band and has therefore great aviation utility. The future GNSS based avionics will be able to leverage a useful property: the influence of ionosphere on signal delay is different at L1 than at L5. Thanks to this property receivers will be able to estimate the full ionospheric delay at each frequency and remove it from the measurements. The errors due to the ionosphere could be for the most part eliminated.

Integrity monitoring architectures will have great benefit from the double frequency: Space Based Augmentation Systems (SBASs) for instance will need much fewer reference stations on the ground, since many of them are now needed to sample and monitor the ionosphere.

The ICAO (International Civil Aviation Organization) ANC (Air Navigation Conference) 12th held in Montreal in November 2012 highlighted that, with the deployment of new constellations and the enhancement of the existing ones, the signals from multi-constellations and multiple frequency bands are becoming available to aviation; these development leads to performance improvement and create potential for achieving significant operational benefits. The conference stressed the need to identify these operational benefits, to enable air navigation providers and aircraft operators to quantify these benefits for their specific operational environment. It also noted that Multi-Constellation, Dual Frequency (MCDF) GNSS brings with it a number of technical and regulatory challenges, beyond the ones associated with current GNSS implementations [88].

Taking into account the operational benefits of MCDF GNSS and the progressive dependency on GNSS for ATM/CNS applications, EUROCONTROL policy on GNSS and the European ATM Master plan set a vision based on the use of signals coming from at least two constellations in dual frequency, that will provide improved performance, robustness and coverage. The SESAR project 15.3.4 made an assessment of the MCDF GNSS technical capabilities on the basis of performance simulations and the analysis of robustness against identified vulnerabilities (ionosphere and interference). The project also assessed the benefits of integrating GNSS sensors with inertial systems. This assessment [88] [44] identified the following improvements in technical capabilities:

- Increased availability and continuity in nominal conditions: as mentioned, GNSS will be able to calculate the ionospheric delay in real time, eliminating this major error source. It will be less likely that scintillation will result in loss of service, thanks to the extra ranging sources available. More and better signals will be available in mountainous terrain/high latitude, making less likely that high terrain or lack of satellites in view would result in loss of service.

- Increased availability and continuity in degraded conditions (increased robustness): MCDF GNSS will guarantee improved robustness and integrity thanks to the increased number of backup modes available in degraded modes (alternative frequency and additional satellites from different constellations). New GNSS signals will be more resistant to interference due to higher power and improved signal design.
- Extended service area: in particular, it will be easier to extend augmentation systems (as EGNOS) services to new areas.
- Independent time reference system: the availability of GNSS timing service will improve, making aviation less dependent on GPS for time distribution and synchronization in an increasing number of CNS/ATM systems and applications (e.g. data-link, ADS-B, terrestrial communication systems, 4D NAV, ...).

As a result of the above listed capability enhancements, the GNSS will be able to support more demanding system performance levels required by new applications or advanced operations that are expected to be developed over time. For example EGNOS V3 is being designed to guarantee system perforance levels with a Vertical Alert Limit (VAL) of 10m, that could be used to certify CAT I auto-land systems.

1.3. INTEGRITY FOR GNSS IN AVIATION

Satellite navigation in aviation is moving more and more to critical applications. We will focus here onwards on CNS/GNC aspects (with less concern to ATM, considering anyway that most concepts are easily transferrable). The application of GNSS started with basic R-NAV, moved into SBAS approaches, is about to move into CAT-II/III GBAS precision approaches, and may ultimately reach auto-land with zero visibility. Thereby the aspect of integrity is getting more and more crucial. But what is integrity exactly?

To be used in aviation, in particular during critical phases of flight as approach and landing, satellite navigation shall provide a very high level of service. Correctness — within tight bounds — of the position solution, shall be guaranteed to extremely high levels of probability. In operating an aircraft, the risk for so-called Hazardously Misleading Information (HMI) due to the navigation system is typically budgeted at the 10^{-7} to 10^{-9} level, as we mentioned in the context of approach operations. Practically, integrity means a guarantee of safety.

More formally, integrity is about the trust that a user can have in the navigation service (and more specifically, the indicated position information). The trust is measured by the probability of HMI (or integrity risk), which is the probability that the position error exceeds a certain tolerance, without being detected and an alert being raised in time. The user should have been warned but is not. The given position information is misleading, as it is not correct within specified bounds, and, as the user or operator is not warned, he or she is not aware of the potentially hazardous situation. The aircraft is somewhere else, than where the navigation system says it is.

The approach in [79] is such that the integrity risk is calculated for a given tolerance, the Alert Limit (AL). The AL is the maximum allowed position deviation, for which no alert needs to be raised. For Safety-of-Life (SoL) applications as aviation, we quote from [79], as an example, the Galileo SoL core system performance requirements (without receiver contribution) with respect to integrity: a risk of $2 \cdot 10^{-7}$ in any 150 seconds, with an AL of 12m for the horizontal, and 20m for the vertical component.

In practice, a navigation system cannot be optimized independently for integrity. Integrity is not a free parameter; it is linked to other navigation performance parameters as accuracy, continuity and availability. Typically, reducing complexity of the problem by making approximations on the safe side on the aspect of integrity, i.e. being conservative with regard to safety, will cost availability of the navigation system/service. This calls for a safe, optimal, but lean approach to integrity of the position solution.

The quality of the position estimator, comprising accuracy and integrity, is one aspect of navigating the aircraft safely from A to B. In consequently flying a designated trajectory, two more aspects come into play. They concern how well a 'to be flown' trajectory can be actually followed, and the intended/planned flight path description/definition itself.

The reference, or desired trajectory can also be corrupted with errors and anomalies (path definition error). One can think of administrative errors (as naming and labelling), in-accuracies of data in the database (also as a result of surveying errors), lack of up-to-date information (such as changes to airport infrastructure and procedural aspects which are not reflected in the database yet), and mis-interpretations of geodetic datums.

The Flight Technical Error (FTE) represents the additional error to, or deviation from, the reference trajectory (additional to the navigation system error), due to the process of physically flying the aircraft under operational circumstances. Due to external circumstances (as wind and turbulence) and aircraft performance, the pilot cannot keep the aircraft exactly on the reference trajectory. This aspect becomes increasingly important as reference trajectories will get more complicated. Flight paths now basically consisting of straight line segments connected by fixed radius turns will be replaced more and more by more sophisticated curved segments.

A formal definition of integrity will be given in Chapter 2. In particular, the problem of single epoch independent integrity monitoring will be addressed in the following and constitute the main body of this work. The last two aspects of integrity in aeronautics mentioned above, trajectory monitoring and FTE, are not part of the scope of this thesis.

1.4. GNSS ANOMALIES

Even though the operation of GPS has been very reliable during its whole life and showed extraordinary performance (sometimes even beyond specification), faults have nevertheless occurred, some man-made and other due to Mother Nature.

For instance, the navigation data broadcast by the GPS satellites may contain significant errors. The GPS satellites are monitored by the ground control network: the measurements taken at the ground control stations are used to predict the orbit of the satellites. These predictions are then uploaded to the satellites and broadcast from the satellites to the user. The estimated orbit is generally accurate up to 1 or 2m [112], but occasionally the broadcast ephemeris may contain rather large errors. Between 1999 and 2007, errors greater than 50m were registered in 24 different occasions [37]. On April 10, 2007, a large outlier occurred in the broadcast ephemeris from Space Vehicle (SV) 54, with an error of at least 350 m [37]. In Delft a sensible anomaly was registered on January 1^{st}, 2004, when the range error from SV 23 grew up to more than 250 km in about two hours, before being detected by the control segment and set to unusable. Figure 1.3 shows the position error registered at a Delft station, which started to grow out of control from about 18:30 Coordinated Universal Time (UTC).

The navigation broadcast from each satellite contains also an estimate of the time offset

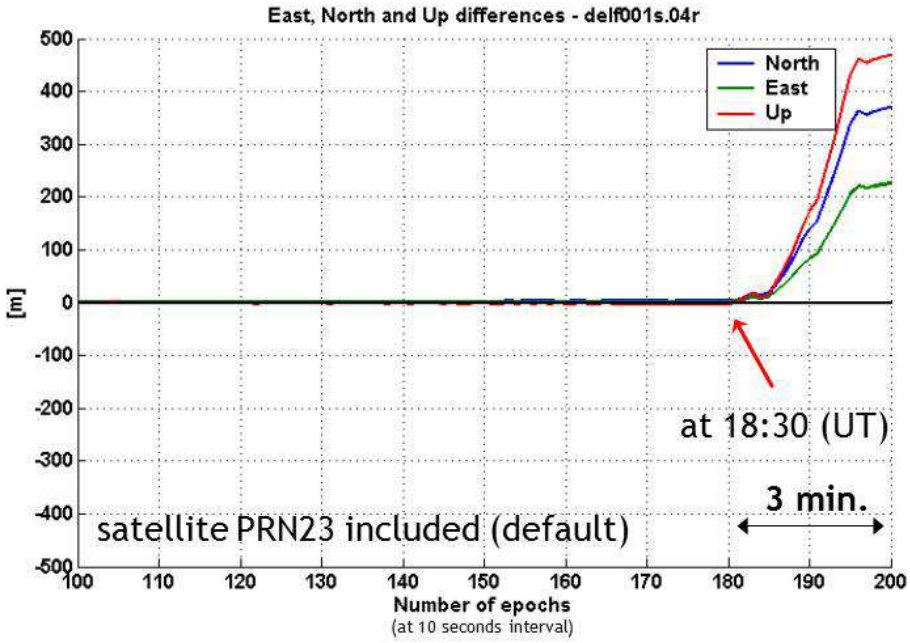


Figure 1.3: GPS anomaly recorded in Delft on January 1^{st}, 2004. Position error in meters as a function of time. In normal situations the position accuracy was about 5 to 10 meters, but the error started to grow suddenly at about 18:30 UTC, reaching the size of hundreds of meters in few minutes.

of the onboard atomic clock with respect to the GPS system time. This estimate is normally accurate to within nanoseconds. Some large clock runoffs were anyway experienced on SV22 on July 28, 2001; SV27 on May 26, 2003; and SV35 on June 11, 2003. These events produced range measurements errors larger than 1 km [29].

A problem with the modulation of the broadcast signal from SV19 was recorded in Fall 1993. The falling edge of the digital modulation was not synchronous with the satellite master clock, and was occurring approximately 30ns later. This lag caused ranging errors of about 3m and position errors of up to 9m. This anomaly was tackled by switching from the active modulation unit to the backup unit available on all GPS satellites [25]. Such peculiar anomaly has been observed only once in the GPS operational life, but it constitutes a clear example of the challenge associated with ensuring integrity at the required service levels.

Nature by itself is sometimes responsible to generate dangerous anomalies in the measurements. The most worrisome source of anomalies is the ionosphere. The residual delay introduced by the ionosphere in the signals amounts nominally to a few meters during the day and about 1m at night. The spatial and temporal variation of this delay is normally quite smooth and easily managed. However, in case of an ionospheric storms the delay in the signal propagation becomes much larger and spatial and temporal gradients are tough to manage. Ionospheric storms do occur in every solar cycle — Datta-Barua lists about 40 significant events in the last solar cycle [22]. Not all of these events resulted in navigation threats, but all must be evaluated in real-time to guarantee the required system integrity.

The introduction of a second frequency available for aviation will obviate most of the ionospheric effects described above, but still second-order ionospheric delays will be present and may require monitoring in specific applications (though unlikely to exceed the size of

some centimeters).

1.5. GNSS INTEGRITY MONITORING SYSTEMS

Integrity of the position solution is concerned with errors that can be due to the user receiver (failure or anomaly), to the system (Signal in Space, SIS), its augmentation, and to signal propagation. Errors and anomalies can thus occur at various stages, and this offers several possibilities of detecting them (such as via the control segment of the satellite navigation system, its augmentation, or the user receiver). Commonly, a distinction is made between *system* level integrity and *user* level integrity. At system level, integrity is monitored directly by the GNSS control segment and can be monitored by additional external augmentation systems. At user level, integrity is monitored directly by the user via statistical algorithms.

Furthermore we can distinguish between a first and a second generation of fault-detection systems [110]: the first generation constituted by the methods developed by the year 2000, currently in use, and the second generation constituted by the future concepts, still under study/development, that will exploit the full functionality of a MCDF GNSS.

The GNSS ground segments (control segments) have themselves a role in the integrity monitoring of the system. They are in fact enrolled to detect failures in satellites, but they are not intended to identify threats within a TTA of few seconds as foreseen in aviation requirements. In [10] for instance a distinction is made between threats that require a TTA of 6s (labeled as High Dynamic Threats, HDT) and threats that do not (Low Dynamic Threats, LDT). The ground segment can thus take responsibility for mitigating part or most of LDTs (more details on types of anomalies are given in Section 2.9).

1.5.1. FIRST GENERATION INTEGRITY MONITORING — PRESENT ARCHITECTURES

Today, in 2015, fault-detection is implemented in two different ways, at system level or at user level. At system level, we distinguish two types of external augmentation systems, Space-Based Augmentation Systems (SBAS) [108], [56] and Ground-Based Augmentation Systems (GBAS) [28], [73]. Both are Differential GPS systems (DGPS).

One category of external augmentation systems is represented by the SBAS. SBASs are constituted by a ground segment and by a space segment. The ground segment is made up of a system of monitoring stations distributed across the area of interest, a master control station and a set of uplink stations, whereas the space segment is constituted by geostationary satellites. The monitoring stations measure the range errors in the satellites observations, which are then processed by the master control station to produce the integrity message. This corrective message is then transmitted by the uplink stations to the geostationary satellites, which role is to re-transmit it to the user. The range errors can be computed because the positions of the monitoring stations are well known, and can be compared to the computed position solutions based on GNSS signals. From these, a corrective message can be generated and forwarded to the user. The SBAS generates a four-dimensional correction for each satellite, which accounts for satellite clock and satellite ephemeris. Furthermore SBAS sends a grid of corrections for ionospheric errors for the region spanned by the SBAS ground system. Since the signal sent by the geostationary satellite is synchronized to the GPS time, it can also augment the normal suite of GPS ranging measurements, beside serving as data link.

Several SBASs have been implemented based on GPS. In particular:

- The US Wide Area Augmentation System (WAAS), covering the continental US (CONUS) and most part of Alaska, Canada and Mexico.
- The European GNSS Navigation Overlay System (EGNOS), covering the European continent, providing integrity and improved accuracy.
- The Multi-functional Satellite Augmentation System (MSAS), implemented by Japan.
- The GPS Aided Geo Augmented Navigation (GAGAN), implemented by India.
- The System for Differential Corrections and Monitoring (SDCM), implemented by Russia, a WAAS-compatible SBAS covering their territory.

A second category of external augmentation systems is represented by the GBAS. These systems are simply constituted by a set of ground monitoring stations, located close to the airports, that transmit correction signals directly to the aircraft via RF links. Again the corrective information is obtained comparing the true distances of the satellites (computable exploiting the known position of the ground stations) with the ranges obtained from the GNSS observations. The GBAS corrections and error bounds are broadcast to the approaching aircraft using a line-of-sight very high-frequency transmitter that is also located on the airport property. The data link is a terrestrial radio.

SBAS and GBAS develop corrections that improve the accuracy of the measurements, but their true purpose is to provide the means to generate real-time error bounds. These bounds are called Protection Levels (PLs) and must overbound the actual error under all conditions and in real-time to a very high probability [23]. SBAS and GBAS are both very powerful means of guaranteeing integrity, and compared to RAIM they are much less sensitive to the strength of the basic GNSS geometry, but they present the drawback of needing a very complex and costly infrastructure. The requirement for warning the user of any loss of integrity within a very little time demands a high amount of computing power and facilities. The reference network must be installed, tested, operated and maintained. The SBAS networks are dense because they must sample the ionosphere at closely spaced intervals to guarantee the detection of sharp gradients with near certainty. The WAAS indeed deploys 38 receivers across North America, EGNOS 39 receivers across Europe and North Africa. Furthermore, both SBAS and GBAS have to transmit the integrity information in high-bandwidth and the ground-to-air data latency must be less than a few seconds [72].

Finally the GNSS integrity can be monitored at user level, by exploiting the redundancy of the GNSS signals as collected at the receiver. This is done by performing calculations within the user equipment itself to check their consistency. This method is called Receiver Autonomous Integrity Monitoring (RAIM). RAIM computations are possible as long as a number of satellites larger than the minimum necessary for a position fix (4 in case of single constellation) is visible, and this occurs most of the time at all latitudes. If one constellation of satellites is employed for positioning, five satellites in view are sufficient to detect a possible anomaly, and six satellites are sufficient to identify/isolate a fault in a single satellite. RAIM is a powerful method because of its autonomy, but strictly relies on the strength of the satellite geometry. With the deployment of the new GNSS constellation many more satellite signals will soon be available and increased redundancy of measurements will be at hand: this will

increase the power of RAIM algorithms — second generation RAIM in particular is designed to exploit the full potential of MCDF GNSS.

1.5.2. SECOND GENERATION INTEGRITY MONITORING — FUTURE CONCEPTS

With second generation integrity monitoring we mean the integrity monitoring concepts currently under investigation, as candidates for the integrity system of the future. Three different architectures are currently investigated [110] [109]: the GNSS Integrity Channel (GIC), Relative RAIM (RRAIM) and absolute RAIM (RAIM or ARAIM[1]). All these architectures employ a ground network that monitors the satellite signals to identify faults and support the integrity monitoring, but they place different fractions of the integrity burden on the aircraft and on the external network.

The GIC architecture is similar to the SBAS, and places all the integrity monitoring responsibility on means external to the aircraft, the space and ground network. One form of GIC could be a worldwide implementation of dual frequency SBAS, with the broadcast to the aircraft of similar capacity (bandwidth of 250 bps), or could have a lighter implementation, broadcasting only a single user range accuracy (URA) per satellite.

The absolute RAIM is instead the heir of current RAIM, and places the minimal burden to external monitors. Differently from standard RAIM though, a ground system is needed to provide a priori failure probabilities for the individual satellites and the associated URA (by means of the so called Integrity Support Message, ISM). This information anyway needs to be updated only every hour or even less frequently. Beside relying on this a priori information provided from the ground, the absolute RAIM will use the same concept of RAIM and will exploit the multiplicity of constellations and the dual frequency available. The main representative of absolute RAIM at the moment is the Advanced RAIM (ARAIM), designed by the Stanford group and currently in testing phase [10]. This dissertation dedicates half of Chapter 4 to the ARAIM review.

RRAIM is a solution that stands in between GIC and absolute RAIM, since it splits the integrity burden between the aircraft and the external monitors [110] [59] [36]. In this method the aircraft uses past carrier smoothed code measurements, previously validated by ground monitors, and projects them forward in time by adding to them the difference between current and past carrier phase measurements. These projections are the new position fixes, and their integrity is ensured by RRAIM, that protects against any anomaly that may occur after the last externally validated data set. Fundamentally RRAIM works as a RAIM algorithm applied on the relative positioning, carried on by means of carrier phase measurements, which is based on reference positions validated through external monitoring. The advantage of RRAIM with respect to GIC is that the latency time of the integrity information coming from the external network can be much longer, of the order of tens of seconds or even minutes, since the main part of the integrity monitoring is carried out internally through carrier phase measurements checks. With respect to absolute RAIM, the advantage lies in the fact that carrier phase measurements are much more precise than code measurements, and this can allow to set very tight detection thresholds for the statistical tests run and consequently reduce the number of false alert and lead to high levels of integrity availability. RRAIM geometry requirements are also relaxed relative to absolute RAIM and the latency time for the external

[1] We will simply refer to it as RAIM because the acronym ARAIM will be used for Advanced RAIM in the following.

1

support message lies in between GIC and RAIM.

1.6. RESEARCH OBJECTIVES

1.6.1. WHAT IS MISSING

Whereas system-level integrity is documented in the MOPS DO-229 document [72], with detailed specification of algorithm and parameter values, and briefly reviewed above, user-level integrity is *not*. According to [72] the equipment shall have a Fault Detection and Exclusion (FDE) capability that uses redundant GPS and SBAS ranging measurements to provide independent integrity monitoring (for en-route and terminal mode). Requirements are put forward in [72], and its Appendix K — Fault Detection and Exclusion references — provides just a list with 9 papers, and nothing more.

Furthermore, let us consider the concept of the Protection Level (PL). The Protection Level (PL) is key to the integrity concept as employed today in SBAS. In earlier versions of the MOPS, the PLs were set based on certain assumed statistical distributions (for the position error) [80]. In the current version [72], the underlying statistical distributions have been abolished, and SBAS service providers must transmit quality indicators with the corrective information, such that the PLs bound the errors within the target probabilities. Responsibility is thus transferred to the service provider, and no direction or guideline is given on how to achieve this.

As previously mentioned, and also outlined in [105], Galileo was at one point planning to use an approach to integrity different from SBAS, namely computation of the overall probability of HMI at the AL, rather than computing the PL for a fixed integrity risk. Different approaches being around show that the community has not reached convergence on the subject of integrity of GNSS for aviation yet, especially on RAIM. The LS RAIM developed in [107], staple of the first generation RAIM, was found to be not completely flawless from a theoretical point of view (see Section 4.1) and is not designed to deal with a multi-constellation system. The ARAIM [8] [11] is currently being tested but has not reached yet a definitive shape and has not fully convinced the community because of its high computational load, its convoluted structure and supposed approximations. New approaches are being proposed, for instance in [50] and [52].

1.6.2. OBJECTIVES

As explained in the previous Section, the MOPS DO-229 document [72] is not definitive in providing methods to perform RAIM, and refers to FDE procedures which constitute only one of the possible approaches to monitor integrity. We believe it is opportune to clarify the mathematical concepts of integrity monitoring and RAIM and provide a sound definition for the latter. The RAIM problem shall be placed in its statistical context, and the tools of statistics be examined and evaluated around the problem.

Once the mathematical problem is defined, the most relevant current and future RAIM concepts are to be reviewed and compared in a systematic way. The FDE procedure developed by the Delft school, the DIA procedure, is the most standard method for gross error detection employed in geodesy. It is natural to wonder how the DIA procedure would perform in a RAIM algorithm, in comparison with the other developed methods.

This work focuses on RAIM. This means that only user level integrity monitoring is analyzed. RAIM is generally less powerful than GBAS and SBAS (GIC), but has the advantage of

not requiring any external aid from ground or space infrastructure, only a contained computing power at the user level. Different integrity concepts, as GIC and RRAIM, are also left aside. Furthermore, RAIM shall be analyzed only in its standard single epoch implementation; smoothing or filtering could possibly bring advancement to the current autonomous integrity monitoring concepts, but they are left aside of the scope of this work.

The most relevant RAIM concepts available in literature shall be reviewed and discussed (in particular Standard LS RAIM and ARAIM). The DIA itselft shall be reviewed and adapted accordingly to the RAIM application requirements; in particular its reliability concept shall be extended to allow the monitoring of the specific RAIM performance parameters. The different RAIM algorithms shall be implemented on software and their performance assessed through numerical simulations. Eventual shortcomings in any of the algorithms presented shall be individuated and opportunities for improvement discussed.

1.7. NEW CONTRIBUTIONS

This dissertation offers to the community a critical review of the most popular RAIM algorithms currently available or under development and highlights their major strengths and shortcomings. The mathematical problem of RAIM is defined and the tools available in statistics to tackle it are reviewed: the RAIM algorithms can as a result be compared on the basis of univocally defined integrity performance parameters.

Next this dissertation proposes the application of the DIA procedure — the reknown method for gross error detection in geodesy developed by TU Delft — to the RAIM problem. A connection is made from the DIA concept of reliability to the integrity risk and a method to evaluate the RAIM performance parameters for a multi-step exclusion/adaptation procedure is proposed (by means of the concept of worst-case bias).

The DIA procedure is then compared to the RAIM methods currently in use (LS RAIM and ARAIM). The study performed shows the viability of the DIA method as a RAIM procedure, and highlights the shortcomings and strengths of the other methods. Simulation results show that in several scenarios the DIA method performs significantly better than the other algorithms. Points of improvement are nevertheless individuated also in the DIA procedure, and recommendations are given for the development of the RAIM of the future.

The DIA procedure was not applied before to the RAIM problem and this step required development of methods to monitor new performance parameters, like the risk of Wrong Detection and the risk of HMI linked to a multiple iterations procedure. A method to test and compare the different RAIM algorithms was also proposed together with a set of parameters to monitor to assess their performance.

1.8. GUIDE FOR READING

In Chapter 2 we attempt to define the RAIM problem, which does not necessarily apply to aircraft navigation only but potentially to any estimation problem. The definitions given by the RTCA are also presented, as well as the GNSS observation model. Chapter 3 presents the possible approaches to the RAIM problem and the general statistical methods available in literature, from a high level perspective. In Chapter 4 the most important RAIM algorithms currently used or in testing phase are reviewed — the standard RAIM as main representative of the first generation RAIM and the ARAIM as a forerunner of the second generation RAIM. Chapter 5 presents instead the TU Delft DIA procedure, the original TU Delft FDE procedure

1

first developed to monitor observations in geodetic networks, and Chapter 6 proposes an adaptation of the DIA procedure to the RAIM problem. Finally in Chapter 7 the algorithms so far reviewed/proposed are tested and compared via numerical simulations. The conclusions are gathered in Chapter 8.

2

THE INTEGRITY PROBLEM

In this chapter we intend to define the problem of integrity monitoring for GNSS based navigation. Most of the concepts that will be presented apply in fact to any type of parameter inference, not just GNSS-based nor navigation oriented. Navigation is based in the first place on parameter estimation, and integrity, as we will see, can refer to any type of parameter estimation/inference.

2.1. NAVIGATION SYSTEM AND PARAMETERS ESTIMATION

A navigation system is any system that provides an operator with information about a vehicle's actual position in a given geometric and time reference frame [78]. The navigation system is enrolled to collect and process some measurements or input data and to deliver a position/state estimation. Based on some input data, that we call observations, the parameters of interest are estimated. The model for this estimation problem can be written as:

$$\underline{y} = G(x) + \underline{e} \tag{2.1}$$

where $\underline{y}$ is a vector of m observables, $G(\cdot)$ is a generic function of the argument, x is the vector (n components) of all the parameters on which the observables depend, among which are the parameters of interest, and $\underline{e}$ is a vector of measurement errors. Collection of observations is in fact a random process, subject to errors, this is the reason why $\underline{y}$ and $\underline{e}$ are random variables.

In the GNSS case, the observables $\underline{y}$ are constituted by the range measurements from each visible satellite. These measurements are a function of the user position, which can be defined by three components x_1, x_2 and x_3, but also of other unknown parameters, as for instance the clock error at the receiver, which can be included in the vector x as additional components. In the GNSS estimation problem therefore we will always estimate the position of the user, which is of interest for navigation, but also 'indirectly' some extra parameters which the observables depend upon.

Since this dissertation is concerned with integrity monitoring, we will mostly focus on the subvector of x that contains only the parameters of interest for integrity (i.e. the position coordinates, the first three components), that we call x_{int}. A simple linear relation can be written that relates x_{int} to x: $x_{\text{int}} = L_{\text{int}}x$ with $L_{\text{int}} = [I_3 \quad O_{3\times(n-3)}]$, where I_3 is the 3×3 identity matrix and $O_{3\times(n-3)}$ is a $3\times(n-3)$ null matrix (assuming, for the aviation case,

that vertical guidance is needed). Since, with this rule, it is easy to transform any requirement/relation/constraint on x_{int} in the corresponding one on x, in the following we will always refer to the general x, bearing in mind that in most cases our interest is focused just on a subset of its components. Furthermore in the following we will always adopt an East-North-Up reference frame, and x_1, x_2 will always be the East and North components respectively, whereas x_3 will be the Up component.

The estimator of x is denoted in the following by $\underline{\hat{x}}$, and is usually a function of the observable $\underline{y}$, $\underline{\hat{x}} = F(\underline{y})$. As we will see in detail in the following, the choice for the function F has a strong influence on the integrity.

In our GNSS case, the model for the estimation problem (Equation (2.1)) is slightly nonlinear, but it is standard practice to transform it into a linear one, of the form:

$$\underline{y} = Ax + \underline{e} \tag{2.2}$$

where the $m \times n$ matrix A can be determined by the geometrical configuration of the satellites in view, and is referred to as the geometry matrix. The GNSS model is treated in more detail in Section 2.8.

2.2. NAVIGATION SYSTEM PERFORMANCE

This section describes how the performance of a navigation system can be expressed. ICAO and RTCA [61] defined four Required Navigation Performance (RNP) parameters, accuracy, integrity, continuity and availability, that are standardly used to describe a navigation system's performance. The definitions of these parameters, based on [61], are reported in the following.

ACCURACY

The accuracy defines how well the estimated or measured position agrees with the true position. For GNSS, accuracy can be defined separately for horizontal and vertical dimension with the two quantiles $l_{hor,0.95}$ and $l_{ver,0.95}$ respectively, such that 95% of the position solutions should present a smaller position error. Accuracy is computed assuming that the system is working in *fault-free* conditions, with standard performance.

$$\begin{aligned} l_{hor,0.95} &= \arg\min_{l_{hor}} P(\sqrt{(\underline{\hat{x}}_1 - x_1)^2 + (\underline{\hat{x}}_2 - x_2)^2} < l_{hor} | \text{No Fault}) \geq 0.95 \\ l_{ver,0.95} &= \arg\min_{l_{ver}} P(|\underline{\hat{x}}_3 - x_3| < l_{ver} | \text{No Fault}) \geq 0.95 \end{aligned} \tag{2.3}$$

INTEGRITY

Integrity defines the level of trust that can be given to the system (accounting for all possible/anticipated system states). This means that the system should provide timely warnings (within the Time-to-Alert, TTA, see Section 2.6) if it cannot be trusted anymore. The probability of Hazardous Misleading Information P_{HMI} (that can be also generally referred to as integrity risk) can be defined as:

$$P_{\text{HMI}} = P(HMI) = P(\underline{\hat{x}} - x \notin \Omega_{\text{AL}} \cap \text{No Alert}) \tag{2.4}$$

where Ω_{AL} is the 'integrity region' around the true position whose boundaries are the Alert Limits (AL), which with the current standard requirements has a cilindric shape. An error

lying outside Ω_{AL} is considered dangerous/not safe for navigation. The ALs are set as requirements for safe navigation by the authorities and differ for each phase of flight. ALs and Ω_{AL} are discussed further in Section 2.6. The integrity (or integrity confidence) is the complement to 1 of the probability of hazardous information (P_{HMI}):

$$Int = 1 - P_{\text{HMI}} \tag{2.5}$$

In the following we will work only with the P_{HMI} as main parameter defining the integrity of navigation.

CONTINUITY

The continuity of a system is the ability of the system to finish an operation, e.g. an approach, without interruption, once it has commenced. This means the system guarantees to stay available during the entire operation duration. Continuity is expressed as the probability that during a certain time interval (e.g. 15 second for a Cat.I approach) the system is providing trustworthy navigation information, assuming it was available at the beginning.

$$c = P(\bigcap_{i=1}^{m}(\text{No Alert at } i \mid \text{System Available at } i = 1) \tag{2.6}$$

where i is the sample epoch considered and m is the number of epochs in the desired time interval, e.g. 15 sec.

AVAILABILITY

Availability is the fraction of time the navigation function is usable, as determined by its compliance with accuracy, integrity and continuity requirements, *before* an approach is initiated:

$$a = \frac{t_{\text{System Available}}}{t_{\text{total}}} \tag{2.7}$$

equivalent to the (frequentist) definition of the probability that the system is available at any given moment/epoch:

$$a = P(\text{System Available}) \tag{2.8}$$

where:

$$\text{System Available} : l_{hor,0.95} \leq req_{a1} \;\;\&\;\; l_{ver,0.95} \leq req_{a2} \;\;\&\;\; P_{\text{HMI}} \leq req_{HMI} \;\;\&\;\; c \leq req_c \tag{2.9}$$

assuming that also vertical guidance is needed. The determination of system availability is based on satellite geometry alone. In a single epoch scenario the system is either available or not available, therefore:

$$a_{se} = \begin{cases} 1 & \text{System Available} \\ 0 & \text{System Not Available} \end{cases} \tag{2.10}$$

2.3. INTEGRITY MONITORING SYSTEMS AND RAIM

A navigation system can include/be backed by an integrity monitoring system. As mentioned in Chapter 1, different types of integrity monitoring systems exist, in particular external (SBAS, GBAS) and internal (RAIM) systems, at user level. The integrity monitoring system

processes input data to determine the health state of the navigation system itself. It can be constituted by extra hardware that provides data extra than that employed for navigation (the case of external system, GBAS and SBAS), or can simply be an algorithm that processes the same observations used for navigation (RAIM).

This dissertation will focus on the latter integrity monitoring system, RAIM. A RAIM algorithm exploits the redundancy in the set of available measurements: once a mathematical model describing the relations between measurements and unknown parameters is assumed, it checks, statistically, the consistency between measurements on one hand, and the model on the other; for instance, with m measurements and n unknown parameters in the model of Equation (2.2), and $m \geq n$, one can (through parameter elimination) formulate $m - n$ conditions which the measurements have to fulfil (or, which the expectations of the observables have to fulfil, to be more precise).

2.4. RAIM PROBLEM DEFINITION

Now that the main navigation performance parameters have been defined, we can introduce the problem of RAIM, i.e. the integrity monitoring that is based only on the observations and does not require any external aid, from ground or space based networks. Specifically, a single epoch snap-shot scenario is considered, so that each epoch can be treated per se and no filtering or smoothing is added to possibly improve the integrity (beside the standard Hatch carrier phase smoothing [39], see Section 2.8.2).

2.4.1. BASIC FORMULATION

Given is a single epoch scenario in which a user at position x (unknown) receives signals from the GNSS satellites, which geometrical configuration with respect to the user position is described by the matrix $A \in R^{m \times n}$ (introduced in Section 2.1)[1]. The observables received by the user are identified by the random variable $\underline{y} \in R^m$, where m is the number of observations.

The RAIM problem is fundamentally an estimation problem, and can be defined in first run as searching for a estimator $\underline{\hat{x}} = F(\underline{y})$ which lies within some bounds around the true position x with a certain probability (larger than the requirements). That is, find:

$$\underline{\hat{x}} = F(\underline{y})$$

such that:

$$P(\underline{\hat{x}} - x \notin \Omega_{\mathrm{AL}}) \leq req \quad \forall x$$

where Ω_{AL} is the integrity region around the true position whose bounds are the Alert Limits (defined by the navigation requirements). This function $F(\underline{y})$ may not exist for all the satellite geometries.

On the other hand, the navigation requirements do not state that such estimation has to be provided necessarily 100% of the time: a requirement on continuity establishes as well the percentage of time that the position estimate must be provided to the user, whilst the remaining time a warning can be provided not to trust the navigation system. This means

[1] Equivalently, in a general estimation problem, the observable $\underline{y}$ can depend upon the unknown parameters through a possibly non-linear relation as in Equation (2.1), in which case the single epoch scenario is defined by $G(x)$.

that after the observation has been taken, it can be checked to determine whether it is safe to trust the navigation system or not, and then provide (or not) a position estimate to the user. A selection can thus be made among the possible observations, based on their coherency, that labels subsets of the observation domain as either trustworthy or not trustworthy. This is anyway possible only when some redundancy of measurements is present.

Therefore, the RAIM problem is finally defined as: for any possibly occurring satellite geometry, to which corresponds a certain statistical distribution of the observable $\underline{y}$ (dependent on x), find an 'acceptance' region $\Omega \in R^m$ (subdomain of R^m) *and* an estimation/detection function $F(\underline{y})$ that to the observable $\underline{y} \in \Omega$ assigns a position estimator $\underline{\hat{x}}$:

$$\underline{y} \in \Omega \rightarrow \underline{\hat{x}} = F(\underline{y}) \tag{2.11}$$

such that:

$$P(\underline{\hat{x}} - x \notin \Omega_{\mathrm{AL}} \cap \underline{y} \in \Omega) = P_{\mathrm{HMI}} \leq P_{\mathrm{HMI}}^{req} \quad \forall x \tag{2.12}$$

and

$$P(\underline{y} \notin \Omega) = P_{FA'} \leq P_{FA'}^{req} \quad \forall x \tag{2.13}$$

where:

- $P_{FA'}^{req}$ is the requirement of False Alert probability, the probability that an Alert is raised by the algorithm and the continuity of the operation is interrupted, without any actual reason. It holds:

  $$P_{FA'}^{req} < 1 - c_0 \tag{2.14}$$

  where c_0 is the continutity requirement per epoch. This means that $P_{FA'}^{req}$ is a sub-allocation of the full continuity requirement $1 - c_0$, which has to account also for *justified* Alert. See also Section 2.5.3 for FA' and $P_{FA'}$ definitions.

- Ω_{AL} is the 'integrity region' around the true position which boundaries are the Alert Limits (AL), which with the current standard requirements has a cilindric shape. In fact two bounds are normally set for the position error, one for the horizontal dimension (defining a circle in the horizontal plane) and the other for the vertical dimension. Fundamentally the position error is required to lie within the boundaries defined by the Alert Limits (therefore inside Ω_{AL}) with an extremely high probability, $1 - P_{\mathrm{HMI}}^{req}$. Note that the region in which the error is required to lie with determined probability $1 - P_{\mathrm{HMI}}^{req}$ is in fact the associated confidence region, therefore Ω_{AL} sets an external bound to the confidence region of the position error (accounting for all possible system states).

- P_{HMI}^{req}, the (maximum allowed) Probability of Hazardous Misleading Information P_{HMI}, is the integrity requirement per epoch or per geometry. This is the probability that the information on the aircraft position provided from the system is wrong by at least a certain amount considered dangerous for the navigation, without any alert or warning on possibly present anomaly being provided along.

The acceptance region Ω fundamentally represents the set of all the measurements y (samples of $\underline{y}$) on which base is possible to determine a safe position estimate $\hat{x}$, i.e. for which the requirement on the P_{HMI} is satisfied.

To be found is therefore the function $\underline{\hat{x}}(\underline{y}) = F(\underline{y})$ that represents the safe position estimator satisfying the integrity requirements, together with the acceptance region Ω, whatever is the actual position x (at least among all plausible possibilities). The distribution of $\underline{y}$ is generally a function of the true position x.

Suppose for now that $\underline{y}$ is fully defined as a random function of x, for instance:

$$\underline{y} = G(x) + \underline{e} \quad \text{cf. Equation (2.1)}$$

where $G(x)$ is a deterministic function of x and $\underline{e}$ is a random variable with known/assumed distribution. Then $\underline{\hat{x}} = F(G(x) + \underline{e})$ is a function of x as well, which distribution can be determined (apart from the unknown x). The quantities in Equations (2.13) and (2.12) can be computed, generally as function of the actual x still, therefore the requirements must be fulfilled by any x. Appendix A provides an explanation of the requirements on integrity and continuity assuming a Bayesian point of view (distribution of observable fully known).

Since $P(\underline{\hat{x}} - x \notin \Omega_{\text{AL}} \cap \underline{y} \in \Omega) = P(\underline{y} \in \Omega) \times P(\underline{\hat{x}} - x \notin \Omega_{\text{AL}} | \underline{y} \in \Omega)$, the integrity requirement in Equation (2.12) could be written as $P(\underline{\hat{x}} - x \notin \Omega_{\text{AL}} | \underline{y} \in \Omega) \leq k$. The first proposed formulation is most generally adopted in literature (as standard formulation of aviation requirement), nevertheless it is possible to work also with this second formulation:

$$P(\underline{\hat{x}} - x \notin \Omega_{\text{AL}} | \underline{y} \in \Omega) \leq k = P^{req}_{\text{HMI}|a} \quad \forall x \tag{2.15}$$

which may allow easier manipulation of the probability distribution functions. Here $P^{req}_{\text{HMI}|a} = P^{req}_{\text{HMI}} / a$ is the probability of hazardous misleading information given (conditioned) the measurements have been accepted as trustworthy.

It is necessary to set the continuity requirement in Equation (2.13) because the requirement on integrity in Equation (2.12) can be satisfied very easily choosing as acceptance region Ω the null set. On the other hand, if it is possible to find a rule $\underline{y} \rightarrow \underline{\hat{x}}$ such that $P(\underline{\hat{x}} - x \notin \Omega_{\text{AL}}) \leq P^{req}_{\text{HMI}}$ for all $\underline{y} \in R^m$, this would mean extending Ω to the whole domain R^m, and it would be of course optimal since the continuity is maximized ($P(\underline{y} \in \Omega) = 1$) . This leads to a simplified definition of the problem, as considered at the beginning of this section, but also means setting a maximum requirement on the continuity, i.e. $P^{req}_{FA'} = 1$ ($P^{req}_{FA'}$ is a sub-allocation of c_0). Therefore for standard continuity requirements, in case $\underline{y} \notin \Omega$ the system would not provide an $\underline{\hat{x}}$ but declare Alert.

There exist geometries for which finding such a rule is impossible, and $\underline{\hat{x}}$ simply cannot be provided, since for instance $P(\underline{y} \notin \Omega) > P^{req}_{FA'}$ when the requirement on P_{HMI} is satisfied (or viceversa). For these geometries the integrity is not available.

In any geometry we can optimize the rule in many different ways. The two extreme approaches would be:

1. Minimizing the P_{HMI} given the requirement on the continuity is satisfied, i.e.:

$$\begin{aligned} &\underline{\hat{x}}(\underline{y} \in \Omega), \Omega = \arg\min_{\underline{\tilde{x}}(\underline{y} \in \tilde{\Omega}), \tilde{\Omega}} P(\underline{\tilde{x}} - x \notin \Omega_{\text{AL}} \cap \underline{y} \in \tilde{\Omega}) \\ &\text{with: } P(\underline{y} \notin \tilde{\Omega}) \leq P^{req}_{FA'} \quad \forall x \end{aligned} \tag{2.16}$$

2. Viceversa minimizing the $P_{FA'}$ (maximizing the continuity c) given the requirement on the P_{HMI} is satisfied, i.e.:

$$\begin{aligned} &\underline{\hat{x}}(\underline{y} \in \Omega), \Omega = \arg\min_{\underline{\tilde{x}}(\underline{y} \notin \tilde{\Omega}), \tilde{\Omega}} P(\underline{y} \in \tilde{\Omega}) \\ &\text{with: } P(\underline{\tilde{x}} - x \notin \Omega_{\text{AL}} \cap \underline{y} \in \tilde{\Omega}) \leq P_{\text{HMI}}^{req} \quad \forall x \end{aligned} \tag{2.17}$$

In general anyway it is possible to minimize (or maximize) a combination of the two monitored quantities, for instance:

$$\begin{aligned} &\underline{\hat{x}}(\underline{y} \in \Omega), \Omega = \arg\min_{\underline{\tilde{x}}(\underline{y} \in \tilde{\Omega}), \tilde{\Omega}} (c_1 P(\underline{\tilde{x}} - x \notin \Omega_{\text{AL}} \cap \underline{y} \in \tilde{\Omega}) + c_2 P(\underline{y} \notin \tilde{\Omega})) \\ &\text{with: } P(\underline{y} \notin \tilde{\Omega}) \leq P_{FA'}^{req}, \; P(\underline{\tilde{x}} - x \notin \Omega_{\text{AL}} \cap \underline{y} \in \tilde{\Omega}) \leq P_{\text{HMI}}^{req} \quad \forall x \end{aligned} \tag{2.18}$$

where c_1 and c_2 are two weighting constants ($c_1 \geq 0, \; c_2 \geq 0$).

A further constraint has still to be added to the RAIM definition: the Time-to-Alert (TTA) constraint. In fact, as from the definition of integrity given in Section 2.2, the integrity monitoring system is required to provide *timely* warnings to the user in case of unavailability, i.e. within the TTA time constraint. This means the RAIM system as well is supposed to carry out its estimator function and satisfy Equations (2.12) and (2.13) *within the TTA*.

An Alert can consequently be defined as:

$$\text{Alert} : y \notin \Omega \tag{2.19}$$

under the assumption that all the computation on y required by the RAIM algorithm can be carried out within the TTA. Note that we distinguish an Alert from an Alarm because we use the latter to refer to the rejection of the null hypothesis in a hypothesis testing contest (see Section 3.6.1).

2.5. RAIM INPUT, OUTPUT AND PERFORMANCE PARAMETERS

Following the definition of the RAIM problem given in the previous Section, we summarize in this Section input, output and performance parameters of any RAIM algorithm. Figure 2.1 presents a schematic representation of a RAIM algorithm.

2.5.1. INPUT PARAMETERS

As from the definition given in previous Section, we can highlight the following set up input for the RAIM algorithm (necessary to define the algorithm equations):

- $P_{FA'}^{req}$, the False Alert requirement per geometry.
- Ω_{AL}, the integrity region delimited by the ALs.
- P_{HMI}^{req}, the maximum allowed P_{HMI} per sample.

The above parameters are sufficient to define the algorithm; next, the input to the algorithm at each epoch is constituted by:

- A, the geometry matrix;
- y, the observation vector;
- $f_{\underline{y}}$, the distribution function of the observable — partial knowledge; in fact this input may differ among the different RAIM algorithms, and each algorithm may make different assumption on this distribution.

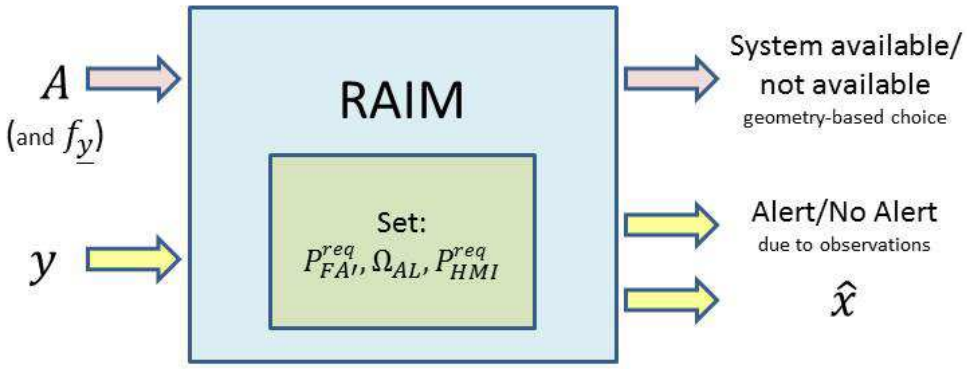


Figure 2.1: RAIM scheme. An Alert can thus be due to the weakness of the geometry itself, i.e. when the estimation rule cannot satisfy Equations (2.12) and (2.13), or to the fact that the measurements taken do not lie in the algorithm acceptance region, $y \notin \Omega$.

2.5.2. OUTPUT PARAMETERS

Outputs of the RAIM algorithm are, at each epoch:

- Alert/No Alert, binary decision whether integrity is guaranteed or not;
- $\hat{x}$, the position estimate (to be used only if no Alert is declared);
- (optional) P_{HMI} or PLs, the Protection Levels.

The raising of an Alert can be due to two different reasons:

1. The satellite geometry is itself too weak: in this case the Alert occurs before even taking the actual measurements, so only the input A is employed. Generally the knowledge of the geometry A and the (partial) knowledge of the $f_{\underline{y}}$ allows the RAIM algorithm to compute the (upperbound to the) P_{HMI} (or the PLs) for the epoch considered, and if this computed P_{HMI} is larger than the maximum allowed an Alert should be raised irrespectively of the measurements.

2. The satellite geometry can guarantee integrity but the sample measurement taken is too inconsistent: this is actually the case $y \in \Omega$ introduced in the definition of RAIM problem. The measurement falls in the subset of measurement that cannot be trusted to guarantee integrity, and no repair/adaptation is possible/foreseen.

On the base of the output of the RAIM algorithm compared with the actual state of nature (actual x), over a long span of time, it is possible to determine the performance of a RAIM algorithm. The performance parameters are discussed in next Section.

2.5.3. PERFORMANCE PARAMETERS

From the comparison of the output (Alert/No alert and $\hat{x}$) with the actual position x (not possible in reality, but possible in simulations or experiments), the following events can be distinguished:

- Positioning Failure (PF), that is defined:

$$PF \equiv \hat{x} - x \notin \Omega_{\text{AL}} \tag{2.20}$$

- Correct Alert (CA), in case of simultaneous Alert and PF (justified Alert):

$$CA \equiv \text{ Alert } \cap PF \equiv y \notin \Omega \cap \hat{x} - x \notin \Omega_{\text{AL}} \tag{2.21}$$

- False Alert (FA'), in case an Alert is not justified by a PF:

$$FA' \equiv \text{ Alert } \cap \text{ No } PF \equiv y \notin \Omega \cap \hat{x} - x \in \Omega_{\text{AL}} \tag{2.22}$$

  FA' is used to indicate a False Alert, because FA will be instead used to indicate a False *Alarm*.

- Hazardous Misleading Information (HMI), in case a PF occurs but is not detected with an Alert:

$$HMI \equiv PF \cap \text{ No Alert } \equiv \hat{x} - x \notin \Omega_{\text{AL}} \cap y \notin \Omega \tag{2.23}$$

As a result the following performance parameters can be computed:

- P_{PF}, the probability of Positioning Failure (PF):

$$P_{PF} = P(PF) = P(\underline{\hat{x}} - x \notin \Omega_{\text{AL}}) \tag{2.24}$$

  If we refer to the RAIM definition in Equations (2.11) to (2.12), we can see that the integrity requirement for the estimation function $\underline{\hat{x}} = F(\underline{y})$ can be rewritten using the PF:

$$P(PF \cap \underline{y} \in \Omega) = P_{\text{HMI}} \leq P_{\text{HMI}}^{req} \quad \forall x \tag{2.25}$$

- P_{HMI}, the probability of Hazardous Misleading Information (HMI), that was defined in Equation (2.4):

$$P_{\text{HMI}} = P(HMI) = P(\underline{\hat{x}} - x \notin \Omega_{\text{AL}} \cap \text{No Alert}) = P(PF \cap \text{No Alert}) \tag{2.26}$$

  which is equivalent to the RAIM definition:

$$P(\underline{\hat{x}} - x \notin \Omega_{\text{AL}} \cap \underline{y} \in \Omega) \tag{2.27}$$

  since No Alert $\equiv y \in \Omega$.

- $P_{FA'}$, the probability of False Alert, defined as:

$$P_{FA'} = P(\text{No PF } \cap \text{ Alert}) = P(\underline{\hat{x}} - x \in \Omega_{\text{AL}} \cap \underline{y} \notin \Omega) \tag{2.28}$$

We want to comment here furthermore on the FA's and on the $P_{FA'}$. FA' and $P_{FA'}$ are not necessarily defined for all RAIM algorithms since a RAIM algorithm does not necessarily provide a position estimate $\hat{x}$ when an Alert is raised (when $y \notin \Omega$). The FA' could be intended as the occurrence of an Alert when the position estimate 'that would have been given in case an Alert did not occur' is not erroneous enough to cause a PF. FA' and $P_{FA'}$ appear to the author controversial quantities, of limited practical use, though they are strongly connected to the quality of the estimation rule and may drive availability. In fact to compute the RAIM availability it is standard practice to consider only the Alerts that occur in the *fault-free* hypothesis, i.e. under the assumption that the system is working with nominal performance (it can be argued in fact that the requirement in Equation (2.13) should be restricted to the fault-free hypothesis). Thus we can approximate $P_{FA'}$ with:

$$P_{FA'} \approx P_{FA'}|\text{No Fault} = P_{FA'}|H_0 = P(y \notin \Omega|H_0) = P_{FA_{tot}} \tag{2.29}$$

where H_0 indicates the fault-free hypothesis and $P_{FA_{tot}}$ stands for total probability of False Alarm (FA). The terms Alarm and FA refers generally to hypothesis testing and FDE procedures (see Section 2.6 and 3.6.1). This is the only reasom why we differentiate the two terms. As outputs of the RAIM algorithm, Alarms and Alert (as well as FA and FA') effectively coincide. We specify *total* probability of FA because a FA standardly refers to a false detection (in the fault-free state) by any of the tests employed, whereas an Alarm in the fault-free state in FDE procedures usually results out of a succession of multiple False Alarms.

2.5.4. NOTE ON THE PLS

To define the PLs the total requirement on the P_{HMI}, the P^{req}_{HMI}, must be split into horizontal and vertical contributes, $P^{req}_{\text{HMI},hor}$ and $P^{req}_{\text{HMI},ver}$. HPL and VPL are defined as the maximum position error size (in the horizontal plane or in the vertical direction) that can pass undetected with probability smaller or equal to the probability requirements, $P^{req}_{\text{HMI},hor}$ and $P^{req}_{\text{HMI},ver}$.

$$\begin{aligned} HPL &= \text{arg}\min_\delta P\left(\sqrt{(\underline{\hat{x}}_1 - x_1)^2 + (\underline{\hat{x}}_2 - x_2)^2} > \delta \,\middle|\, \text{No Alert}\right) \leq P^{req}_{\text{HMI},hor} \\ VPL &= \text{arg}\min_\delta P\left(|\underline{\hat{x}}_3 - x_3| > \delta \,\middle|\, \text{No Alert}\right) \leq P^{req}_{\text{HMI},ver} \end{aligned} \tag{2.30}$$

with $P^{req}_{\text{HMI},hor} + P^{req}_{\text{HMI},ver} = P^{req}_{\text{HMI}}$. To satisfy the navigation availability requirement it has to be:

$$\begin{aligned} HPL \leq HAL \text{ and} \\ VPL \leq VAL \end{aligned} \tag{2.31}$$

If those equations are satisfied integrity is maintained for the epoch under consideration. Instead of computing the PLs, the integrity monitoring system can simply compute the actual P_{HMI} or an upperbound, and compare it to the requirement P^{req}_{HMI}, as mentioned. If $P_{\text{HMI}} \leq P^{req}_{\text{HMI}}$, integrity is maintained.

PLs are computed by ARAIM and by many other RAIM algorithms. In this dissertation nevertheless preference will be given to the simpler direct computation of the P_{HMI}, and the comparison between the different RAIM algorithms will be based on the computed P_{HMI} or on a computed upperbound on it. The PLs provide extra information to the user, but this information is not vital for safety warranty.

2.6. COMMENTS ON RTCA DEFINITIONS

Working Group 5 of RTCA Special Committee (SC)-159 spent several years working on the problem of GPS Fault Detection and Exclusion (FDE) for a primary-means navigation system. The results of this working group activities have been incorporated in [72] and are summarized in [61]. In this work performance parameters strictly related to RAIM other than the RNP parameters were defined. An FDE algorithm is a possible approach to tackle the RAIM problem, but it is not the only one possible. In Chapter 3 other possible ways to approach the problem will be considered. Nevertheless, as today almost all RAIM algorithms in use or in testing phase employ an FDE procedure.

Here we report the definitions given in the RTCA that refer strictly to FDE procedures. FDE procedures are all based on statistical hypothesis testing [101]. They perform statistical tests to detect whether any fault is affecting the system or not, and to identify and exclude the fault. Often the detection step and the identification/exclusion step are distinct and employ different tests, but in some cases the same tests can perform detection and identification at the same time. A hypothesis is describing a possible state of the (navigation) system. As in standard hypothesis testing, the FDE algorithm tests a null hypothesis (H_0) against one or multiple alternative hypotheses (H_a or H_i). The null hypothesis (H_0) reads: no faults, healthy system, nominal operations (standard performance). The alternative hypotheses (H_i) read: presence of failure or anomaly in the positioning system.

- **False Alarm/Detection (FA)** The term False Alarm (FA) is associated to a Hypothesis Testing procedure. In the RTCA document it is referred to as False Detection. This event occurs when a statistical test rejects the null hypothesis when instead the null hypothesis holds true. An FA occurs when the test statistic exceeds the critical value (the system detects a fault), while there is actually no failure (the null-hypothesis H_0 holds):

  $$T > k \cap H_0 \tag{2.32}$$

  where T is the test statistic and k is its critical value. A FA might result in an False Alert (see the following), and thus loss of continuity. A (False) Alarm can refer to any single test run in the algorithm, therefore if the full procedure is made up of multiple tests and multiple steps, an Alarm does not necessarily results in an Alert, as final output of the monitoring algorithm.

- **Missed Detection (MD)** Also the term Missed Detection (MD) is associated to a Hypothesis Testing procedure. A Missed Detection occurs if a failure is affecting the system (an alternative hypothesis H_a holds true) and it is not detected. This coincides to the event of a test accepting the null hypothesis when in fact an alternative hypothesis holds true:

  $$T < k \cap H_a \tag{2.33}$$

  where T is the test statistic and k is its critical value. An MD may coincide to a Missed Alert depending on the actual Positioning Error and on the result of other tests eventally run in the algorithm.

- **Wrong Detection/Exclusion (WD)** A Wrong Detection (WD) and Exclusion occurs when a satellite failure is detected but the exclusion algorithm declares the wrong satellite or the wrong set of satellites to be in faulty and thus excludes from the model the wrong

satellite/satellites. In the RTCA document the term Wrong Exclusion is employed only when after the exclusion of the wrong satellites a Positioning Failure occurs. If after the exclusion there is no Positioning Failure, the event is called Incorrect Exclusion. In the following anyway we will employ the same term WD for both cases.

2.7. GENERAL DISTRIBUTION OF THE OBSERVABLE

One big issue of the RAIM problem is the distribution of the observable $\underline{y}$. The pdf of $\underline{y}$ is generally supposed to be known in standard fault free conditions, but not really known in case some anomalies would occur; it is furthermore generally assumed that anomalies in the systems will occur with a certain failure rate.

Different hypotheses can be defined to represent the state of the system: a fault free (null) hypothesis H_0 and N_{H_a} alternative hypothesis H_i, representing the different possible types of anomalies affecting the system, with $i = 1,\ldots,N_{H_a}$. In this dissertation we consider only observation systems the states of which can be modelled or approximated by a linear model, of the type shown in Equation (2.2). In fact in the null hypothesis H_0 the observation model is assumed to be of the form:

$$H_0: \ \underline{y} = Ax + \underline{e} \ \text{ cf. Equation (2.2)}$$

whereas in an alternative hypothesis H_i we assume:

$$H_i: \ \underline{y} = Ax + \nabla y_i + \underline{e} \tag{2.34}$$

where $\nabla y_i = C_{y_i}\nabla_i$ is the bias (due to an anomaly) affecting the observations, and is characterized by a signature/structure matrix $C_{y_i} \in R^{m\times q}$, which is a characteristic of the corresponding alternative hypothesis, and by a bias size vector $\nabla_i \in R^q$, which is *unknown*. Furthermore we assume:

$$\underline{e} \sim N(0, Q_{yy}) \tag{2.35}$$

that is, the distribution of the random errors is normally distributed with zero mean and known variance, in all null and alternative hypotheses. This model of alternative hypothesis is the so-called mean shift (or slippage) model, by far the most popular in geodesy. Different models exist for the alternative hypotheses, as the variance inflation model and mixed models (see for instance [63], [76], [41]). The choice of the mean shift model is justified in Section 2.9.

The distribution of the observable $\underline{y}$, as a result, depends on the state of the system. Under each hypothesis $\underline{y}$ is assumed to be distributed as a multivariate normal distribution (this follows from Equations (2.2), (2.34) and (2.35)). In particular, we can write the conditional distributions of $\underline{y}$ under each hypothesis as:

$$\begin{cases} f_{\underline{y}|H_0} = N(Ax, Q_{yy}) \\ f_{\underline{y}|H_i} = N(Ax + \nabla y_i, Q_{yy}) \quad \forall i = 1,\ldots,N_{H_a} \end{cases} \tag{2.36}$$

Beside defining multiple hypotheses to describe the distribution of the observable, we need to introduce as well the prior assumption about the probability of occurrence of each hypothesis. The approach described in the following is equivalent to the one adopted in ARAIM [8]. We can introduce the random variable $\underline{H}$ representing the state of the system,

which possible realizations are the previously defined hypotheses H_0 and H_i, with $i = 1,\ldots,$ N_{H_a}. It is possible to associate prior probabilities to the occurrence of the different hypotheses, in such a way that the variable $\underline{H}$ has a prior Probability Mass Function (PMF):

$$\underline{H} \sim \begin{cases} P(\underline{H} = H_0) = p_0 \\ P(\underline{H} = H_1) = p_1 \\ \vdots \\ P(\underline{H} = H_{N_{Ha}}) = p_{N_{Ha}} \end{cases} \tag{2.37}$$

The marginal distribution of $\underline{y}$ is thus obtained as:

$$\underline{y} \sim p_0 \cdot f_{\underline{y}|H_0} + \sum_{i=1}^{N_{Ha}} p_i \cdot f_{\underline{y}|H_i} \tag{2.38}$$

At this point the uncertainty about the $\underline{y}$ distribution is expressed by its dependence on the unknown variable ∇_i beside x. There exists the possibility to assume some prior distributions also for the bias sizes ∇_i, for instance special heavy tailed distributions commonly used for outliers modelling (as the Cauchy distribution, which has infinite standard deviation [96]). Other robust approaches are also possible (assumption of worst-case scenarios etc.). Adopting any of these approaches allows writing the observable distribution as a function of the sole position x, in such a way that any (estimation) function of the observable can be written as a function of x.

2.8. THE GNSS MODEL

In the previous Section (2.7) the general observable distribution that is considered in this dissertation was presented. In this Section we focus on the GNSS case. The observation model in this case can be linearized and reduced to the same general form introduced in the previous Section; the specific matrix A and error variance Q_{yy} have to be opportunely defined. This Section describes how to construct the geometry matrix A (which defines the GNSS functional model) and how to compute/approximate the variance matrix Q_{yy} (at the basis of the stochastic model).

2.8.1. GNSS FUNCTIONAL MODEL

Suppose receiver r simultaneously tracks pseudo-range measurements of m satellites on one frequency, say L1, then we can formulate a complete system of equations:

$$E\begin{bmatrix} \underline{\rho}_r^1 \\ \underline{\rho}_r^2 \\ \vdots \\ \underline{\rho}_r^m \end{bmatrix} = \begin{bmatrix} \sqrt{(X_1 - X_r)^2 + (Y_1 - Y_r)^2 + (Z_1 - Z_r)^2} + cdt_r^g - cdt^1 + a_r^1 + b_0^1 \\ \sqrt{(X_2 - X_r)^2 + (Y_2 - Y_r)^2 + (Z_2 - Z_r)^2} + cdt_r^g - cdt^2 + a_r^2 + b_0^2 \\ \vdots \\ \sqrt{(X_m - X_r)^2 + (Y_m - Y_r)^2 + (Z_m - Z_r)^2} + cdt_r^g - cdt^m + a_r^m + b_0^m \end{bmatrix} \tag{2.39}$$

where:

- $\underline{\rho}_r^i$ is the pseudorange observable from satellite i at receiver r;
- X_i, Y_i, Z_i are the actual coordinates of satellite i;

- X_r, Y_r, Z_r are the actual coordinates of receiver r;
- c is the speed of light;
- dt_r^g is the time offset of receiver r with respect the clock of constellation g, to which the satellite i belongs;
- dt^i is the clock offset of satellite i with respect to the clock of its own constellation;
- a_r^i is the atmospheric delay at receiver r from satellite i;
- b_0^i is the nominal bias affecting the measurement from satellite i in fault-free conditions (considered in the ARAIM algorithm [111]). Addition of these nominal biases is made to take into account the non-Gaussianity of the observations (due mainly to nominal code correlation peak deformations) by means of the concept of paired bounding [84] [85].

The parameters of interest are X_r, Y_r, Z_r, the coordinates of the receiver. The model in Equation (2.39) is a nonlinear model. Given $\underline{y}$ is the full vector of observables (ρ_r^i) and x is the vector of unknown parameters, this model can be written in the form $E(\underline{y}) = G(x)$.

For our GNSS positioning problem, the unknown parameters are X_r, Y_r, Z_r, and dt_r^g (one for each constellation). In fact satellites positions and satellites clock offsets are known (though with limited accuracy). It is possible to correct the measurements for atmospheric delays a_r^i (using tropospheric models and exploiting the dual frequency to remove the ionospheric delay, as explained in the following), and the nominal biases, as introduced in the ARAIM algorithm, can be bounded by conservative estimates [111]. Therefore parameters X_i, Y_i, Z_i, dt^i, a_r^i and b_0^i in Equation (2.39) are not to be estimated.

An identical system can be written for the code observations collected in the second frequency L5. For each pair of measurements coming from the same satellite i, in the two frequencies L1 and L5, it is possible to write the ionosphere-free combination:

$$\underline{\rho}_{IF}^{\,i} = \frac{f_{L1}^2 \underline{\rho}_{L1}^i - f_{L5}^2 \underline{\rho}_{L5}^i}{f_{L1}^2 - f_{L5}^2} \tag{2.40}$$

As a result of this combination, the same functional model as in Equation (2.39) can be rewritten using the ionosphere-free pseudorange measurements; the same relations remain valid, the only difference is that the ionospheric delay is canceled from the atmospheric delay (it was included in the terms a_r^i, which after the combination account only for the tropospheric delay). The new observables obtained have reduced precision (compared to the original pseudoranges), this is accounted for in the definition of the stochastic model (next Section).

The fault free nominal biases b_0^i introduced in the ARAIM algorithm are assumed (for the moment) to be always absent for any code measurement. These biases are introduced to tackle the non-Gaussianity of the actual observations, mainly due to nominal code correlation peak deformations: the underlying idea is that the actual distribution can be overbounded by a pair of Gaussian distributions displaced from the zero mean, better than by a variance inflated single Gaussian distribution. For simplicity, and given that their influence on the results is often very small, in the following these biases are neglected.

The functional model in Equation (2.39) represents the state of standard operations, that is the case in which the system is working properly without any fault. In the case of a fault affecting the system we can expect some satellite measurements to be biased. For our simulations in Chapter 7 we will consider mainly the case of a single satellite measurement to be biased or the case of two satellites measurements to be simulateneously biased.

If the measurement from satellite i is affected by anomaly, we have:

$$E(\underline{\rho}_r^i) = E(\underline{\rho}_r^i \mid \text{No Fault}) + \nabla_i \tag{2.41}$$

where $E(\underline{\rho}_r^i \mid \text{No Fault})$ stands for the expectation in the fault-free hypothesis, the right hand side of each line in Equation (2.39), and ∇_i is the bias (positive or negative) affecting the observation.

To solve the nonlinear problem of Equation (2.39), the observation equations are linearized: the vector function $G(x)$ is approximated as $G(x) \approx G(x_0) + \partial_{x^T} G(x_0)(x - x_0)$, where x_0 is a vector of approximate values for the unknown parameters and $\partial_{x^T} G(x_0)$ is the $m \times n$ matrix of partial derivatives. When the higher-order terms in the approximation of the vector function $G(x)$ may be neglected, the linearized model to solve becomes $E(\underline{y}) = G(x_0) + \partial_{x^T} G(x_0)(x - x_0)$. This can be written as $E(\Delta\underline{y}) = \partial_{x^T} G(x_0)\Delta x$, where $\Delta\underline{y} = \underline{y} - \underline{y_0}$ (with $y_0 = G(x_0)$) is known as the vector of incremental observations and $\Delta x = x - x_0$ is known as the vector of incremental parameters. Redefining the observation vector $\underline{y} = \Delta\underline{y}$ and the unknown parameters vector $x = \Delta x$, and defining the geometry matrix $A = \partial_{x^T} G(x_0)$, we can write this linear model as in Equation (2.2):

$$E(\underline{y}) = Ax \quad \text{or} \quad \underline{y} = Ax + \underline{e}$$

with $E(\underline{e}) = 0$.

2.8.2. STOCHASTIC MODEL

As described in previous Section, the expectation of the pseudo-range code measurements at each epoch is expressed by Equation (2.39), the functional model of GNSS code observations. After the above-mentioned information available on atmospheric (tropospheric) delay, position and clock of the satellites has been exploited, the expectation of the measurements is in fact a function of $3 + N_{const}$ variables (receiver position and clock offsets), where N_{const} is the number of constellations. The collection of the measurements is a stochastic process, affected by random errors generated by thermal noise at the receiver, multipath, residual tropospheric error and satellites clock and orbit errors (see Table 7.1). We write therefore:

$$\underline{y} = E(\underline{y}) + \underline{e} \tag{2.42}$$

The error vector $\underline{e}$ is modelled to have a multivariate normal distribution:

$$\underline{e} \sim N(0, Q_{yy}) \tag{2.43}$$

with Q_{yy} a diagonal matrix with the $\sigma_i^{GPS^2}$ and $\sigma_i^{Gal^2}$ as diagonal elements (depending whether the satellite i belongs to GPS or Galileo constellation). The standard deviations σ_i^{GPS} and σ_i^{Gal} are defined in Table 7.1. As reported in Table 7.1, the total standard deviation of each ionosphere-free observation is made up of three contributions, $\sigma_i = \sqrt{\sigma_{URA}^2 + \sigma_{tropo}^2 + \sigma_{user}^2}$,

of which the last one, due to the receiver thermal noise and by multipath, is affected by the combination across the frequencies. Carrier phase smoothing is assumed to be performed over 100 seconds as recommended in [72], with the standard Hatch algorithm [39]. *Note that with this approach the risk of occurrence of an anomaly (as for instance a cycle slip) affecting the carrier phase measurements is disregarded. It is assumed therefore that a separate integrity monitoring procedure assures the integrity risk connected with the carrier phase smoothing is below a certain threshold. For the moment the approach is followed as it is commonly employed in literature but we stress the absence of an assessment of the risk related to cycle slip detection algorithms.*

The stochastic model parameters that are used in the simulations in Chapter 7 are listed in Table 7.1; they are based on [48] and [49] and are employed also in other recent simulations as in [50] and [11].

2.9. GNSS ANOMALIES AND THEIR MODELS

The observation model described in Sections 2.7 and 2.8 implies the adoption of the *mean shift model* (see Section 2.7) to represent the occurrence of anomalies in the navigation system. This means it is assumed that in case of anomaly *only* the mean of one or more observations will be biased (by the unknown amount ∇_i in Equation (2.34)), but not their standard deviation. The model in Section 2.7 furthermore assumes that each anomaly (alternative hypothesis) is uncorrelated to any other (the probabilities p_i are fixed). These assumptions are justified by the record of system anomalies registered and by the type of most common anomalies affecting GNSS measurements. Fundamentally, it is assumed that the occurrence of both mean shift and variance inflation at the same time is extremely more unlikely than the occurrence of just one of the two, so that such an event can safely be neglected. Furthermore it is reckognized that the occurrence of variance inflation is a less critical event than a mean shift, and can be protected against by taking care of the latter.

We do not provide here a detailed description of the GNSS anomalies, but we just list the main categories of High Dynamics Threats (HDTs) and provide references for more in-depth study. As previously mentioned, the HDTs are threats that cannot be monitored by the ground control system, as opposed to the Low Dynamics Threats (LDTs). The most common HDTs affecting a single measurement at a time can be categorized in:

- Clock and ephemeris estimation errors, see [38], [91], [86], [37], [42]
- Signal deformations, see [46]
- Code-carrier incoherency, see [35], [70]

From the snap-shot perspective (thus considering a single epoch of time), and working with just a single type of measurement per satellite, an outlier in a single satellite is believed to be the main threat (in terms of probability of occurrence). The satellites are to a large extent independently operating devices, typically at huge distances from each other (we mentioned an example of clock runoff in Section 1.4). In the receiver, a tracking channel is assigned to each satellite in view, and these channels work (in high-end equipment) in parallel, typically independently from each other (signal tracking is done on a per satellite basis). Of course multiple (different) outliers (for different satellites) can occur at the same time. With reference to the alternative hypotheses defined in Equation (2.34), with $\nabla y_i = C_{y_i} \nabla_i$, this means

that the main C_ys to consider will be the canonical unit vectors of R^m or $m \times q$ matrices made up of different canonical unit vectors of R^m , e.g.:

$$C_{y_1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ or } C_{y_{m+1}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}$$

in case the first satellite is believed to be affected by an anomaly or the first two satellites are believed to be affected by two different anomalies.

Anomalies that specifically affect more than one satellites at a time (wide failure errors) are furthermore described in [10]. Among these, constellation faults may happen (e.g. through upload of incorrect navigation messages, and may impact a full constellation), see [10], [103]. Finally we shall consider common errors/anomalies in propagation as ionosphere, troposphere and multipath: the tropospheric delay is typically a small effect (and one can correct sufficiently well for this error source), ionosphere gradients/fronts effects are supposed to cancel out with the use of ionosphere-free combination (only second order delay effects would be of concern, but significant second order effects are regarded as extremely unlikely events), and multipath very much depends on the local satellite-receiver geometry and can thereby be considered on a per satellite basis (typically outlier-like).

2.10. INTEGRITY REQUIREMENTS

Depending on the aircraft's approach mode, it must operate with different ALs. As previously defined, the AL is defined as the error tolerance not to be exceeded without issuing an alert, whereas the PL defines the error bound provided by the integrity system. The PL must always be smaller than the AL, so the requirements become more stringent with decreasing AL. Table 2.1 lists the most common approach modes with the associated ALs, according to the International Civil Aviation Organization (ICAO) standards and recommended practices (SARPs) [46].

SBAS systems are aiming at providing integrity down to the LPV-200 level. LPV-200 is a recently introduced approach mode that provides lateral performance with vertical guidance down to a decision height of 200 feet. At the moment, GBAS is the only GNSS based system that can provide the integrity performance necessary for precision approach up to CAT-III, but is rarely implemented by civilian airports because does not provide the advantages of a reduced ground segment that SBAS provides (and is still under development).

2.11. SUMMARY AND CONCLUSIONS

In this Chapter we introduced the concept of integrity in navigation and of RAIM, and we provided a mathematical definition of the RAIM problem (Equations (2.11)-(2.15)). We outlined the functions that a RAIM algorithm should fulfil, and defined its input, output and performance parameters. These definitions constitute a starting point for the development of a RAIM algorithm, and can help to understand the different approaches that can be taken to tackle the problem and the different algorithms already available in literature, that will be presented in the following of the dissertation. Later in the Chapter we introduced the general

Table 2.1: Performance requirements for landing of civil aircraft (ICAO SARPs [46] and [27]).

Aircraft Phase of Flight	Accuracy		Integrity			Maximum Probabilities of Failure	
	(2σ or 95%)		Alert Limits (4-5σ)		Time to Alert	Integrity	Continuity
	Vertical	Horizontal	Vertical	Horizontal			
NPA, Initial Approach, Departure	N/A	0.22-0.74 km	N/A	1.95-3.7 km	10-15 s	10^{-7}/hr	10^{-4}/hr
LNAV/VNAV	20 m	220 m	50 m	556 m	10 s	1.2×10^{-7} / 150 s	4.8×10^{-6} / 15 s
LPV		16 m		40 m			
APV I			35 m				
APV II	8 m		20 m		6 s		
LPV 200	4 m		35 m				
Precision Approach CAT I			10 m				
Precision Approach CAT II/III	< 2.9 m	< 6.9 m	5.3 m	< 17 m	< 2 s	$< 10^{-9}$ / 150 s	$< 4 \times 10^{-6}$ / 15 s

observation model that will be considered here onwards, in particular the GNSS observation model and its linearized equivalent. The GNSS anomalies' models were introduced and justified, and a set of typical integrity requirements for aviation was provided.

3

POSSIBLE APPROACHES TO RAIM

On the basis of the definition given in Section 2.4.1, the RAIM problem is for large part an estimation problem. In fact we want to find a function $\hat{\underline{x}}(\underline{y})$ that maps the measurements $\underline{y}$ into the position domain. On the other hand it is not just an estimation problem since we want to determine also a subdomain Ω of the full measurement domain for which the estimation must work, guaranteeing some properties for the estimator distribution.

Since we must perform an estimation, we can distinguish three main approaches:

- **FDE procedure:** we adopt the standard estimation rule, for instance the Best Linear Unbiased Estimation (BLUE, characterized by highest accuracy in fault free conditions), and we implement a Fault Detection and Exclusion (FDE) algorithm: in case the BLUE is not satisfying the integrity requirements (too large P_{HMI}), we can switch to different estimators, normally still BLUE for a modified set of measurements (for instance a subset of the original measurements set). Practically a different BLUE is applied depending on which hypothesis H is likely to hold true. The DIA algorithm, covered in Chapter 5, is an example of this type of approach.

- **Robust estimation:** we adopt/develop a new estimation rule tailored to integrity. Instead of employing the BLUE we can give up on some accuracy in fault-free conditions in change of enhanced integrity (see Sections 3.4 and 3.5).

- **Robust estimation & FDE:** a combination of both methods above is possible.

In this Chapter we consider separately the two main approaches (the third option is not analyzed in this work). First we consider the simple case of no fault possibly occurring, to show that even in this case the BLUE is generally not the best solution in terms of integrity, and that the RAIM problem is strictly related to the estimation problem.

The last Section of the Chapter presents furthermore the main methods used in statistics to tackle the same or similar problems that show up in RAIM — e.g. detection and identification of outliers — and compare them to the different RAIM approaches.

3.1. FAULT-FREE CASE

We defined the observable model of interest in Section 2.7. In this Section we restrict the analysis only to the special case in which only the null hypothesis H_0 can hold true. The

standard estimator employed under this assumption in aviation, as well as in most other disciplines, is the BLUE (formula given in Equation (3.7)). The BLUE has the highly desirable qualities of being unbiased and of maximizing the accuracy. On the other hand, maximizing the accuracy does not imply maximizing the integrity. We show here that the BLUE does not necessarily guarantee maximization of the integrity, in the simple fault-free case.

Suppose the model describing the measurements is simply the fault-free model ($P(\underline{H} = H_0 = 1)$:

$$\underline{y} = Ax + \underline{e} \quad \text{cf. Equation (2.2)}$$

with $\underline{e} \sim N(0, Q_{yy})$, x unknown but Q_{yy} known. The distribution of the observable $\underline{y}$ is simply:

$$f_{\underline{y}} = N(Ax, Q_{yy}) \tag{3.1}$$

The BLUE applied to the model relating to the null hypothesis reads:

$$\underline{\hat{x}}_0 = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} \underline{y} \tag{3.2}$$

We want now to solve the RAIM problem for this case. Suppose to use a linear function for $\underline{\hat{x}} = F(\underline{y})$:

$$\underline{\hat{x}}(\underline{y}) = L\underline{y} \tag{3.3}$$

Then the distribution of the estimator is easily obtained:

$$\underline{\hat{x}} \sim N(LAx, LQ_{yy}L^T) \tag{3.4}$$

The problem to solve is finding L and Ω that satisfy Equations (2.13) and (2.12).

Suppose now to further simplify the problem such that, at least for some special A, Ω can be chosen as the full domain of the observable; the problem becomes to minimize $P(\underline{\tilde{x}} - x \notin \Omega_{\text{AL}} | \underline{y} \in \tilde{\Omega}) = P(\underline{\tilde{x}} - x \notin \Omega_{\text{AL}})$, as with the approach defined in Equation (2.16). It can be easily seen that the solution will very much depend on the shape of the required Ω_{AL}.

We are in fact looking at finding the matrix L_I for which the integral of the multivariate distribution $N(LAx, LQ_{yy}L^T)$ outside the region Ω_{AL} is minimized:

$$L_I = \arg\min_L \int_{\mathbb{R}^n \setminus \Omega_{AL}} N(LAx, LQ_{yy}L^T) \tag{3.5}$$

whereas in case of using an unbiased estimator we would have:

$$L_I = \arg\min_{L_u} \int_{\mathbb{R}^n \setminus \Omega_{AL}} N(0, L_u Q_{yy} L_u^T) \quad \text{with } L_u A = 0 \tag{3.6}$$

The BLUE, which reads:

$$\underline{\hat{x}}_0 = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} \underline{y} \tag{3.7}$$

minimizes the trace of $L_u Q_{yy} L_u^T$, but not necessarily the integral in the equation. Therefore even choosing to use an unbiased estimator the solution to minimizing the integral over the multivariate distribution will generally *not* be the BLUE.

3.2. MODEL WITH FAULTS — BLUE APPLICATION

We consider here the application of standard BLUE to the general case that foresees the occurrence of possible anomalies. The model for this general case was given in Section 2.7. The BLUE applied to the model relating to the null hypothesis reads:

$$\underline{\hat{x}}_0 = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} \underline{y} \quad \text{cf. Equation (3.7)}$$

Since $\underline{\hat{x}}_0$ is a linear function of $\underline{y}$, it is possible to propagate the distribution of the measurements $\underline{y}$ as in Equation (2.38) into $\underline{\hat{x}}_0$, in the position domain. It is in fact:

$$\underline{\hat{x}}_0 \sim p_0 \cdot f_{\underline{\hat{x}}_0|H_0} + \sum_{i=1}^{N_{Ha}} p_i \cdot f_{\underline{\hat{x}}_0|H_{a_i}} \tag{3.8}$$

with

$$\begin{cases} f_{\underline{\hat{x}}_0|H_0} = N(x, Q_{\hat{x}_0\hat{x}_0}) \\ f_{\underline{\hat{x}}_0|H_{a_i}} = N(x + \nabla\hat{x}_i, Q_{\hat{x}_0\hat{x}_0}) \quad \forall i = 1, \dots, N_{Ha} \end{cases} \tag{3.9}$$

$Q_{\hat{x}_0\hat{x}_0}$ can be obtained by the propagation law of variance, $Q_{\hat{x}_0\hat{x}_0} = (A^T Q_{yy}^{-1} A)^{-1}$.

Since the estimation function has been chosen, the integrity problem becomes now the determination of the acceptance region Ω. In general once we have the joint distribution $f_{\underline{\hat{x}}_0,\underline{y}}(\hat{x}_0, y)$, from Equations (2.38) and (3.8) (still with the unknown parameters ∇) and the region Ω it is possible to compute the quantities in Equations (2.13) and (2.12), that will be therefore function of the unknown bias sizes ∇. As previously mentioned, it is possible to assign a prior distribution to ∇, or just employ a conservative approach and perform a maximization of the risk over the possible ∇ values to consider a worst-case scenario. Once one of these possible approaches has been adopted, the quantities in Equations (2.13) and (2.12) will be fully computed. Ω must be chosen based on this computation, to at least satisfy the requirements on continuity and integrity.

Instead of using a simple estimator as the BLUE for the null hypothesis, it is possible to switch to different estimators depending on the observation taken. This is the principle of the FDE algorithms, explained in Section 3.3.

3.3. FDE PROCEDURE

The observation model defined in Section 2.7 foresees the occurrence of different hypotheses, assigned to the fault-free case (null hypothesis H_0) and to the occurrence of different types of anomalies (alternative hypotheses H_i). In this set-up, an FDE can naturally apply. As the name suggests, the scope of an FDE procedure is to detect whether an anomaly is affecting the system (detect whether the null hypothesis or an alternative is holding true), and, in case of detection, exclude/isolate the anomaly, adapting the observation model in such a way that the navigation is still available.

In a common FDE procedure, typically the BLUE is applied to the model corresponding to the hypothesis H that is more likely (or safer to use). Once it has been decided which hypothesis is most likely to hold true (this decision is made through a statistical testing procedure), the estimator to be used is the BLUE for the model corresponding to that hypothesis.

Fundamentally $\underline{\hat{x}}(\underline{y})$ in this approach will be constituted by different linear functions of the observable, each corresponding to different subdomains of the full Ω domain, assuming

the form in Equation (3.7) when the null hypothesis H_0 is considered most likely, or conversely different forms $\underline{\hat{x}}_i$ (corresponding to one of the N_{H_a} alternative hypotheses) when they are designated to be most likely.

The model under an alternative hypothesis H_i, as from Equation (2.34), reads:

$$H_i: \ \underline{y} = Ax + C_{y_i}\nabla_i + \underline{e} \tag{3.10}$$

and, defining $A_i = [A\ C_{y_i}]$, the BLUE for x and ∇_i under H_i is:

$$\begin{bmatrix} \hat{x}_i \\ \hat{\nabla}_i \end{bmatrix} = (A_i^T Q_{yy}^{-1} A_i)^{-1} A_i^T Q_{yy}^{-1} \underline{y} \tag{3.11}$$

In case C_{y_i} is a canonical unit vector or a matrix which column are canonical unit vectors, i.e. only single satellite faults and combinations of satellite faults are considered (as from Section 2.9, this is most often the case in GNSS applications), $\underline{\hat{x}}_i$ can be computed employing the BLUE formula in Equation (3.7) but removing the faulty observations from the measurement vectors and ignoring the corresponding column from matrix A, with a result equivalent to Equation (3.11). Equivalently, the following formula could be applied:

$$\underline{\hat{x}}_i = (A^T Q_{yy_i}^{-1} A)^{-1} A^T Q_{yy_i}^{-1} \underline{y} \tag{3.12}$$

with $Q_{yy_i}^{-1}$ obtained from Q_{yy}^{-1} substituting the diagonal elements corresponding to the faulty observations with zero. Using this formula means completely de-weighting the observations corresponding to the (assumed) faulty measurements. The equivalence between the above listed different formulae for $\underline{\hat{x}}_i$ is demonstrated in Appendix C.

If we imagine Ω to be subdivided in subdomains, Ω_0 corresponding to the null hypothesis, $\Omega_1, \Omega_2, \ldots, \Omega_{N_{H_a}}$ corresponding to the alternative hypotheses (but for which still a solution that satisfies the integrity requirement can be found), for $\underline{y}$ belonging to each of these subdomains a different linear function $\underline{\hat{x}}(\underline{y})$ can be defined. To summarize, we have:

$$\underline{y} \in \Omega: \begin{cases} \underline{y} \in \Omega_0 & \underline{\hat{x}} = \underline{\hat{x}}_0 \\ \underline{y} \in \Omega_1 & \underline{\hat{x}} = \underline{\hat{x}}_1 \\ \vdots & \\ \underline{y} \in \Omega_{N_{H_a}} & \underline{\hat{x}} = \underline{\hat{x}}_{N_{H_a}} \end{cases} \tag{3.13}$$

$$\underline{y} \notin \Omega: \text{ Alert}$$

Ω corresponds to the realizations y that 'relate' to hypotheses on the model for which still the corresponding estimator distribution satisfies the integrity requirement, hypotheses that can be either fault-free or faulty. In this approach the switching to a different $\underline{\hat{x}}_i$ is based on the capability of identifying which hypothesis is most likely true — for this scope we recur to hypothesis testing. Figure 3.1 represents schematically the FDE approach.

Again, as explained in the previous section, once the estimation rule has been chosen (together with the approach to treat the bias size), it is possible to compute the integrity risk as function of only Ω, for any geometry (and Q_{yy}). In this case Ω is split into different subdomains, one for each different hypothesis on the state of the system, therefore the problem becomes how to choose the best subdivision and size of these domains to maximize the continuity and minimize the integrity risk.

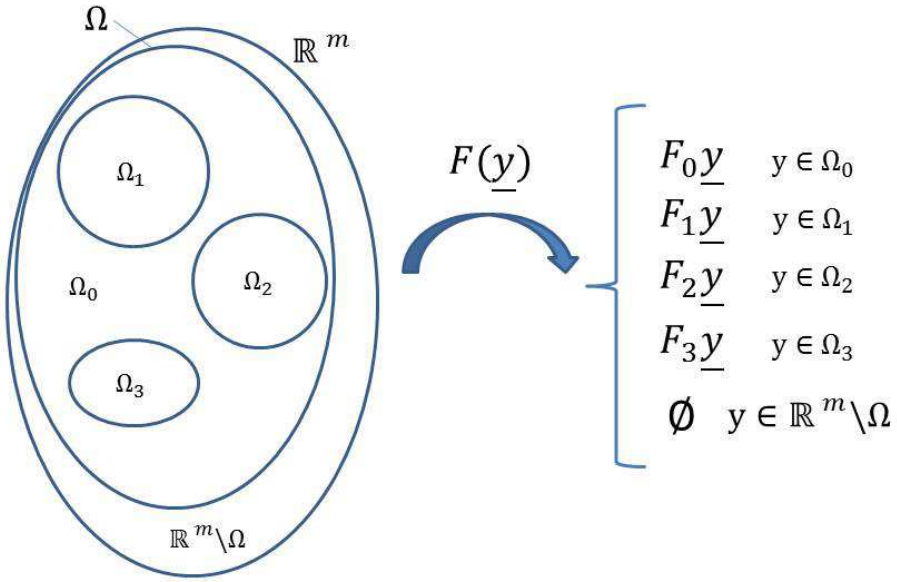


Figure 3.1: FDE scheme: observation space R^m on the left, estimators on the right. The acceptance region Ω, subset of R^m, is subdivided in different subsets to each of which corresponds a different estimation rule.

3.4. ROBUST ESTIMATION

We have seen how in an FDE procedure the monitoring system selects the hypothesis (on the state of the system) that is judged most trustworthy and determines the position solution on the basis of the corresponding model. A drawback of this approach is that there is not much room left for indecisiveness among different hypotheses: if for instance two (or more) hypotheses have similar likelihood of occurrence, with a FDE procedure we may be forced to choose only one of them. Robust estimation may allow to go for solutions that result as a 'compromise' between different hypotheses.

Robust estimation is a relatively new field of statistics and its main aim is to provide estimators that are not unduly affected by outliers in the measurements and are resistant to deviations from the nominal observation model, while still maintaining a good accuracy under this model [78]. In a RAIM algorithm, we can consider to employ an estimator different from standard BLUE, and optimized from the integrity point of view. If we allow the use of any function for the estimation $\underline{\hat{x}}(\underline{y})$, the distribution of the estimator has to be found by the pdf transformation rule. The problem will traduce into looking for the function $\underline{\hat{x}}(\underline{y})$ and for the acceptance region Ω that minimize the probability of hazardous misleading information P_{HMI} (Equation (2.12)) satisfying the continuity requirement in Equation (2.13) or viceversa.

This approach is adopted for instance in [50], where they try to look for an estimator that maximizes integrity, with the only constraint of unbiasedness. This estimation is made adding to the standard BLUE a vector lying in the left null space of the geometry matrix A. For each geometry, finding the optimal modifier vector constitutes a multi-dimensional optimization problem and can be solved using a numerical method (a modified Newton method for instance). In this approach the optimization is made regardless of the observation that will be taken, and it is based on a worst-case assumption for the bias size affecting the mea-

surements — therefore the estimator is not optimal for all the possible observations. An optimization that takes into account all the possible observations as well would result even more complex and of difficult practical implementation (high computational load).

An alternative way to look at this problem is by recurring to a Bayesian perspective, and is discussed in the following section.

3.5. ROBUST ESTIMATION VIA BAYESIAN APPROACH

When searching for an optimized estimation rule, the adoption of a Bayesian perspective[1] is an alternative that allows to look at the problem in a reverse perspective. In the Bayesian approach a fundamental step is added to the problem, at the very beginning: in our classical approach the actual position x of the aircraft is a deterministic unknown parameter, and we work on probability distributions as if they were conditioned on the actual value x of the position; with a Bayesian approach instead the actual position is considered as an unknown parameter (state of nature), to which a probability distribution can be assigned, prior and post the measurements (so x can be considered a random variable $x \rightarrow \underline{x}$). A full procedural but also theoretical step, of assignment of prior knowledge to the unknown parameter, is therefore added.

With a Bayesian perspective it is possible to obtain a *posterior* probability for the state of nature, when the measurements have been drawn: $f_{\underline{x}|\underline{y}}(x|y)$. This means that after the measurements we directly infer on the probability of the state of nature being a certain value. With the classical approach instead the results have a different meaning: considering the confidence regions for instance (supposing one-dimensional), $[\underline{a}, \underline{b}] : P(\underline{a} \geq x \geq \underline{b}) \geq k$, the probability statement is always only made on the confidence bounds, not on the parameter x. The results are conditioned on a statement about the true (but unknown) position x: if the true position is x, then the confidence region will come to be within these values around the unknown x with a certain probability.

Sometimes a confidence region is interpreted (wrongly) in reverse way: once the realizations of $[a, b]$ are determined, one would say that the true parameter x is inside those bounds with probability k. As mentioned this is *not* the meaning of the result from a classical approach; nevertheless this result is normally obtained by assuming a Bayesian perspective and assigning a non-informative prior to the parameter $\underline{x}$. With a non-informative prior any value for the parameter would be equally likely, therefore the (almost[2]) full information on the parameter will be obtained through the measurements.

The possible advantage of the adoption of a Bayesian perspective in our integrity problem lies in the fact that it is possible to obtain the posterior distribution $f_{\underline{x}|\underline{y}}(x|y)$ directly after the measurements, and decide for an estimate $\hat{x}$ therefore based on this distribution. Given in fact the distribution of the measurements as in Equation (2.38) and assuming a non-informative prior for the unknown parameters $\underline{x} \sim 1$ ($U(-\infty, +\infty)$), the posterior distribution of $\underline{x}$ is obtained by direct application of the Bayes rule, and the *posterior* probability of PF, conditioned on the measurements:

$$P_{PF|\underline{y}} = P(\hat{\underline{x}} - \underline{x} \notin \Omega_{\mathrm{AL}} | \underline{y}) \tag{3.14}$$

[1] An alternative can be the use of the inversion problem theory as developed by Tarantola [96], which also makes use of priors (non-informative or informative) and leads to the same results.

[2] In some cases non-informative priors actually provide prior information to the problem — special care has to be taken in the choice of these priors [7].

can be determined as well as a function of $\underline{\hat{x}}$. Since the $P_{PF|\underline{y}}$ can be written as a function of $\underline{\hat{x}}$, an estimation rule that minimizes this probability can be obtained. Once we have the rule we can go back to determine an acceptance region Ω and the *prior* quantities of Equations (2.13) and (2.12). Nevertheless this Bayesian approach results quite complex and computationally expensive. Furthermore it requires to assign prior distributions to the unknown ∇_i parameters (either non-informative priors or special priors assuring robustness of the procedure), which are sometimes difficult to justify. Proposals for a Bayesian approach to RAIM are presented in Appendix D, [82] and [19].

The Bayesian approach can be applied also in an FDE procedure. The extension of the Bayesian theory to fault detection (through hypothesis testing) is quite straightforward, since it reduces to determining the posterior probabilities for occurrence of faults (or any hypothesis considered possibly occurring). As mentioned also in the following, the Bayesian theory has been developed also in the fields of hypothesis testing, multiple comparisons and subset selection [7], [67].

3.6. STATISTICAL TOOLS OF AN FDE PROCEDURE

The FDE approach is by far the most widely adopted in the current RAIM algorithms. In the rest of this dissertation we will in fact restrict our analysis to FDE procedures. In this Section we present the tools available in statistics to deal with the RAIM problem and develop an FDE procedure. The basic theories of statistical hypothesis testing, Multiple Comparisons and Subset Selection, topics tightly connected among each others, are presented.

3.6.1. STATISTICAL HYPOTHESIS TESTING IN LINEAR MODELS

FDE procedures (introduced in Section 3.3) are based on statistical hypothesis testing, the theory of which goes back to Neyman-Pearson (1933) [77]. In fact in an FDE procedure tests statistical tests are normally performed to determine which hypothesis (fault-free/faulty) on the system state is most likely to hold. The theory on hypothesis testing in linear models presented in the following is based on [101].

A statistical hypothesis is an assertion or conjecture about the probability distribution of one or more random variables, for which a random sample will be available (mostly through measurements) [101]. The hypotheses defined in Section 2.7 are an example of statistical hypotheses. We keep using here the capital H to denote an hypothesis. The general structure of a statistical hypothesis therefore is as follows:

$$H: \underline{y} \sim f_{\underline{y}}(y|x) \tag{3.15}$$

with x fully or partially specified.

In standard (binary) hypothesis testing two hypotheses are set ahead: a null hypothesis H_0 and an alternative H_a; following the hypotheses are tested to determine which of them best fits the observations. A test of a statistical hypothesis is a rule or procedure in which a sample of observable $\underline{y}$ is used to decide whether or not to reject H_0. A test of a statistical hypothesis is completely specified by the critical region in the observation space R^m (m the dimension of $\underline{y}$). The critical region of a test is the set of sample values of $\underline{y}$ for which H_0 is to be rejected. If we denote the critical region by $U \subset R^m$, we can define the test by the rule:

$$\text{reject } H_0 \text{ if } y \in U \tag{3.16}$$

and do not reject H_0 if $y \notin U$. The decision rule, in binary hypothesis testing, can be aggregated to a single dimension. This means we can define a scalar test statistic, say $\underline{T}$, function h of the observable (therefore a mapping from R^m to R), $\underline{T} = h(\underline{y})$, that contains all the information necessary to take the decision of choosing between H_0 and H_a (a property called sufficiency of the test statistic). The critical region U is mapped accordingly on the real line R, into the new critical region (that we denote with K). The test reads then:

$$\text{reject } H_0 \text{ if } T \in K \tag{3.17}$$

and do not reject H_0 if $T \notin K$. We can define furthermore the acceptance region of the test, $\Omega_T = R - K$ (or $\Omega_T = R \backslash K$), complement of K to R, .

In a binary decision scenario as the one described, there are two possible states of the system, H_0 and H_a, and two possible decisions (accept H_0 or reject H_0 and go for H_a). This yields a total of four combinations, which are listed in Table 3.1.

Table 3.1: Type of errors in statistical hypothesis testing, case of binary decision between null (H_0) and alternative hypothesis (H_a).

	Result of the test	
Unknown reality	H_0	H_a
H_0	OK	FA (type I error)
H_a	MD (type II error)	OK

In two out of four combinations, the decision made is correct (marked by OK in the Table). In the other two, the decision is incorrect and a testing error is made. We distinguish two types of errors:

- Type I error, or False Alarm (FA): rejection of H_0, when in fact H_0 is true
- Type II error, or Missed Detection (MD): acceptance of H_0, when in fact H_0 is false

The probability P_{FA} or α of a type I error, also called *significance* of the test, is:

$$P_{FA} = \alpha = P(\underline{T} \in K | H_0) = \int_K f_{\underline{T}|H_0}(T|H_0)\mathrm{d}T \tag{3.18}$$

The probability P_{MD} or β of a type II error is:

$$P_{MD} = \beta = P(\underline{T} \notin K | H_a) = 1 - \int_K f_{\underline{T}|H_a}(T|H_a)\mathrm{d}T \tag{3.19}$$

Instead of β, it is also possible to monitor its complement to 1, known as the *power* of the test, γ:

$$\gamma = 1 - \beta \tag{3.20}$$

Once the distribution of the test statistic is known, at least under the null hypothesis H_0, a critical region can be defined by setting a value for the significance of the test α.

In this work only linear(ized) models are analyzed, therefore we focus here on statistical hypothesis testing in linear models. The problem is, once observations have been made, to decide between competing linear models that could describe the observed phenomenon

or process. Furthermore all observables are assumed to have a *normal* distribution, and different hypotheses differ only in the specification of the functional model. The models considered are all Gauss-Markov models [101].

Given is a linear model (to which our GNSS model can be simplified):

$$\underline{y} = Ax + \underline{e} \quad \text{cf. Equation (2.2)}$$

with

$$\underline{e} \sim N(0, Q_{yy}) \quad \text{cf. Equation (2.35)}$$

where $\underline{y}$ is the vector of measurements (m entries), x is the unknown position vector (n entries), A is the $m \times n$ geometry matrix and $\underline{e}$ the error vector (m entries).

The above system represents the state of standard or nominal operations, that is the case in which the system is working properly without any fault. This state is considered as the null hypothesis H_0. The case of a fault affecting the system constitutes instead a different state, described by an alternative hypothesis H_a, under which the linear model assumes a different form. We write therefore:

$$\begin{aligned} H_0 &: \quad \underline{y} = Ax + \underline{e} \\ H_a &: \quad \underline{y} = Ax + C_y\nabla + \underline{e} \end{aligned} \tag{3.21}$$

where C_y is a $m \times q$ matrix which represents the 'signature' of the errors in the measurements and ∇ is a q sized vector that contains the sizes of the biases in each degree of freedom (q) of C_y.

To test H_a against H_0, the Delft school developed the Uniformly Most Powerful Invariant (UMPI) test statistic (through application of the Generalized Likelihood Ratio (GLR) criterion), which reads:

$$\underline{T}_q = \underline{\hat{e}}_0^T Q_{yy}^{-1} C_y (C_y^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} C_y)^{-1} C_y^T Q_{yy}^{-1} \underline{\hat{e}}_0 \tag{3.22}$$

where $\underline{\hat{e}}_0 = \underline{y} - A\underline{\hat{x}}_0$ is the vector of residuals computed considering the null hypothesis holding true ($\underline{\hat{x}}_0$ being the position estimator under the null hypothesis, Equation (3.7)). The vector of residuals is obtained through Best Linear Unbiased Estimation (BLUE). More details on the Delft hypothesis testing theory are given in Chapter 5.

The statistic $\underline{T}_q$ can be also written in alternative formulations, as shown in [99], which are given in Section 5.1.1 (Equation (5.1)), where the Detection Identification and Adaptation (DIA) procedure is introduced.

The test statistic $\underline{T}_q$ is χ^2 distributed:

$$H_0 : \underline{T}_q \sim \chi^2(q, 0) \text{ and } \quad H_a : \underline{T}_q \sim \chi^2(q, \lambda) \tag{3.23}$$

with noncentrality parameter:

$$\lambda = \nabla^T Q_{\hat{\nabla}\hat{\nabla}}^{-1} \nabla \tag{3.24}$$

where $Q_{\hat{\nabla}\hat{\nabla}}^{-1} = C_y^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} C_y$, $Q_{\hat{e}_0\hat{e}_0} = P_A^{\perp} Q_{yy} P_A^{\perp T}$ and $P_A^{\perp} = I - A(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$.

Knowing (though only partially in case of H_a holding true) the distributions of the test statistic under the different hypotheses, we can define a critical region K (to reject the null hypothesis) on the basis of the error probabilities defined in Equations (3.18) and (3.19). For this test statistic, the critical region is one sided, of the type:

$$K : T_q > k \tag{3.25}$$

with k the test threshold (or critical value). For instance, requiring a determined (maximum) FA rate (or significance, α) for the test, the critical region K (in fact the threshold k) can be accordingly set.

The theory above constitutes the basis of statistical hypothesis testing.

3.6.2. MULTIPLE COMPARISONS

The theory on statistical hypothesis testing presented in the previous Section applies to the simple case of testing a null hypothesis against *one* alternative hypothesis. In our GNSS case, as in many other statistical testing problems, it is necessary to compare the null hypothesis of fault-free state against *multiple* alternative hypotheses, corresponding to the different types of anomaly that can affect the system — the different C_{y_i} in the model in Equation (2.36). The statistical hypothesis testing theory described so far allows to develop for instance the UMPI test for each binary case of testing the null hypothesis against any of the alternatives, but how to combine the results of each different UMPI test? We are facing a problem of Multiple Comparisons (MC).

We present in the following an example of what type of problem is to be solved. Suppose a simple case in which a decision has to be made among 4 different alternatives: a fault-free case, two one-dimensional anomalies and one bi-dimensional anomaly combination of the two one-dimensional ones. Such a scenario can result out of a linear model as in Equation (2.2) with two observables $\underline{y}_1$ and $\underline{y}_2$ ($m = 2$), and no unknown parameter ($n = 0$). Each one-dimensional alternative hypothesis stands for the presence of a bias in one of the measurements, whereas the bi-dimensional alternative hypothesis represents the case of both measurements being biased. The model in the null hypothesis is:

$$\underline{y} = c_0 + \underline{e} \quad \underline{e} \sim N(0, Q_{yy})$$

For simplicity we can assume $c_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $Q_{yy} = I_2$.

COMBINATION OF ONE-DIMENSIONAL TESTS

For each of the two one-dimensional alternative hypothesis a UMPI test can be written. The UMPI test for a one-dimensional hypothesis, obtained from Equation (3.22) with $q = 1$, is the so-called w-test ($\underline{w}$) in the TU Delft theory. More details on the testing theory are given in Chapter 5. Suffice it to know for the moment that the w-test statistic is in fact the projection of the residual vector $\underline{\hat{e}} = P_A^{\perp}\underline{y}$ onto the anomaly-specific signature vector C_{y_i}. In the special case considered, the w-tests simply coincide with the measurements themselves, $\underline{w}_1 \equiv \underline{y}_1$ and $\underline{w}_2 \equiv \underline{y}_2$.

In Figure 3.2 the decision regions for the one-dimensional UMPI tests (w-tests) are shown for the simple case considered. The test statistics can normally be fully represented on the space $R(A)^{\perp}$, which in this case is simply R^2, since matrix A is rank 0. In testing linear models in fact only the projection of $\underline{y}$ onto $R(A)^{\perp}$ is of interest, to discriminate between hypotheses. The space $R(A)^{\perp}$ represents the space of the redundancy of the system, and the coherency of the redundant measurements can be measured on this space. For the definition that we gave to the w-tests, the axes in figure represent the two different projections of C_{y_i} (C_{y_1} and C_{y_2}) on $R(A)^{\perp}$, which in this particular example are perpendicular, and along which the two test statistic w_1 and w_2 are measured.

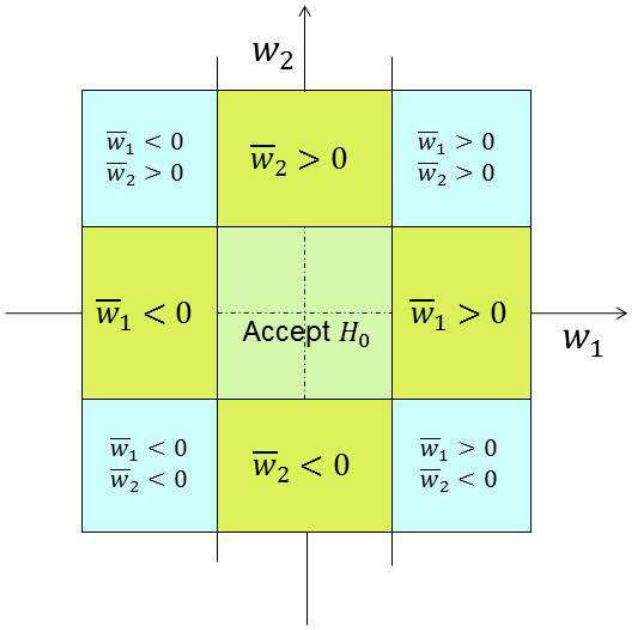


Figure 3.2: Testing of null hypothesis against two one-dimensional alternative hypotheses: Tukey-like detection regions, based on the use of one-dimensional tests only (w-tests). The overline stands for the mean of the random variable: the null hypothesis reads $E(\underline{w}_i) = 0$ (for any i) whereas the alternative hypotheses read $E(\underline{w}_i) \neq 0$.

For each w-test, the acceptance region is a strip centered at the origin and extending indefinitely in the direction perpendicular to the corresponding w-test statistic axis. In Figure 3.2 the rejection regions for the first w-test, with test statistic $\underline{w}_1$, are denoted by $\bar{w}_1 > 0$ and $\bar{w}_1 < 0$ (with $\bar{w}_1 = E(\underline{w}_1)$, the overline standing for the variable mean), and the rejection regions for the second w-test are denoted by $\bar{w}_2 > 0$ and $\bar{w}_2 < 0$; this signifies that in those areas the alternative hypotheses, which read in fact $\bar{w}_1 \neq 0$ and $\bar{w}_2 \neq 0$, are to be accepted.

Just putting together the detection regions for the one dimensional anomalies, we can determine the rejection regions for the two dimensional anomaly involving the combination of the two one-dimensional ones by simply intersect the rejection regions of the w-tests (in figure, the areas denoted by both $\bar{w}_1$ and $\bar{w}_2$ different from zero ($\bar{w}_1 \gtrless 0$ and $\bar{w}_2 \gtrless 0$). These represent the standard detection regions employed in methods such as Tukey (only for observations characterized by $Q_{yy} = \sigma^2 I$) or Bonferroni [68].

MULTI-DIMENSIONAL TESTS

When we want to add instead a test that specifically addresses the multi-dimensional anomalies (in this case bi-dimensional), this would result in a circular detection region, as in Figure 3.3. In fact the UMPI test statistic that monitors these general anomalies, in this particular case, is $\underline{T}_2 = \|\underline{w}\|^2$ (in the general case, the test statistic for the most comprehensive alternative hypothesis is equivalently the length, in the metric defined by Q_{yy}, of the projection of the measurements on $R(A)^{\perp}$, $\underline{\hat{e}}$). The use of the multidimensional test is standard practice in ANOVA procedures also to control the total significance α of the tests [3]. The use of the ANOVA F-test can be compared to the use of the Overall Model Test (OMT) in the De-

tection Identification and Adaptation (DIA) procedure (the TU Delft original FDE procedure described in detail in Chapter 5): in fact it coincides with the UMPI test in case of unknown variance for the maximum q that is monitored in the problem at hand ($q < m - n$ and $q = 2$ in the example of Figure 3.2). The test is first run to check for a general inconsistency in the measurements, then single w-tests are run to identify the specific one-dimensional inconsistencies. With reference to the figure, first the OMT is run, and this means first we check whether the realization lies inside the circle in the centre. If this is the case, the null hypothesis is not rejected. Otherwise, we run in addition the single w-tests separately, and depending on how many are rejected we make our choice on the observations to exclude (same areas as in Figure 3.2).

Depending on the choice of the threshold for the OMT with respect to the w-test threshold, it may happen that the circle intersects the detection regions of the w-tests, as in Figure 3.3. This is the case for instance when the same significance α is chosen for both OMT and w-tests (this approach is used in the Least Significant Difference (LSD) method, see Section 5.2.2), but also when other criteria for the thresholds setting are used (for instance in the B-method, see Section 5.2.2). Note that there are 4 small regions in red in Figure 3.3 outside the inner cirle but still inside the acceptance regions for the single w-tests for which the decision to take according to this procedure is not clear. They represent, to say, a dimensional inconsistency in the procedure.

In multiple comparison approaches this general consistency test is usually run only for anomalies of the highest dimension allowed, that is, as in the OMT, for anomalies with $q = m - n$ degrees of freedom, where m is the total number of observations and n the number of unknown parameters, as in model (2.2). This means only one-dimensional tests are run for identification purposes; the reason for this is that in most general cases m can be a large number and the dimension of the 'anomaly' (in general multiple comparisons problems), say q, can be large as well, so that the total number of tests to be run, $\binom{m}{q}$ only for the dimension q, would become unacceptably large. In the GNSS positioning case instead, given the dimensions of the model, the use of multidimensional tests can be still acceptable. In Figure 3.4 the procedure for standard OMT followed by w-tests is shown.

Using the simple regions in Figure 3.2 is a possible and quite natural choice — but note that the representation refers to uncorrelated one-dimensional tests, in most practical cases (as the common w-tests case in GNSS positioning) correlation should be taken into account. Such type of detection regions are employed in the Tukey and Bonferroni methods, which are in fact widely adopted especially in the cases when the alternative hypotheses (comparisons) to be considered are well-defined already before collecting the measurements.

As well it is possible to combine the circular acceptance region for the bi-dimensional (multi-dimensional) hypothesis testing with the detection regions in Figure 3.3. This method requires as mentioned to run a larger number of tests, since the number of bi-dimensional alternative hypotheses to test for is likely the number of combinations of any pair of one-dimensional alternative hypotheses, but it is more powerful in detecting bi-dimensional (multi-dimensional) alternative hypotheses. As a general rule, beside all the tests for one-dimensional, two-dimensional or even higher-dimensional specific alternative hypotheses, running an OMT for alternative hypotheses of the highest dimension is always recommended, as safeguard against the most generic anomalies.

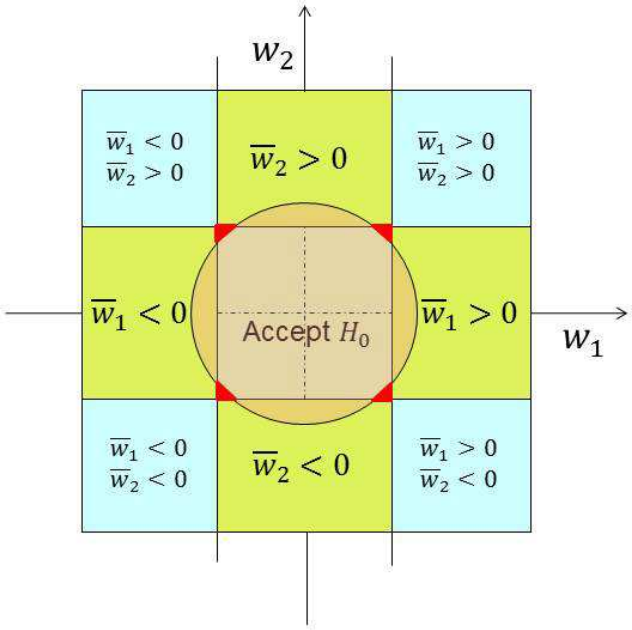


Figure 3.3: Detection regions for one-dimensional anomaly testing combined with the detection region for multi-dimensional anomaly testing.

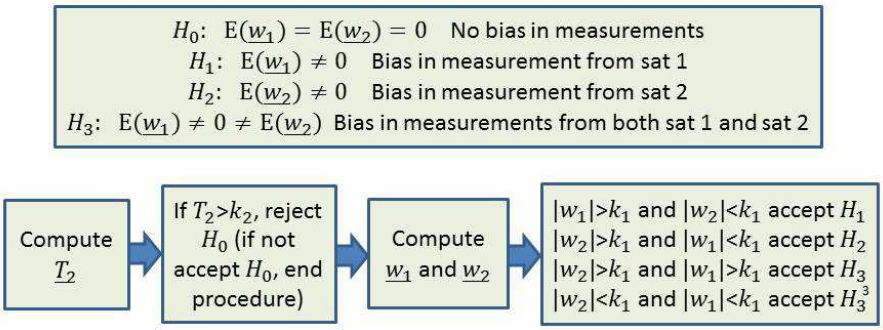


Figure 3.4: Multiple-steps procedure for a simple example with $m = 2$ and $n = 0$, in which first a multi-dimensional test is run and afterwards (in case of rejection) uni-dimensional tests are employed for identification. Examples of these procedures are the Fisher's LSD method [68] or the Baarda's OMT+w-tests method [5]. In these procedures the multi-dimensional test is the OMT, relating to the highest dimension possible for the fault; hence in case of OMT rejection but no rejection for any of the w-tests (last line of last block on the right), the presence of a general anomaly not specifically identifiable would be declared[3].

P-VALUES

The use of p-values as discriminants between 1 or 2 dimensional more likely alternatives is as well possible [33]. The p-value is defined as the probability, under the assumption of a hypothesis H, of obtaining a result equal to or *more extreme* than what was actually observed.

[3] In the scheme in Figure 3.4 the case of OMT rejection but no rejection of any of the w-tests leads to acceptance of H_3 because in this special simple case H_3 represents the occurrence of the most general type of anomaly.

Considering for instance a test characterized by a *right-sided* rejection region (of the type $K: T > k$), for any realization T_0 of the test statistic $\underline{T}$, the corresponding p-value is defined as:

$$p(T_0) = \int_{T_0}^{+\infty} f_{\underline{T}|H}(T|H)\mathrm{d}T \tag{3.26}$$

When different tests are employed to test against hypotheses with different dimensions, the p-values can be used as discriminants to choose which of the alternative should be preferred: the test statistic that realizes the smallest p-value indicates the alterative hypothesis that *challenges most* the null hypothesis, since its realization is the most unlikely to occur in H_0.

The use of p-values in the particular example considered would lead to decision regions indicatively as in Figure 3.5 (the detection regions are bounded by the circle and by the purple lines). The corresponding procedure is given in Figure 3.6. As Figure 3.5 suggests, with this method the choice for a higher dimensional anomaly would come slightly easier than with the previous method, which therefore privileges the choice for lower dimensional anomalies. In fact as visible in figure the area corresponding to $\bar{w}_1 > 0 \cap \bar{w}_2 > 0$ is bigger and includes some part of the regions that in the Tukey approach foresee only single detection. With the Tukey method in fact the threshold (with associated significance α_1) to be exceeded for detection of one-dimensional anomaly must be reached by all the w-tests regardless the realization of the other ones, but of course simultaneous occurrence of high values of the w-test statistics is less likely under the null hypothesis. This p-value criterion seems especially reasonable noticing the areas in red in Figure 3.3, for which with the Tukey method the decision to take would be unclear (whereas instead the p-values method leads to a choice for bi-dimensional anomaly present). There is anyway controversy about the use of p-values as measure of support of different hypotheses, see [75] and [6].

TEST-RATIOS

An alternative criterion to discriminate between hypotheses of different dimensions is the use of the value of the UMPI test statistic divided by its threshold (so called 'test-ratio'). Fundamentally the ratio between each test statistic and its threshold is computed, for each alternative hypothesis, and the largest of these ratios indicates the most likely alternative hypothesis. This method has been proposed in [24] and implemented also more recently (for instance in [54] and [20]), mainly on the basis of empirical reasons. The thresholds for the different test statistics are chosen to obtain the same power in detecting a bias of a specific size (Minimum Detectable Bias MDB), following the same approach as for the Baarda method [5] (see also Section 5.2.2, the B-method).

We want to point out that, if the thresholds for the different tests were chosen to have the same α instead, this method would result as an approximate form of the p-values method. To explain this, suppose that the realization of a test is exactly equal to the threshold: then the corresponding p-value is exactly α and the test-ratio is 1. If another test (for an alternative hypothesis to compare with the first) results in a p-value smaller than α, its corresponding test-ratio value is larger than 1 and both p-value and test-ratio methods would select this last alternative hypothesis. This suggests that in the 'vicinity' of the threshold value (i.e. for realizations of the test statistics that lie close to the threshold) the two methods should yield the same result.

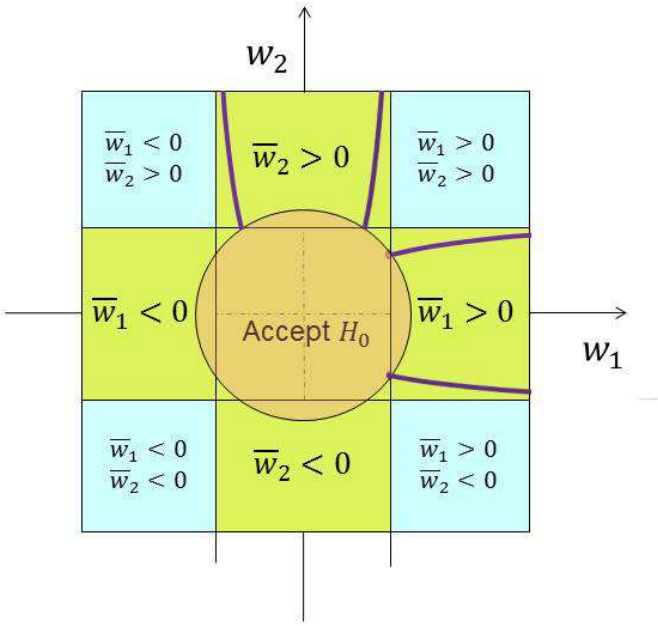


Figure 3.5: Detection regions relative to the p-values method (purple lines) in comparison with other methods detection regions. Note that the purple line start at the intersection of the edges of the detection regions of OMT and w-tests, since in those points the p-values for both tests are the same, and tend then to align with the edges of the detection regions of the single w-tests, since from the H_0 acceptance region the OMT loses its influence.

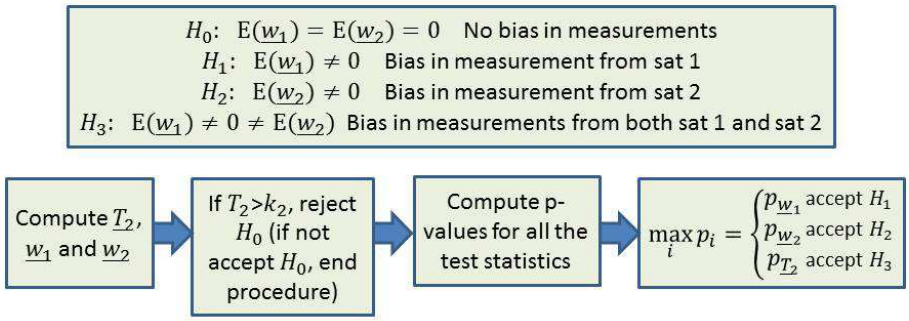


Figure 3.6: Procedure accordingto the p-values method. It is also possible, alternatively than here in the second step, to reject the null hypothesis if *any* of the test statistics exceeds its threshold.

3.6.3. SUBSET SELECTION THEORY

In this Section we present the basics of the Subset Selection theory, which is a topic very much related to Multiple Comparisons (in fact, some Subset Selection methods employ tools borrowed from MC theory and from hypothesis testing theory). Subset Selection deals with the problem of choosing a subset of parameters (from an original pool of candidates) that is able to describe 'thoroughly' a certain phenomenon/observation. This subset should include the parameters that are actually involved in the physical process behind the obser-

vation taking, parameters that should be included in the model that describes the physical process. If we consider a linear model as in Equation (2.2), the problem of selecting how many and which parameters should be part of x is a linear Subset Selection problem.

In our RAIM problem, the parameters x in Equation (2.2) are already fixed — they 'have to' be chosen to build up the model. On the other hand, we have seen defining our GNSS observation model as in Equation (3.21) that extra parameters can show up in the distribution functions of the observables, parameters describing the effect of a possible anomaly. As in statistical hypothesis testing, a decision has to be made whether including or not extra parameters in the definition of the model, choosing a particular hypothesis instead of another (see Equation (3.21)). Differently from hypothsis testing though, Subset Selection procedures usually do not start from a predetermined 'favored' hypothesis, the null hypothesis H_0, and test then to check if there is enough evidence to reject it, but instead they lay all the alternative hypotheses/models (possible subsets of parameters) on the same level, without any predetermined preference. An exception is the Spjøtvoll method, which instead employs explicitly the tools of hypothesis testing, and Bayesian methods in general, that explicitly set prior probabilities for the adoption of each alternative model [67].

The problem of selecting the correct subset of satellites to exclude from the full set of satellites in view is equivalent to the problem of starting from the pool of parameters x and ∇ for *every* observation and determining which subset of these parameters fits best some measured data, as in a linear regression problem. With reference to the model presented in Equation (3.21), adding an extra parameter to the original set x adopted in the linear model is equivalent to excluding the observation from one satellite from the model. This equivalence is proved in Appendix C. Excluding q satellites means adding q extra regression parameters (for a total of $n+q$), represented by the vector ∇. The parameters to be estimated are therefore the n standard unknowns (dimensions of x) plus eventually q biases in the measurements.

STANDARD MODEL

The general model adopted in linear regression is:

$$\underline{y} = A_i x_i + \underline{e} \tag{3.27}$$

where $\underline{y}$ is the observable vector of m measurements, A_i an $m \times n$ matrix (n unknown), x_i an n-dimensional vector of unknown regression parameters and $\underline{e} \sim N(0, \sigma^2 I_n)$, with σ often unknown. The case $\underline{e} \sim N(0, \sigma^2 I_n)$ is the case of homoskedastic errors, whereas in case of Q_{yy} simply diagonal we would have the heteroskedastic case.

With comparison to the model in Equation (3.21), we can see that in this most general case there is not a fixed geometry matrix A, i.e. there are no parameters to be necessarily included in the regression. The actual A_i is unknown, though in most cases many competitive models are proposed, each characterized by a peculiar A_i (that could be associated to different hypotheses H_i). There is normally no distinction among a null and the alternative hypotheses, all models represent in fact competitive alternative hypotheses.

Another standard model, usually adopted in Analysis of Variance (ANOVA) problems, is

the following:

$$\underline{y} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} x_0 + C_{y_i} \nabla + \underline{e} \tag{3.28}$$

where x_0 is the expected value of the measurements and again $\underline{y}$ is the observable vector of m measurements, C_{y_i} an $m \times q$ matrix (q unknown) and $\underline{e} \sim N(0, \sigma^2 I_m)$, with σ often unknown. In this case, differently from the previous model, the parameter x_0 has to be included among the regression parameters. With comparison to the model in Equation (3.21), we can see that we have the same structure as in the alternative case, with A being in this case a simple column vector of ones ($n = 1$), and Q_{yy} having a simpler form. As said there is standardly no distinction between null and alternative hypothesis. q is unknown, the parameters x_0 and ∇ have to be estimated for many different alternative C_{y_i} with different q, and the objective is to find the best subset of parameters apt to estimate the observable. In fact this is equivalent to having many different alternative hypotheses, one for each possible subset of parameters (and hence C_{y_i}), among which to choose the most fit to the measurements collected.

MINIMAL WSSE

With reference to the statistical Subset Selection theory [67], looking at the Weighted Sum of Squared Errors (WSSE) is the typical approach when trying to find subsets that fit well a set of data. Choosing the subset with lowest WSSE (also referred to as weighted Residuals Sum of Squares RSS) is the common method of selection among subsets with the same number of variables.

The WSSE ($\underline{WSSE} = \underline{\hat{e}}^T Q_{yy}^{-1} \underline{\hat{e}}$) is equivalent to the OMT statistic for the model under consideration, which distribution is a $\chi^2(0, m-n-q)$ under hypothesis of correct choice of the model, therefore small values or values close to the expectation $m-n-q$ represent more likely realizations under such hypothesis.

For our application we consider only small numbers for the value q of satellites possibly simultaneously failing (it is very unlikely to have $q > 3$ in practice), and the total number of combinations of $n+q$ subsets out of the total m (with double constellation $m < 30$) never grows too large ($< 10^5$), therefore it is always possible to compute the WSSE for all the possible subsets candidate to fit the observations[4]. For each number of variables ($n+q$), therefore for each dimension of the anomaly q, the best candidate subset is thereby found as the one generating the minimum WSSE. The problem is that when the number of parameters used for regression increases, the WSSE always decreases. If the WSSE was used in 'absolute' sense, it would always go for the most relaxed model (with largest number of parameters), therefore some additional criteria are necessary to select among subsets with different q. In the following some possible approaches are presented, while more details on each method can be found in Appendix E.

SPJØTVOLL METHOD

Once the best candidates for exclusion for each failure dimension q have been identified, these subsets can be directly compared across different dimensions using the Spjøtvoll test

[4] In case it is unfeasible to perform an exhaustive search among all the subsets there exist methods to speed up the search among the subsets, see [67].

[94], [67]. The Spjøtvoll test is a significance test that can be used to test whether a larger subset of variables or a subset with extra parameters performs significantly better than another smaller subset or not. This method is designed *to guarantee a maximum total significance for **any** possible comparison with other subsets*, therefore, similarly to the Scheffé method, it is quite conservative, i.e. tends to favor the null hypothesis. More details on this method are in Appendix E.1.

STEPWISE REGRESSION

The Spjøtvoll method is characterized by great generality since it is apt to test an original model against any alternative one, without any particular restriction. The null model can either be *nested* to the alternative or not — two models are said to be nested in case one is a special case of the other, or, in terms of corresponding hypothesis, one hypothesis implies the other (e.g. a pair of satellites failing and only one of the two failing). To compare *nested* models, i.e. subsets of parameters contained or containing each other, it is possible to employ the UMPI test, which corresponding test statistics in this case have a known distribution (differently from the case of non-nested models). Starting from the candidate subsets it is possible to perform tests only on other subsets which are contained or contain the original candidates; that means subsets that have either extra parameters added or removed from the original considered. In this case the distribution of the single test statistics (formulae in Appendix E.2) is well known, a χ^2 distribution in case of known variance and an F distribution in case of unknown variance. The thresholds for each test can be set on the basis of such distribution, and acceptance rejection regions of the type of Figure 3.2 can be employed. Since the unknown variance case is the most common in linear regression, the tests thresholds are the so called F-to-enter (and F-to-exit) values in [67]. This test can be used to check whether or not one should include a further variable in the original candidate subset (forward selection method) or to take out instead a variable from the original candidate subset (backward elimination method, parameter significance test in our theory). The total significance of each step of the method can be approximated conservatively using Bonferroni or Šidák methods. Forward selection and backward elimination methods are described in Appendix E.2. The combination of forward selection and backward elimination is usually referred to as stepwise regression.

MSEP AND MALLOWS'S $\underline{C}_p$

Another common approach in subset selection, to decide also the dimension of the subset of variables that best models the observations, is looking at the Mean Squared Error of Prediction (MSEP) [67], that in our terminology is the Mean Squared Error of the observable estimation, instead of the WSSE. The MSEP is the expectation of the actual error we make when estimating the expected value of the observable ($E(y)$). The best candidate subset will be found by minimizing the MSEP or, equivalently, the Mallows's $\underline{C}_p$ statistic. The MSEP can be estimated only under the assumption that the subset being considered does not contain already wrong parameters that would lead to a bias in the estimation. Instead a bias in the estimated parameters or observations can only be due to omission of necessary parameters. The idea behind this method is that when the number of variables used for estimation increases, the random error (the variance of the estimator) will increase but the bias due to possible omission of important variables will instead decrease; therefore we would look for minimizing the sum of the two contributions, finding the point at which the influence from the omission of variables becomes small. More details on the method are in Appendix E.3.

We consider the Mallows's $\underline{C}_p$ method not completely adequate to the RAIM problem: we would suggest this method for cases in which we have from the onset no clue of what model we should choose for our null hypothesis. This method can give a good indication on the subset of variables that can model the observations when there is no prior information at all about which variables or which number of variables are best suited to form the observation model. For our problem we usually strongly favor the models with smaller number of variables; therefore we would suggest procedures that employ tests to accept/reject our predetermined null hypothesis (of fault free state or fault of small dimension).

OTHER METHODS

There exist other methods that foresee the choice of the subset of parameters that minimizes or maximizes certain quantities describing the quality of the fit. The Akaike's Information Criterion (AIC) is the equivalent of the Mallows's $\underline{C}_p$ developed in information theory, and it is based on the minimization of the expectation of the Kullback-Leibler mean information [1]. The final result is the same as Mallows's.

Also the R^2_{adj} statistic, the adjusted coefficient of multiple determination, can be used as a measure of fit of the models [67], [71]. Similarly to the $\underline{C}_p$ and AIC criteria, this method foresees the selection of the model that maximizes a certain quantity (the R^2_{adj}) that measures the goodness of the fit. The use of the R^2_{adj} statistic is fundamentally equivalent to the use of the F statistic described in the previous Section, which resulted to be quite liberal as a method. In fact the R^2_{adj} method tends to select bigger subsets than the $\underline{C}_p$ method. More details on these methods are given in Appendix E.3.

Other methods from MC theory can be employed as well, based on the assumption of a null hypothesis and the adoption of detection and identification tests that may reject the null hypothesis and select one of the alternatives. For instance the p-values, as introduced in Section 3.6.2, can be used to choose among the alternative hypotheses. More details on this method are in Appendix E.3.

Furthermore, there exist Bayesian methods for subset selection that allow to weigh the different models on the basis of a-priori probabilities assigned to each [67]. These methods are not discussed in this dissertation. More details on Bayesian inference and hypothesis testing in linear models are in Appendix D.

3.7. FDE-BASED RAIM PERFORMANCE COMPUTATION

In this Section we discuss the computation of the RAIM performance parameters c and P_{HMI} defined in Section 2.4 in Equations (2.8) and (2.27). In Section 3.3 we have seen that an FDE procedure foresees the division of Ω, the acceptance region of the full RAIM algorithm, in multiple decision regions Ω_is. Furthermore in an FDE procedure it is standard practice to monitor some performance parameters strictly related to the underlying hypothesis testing theory, as the probabilities P_{FA} and P_{MD} defined in Equations (3.18) and (3.19) and introduced also in Section 2.6. In this Section we try to define some relationships existing between the RAIM performance parameters and the FDE related performance parameters, that may ease the computation of the RAIM performance when employing an FDE procedure.

Since all the FDE related performance parameters are probabilities conditioned on the occurrence of a particular hypothesis, we determine here relationships between the $P_{\text{HMI}|H_i}$ and $P_{FA'}|H_i$, i.e. the RAIM performance parameters under each particular hypothesis. In

case the prior probability of occurrence of the hypotheses is known as in the model defined in Section 2.7, the actual P_{HMI} and $P_{FA'}$ can be determined as:

$$P_{\text{HMI}} = p_0 P_{\text{HMI}|H_0} + \sum_{i=1}^{N_{Ha}} p_i P_{\text{HMI}|H_i}$$
$$P_{FA'} = p_0 P_{FA'}|H_0 + \sum_{i=1}^{N_{Ha}} p_i P_{FA'}|H_i$$

with the same notation adopted in Section 2.7.

Let us consider first the simple case in which an Exclusion (or Adaptation) of the model is not foreseen by the procedure (only Fault Detection is performed) and therefore a Detection of a fault results directly in an Alert. In this case the regions $\Omega_1, \Omega_2, \ldots, \Omega_{N_{Ha}}$ do not exist, but there is only one simple acceptance region $\Omega \equiv \Omega_0$. In this case we have:

$$\begin{gathered} P_{\text{HMI}|H_0} = P(PF \cap NO|H_0) = P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} \cap \underline{y} \in \Omega_0|H_0) \\ P_{\text{HMI}|H_a} = P(PF \cap MD|H_a) = P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} \cap \underline{y} \in \Omega_0|H_a) \\ P_{FA'}|H_0 = P(FA|H_0) = P_{FA} = P(\underline{y} \notin \Omega_0|H_0) \\ P_{FA'}|H_a = P(CD|H_a) = P(\underline{y} \notin \Omega_0|H_a) \end{gathered} \tag{3.29}$$

where NO stands for Nominal Operation, i.e. no detection when H_0 is holding true, and CD stands for Correct Detection, i.e. detection when an alternative hypothesis H_i holds true (NO and CD therefore stand for the events designated with OK in Table 3.1, the correct outputs of the tests). In terms of test statistics, we have:

$$\begin{gathered} P_{\text{HMI}|H_0} = P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} \cap \underline{T} \in \Omega_T|H_0) \\ P_{\text{HMI}|H_a} = P(PF \cap MD|H_a) = P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} \cap \underline{T} \in \Omega_T|H_a) \\ P_{FA'}|H_0 = P(FA|H_0) = P_{FA} = P(\underline{T} \notin \Omega_T|H_0) \\ P_{FA'}|H_a = P(CD|H_a) = P(\underline{T} \notin \Omega_T|H_a) \end{gathered} \tag{3.30}$$

where we indicate with $\underline{T}$ the set of the generic test statistics employed by the algorithm and again with Ω_T the generic acceptance region of the full set of tests.

Most often the $P_{\text{HMI}|H_0}$, the risk under the null hypothesis, is considered negligible. In fact by design this quantity is always many orders of magnitude smaller than the required P^{req}_{HMI}, and many RAIM algorithms neglect its computation. On the other hand in the case of the $P_{FA'}$, as mentioned already in Section 2.5.3, the contributions under the alternative hypotheses, the $P_{FA'}|H_a$, are often neglected. This is because these contributions are related to the occurrence of the anomalies, the H_i, and therefore are not a full 'responsibility' of the RAIM algorithm. An Alert that occurs when an anomaly is present, even though the RAIM algorithm *could have* detected and successfully excluded it, is normally not considered the result of a deficiency of the algorithm. It can be argued in fact that the requirement on continuity for the RAIM algorithm in Equation (2.13) should only address the fault-free hypothesis. We will see in the following that many RAIM algorithms employ the approximation:

$$P_{FA'} \approx P_{FA'}|H_0 \equiv P_{FA} \tag{3.31}$$

In case of a 'full' FDE procedure, which foresees possible exclusions of faulty measurements and multiple iterations of the Detection and Exclusion steps, the computation becomes more complex. First of all we note that the regions $\Omega_1, \Omega_2, \ldots, \Omega_{N_{Ha}}$ introduced in Section 3.3 are usually not explicitly defined in an FDE procedure, but they result as combinations/intersections of the acceptance/rejection regions of the statistical tests that are run

during the procedure. The main complication is that after a Detection and Exclusion the RAIM algorithm could (and should) attempt to give a new position solution, but with the positioning being again available a new risk is generated (the possibility that the new position solution is not correct). This new risk is generated both by a Correct Detection and Exclusion and (especially) by a *Wrong* Detection and Exclusion (WD and WE). The $P_{\text{HMI}|H_a}$ in this case is not given only by the second of Equations (3.29), i.e. by the probability of simultaneous MD and PF, but also by a term including the cases of CD and WD (followed by exclusion/adaptation and no further Alarm). The computation of the P_{HMI} in case of exclusion/adaptation of the model is treated in more detail in Section 6.5, where multiple iterations of an FDE procedure are analyzed.

Similar considerations have to be made also for the $P_{FA'}$ computation. Considering for instance the $P_{FA'}|H_0$ computation, in fact, it is necessary that a Detection occurs multiple times through the different iterations of the algorithm, so that the exclusion/adaptation of the model is not successful. This means that $P_{FA'}|H_0$ results as the probability of occurrence of a combination of multiple FA (from multiple tests). This probability was introduced already in Section 2.5.3, as total probability of FA, $P_{FA_{tot}}$:

$$P_{FA'}|H_0 \equiv P_{FA_{tot}}$$

More details on its computation are in Section 6.5.

3.8. SUMMARY AND CONCLUSIONS

In this Chapter we have given an overview of the most relevant tools available in statistics to deal with the RAIM problem. In the following of this dissertation, as mentioned, we will focus specifically on some FDE procedures. Robust estimation is left aside; many recent RAIM algorithms including ARAIM are borrowing though some elements from robust estimation theory to optimize the positioning from an integrity point of view, with promising results. The main issue with robust estimation techniques is the resulting complication in obtaining performance parameters such as the P_{HMI}, since often the distribution of the position estimator cannot be determined in closed form. Furthermore current RAIM algorithms need anyway to include some form of fault detection, because the ability of detecting anomalies and isolate malfunctioning satellites, not to be used in the following epochs, is considered a fundamental function of the integrity monitoring algorithm. Therefore at the moment robust estimation can possibly be applied, but always combined with an FDE procedure.

The FDE procedures that we are going to analyze in the following chapters make fundamentally use of the Multiple Comparisons and Subset Selection theories presented in this Chapter, sometimes with minor variations. The LS RAIM (Section 4.1) employs a simple binary hypothesis testing method (null hypothesis against one single alternative hypothesis). The ARAIM (Section 4.2) employs a single step Bonferroni Multiple Comparison method, as described in Section 3.6.2 (though tests of higher dimensionality are combined with one-dimensional tests, not a common procedure in literature); the exclusion mechanism shares the minimal WSSE principle with the Subset Selection theory, but presents also original elements (exclusion confirmation tests). The DIA procedure (Chapter 5) employs a two steps Multiple Comparison procedure, with a multi-dimensional test preceding the one-dimensional tests, as described in Section 3.6.2; the exclusion mechanism is then based on the forward selection method (stepwise regression), described in Section 3.6.3.

Subset Selection methods other than forward selection and minimal WSSE methods are not employed in current RAIM algorithms. Nevertheless, it is opinion of the author that they could constitute valid complements for the Exclusion mechanism. In fact Subset Selection methods can push (when opportune) towards simultaneus detection and exclusion of multiple observations at a time. This can be a desirable quality because it reduces the risk of detecting and excluding wrong measurements leaving the faulty ones undetected — it will be shown in the following (Section 6.5 and Chapter 7) that wrong exclusion are particularly detrimental for integrity. It has to be noted anyway that many methods are fundamentally equivalent to each other, as for instance the minimal $\underline{C}_p$ and minimal $\underline{R}^2_{adj}$ criteria are equivalent to employing UMPI tests, but with different choices for the thresholds (different significance). Even the p-values method gives results similar to the application of the UMPI tests. Some equivalences between Subset Selection methods are discussed in Appendix E.

One issue with the RAIM application is that the a-priori probabilities of the alternative hypotheses should better be taken into account when selecting the subsets of satellites to exclude, whereas they are mostly disregarded by the Subset Selection methods described. Furthermore the specific RAIM application requires to monitor and contain the P_{HMI}, and this need may overcome the desire of individuating the most likely alternative hypothesis. Multiple Comparison methods that allow easier computation of the probabilities of correct or wrong identification of a fault (monitored FA and MD probabilities) are preferred compared to other methods such as minimal $\underline{C}_p$ or p-values.

In the following of the dissertation we will focus on FDE procedures employed in the most popular RAIM algorithms currently in use or development. These FDE procedures are based on statistical testing (MC procedures), for which computation of fundamental performance parameters as FA and MD rate is quite straightforward. All RAIM algorithms need to focus on the reliability of their FDE procedure — monitoring the impact of faults on the solution domain. It becomes also practically important that the FDE methods employed are able to compute the risk (P_{HMI}) involved in the procedure without a too high computational effort. This is the main reason why most Subset Selection methods have not been taken into consideration for most RAIM algorithms — though in future availability of higher computation capabilities may steer the choice towards that directions.

The overview of the statistical methods given in this Chapter will serve as backgroung and point of reference for the RAIM algorithms that are described in the following Chapters. In those algorithms the hypothesis testing tools are applied to the specific RAIM problem and tailored to its specific requirements. Some methods are preferred to others because of the ease of computation of some performance parameters, as mentioned earlier: therefore it is opinion of the author that some of the methods described in this Chapter (for instance, the Bayesian approach or the p-values method), for which there is no room in current RAIM approaches, may deserve more attention and study in the future.

4

OVERVIEW OF APPROACHES TO RAIM IN AVIATION

In this Chapter the RAIM algorithms currently in use or in testing phase in aviation will be reviewed. As mentioned in the Introduction (Chapter 1), we can subdivide these RAIM algorithms in 'first generation' algorithms, in use today, and 'second generation' algorithms (absolute RAIM), under development/testing and proposed as algorithms of the future.

RAIM algorithms have been investigated since the late 1980s, starting with publications by Lee [58], Brown et al. [16], [17], and Brenner [15]. As main representative and reference of the first generation RAIM algorithms we present here the Weighted RAIM algorithm, also referred to as Least-Squares-Residuals (LS) RAIM or (in the following) Standard RAIM, proposed by Walter and Enge [107]. This algorithm is still in use today, typically implemented in aviation grade GPS receivers, to provide low-precision lateral integrity only [92]. As of today no RAIM implementation exists for any application requiring integrity in the vertical plane (i.e. precision approaches), which has more stringent certification requirements.

Following, we present the Advanced RAIM (ARAIM), also referred to as Multiple Hypothesis Solution Separation (MHSS) RAIM, which concepts were first introduced in [17] and [81]. This is the (absolute) RAIM algorithm currently being mostly investigated for aviation applications, though it has not yet assumed a definitive shape. The ARAIM algorithm is designed to provide protection also against vertical errors and fully exploit MCDF GNSS capabilities. It needs nevertheless support from an ARAIM ground-monitoring network compliant with the necessary safety requirements (through the Integrity Support Message, ISM).

4.1. STANDARD RAIM — SLOPE-BASED METHOD

Todd Walter and Per Enge [107] proposed the method of the Weighted Receiver Autonomous Integrity Monitoring (Weighted RAIM), based on the previous work of Lee [58], Brown et al. [16], [17], and Brenner [15], which is a simple yet powerful method to check the validity of the observations. The Weighted RAIM algorithm, to which we will refer also as Standard RAIM since represents the latest update of the baseline methods most cited in the MOPS [72], checks the consistency of the Weighted Least Square Estimation (WLSE) to detect errors from measurements. It processes all satellites range measurements and computes a test statistic, a single scalar carrying information on the consistency or 'health' of the measurement system. When the test statistic exceeds a determined threshold, a failure is declared.

This Section critically reviews the Weighted RAIM approach developed in the above mentioned key paper [107] in aviation integrity, and proposes possible modifications. A brief summary of the Walter and Enge's paper is provided in Section 4.1.1, together with a brief description of the Weighted Least Squares Estimation (WLSE) for the resolution of a redundant linear system. This is the method commonly used to solve the linearized observabtion model of GPS range measurements and constitutes the background for the development of the integrity theory.

The review of the method is carried on through Section 4.1.2, which is organized in five subsections. First the set up of the test statistic and its threshold is analyzed, and a small error in [107] in the derivation of the test statistic's Cumulative Distribution Function is highlighted. Following, the method to determine whether a satellite geometry can guarantee sufficient navigation integrity or not, based on the calculation of a simple quantity, the V_{slope}, is reviewed. An alternative to this approach is then proposed: a simulation based on actual GPS data has been run to validate the theoretical analysis and the results are presented. Next subsection — as an extension of the method — focuses on the choice of the test statistic for the Identification algorithm. The advantage of using a different type of test, the w-test, rather than the Overall Model Test (OMT) is demonstrated (these tests are covered in detail in Chapter 5. Then the definition of the quantity V_{slope} and its threshold are reviewed, and possible ways to improve the algorithm and in particular to compute the P_{HMI} are proposed. Finally we present the standard RAIM algorithm in detail.

4.1.1. STANDARD RAIM BASIC CONCEPTS

In Walter and Enge's work [107] a Weighted RAIM implementation is described. The method proposed consists of a first algorithm to detect errors in the range measurements, to warn timely the user when the navigation information cannot be relied on, and of a second algorithm to determine whether the satellite geometry at hand can provide sufficient integrity, i.e. determine when integrity is available.

First the Weighted Least Squares Estimation (WLSE) is introduced as the common method to obtain a positioning solution from a redundant set of measurements. The WLSE of user position is based on the linearized observation model of GNSS range measurements. After the linearization, the general observation model is:

$$\underline{y} = Ax + \underline{e} \tag{4.1}$$

where $\underline{y}$ is the $m \times 1$ measurement vector; x is the 4×1 unknown parameter (position and clock bias); A is the observation matrix and $\underline{e}$ is the $m \times 1$ Gaussian measurement noise. The measurements weight matrix W is commonly chosen diagonal, as these measurements are assumed to be uncorrelated. The WLSE of x is:

$$\underline{\hat{x}} = (A^T W A)^{-1} A^T W \underline{y}$$

and the residuals vector $\underline{\hat{e}}$ is:

$$\underline{\hat{e}} = \underline{y} - A\underline{\hat{x}}$$

Since it is impossible to evaluate the positioning error $\underline{\hat{x}} - x$ directly (since we do not know the true position x), we have to assess it by examining other quantities. The OMT method, which computes the weighted squared norm of the measurements residuals, is used

to judge the goodness of the WLSE of the position [107]. The weighted squared norm of the measurements' residuals, or the Weighted Sum of Squared Errors (WSSE), is defined as:

$$\underline{WSSE} = \underline{\hat{e}}^T W \underline{\hat{e}} \tag{4.2}$$

and it is equivalent to

$$\underline{WSSE} = \underline{y}^T W (I - P_A) \underline{y} \tag{4.3}$$

where $P_A = A(A^T W A)^{-1} A^T W$ is the projector matrix onto the space $R(A)$. Taking Q_{yy}^{-1} as a special Weighted matrix W, with Q_{yy} the variance matrix of the measurements, gives us Best Linear Unbiased Estimation (BLUE). Alongside the reviewed paper treatment, in the following only the case $W = Q_{yy}^{-1}$ will be considered, since this choice yields the most precise results for position estimation.

In [107], the variable $\sqrt{\underline{WSSE}}$ is used as test statistic rather than $\underline{WSSE}$ itself to detect a big bias in the range domain. We denote this test statistic with $\underline{T}$, $\underline{T} \equiv \sqrt{\underline{WSSE}}$. If this statistic exceeds a certain threshold k, the estimated position is assumed badly biased; otherwise, it is assumed acceptable. This threshold is chosen to meet the probability of False Alert requirement, which coincides with the probability of False Alarm since the algorithm is made up of a single step. $\underline{WSSE}$ is distributed as χ^2 with $m-4$ degrees of freedom (in case of single constellation), assuming $E(\underline{e}) = 0$.

If a range error from one measurement occurs, the expected value of the test statistic will increase along with an increase in the expected position error. It depends on the satellite geometry how the error in the range domain propagates into the position domain. In [107] is stated that a relation of simple linear proportionality holds between the test statistic $\underline{T}$ and the absolute position error, a relation that turned out not to hold exactly true. In a simple two-dimensional graph, plotting $\underline{T}$ on the horizontal axis and the absolute position error on the vertical axis, the relation is represented by a straight line (see Figure 4.1, dashed lines), with a steepness given by:

$$V_{slope} = \frac{|S_{3i}|\sigma_i}{\sqrt{1 - P_{A_{ii}}}} \tag{4.4}$$

with $S = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$ and $\sigma_i = \sqrt{Q_{yy_{ii}}} = \sigma_{y_i}$. The value of the steepness V_{slope} can be used to decide whether the ongoing satellite geometry can guarantee integrity. A threshold for the V_{slope} can be set and, if it is exceeded, integrity is declared unavailable. The maximum allowable V_{slope} is set based on AL, P_{HMI} and decision threshold of the test statistic. The use of V_{slope} is the main concern of this review and is addressed from Section 4.1.2 onwards; the setting of a threshold for the V_{slope} is referred as an algorithm to prune out bad geometries and treated in Section 4.1.2.

4.1.2. ANALYSIS AND REVIEW

TEST STATISTIC THRESHOLD

In [107], the CDF (Cumulative Distribution Function) of $\underline{T} \equiv \sqrt{\underline{WSSE}}$ is presented as:

$$F_{\underline{T}}(k) = \int_0^{k^2} \frac{1}{\Gamma(\frac{m-4}{2})} s^{\frac{m-4}{2}-1} exp(-s) ds$$

where Γ is the *gamma function*[1], m is the number of satellites in use, k is the threshold of the test (that can assume any value) and $F_{\underline{T}}$ is the CDF of $\underline{T}$.

[1]The gamma function is defined as $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} \mathrm{d}x$.

Yet it is not fully correct. By applying the CDF transformation law, the integral upper limit in the above equation reduces to its half. Thus the CDF of the random variable $\sqrt{\underline{WSSE}}$ is found to be:

$$F_{\underline{T}}(k) = \int_0^{\frac{k^2}{2}} \frac{1}{\Gamma(\frac{m-4}{2})} s^{\frac{m-4}{2}-1} exp(-s) ds \tag{4.5}$$

Suppose we have chosen a threshold k, the probability of false alarm P_{FA} can be obtained by

$$P_{FA} = Q(k) = 1 - F_{\underline{T}}(k)$$

Reversely, k can be computed as:

$$k = Q^{-1}(P_{FA}).$$

where Q is defined as $Q(k) = 1 - F_{\underline{T}}(k)$, and Q^{-1} is the inverse function of Q.

For convenience, several typical values have been calculated beforehand and stored for later use (see Table 4.1). This table in [107] is correct.

A requirement on the P_{FA} is set for the algorithm determined from the requirement $P_{FA'}^{req}$ defined in Section 2.4.1. Since in the standard RAIM only one χ^2 test is run, an Alarm from the test corresponds to an Alert from the full algorithm, therefore P_{FA} can be used effectively to approximate $P_{FA'}$.

Table 4.1: Critical values for given probabilities of FA and number of satellites

$m \setminus P_{FA}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
5	2.5758	3.2905	3.8906	4.4172	4.8916	5.3267	5.7307	6.1094
6	3.0349	3.7169	4.2919	4.7985	5.2565	5.6777	6.0697	6.4379
7	3.3682	4.0331	4.5943	5.0894	5.5376	5.9503	6.3348	6.6964
8	3.6437	4.2973	4.8490	5.3360	5.7773	6.1838	6.5629	6.9195
9	3.8841	4.5293	5.0739	5.5548	5.9907	6.3924	6.7672	7.1198
10	4.1002	4.7390	5.2779	5.7539	6.1853	6.5831	6.9543	7.3037
11	4.2983	4.9317	5.4660	5.9379	6.3657	6.7602	7.1283	7.4749
12	4.4822	5.1112	5.6416	6.1100	6.5346	6.9262	7.2917	7.6359

RANGE ERROR TO POSITION ERROR PROPAGATION

During normal operation, the test statistic, given by the $\underline{WSSE}$, is small and also the position error is small. If, however, a range error in one of the satellites occurs, the expected value of the test statistic will become bigger while also the expected position error will become bigger. It depends on the satellite geometry how the error in the range domain propagates into the position domain. In [107] it is stated that this relation is a straight line, with a steepness given in Equation (4.4):

$$V_{slope} = \frac{|S_{3i}|\sigma_i}{\sqrt{1 - P_{A_{ii}}}} \quad \text{cf. Equation (4.4)}$$

if using the square root of the $\underline{WSSE}$ on the horizontal axis and the absolute position error on the vertical axis. On the horizontal axis, the decision threshold is set by the probability of false alarm. The maximum allowable V_{slope} is set by alert limit and the probability of misleading information. Under the acceptable probability of false alarm and misleading information, for any geometry with a V_{slope} larger than the maximum allowable V_{slope},

the integrity is declared unavailable. The geometrical interpretation of V_{slope} is given in Appendix F.

The straight line obtained from Equation (4.4), however, is *only* valid for (infinitely) large range errors, see Figure 4.1. To find the actual relation between the $\underline{WSSE}$ and the absolute position error $|\underline{\hat{x}}_3 - x_3|$, with x_3 the true, but unknown position, we have to calculate the expectations of both the test statistic and the position error, for a growing range error in a satellite. It is assumed that only one range at a time might suffer a range error.

If we assume, like before, that the range measurements are normally distributed, than the $\underline{WSSE}$ has a $m-4$ dimensional Chi-squared distribution $\underline{WSSE} \sim \chi^2(m-4,\lambda)$. The expectation for the $\underline{WSSE}$ is then given by:

$$E(\underline{WSSE}) = m - 4 + \lambda \tag{4.6}$$

with λ the non-centrality parameter. Under the null-hypothesis λ will be zero while if there is a range error/bias the non-centrality parameter becomes:

$$\lambda = \nabla y^T W (I - P_A) \nabla y \tag{4.7}$$

where ∇y is the range error/bias vector, which contains all but one zeros. This only holds if we assume that only a single satellite might fail at a time. To get the expectation of $\underline{T} = \sqrt{\underline{WSSE}}$, it is not allowed to simply take the square root of $E(\underline{WSSE})$ because:

$$E(\sqrt{\underline{WSSE}}) \neq \sqrt{E(\underline{WSSE})} \tag{4.8}$$

So, the non-linear mean propagation law should be used. Suppose:

$$\underline{v} = G(\underline{u}) \tag{4.9}$$

then the mean propagation law is given by the approximation:

$$\bar{v} \approx G(\bar{u}) + \frac{1}{2}\,\mathrm{trace}(\partial^2_{uu^T} G(\bar{u}) Q_{uu}) \tag{4.10}$$

where $\bar{u}$ and $\bar{v}$ are the mean of $\underline{u}$ and $\underline{v}$ respectively.

Applying the non-linear propagation law, the expectation of $\underline{T}$ becomes:

$$E(\underline{T}) \approx \sqrt{m-4+\lambda} - \frac{(m-4)+2\lambda}{4(m-4+\lambda)^{\frac{3}{2}}} \tag{4.11}$$

For the expectation of $|\underline{x}| = |\underline{\hat{x}}_3 - x_3|$, we assume, in line with earlier assumptions, that the distribution of $\underline{x}$ is a normal distribution $\underline{x} \sim N(\mu,\sigma^2)$. The expectation of the absolute value $|\underline{x}|$ can then be calculated by:

$$E(|\underline{x}|) = \int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx \tag{4.12}$$

Where μ is calculated from the range error vector ∇y by:

$$\mu = S_{3i} \nabla y \tag{4.13}$$

Evaluation of this integral leads to the following expression:

$$E(|\underline{x}|) = \frac{2\sigma}{\sqrt{2\pi}} \exp\left\{-\frac{\mu^2}{2\sigma^2}\right\} + \mu\left[\Phi\left(\frac{\mu}{\sigma}\right) - \Phi\left(-\frac{\mu}{\sigma}\right)\right] \tag{4.14}$$

where $\Phi(z)$ is the normal distribution function of a standard normal random variable $\underline{z}$.

Proof. The mean of the distribution can be calculated as the mean of $|\underline{x}|$ with $\underline{x} \sim N(\mu, \sigma^2)$:

$$E(|\underline{x}|) = \int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\} \mathrm{d}x =$$

$$\int_0^{+\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\} \mathrm{d}x - \int_{-\infty}^{0} x \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\} \mathrm{d}x$$

and with the change of variable $\frac{x-\mu}{\sigma} = z$:

$$E(|\underline{x}|) = \int_{-\frac{\mu}{\sigma}}^{+\infty} \frac{1}{\sqrt{2\pi}} (\sigma z + \mu) \exp\{-\frac{1}{2}z^2\} \mathrm{d}z - \int_{-\infty}^{-\frac{\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} (\sigma z + \mu) \exp\{-\frac{1}{2}z^2\} \mathrm{d}z$$

This finally leads to:

$$E(|\underline{x}|) = \frac{2\sigma}{\sqrt{2\pi}} \exp\{-\frac{\mu^2}{2\sigma^2}\} + \mu\left(\Phi\left(\frac{\mu}{\sigma}\right) - \Phi\left(-\frac{\mu}{\sigma}\right)\right) \tag{4.15}$$

This equation clearly shows that even if $\mu = 0$ (i.e. $E(\underline{x}) = 0$), we have $E(|\underline{x}|) = \frac{2\sigma}{\sqrt{2\pi}} \neq 0$. □

In Figure 4.1, the two expectations given by Equations (4.11) and (4.14) are plotted (solid lines) together with the V_{slope} given by Equation (4.4) (dashed lines), which is the equation given in paper [107]. On the horizontal axis is the expectation of the test statistic, while on the vertical axis is the absolute value of the expectation of the error in the vertical direction. The approximation by Equation (4.4) reproduces the actual curves for very big errors. As can be seen in the graph, the approximation does *not* overbound the real error for all the satellites and hence may pose an extra, unaccounted integrity risk.

A LINEAR RELATION BETWEEN SQUARED RANGE ERROR AND SQUARED POSITION ERROR

As shown in the previous section, the relation between the expected value for the test statistic and the expectation for the vertical position error, is *not* a linear function, as was suggested by [107]. It is, however, possible to describe a linear relation between the test statistic and the position error. In this case we use the $\underline{WSSE}$ as test statistic and not $\sqrt{\underline{WSSE}}$. This simplifies the calculation of the expected value for the test statistic. For the vertical position error, instead of using $|\underline{\hat{x}}_3 - x_3|$, which has an unfamiliar distribution function, we use the normalized squared vertical position error, $(\underline{\hat{x}}_3 - x_3)^2/\sigma^2_{\hat{x}_3}$. This results in a Chi-squared distribution for both the vertical position error and for the test statistic. The expectation for the test statistic is as in Equation (4.6):

$$E(\underline{WSSE}) = m - 4 + \lambda \tag{4.16}$$

with λ as given by Equation (4.7). The expectation of the normalized squared position error becomes:

$$E\left(\frac{(\underline{\hat{x}}_3 - x_3)^2}{\sigma^2_{\hat{x}_3}}\right) = 1 + \lambda_{\hat{x}_3} \tag{4.17}$$

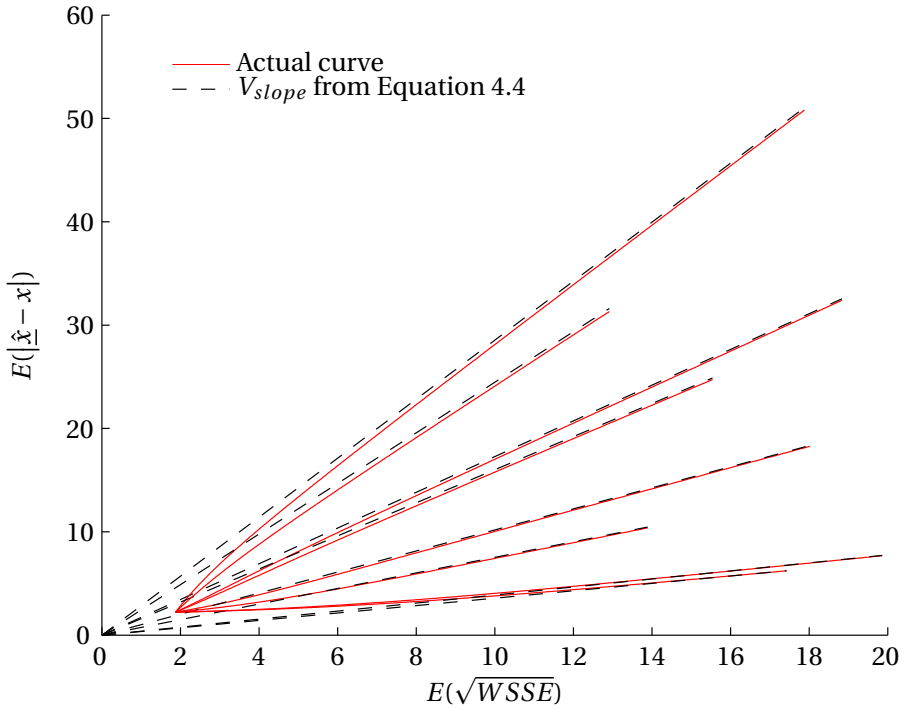


Figure 4.1: This figure shows the V_{slope}'s as proposed in [107] and the lines as calculated from the expectations given by Equations (4.11) and (4.14), for a simulated satellite constellation. The range error in ∇y varies between 0 and 50m for each satellite. This range error can be translated into a non-centrality parameter by Equation (4.7).

where:

$$\lambda_{\hat{x}_3} = \frac{(E(\underline{\hat{x}}_3) - x_3)^2}{\sigma^2_{\hat{x}_3}} \tag{4.18}$$

If we now draw the same picture as given in Figure 4.1, but in this case the $E(\underline{WSSE})$ on the horizontal axis and the $E((\underline{\hat{x}}_3 - x_3)^2/\sigma^2_{\hat{x}_3})$ on the vertical axis, an error will propagate along a line with the slope:

$$V'_{slope} = \frac{\lambda_{\hat{x}_3}}{\lambda} = \frac{S^2_{3i}\sigma^2_{y_i}}{(1 - P_{A_{ii}})\sigma^2_{\hat{x}_3}} \tag{4.19}$$

The given line will start from $[m-4, 1]$, for a range error/bias equal to zero.

In Figures 4.3 to 4.7, the new V'_{slope}'s as calculated by Equation (4.19) are given by the dashed black lines, together with simulated data. At the end of each V'_{slope}'s the associated PRN number is given. The horizontal line indicates the Alert Limit and the vertical the threshold for the test statistic (the $\underline{WSSE}$). In these simulations, the threshold is set such, that the probability of false alarm is 10^{-4}, and the vertical Alert Limit (VAL) for $|\underline{\hat{x}}_3 - x_3|$ is set to 20m. This alert limit can be translated to a critical value for a χ^2 distribution, thus for $(\underline{\hat{x}}_3 - x_3)^2/\sigma^2_{\hat{x}_3}$, by:

$$AL_{\chi^2} = \frac{VAL^2}{\sigma^2_{\hat{x}}} \tag{4.20}$$

The simulated data consist of 100,000 measurements at a single epoch, and are given by the red dots. Along the horizontal axis, the $\underline{WSSE}$ and along the vertical axis the squared posi-

Bias	NO	HMI	CD	FA'	Figure
0m	0.99987	0	0	0.00013	Figure 4.3
15m @ PRN 10	0.69928	0.02777	0.01018	0.26277	Figure 4.4
25m @ PRN 10	0.011	0.01405	0.55385	0.4211	Figure 4.5
15m @ PRN 2	0.38064	$2e^{-5}$	$9e^{-5}$	0.61925	Figure 4.6
25m @ PRN 2	0.00025	0	0.00132	0.99843	Figure 4.7

Table 4.2: Result of the 5 simulation cases: fraction of NOs, HMIs, CDs and FA's of the total measurements in case of fault-free case or presence of a bias in the measurement from PRN10 or PRN2. The results for each of the five cases are shown in more detail in the dedicated figures.

tion error are given. The expectation is given by the green dot. Along both axes a separate plot is given for the marginal probability distribution (along that axis), the distribution is given by the blue line and the expected value by the green line. The simulations are run for 5 cases, one nominal case without error, and four cases where an error of 15m or 25m is given to either PRN10 or PRN2, which represent the two extreme cases (largest and smallest values) for the V'_{slope}. In Table 4.2, the fraction of measurements falling in the four different regions of the plot shown in Figure 4.2, i.e. Nominal Operation (NO), Hazardly Misleading Information (HMI), Successful Detection (SD) and False Alert (FA'), are given. Figures 4.3 to 4.7 correspond to the 5 cases considered. In the figures it is clearly visible that, as expected, the expected value travels along the V'_{slope}. One obvious conclusion one can draw from the results given in Table 4.2 is that with a very shallow V'_{slope}, such as PRN2, a range error will result in a very high number of correct detections in the sense that the test statistic is too high. However, the actual position error is not exceeding the AL. So there is a correct detection in the range domain, but it is a FA' in the position domain (but not a False Alarm, FA). This is caused by the fact that errors in ranges with low V'_{slope} contribute very little to the positioning error. To solve this problem, each satellite could be tested on its own, e.g. the w-test, see the next section. Also, the threshold could be set to each individual case. In that case, a satellite with a small V'_{slope} will have a higher threshold for the test statistic than satellites with a large V'_{slope}. In this case the probability of a measurement exceeding the alert limit or the test statistic threshold is equal for each satellite. In a sense the integrity is traded for continuity, all within the safety bounds of course.

POWER DIFFERENCE BETWEEN OMT AND W-TEST

RAIM mainly consists of two algorithms, a detection algorithm and an identification (or exclusion) algorithm. The standard algorithms employed are the Overall Model Test (OMT) and the w-test. A brief introduction to these two methods is presented later in this section. Many other algorithms for fault detection or identification have been proposed in literature (e.g. Parity Method and Range Comparison Method), but their equivalence to the aforementioned two methods has been demonstrated (see [18]).

In general testing of linear models, if the possible occurring error signatures are unknown, it is advisable to use the OMT as a means of detection. The OMT guarantees protection against any type of faults possibly affecting the system.

On the other hand, when, as in the analyzed GPS case, we have a clue beforehand on the error signatures that can occur, it is advisable to run 'specialized' tests, such as the w-tests. In the case at hand it is assumed that single satellite faults are highly more likely than others

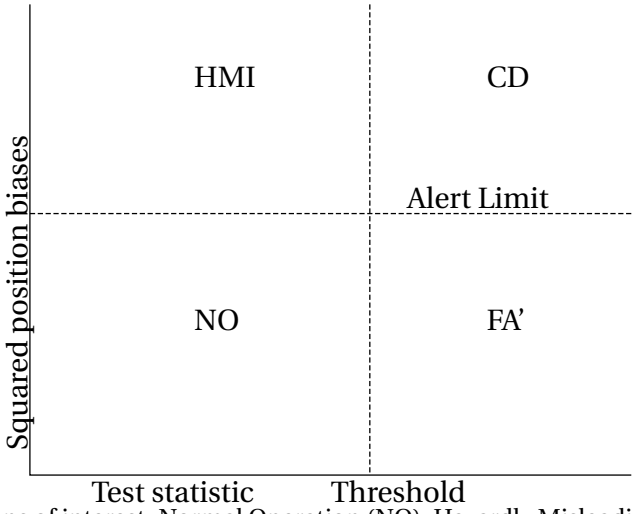


Figure 4.2: The four regions of interest: Normal Operation (NO), Hazardly Misleading Information (HMI), Correct Detection (CD) and False Alert (FA').

(for instance multiple satellite simultaneous faults), therefore it is advisable to run m (as the number of satellites in view) w-tests as detection/identification algorithm.

As outlined in the summary, Walter and Enge propose to run the OMT as a complete and optimal detection algorithm. Instead the method proposed in this chapter is to run multiple tests. First an OMT is performed as a general check, always necessary to control the overall consistency of the data. Afterwards, since it is highly probable that a fault, if present, affects only one satellite rather than a whole set of measurements (this is a conventional hypothesis), m w-tests should be run (see Chapter 5). The choice of multiple tests generates the issue of choosing a threshold for each test: the overall probability of False Alarm/False Alert of the full procedure will have to comply with the requirements.

Detailed analysis of OMT and w-test can be found in [101] and in [99]. They were introduced in Section 3.6.1, and they are discussed in much more detail in Chapter 5, as they are the main tests run in the Delft DIA. Here a brief overview is provided.

In hypotheses testing of linear models (Section 3.6.1) the null hypothesis H_0 usually reads 'no fault' (therefore the measurements $\underline{y}$ are distributed as expected), while the alternative hypothesis H_a reads 'measuring fault' and $E(\underline{y}) = Ax + \nabla y$, as described by Equation (3.21). $\nabla y = C_y \nabla$ represents the error (or bias) vector, with C_y a $m \times q$ matrix that describes the bias signature and the vector ∇ that states the size of the bias. q depends on the choice for the alternative hypothesis, and represents the degrees of freedom of the error.

The application of the Generalized Likelihood Ratio test (see [101]) to linear models leads to the choice of the test statistic reported in Equation (3.22), which can be written also:

$$\underline{T}_q = \underline{\hat{\nabla}}^T Q_{\hat{\nabla}\hat{\nabla}}^{-1} \underline{\hat{\nabla}} \tag{4.21}$$

where $\underline{\hat{\nabla}}$ is the estimator of the bias vector ∇ and $Q_{\hat{\nabla}\hat{\nabla}}$ is its variance matrix. Under the null hypothesis H_0 (no faults):

$$\underline{T}_q \sim \chi^2(q, 0)$$

Under the alternative hypothesis H_a (satellite failure) instead:

$$\underline{T}_q \sim \chi^2(q, \lambda)$$

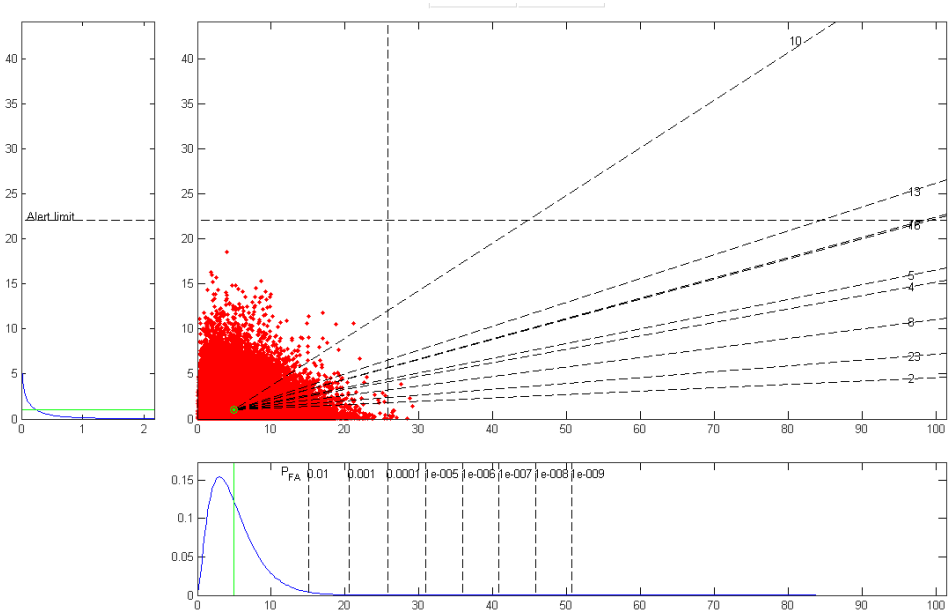


Figure 4.3: Simulation result for the nominal case, without errors.

In this case the non-centrality parameter λ assumes the value:

$$\lambda = \nabla^T Q^{-1}_{\hat{\nabla}\hat{\nabla}} \nabla = \nabla^T C_y^T Q_{yy}^{-1} Q_{\hat{e}_0 \hat{e}_0} Q_{yy}^{-1} C_y \nabla \tag{4.22}$$

$$\lambda = \| P_A^{\perp} C_y \nabla \|^2_{Q_{yy}^{-1}} = \nabla y^T Q_{yy}^{-1} P_A^{\perp} \nabla y \tag{4.23}$$

with $\nabla = E(\hat{\underline{\nabla}})$ and $P_A^{\perp} = I - P_A$, and where the norm of a vector v in the metric defined by the positive definite matrix M is defined as $\|v\|_M = \sqrt{v^T M v}$. As a special case, in the OMT the parameter q assumes the maximum value allowable by the observations redundancy, i.e. $m-n$, with $n=4$ in case of single constellation GNSS. In case $q=1$ (w-test), instead, the expression for the test can be simplified:

$$\underline{T}_1 = \frac{\hat{\underline{\nabla}}^2}{\sigma^2_{\hat{\nabla}}} \tag{4.24}$$

with distribution under H_0:

$$\underline{T}_1 \sim \chi^2(1,0)$$

and under H_a:

$$\underline{T}_1 \sim \chi^2(1,\lambda)$$

The main difference between the two tests is the degree of freedom of the χ^2 distribution under the two hypotheses: unity in the w-test versus $m-4$ for the OMT. This translates into higher power for the w-test, when testing for that specific single outlier, given the same absolute value of the bias in the observation (∇ is fixed and λ is the same).

In Figure 4.8 an example is shown of how the power varies with the number of degrees of freedom of the χ^2 distribution, given λ. These results have been obtained from the cumulative distribution function of the non-central χ^2 for a significance level α of 10^{-5}.

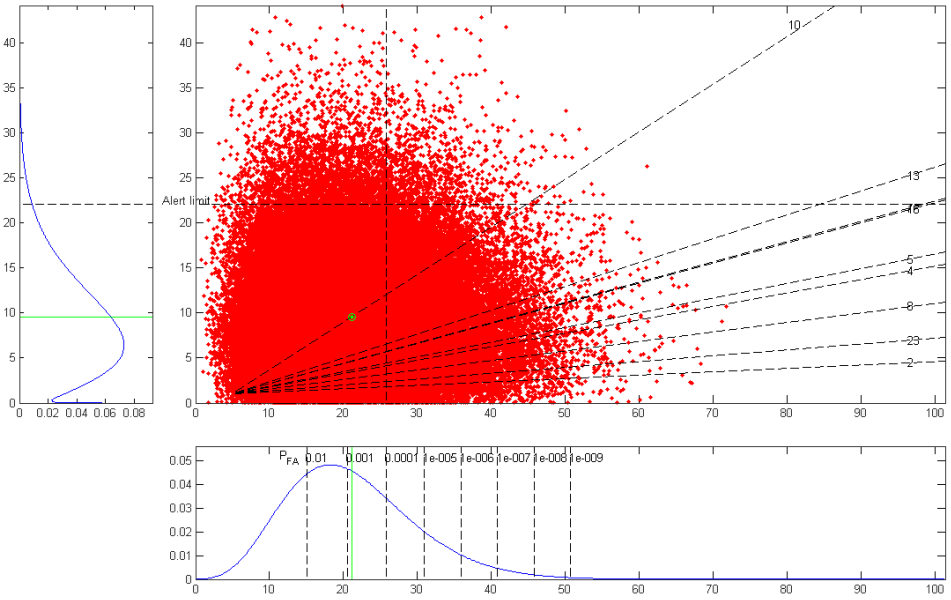


Figure 4.4: Simulation result for a 15m error on PRN10.

The w-test corresponds to the case $q = 1$ and exhibits the highest power (in case of equal non-centrality parameter) when detecting an error with a single degree of freedom, being unable to detect errors in other dimensions.

As an example, if nine satellites are in view ($m = 9$) and λ as the non-centrality parameter is chosen to be equal to the value 40, the resulting χ^2-distribution will have $q = 9-4 = 5$ degrees of freedom. The power of the OMT in this case can be easily retrieved from the graph, slightly less than 0.87, while the w-test ($q = 1$) exhibits a power of approximately 0.97.

A formal demonstration of the optimality (in detecting specific errors) of the w-test is in Appendix G.

PRUNING OUT OF BAD GEOMETRIES

In [107] the V_{slope} defined by Equation (4.4) is chosen as a discriminant to make the decision whether the current satellite geometry can guarantee integrity to the required level or not. As previously stated, Equation (4.4) can be corrected and replaced with Equation (4.19) for more proper results. Still this way only the OMT is implemented as a detection test. Following the reasoning in the previous section, as a next step a V_{slope} should be developed but by applying the w-test.

In any case, neglecting for the moment how it is computed, the V_{slope} can be used to assess the availability of integrity for a particular geometry. A threshold can be set for the V_{slope}: when the threshold is exceeded integrity is declared unavailable for the satellite geometry at hand. This way an algorithm to prune out bad geometries can be implemented.

In [107] the formula for the Vertical Protection Level (VPL) is set by the maximum value reached by the slope (here is reported the formula as it appears in Walter and Enge's paper):

$$VPL \equiv \max_i(V_{slope_i})k + k_{MD}\sigma_{\hat{x}_3} \quad i = 1,2,\ldots,m \tag{4.25}$$

A V_{slope_i} is in fact computed for each satellite and the maximum value is considered. The formula can be connected to Figure 4.9. The diagram shows the distribution of the test statistic

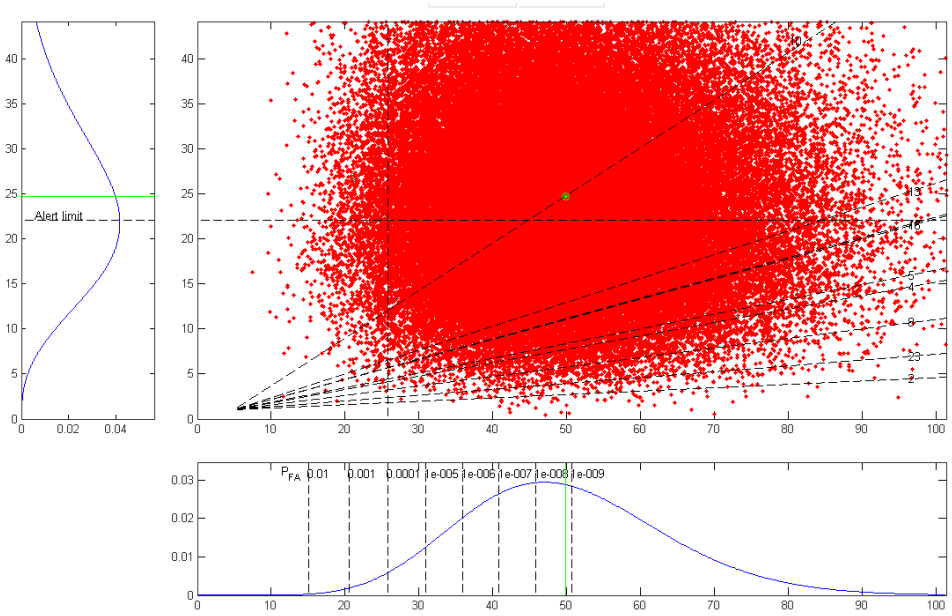


Figure 4.5: Simulation result for a 25m error on PRN10.

and vertical position error and linked to it the quantities that show up in the above equation. The PL defines the confidence interval, symmetric around the position estimator, for the actual aircraft position. k is the threshold of the test statistic, and $\max_i(V_{slope_i})k$ is the corresponding bias in the vertical position estimation. Hence $k_{MD}\sigma_{\hat{x}_3}$ represents the margin between the expectation of the error in the vertical position corresponding to the expectation of the test statistic equal to its critical value, and the threshold set by the PL. k_{MD} stands for the number of standard deviations of normal distribution that yield a tail probability of P_{MD}. As long as $PL < AL$ (Alert Limit), the integrity is guaranteed. Therefore, the Alert Limit defines a threshold for the VPL and, using formula 4.25, a threshold for the $\max_i(V_{slope_i})$.

Note first of all that P_{MD} should be set a-priori, and no method to do it is given in [107]. The standard definition of P_{MD} as in [95]:

$$P_{MD} = \beta = P(\text{No Alarm} \mid \text{Fault}) \tag{4.26}$$

equivalent to the definition given in Equation (3.19), cannot apply correctly in this case, since in Figure 4.9 this quantity seems to mainly relate with the position error distribution.

In view of successive implementation of the Standard RAIM algorithm (full algorithm given in Section 4.1.4) we give here an interpretation for this P_{MD} and how it should be computed. Let us refer to it with P'_{MD}, to distinguish it from the actual P_{MD}, defined as in Equation (4.26). The P'_{MD} can be obtained as:

$$P'_{MD} = P^{req}_{\text{HMI}} / \sum_{1}^{N_{sat}} P_{sat_i} \tag{4.27}$$

with P_{sat_i} a-priori probability of hazardous fault in satellite i as given in Table 4.4 (ARAIM input). The rationale for this choice is given by the following approximated relation:

$$P_{\text{HMI}} \approx p_a \cdot \beta \cdot P_{PF|H_a}$$

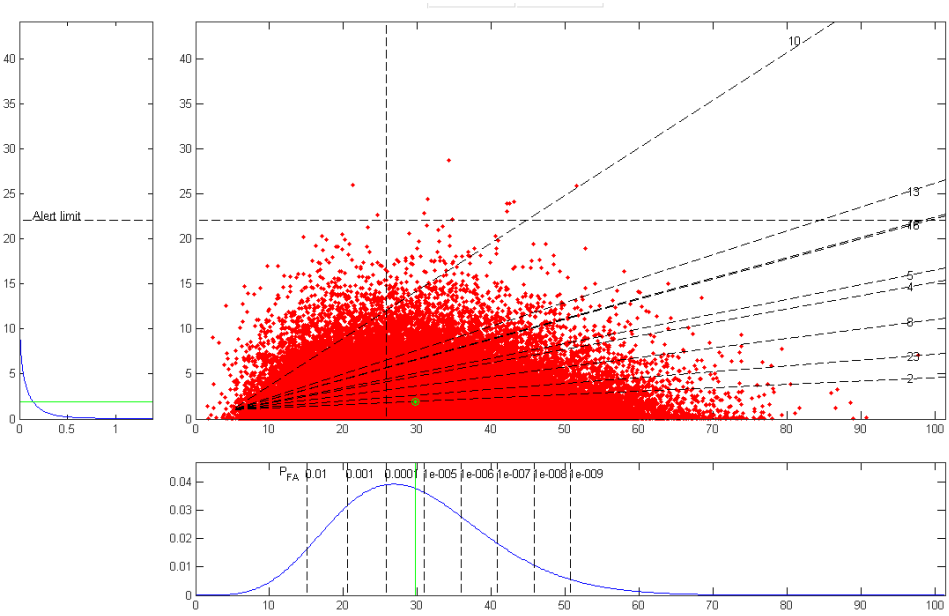


Figure 4.6: Simulation result for a 15m error on PRN2.

where p_a is the prior probability of occurrence of a fault, that represents the alternative hypothesis H_a, $\beta = P_{MD}$ is the (actual) probability of Missed Detection and $P_{PF|H_a}$ is the Probability of Positioning Failure, that is the probability of the error in the solution exceeding the AL (uncorrelated with the event of non-detection), under the assumption of a failure occurring. This approximation, already proposed in Section 3.7, neglects the risk that is present in the fault-free hypothesis (which is very small), therefore it is actually a lower bound for the P_{HMI}. The correct relation would be:

$$P_{\text{HMI}} = p_a \cdot \beta \cdot P_{PF|H_a} + p_0 \cdot (1-\alpha) \cdot P_{PF|H_0}$$

where $p_0 = 1 - p_a$ is the probability that no failure would occur (null hypothesis H_0) and $(1-\alpha)$ is the probability of no detection in that case; the second term is therefore assumed to be negligible. Assuming that only single satellite failures are possibly occurring, we have $p_a \leq \sum_i P_{sat_i}$ (with the assumption of failures being mutually exclusive events, and that only single satellite faults are possible). In the following the VPL and the P_{HMI} are computed in such a way that $P'_{MD} P_{PF} \leq P^{req}_{\text{HMI}}$ (Equation (4.31)), so with the choice for P'_{MD} in Equation (4.27) such inequality is equivalent to $P_{\text{HMI}} \leq P^{req}_{\text{HMI}}$.

Another, and probably the main, comment to Figure 4.9 is that the distribution of test statistic and vertical position error is shown as an ellipsoid centered at a value of the test statistic equal to the threshold k. This is in fact the only case to which the relation in equation can apply. Considering the one-dimensional case as in the Weighted RAIM paper, x_3 standing for the vertical position, the probability of occurrence of HMI (P_{HMI}) is defined:

$$P_{\text{HMI}} = P(\underline{T} < k \cap |\underline{\hat{x}}_3 - x_3| > \text{AL}) \tag{4.28}$$

and here we only base the computation on:

$$P_{\text{HMI}|H_a} = P(\underline{T} < k \cap |\underline{\hat{x}}_3 - x_3| > \text{AL} | H_a) \tag{4.29}$$

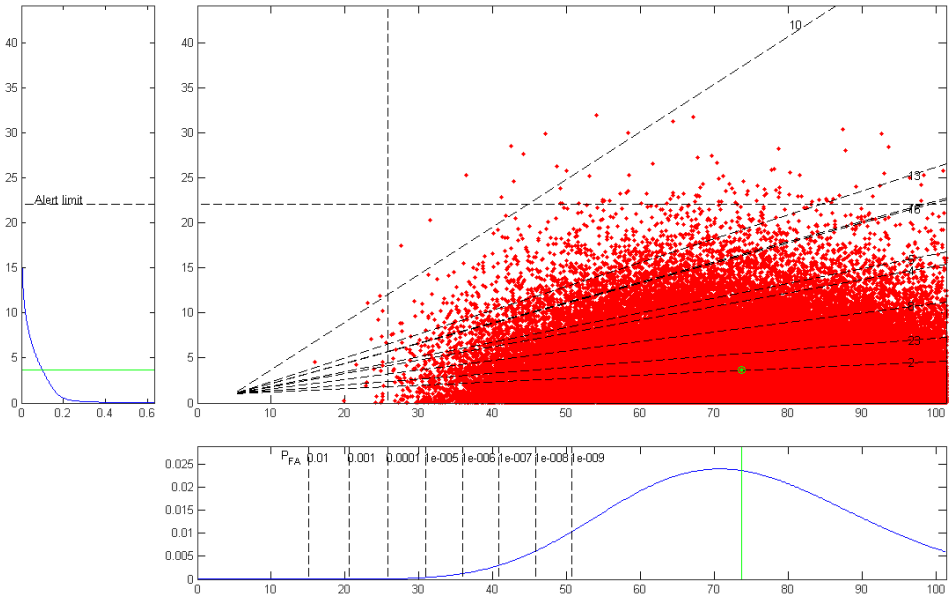


Figure 4.7: Simulation result for a 25m error on PRN2.

with the alternative hypothesis H_a reading: $\nabla\hat{x}_3 \neq 0$ and fixed to a specific value. That is, the bias in the vertical measurement is fixed to the value corresponding to a test statistic equal to the threshold k in the absence of noise. We show in Appendix H that Equation (4.25) can in fact be obtained only in this particular scenario (and with further approximations).

Figure 4.10 instead shows an example of an actual case: the simulated realizations of the vertical position error (on the vertical axis) against the test statistic (on the horizontal axis), in the case of a bias present in the measurements from satellite PRN10. Occurrences of Hazardous Misleading Information can be visualized on the graph: such events are the dots falling in the colored area labelled with 'HMI', in the left-up corner. The situation can thus be much different from the particular case in Figure 4.9. Note furthermore that the error $|\underline{\hat{x}} - x|$ is not Gaussian distributed but half-normal, as discussed previously in this Section.

Two main remarks can thus be made on the approach described:

- In this approach only one size of all the possible measurement range biases is considered, i.e. the bias corresponding to the case that the expected value of the test statistic equals the threshold k. There is no proof this is the worst case scenario or the most probable occurrence.
- No proper explanation on which value to adopt for P_{MD} is given, and what meaning is given to this quantity.

Alternative approaches that could possibly solve these issues are the following:

- A calculation similar to the previous, applied to a single bias size, could be made for all (or many) possible bias sizes; the one that presents the worst results (highest P_{HMI}) would then be taken in consideration (worst case scenario).
- It would be even possible to consider probability distribution functions for the range biases, especially if any assumption can be made on the effects of occurring faults.

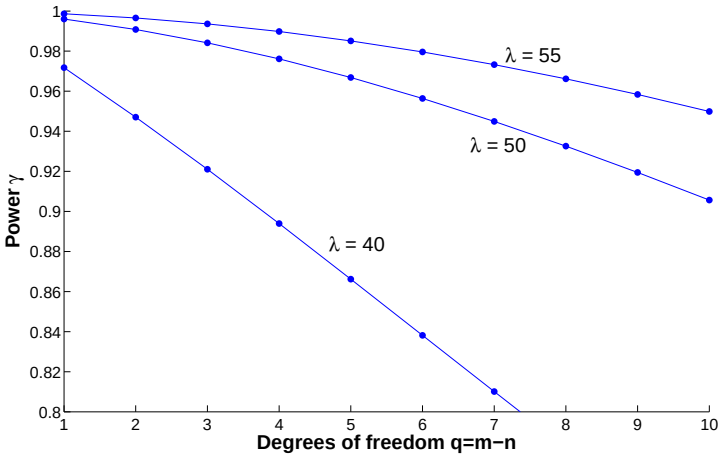


Figure 4.8: Detection power against number of degrees of freedom of the test distribution $q = m - n$, for different values of the non-centrality parameter λ

This approach would allow to weight differently the possible occurring faults, on the basis of a priori assumptions, but increases by far the computation complexity.

Another point to stress was already discussed in previous section: the test used to calculate the V_{slope} for each satellite is the OMT; instead the w-test should be used. Each w-test can generate, for each satellite, one V_{slope}, that is to be preferred in the algorithm to prune out bad geometries, to determine integrity availability.

4.1.3. SUMMARY AND CONCLUSIONS

The standard and most referenced [72] approach to integrity developed in [107] and implemented in current integrity algorithms (for low precision lateral navigation only [92]) has been described and reviewed. The main shortcomings have been pointed out and possible solutions have been proposed. In particular:

- An error in the derivation of the Cumulative Distribution Function for the test statistic used as OMT has been corrected in Section 4.1.2, Equation (4.5).
- Some approximations for the calculation of the relation between range error and position error (V_{slope} in Equation (4.4)) have been revisited and the use of V_{slope} has been inserted in the general theory of detection and validation of linear models developed by TU Delft (Section 4.1.2). Also in Section 4.1.2 a simple relation between range error and position error together with a new definition for the V_{slope} (Equation (4.19)) have been proposed. The validity of the relation has been confirmed by the results of a simulation based on an actual GPS geometry.
- It has been pointed out that the Overall Model Test (OMT) is not the most powerful test statistic for the GPS application (Section 4.1.2). The assumption that only one satellite would fail at a time justifies the use of w-tests as detection algorithm. The advised testing procedure is to run in parallel an OMT and a w-test for each satellite possibly failing. The choice for the thresholds for each test has to be made coherently with the requirement of False Alarm probability.

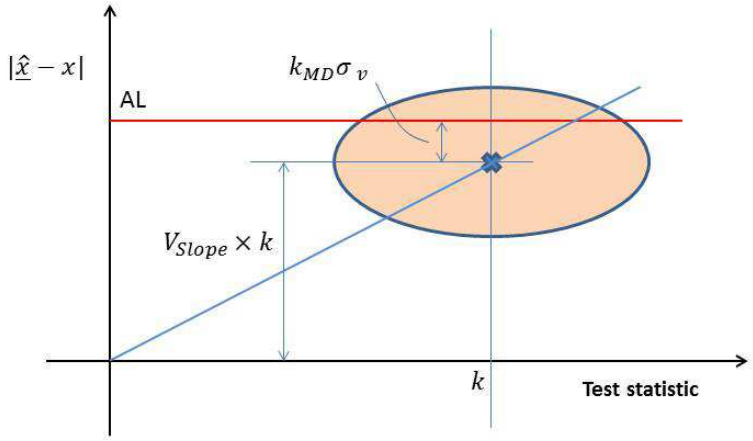


Figure 4.9: Visual representation of the V_{slope} and of the method to determine the VPL, with reference to Equation (4.25).

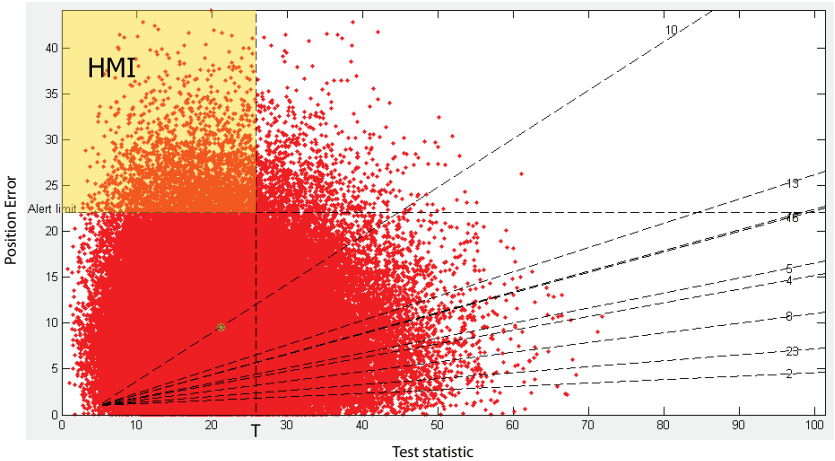


Figure 4.10: Simulated results for vertical position errors (on the vertical axis) and test statistic realizations (on the horizontal axis) in the case of bias present in the measurements from satellite PRN10.

- Finally the fundaments of the algorithm to prune out bad geometries by means of the V_{slope} have been analyzed, and its flaws pointed out (Section 4.1.2). In particular some lack of clarity has been found in the use of a fundamental term (Probability of Missed Detection), and the assumption on the bias size affecting the faulty system has been found unjustified. A definition for the P_{HMI} has been provided and alternatives to the previous use of the V_{slope} have been proposed.

Table 4.3: Specific input to Standard RAIM (complement to Table 7.3).

Parameter	Description	Value
P'_{MD}	Maximum allowed probability of non-detecting a hazardous fault	A-priori

4.1.4. STANDARD RAIM ALGORITHM

The full Standard RAIM algorithm is here reported in detail. The baseline input for the algorithm, common to all the other RAIM algorithms, is given in Tables 7.1, 7.3 and 7.4 (the values reported are the ones employed in the numerical simulations in Chapter 7).

1. Input (extra over Tables 7.1, 7.3 and 7.4) is in Table 4.3. *This P'_{MD} is computed in the author's implementation (for the simulations in Chapter 7) as:*

$$P'_{MD} = P^{req}_{\text{HMI}_{ver}} / \sum_{1}^{N_{sat}} P_{sat_i}$$

with P_{sat_i} a-priori probability of hazardous fault in satellite i as given in Table 4.4 (ARAIM input). This P'_{MD} is only employed for the computation of the VPL, and not for the $P_{HMI_{ver}}$ computation.

2. Check if the redundancy is sufficient:

$$\begin{cases} \text{Continue procedure if } N_{sat} > 3 + N_{const} \\ \text{Declare Alert if } N_{sat} \leq 3 + N_{const} \end{cases} \tag{4.30}$$

3. Compute the threshold for the $\underline{T} = \sqrt{\underline{WSSE}}$ statistic:

$$k = \sqrt{\chi^{2inv}_{N_{sat}-3-N_{const}}(P^{req}_{FA'_{ver}})}$$

where $\chi^{2inv}_{N_{sat}-3-N_{const}}$ is the inverse of the cumulative distribution function of a central χ^2 distribution with $N_{sat}-3-N_{const}$ degrees of freedom and $P^{req}_{FA'}$ is defined in Table 7.3.

4. Compute a Missed Detection threshold k_{MD}:

$$k_{MD} = \Phi^{-1}(1 - P'_{MD})$$

with Φ being the cumulative distribution function of a zero mean standard Gaussian distribution. Use of just a one-sided integral means neglecting the probability of occurrence of a dangerous position estimation error on the side opposite to the measurement bias direction influence, see the following Equation (4.31).

5. Determine geometry integrity availability:

   (a) Compute V_{slope_i} for each satellite i:

$$V_{slope_i} = \frac{|S_{3i}|\sigma_{y_i}}{\sqrt{1 - P_{Aii}}}$$

   with $S = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$, $P_A = A(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$ and $\sigma_{y_i} = \sqrt{Q_{yy_{ii}}}$.

   (b) Compute Vertical Protection Level (VPL) and/or the $P_{HMI_{ver}}$:

$$\begin{aligned} VPL &= \max_i(V_{slope_i})k + k_{MD}\sigma_{\hat{x}_3\hat{x}_3} \\ P_{\mathrm{HMI}_{ver}} &= (\textstyle\sum_i P_{sat_i})\left(1 - \Phi\left(\frac{VAL - \max_i(V_{slope_i})k}{\sigma^2_{\hat{x}_3\hat{x}_3}}\right)\right) \end{aligned} \tag{4.31}$$

   where $\sigma_{\hat{x}_3\hat{x}_3}$ is the standard deviation of the vertical component (third component) of the parameter estimator $\underline{\hat{x}} = S\underline{y}$ and Φ being the cumulative distribution function of a zero mean standard Gaussian distribution. See Section 4.1.2 and [107] for an explanation of these formulas. The formula for the $P_{\mathrm{HMI}_{ver}}$ is obtained directly from the VPL equation.

   (c) Check if the integrity is guaranteed:

$$\begin{cases} \text{Continue procedure if } VPL \leq VAL \\ \text{Declare Alert if } VPL > VAL \end{cases} \tag{4.32}$$

   or equivalently:

$$\begin{cases} \text{Continue procedure if } P_{HMI_{ver}} \leq P^{req}_{HMI_{ver}} \\ \text{Declare Alert if } P_{HMI_{ver}} > P^{req}_{HMI_{ver}} \end{cases} \tag{4.33}$$

6. Compute test statistic T:

$$\underline{T} = \sqrt{\underline{\hat{e}}^T Q_{yy}^{-1} \underline{\hat{e}}}$$

7. Check consistency of the measurements: compare the computed $\underline{T}$ statistic with the threshold k. If:

$$T > k$$

then an Alert is declared. Otherwise standard operations can continue with the parameter estimate $\hat{x} = Sy$.

4.2. ARAIM

As an advanced version of standard RAIM, ARAIM is intended to provide integrity monitoring at user level, that is simply exploiting the GNSS signals available at the receiver with minimum aid from ground stations or other auxiliary means. The redundancy of the GNSS is exploited to self-check the 'health' of the system. The integrity monitoring provided by the RAIM would finally increase the availability and continuity of the navigation service.

As of today, the main reference in the topic of integrity for aviation is [107], with the Standard RAIM reviewed in the previous Section. As we pointed out, Standard RAIM was developed for the single GPS constellation and has been found generally suboptimal, even though presenting a very practical and efficient approach. With respect to Standard RAIM, according to [11] ARAIM provides the following improvements:

- Standard RAIM is not proven to be always conservative. ARAIM instead provides a proof of safety ([11],[10]).
- Standard RAIM is not really tailored for the explicit computation of the integrity risk (it is based mainly on probability of False Alarm and Missed Detection), whereas ARAIM is ([81]).
- Standard RAIM does not employ the best test to identify possibly faulty satellites, resulting in lower detection power than theoretically possible. ARAIM employs different tests, which are more tailored to detecting faults that have sensible impact on the position estimate ([17],[13]).
- Standard RAIM accounts only for faults in one satellite per time (single failure), whereas ARAIM is designed to account for multi-dimensional faults as well ([11],[10]).
- Standard RAIM is designed only for GPS whereas ARAIM is designed to fully exploit the potential of the multi-constellations GNSS available ([11],[10]).
- Standard RAIM is based on single frequency observations, whereas ARAIM foresees the use of dual-frequency observations to remove first order ionospheric delay ([11],[10]).

Finally, ARAIM aims at fulfilling the requirements for vertical guidance in demanding approaches as LPV200 or Cat I and replacing SBAS-GBAS. This review of ARAIM is mostly based on [81], [11], [8], [10], [12], [26], [17], [13].

4.2.1. ARAIM BASIC CONCEPTS

As mentioned in the introduction, ARAIM is based on the use of Multi-Constellation GNSS and on dual frequency measurements. ARAIM also relies on a ground system that provides nominal performance and fault rates for all the constellations (Integrity Support Message, ISM), as described in [11], [8]. The use of this further information from a ground system is foreseen by second generation RAIM concepts, and is deemed necessary to comply with the safety requirements. The ISM can furthermore have quite long latency times (one message per several minutes [8]).

From a statistical point of view, ARAIM is based on the following concepts:

- **Multiple Hypothesis approach with a-priori probabilities**: the measurement system is supposed to be in one out of a set of different possible states described by multiple hypotheses, to each of which is assigned an a-priori probability of occurrence. The fault-free state is the null hypothesis, whereas an alternative hypothesis is considered for each anticipated failure mode, that is for each combination of satellites failures to which is assigned a significant prior probability. The P_{HMI} is computed by the sum of the P_{HMI}s under the different hypotheses weighted on the base of their prior probabilities.

- **Solution Separation (SS) as test statistics**: to discriminate between hypotheses, to eventually exclude faulty measurements, the difference between the position solutions under the different alternative hypotheses and the null hypothesis is computed and used as a test statistic. For each alternative hypothesis considered a difference vector (SS) is computed and a test is run for each of the position components of the vector.

These two basic concepts are explained in the following.

MULTIPLE HYPOTHESIS APPROACH

The ARAIM Multiple Hypothesis approach is based on the type of distribution for the observable introduced in Section 2.7. With respect to that baseline model, further assumptions are made on the type of anomalies that can affect the system (a maximum number of satellites simultaneously failing is set, and constellation faults are introduced), and specific approximations are introduced to simplify the computation of the P_{HMI}.

As previously mentioned, an alternative hypothesis is considered for each combination of satellites failures. First it is necessary to compute the maximum number $N_{sat,max}$ of satellites simultaneously failing to be monitored, then an alternative hypothesis will be raised for each combination of 1 to $N_{sat,max}$ satellites. The number $N_{sat,max}$ is computed on the basis of the a-priori probability assigned to the fault of a single satellite.

In fact an a-priori probability $P_{sat,j}$ is assigned to the event of satellite j failing, and the probability of a combination — let us refer to it as combination i, which subset of failing satellites is indexed by vector idx_i — of r satellites failing is computed as $p_i = \prod_{j \in idx_i} P_{sat,j}$ (the failures of the satellites are treated as independent events). When r increases the probability of the corresponding combination of faults decreases, therefore after a certain value it can be considered negligible.

Furthermore, different alternative hypotheses are also considered for the events of constellation failures (one alternative hypothesis for each constellation part of the system), and to each of these hypotheses an a-priori probability of fault, P_{const}, is set ahead.

The P_{HMI} can be computed under each different hypothesis. The total integrity risk results as a weighted sum of the $P_{\text{HMI}|H_i}$ under the different hypotheses, where the weights are the prior probabilities assigned to the different hypotheses. This approach was described in Section 2.7. We defined the random variable $\underline{H}$ representing the state of the system, and its possible realizations: H_0 the fault-free (null) hypothesis and H_i the alternative hypotheses, representing the different possible types of anomalies affecting the system, with $i = 1, \ldots, N_{H_a}$ (N_{H_a} the total number of alternative hypotheses considered). Once prior probabilities are associated to the occurrence of the different hypotheses, the variable $\underline{H}$ has a

prior PMF:

$$\underline{H} \sim \begin{cases} P(\underline{H} = H_0) = p_0 \\ P(\underline{H} = H_1) = p_1 \\ \vdots \\ P(\underline{H} = H_{N_{Ha}}) = p_{N_{Ha}} \end{cases} \quad \text{cf. Equation (2.37)}$$

The distribution of the observable $\underline{y}$ depends therefore on the state of the system, so that we can write for the marginal distribution of $\underline{y}$ (equivalent to Equation (2.38)):

$$\underline{y} \sim p_0 \cdot f_{\underline{y}|H_0} + \sum_{i=1}^{N_{Ha}} p_i \cdot f_{\underline{y}|H_i} \tag{4.34}$$

with the conditional distributions being multivariate normal distributions. In particular:

$$\begin{cases} f_{\underline{y}|H_0} = N(Ax, Q_{yy}) \\ f_{\underline{y}|H_i} = N(Ax + \nabla y_i, Q_{yy}) \quad \forall i = 1, \ldots, N_{Ha} \end{cases} \tag{4.35}$$

where ∇y_i is the bias (due to an anomaly) affecting the set i of faulty observations, and is unknown. The distribution of $\underline{y}$ is therefore a mixture of Gaussian distributions, weighted by the prior probabilities of faults.

The position solution given to the pilot in standard operation is:

$$\underline{\hat{x}}_0 = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} \underline{y} \tag{4.36}$$

where A is the satellite geometry matrix relative to the epoch considered and Q_{yy} is the variance matrix of the observable used for integrity[2] This is in fact the BLUE for the unknown position x in the fault-free hypothesis. Under each hypothesis the $P_{\text{HMI}|H_i}$ (the P_{HMI} under the hypothesis H_i, see Equation (2.27) for the P_{HMI} definition) is computed as:

$$P_{\text{HMI}|H_i} = P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} \cap \underline{y} \in \Omega | H_i) \tag{4.37}$$

where x is the unknown true position, Ω_{AL} is the 'integrity region' around the true position whose boundaries are the Alert Limits (AL) and Ω is the acceptance region for the observable, that includes all the measurements for which the tests run do not lead to exclusion or Alert. Due to the unknown nature of the $\underline{y}$ distribution as assumed in Equation (2.36), this $P_{\text{HMI}|H_i}$ *cannot* be computed. It is possible nevertheless to compute approximations or upper bounds for it, making some assumptions on the size of the biases affecting the measurements ∇y.

In ARAIM the $P_{\text{HMI}|H_i}$ in Equation (4.37) is computed as:

$$P_{\text{HMI}|H_i} = P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} \cap \underline{T}_{SS_i} \in \Omega_{T_i} | H_i) \tag{4.38}$$

where $\underline{T}_{SS_i}$ is the vector of SS test statistics run for the alternative hypothesis H_i and Ω_{T_i} is its acceptance region, bounded by the thresholds k_i, which are the thresholds for the first iteration of the algorithm. Use of this formula is not fully correct and it is interpreted as an approximation, in addition to the one necessary to tackle the unknown nature of the bias. In fact Equation (4.38) is not precise for two reasons:

[2] In the ARAIM algorithm two different variance matrices for the observable are employed, one for integrity computation and one for accuracy computation. Nevertheless in this dissertation only one variance matrix will be employed, for both purposes. This approach simplifies the problem and is in line with the standard TU Delft approach to the testing problem.

1. all the possible iterations of the algorithm should be taken into account in the computation of the $P_{\text{HMI}|H_i}$ (which is an a-priori probability), because each new iteration adds a further contribution to the risk (after an exclusion it is possible that no Alert will be given). This consideration was made already in Section 3.7. This issue is also addressed in [51].

2. all the tests, for all the alternative hypotheses, should be taken into account in the computation of $P_{\text{HMI}|H_i}$ (say $\underline{T}_{SS_s} \in \Omega_{k_s}$ for all $s = 1, 2, \ldots, N_{H_a}$ and not just the specific index i). What happens for instance if a wrong detection occurs, i.e. a test different than $\underline{T}_{SS_i}$ leads to rejection? A decision to reject the null hypothesis and go for hypothesis H_i can also come from testing for $H_{s\neq i}$. The issue of taking into account the following steps of exclusion and further detection rises again.

More comments on these shortcomings are in Section 4.2.4.

At this point the approximation to tackle the unknown nature of the bias sizes comes into place. An upperbound for Equation (4.38) is derived in ARAIM as described in Appendix J. The upperbound is fundamentally based on the following two inequalities:

$$P_{\text{HMI}|H_i} = P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} \cap \underline{T}_{SS_i} \in \Omega_{T_i} | H_i) \leq P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} | \underline{T}_{SS_i} \in \Omega_{T_i}, H_i)$$

$$P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} | \underline{T}_{SS_i} \in \Omega_{T_i}, H_i) \leq P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} | \underline{T}_{SS_i} = k_i, H_i) \tag{4.39}$$

The equality in the first line would hold in case $P(\underline{T}_{SS_i} \in \Omega_{T_i}) = 1$. This is equivalent to assume that the tests are always accepted. The second line instead means assuming that the tests are *not only* always accepted, but the test statistics always assume the most extreme values that do not lead to rejection. These approximations, especially the second one, are quite crude (for any possible actual bias size the probability of occurrence of a particular value of the test statistic is infinitesimal), and result in quite loose upper bounds for the P_{HMI}. I believe these approximations are made because of the complexity of computing the probability of Missed Detection for all the tests run and for all possible bias sizes. The example in Appendix K shows how these approximations lead to more conservative results than the assumption of a worst-case bias scenario.

Figure 4.11 describes the Multiple Hypothesis concept, as proposed in [81], [8], for a single position dimension (for instance the vertical, the concept being the same for the other dimensions). The position estimator $\underline{\hat{x}}_0$ is supposed to be distributed around the true position x with a Gaussian distribution under the null hypothesis H_0. The other Gaussian distributions in the figure are instead the distributions of $\underline{\hat{x}}_0$ in each case of the corresponding alternative hypothesis holding true, they are centered at a distance from x corresponding to the threshold value of each test statistic (SS statistic) and are weighted with the a-priori probability assigned to the corresponding alternative hypothesis. In the figure the threshold value is indicated as $|\hat{x}_a - \hat{x}_0|_{max}$ because it is the maximum value that the SS statistic can assume without leading to an exclusion or an Alert. The P_{HMI} is computed as the sum of the integrals of all $\underline{\hat{x}}_0$ distributions outside the ALs (around the true position x), i.e. as the integral of the full (multimodal) marginal distribution of $\underline{\hat{x}}_0$, not conditioned on the hypothesis, which results out of the weighted sum of the conditional distributions. The biases in the positioning that correspond to the displacement of the Gaussian distributions in the alternative hypotheses are not known a-priori, but the use of $|\hat{x}_a - \hat{x}_0|_{max}$, the threshold value of the corresponding SS test, is justified by the approximations made in Equation (4.39), that

allow to replace the prior distribution with the one conditioned after the measurements. As mentioned more details on this approximation are in Appendix J.

The P_{HMI} for a single component, vertical for instance, is therefore computed in ARAIM as:

$$P_{\text{HMI}_{ver}} = 2\Phi\left(\frac{-VAL}{\sigma_{\hat{x}_0,3}}\right) + \sum_{i=1}^{N_{Ha}} p_i \Phi\left(\frac{-VAL + k_{i,3}}{\sigma_{\hat{x}_i,3}}\right) \tag{4.40}$$

where $k_{i,3}$ is the threshold value for the SS test for the i^{th} alternative hypothesis and for the vertical component (3^{rd}), Φ is the cumulative distribution of a standard normal distribution, $\sigma_{\hat{x}_0,3}$ and $\sigma_{\hat{x}_i,3}$ are the standard deviation of the vertical component respectively of the position solution $\underline{\hat{x}}_0$ under the null hypothesis and $\underline{\hat{x}}_i$ under the i^{th} alternative hypothesis. Note that this formula also contains a small approximation, since the contributions to the risk coming from the tails of the distributions under the alternative hypotheses on the far side from the closest AL are neglected. This formula is a little simplified with respect to the final one adopted in the algorithm (given in Section 4.2.5) which takes into account also nominal biases in the measurements and P_{HMI} allocated to un-monitored faults.

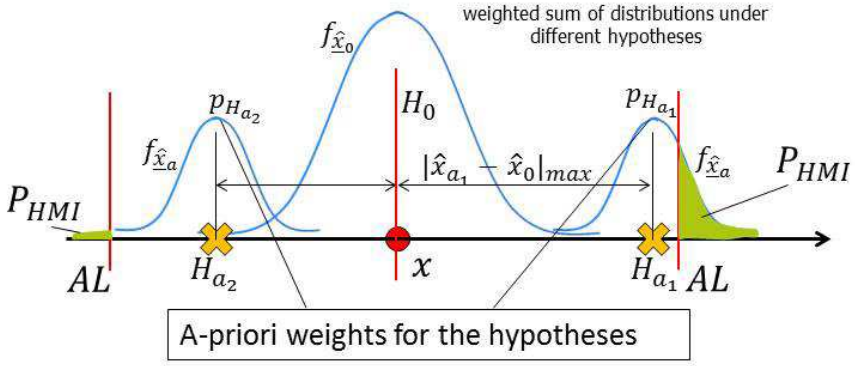


Figure 4.11: Sketch of the ARAIM Multiple Hypothesis concept, from [81], [8]: probability distribution (approximated) of the position estimator in the null hypothesis $\underline{\hat{x}}_0$ once the measurements have been collected. The P_{HMI} is computed as the integral of this distribution over the position domain outside the ALs.

Solution Separation tests

The tests employed in ARAIM are designed to monitor the effect of the faults only in the position domain. These tests were first presented in [17] and justified mainly on the basis of heuristic and empirical reasons. A comparison between SS tests and standard UMPI tests is proposed in [47]: whereas the UMPI test can detect any type of anomaly, the SS focuses only on anomalies that affect the position estimation; nevertheless the results of the two tests are equivalent in case of one-dimensional anomalies.

In the SS approach, the difference between the position solutions under the different alternative hypotheses and the null hypothesis is computed for every alternative hypothesis. Each of these differences constitutes a different SS vector. If we characterize each alternative hypothesis by a different subscript i, any of the Solution Separation vectors can be written as:

$$\underline{T}_{SS_i} = \underline{\hat{\nabla} \hat{x}_i} = \underline{\hat{x}}_0 - \underline{\hat{x}}_i \tag{4.41}$$

where $\underline{\hat{x}}_0$ and $\underline{\hat{x}}_i$ are the position solutions obtained respectively employing the null and the alternative model:

$$\begin{aligned} \underline{\hat{x}}_0 &= (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} \underline{y} = S\underline{y} \\ \underline{\hat{x}}_i &= (A^T Q_{yy_i}^{-1} A)^{-1} A^T Q_{yy_i}^{-1} \underline{y} = S_i \underline{y} \end{aligned} \tag{4.42}$$

where Q_{yy} is the diagonal variance matrix of the observable and $Q_{yy_i}^{-1}$ is obtained from Q_{yy}^{-1} replacing the diagonal elements corresponding to the faulty satellites in hypothesis H_i with 0 (this means giving zero weight to such observations). In fact the position solution in the alternative hypotheses considered by ARAIM can be computed simply removing the faulty observations from the measurement vectors, with a result equivalent to applying the BLUE to the model in Equation (2.34), leading to the formula in Equation (3.11). This is demonstrated in Appendix C.

For each of these SS vectors, the three position components (North, East, Up) are treated separately as different test statistics. The SS tests are run comparing the test statistics just introduced to a corresponding threshold (different for each test). If the threshold is exceeded, the system is believed to be in the corresponding alternative hypothesis, and an adaptation of the model is performed (the suspected faulty satellite(s) is/are excluded from the set of measurements, for instance). The statistical distribution of the test statistics is known and thresholds for the tests statistics are set on the basis of continuity considerations: each test draws from the total continuity budget. In the basic ARAIM algorithm the allocation of continuity among the different alternative hypotheses is not optimized: the continuity budget is equally split among all the possible failure modes. If say $P_{FA_{ver}}$ is the continuity budget (False Alarm rate) allocated for the vertical component, and N_{H_a} is the number of alternative hypotheses considered, then:

$$P_{FA_{ver-test}} = \frac{P_{FA_{ver}}}{N_{H_a}}$$

is the FA rate required to each test for the vertical component. This sub-optimal continuity allocation is pointed out already in [11] and methods to improve the allocation are described in [9]. Simply splitting evenly the overall FA budget among the faults can be quite inefficient, since some faults are highly more likely than others, and may deserve a larger portion of total continuity budget.

If any of the SS tests lead to rejection of the fault-free hypothesis, a measurements exclusion has to be made. To decide how many and which measurements to exclude from the model, a selection method is applied, regardless of which SS test rejected the null hypothesis. To select the subset of satellites most likely to be faulty, a χ^2 statistic is computed for all the subsets i:

$$\chi^2_{ex,i} = \underline{y}^T (Q_{yy_i}^{-1} - Q_{yy_i}^{-1} A (A^T Q_{yy_i}^{-1} A)^{-1} A^T Q_{yy_i}^{-1}) \underline{y} \tag{4.43}$$

where $Q_{yy_i}^{-1}$ is obtained from Q_{yy}^{-1} replacing the diagonal elements corresponding to the satellites reputed faulty in the i^{th} alternative hypothesis with 0. For each size of the subsets, the best candidate for exclusion is the one that realizes the smallest value of the above χ^2 test statistic. The details of theexclusion algorithm are given in the algorithm description (Section 4.2.5). The fact that the SS tests do not automatically individuate the best candidates for exclusion is a peculiar choice of the ARAIM algorithm.

4.2.2. COMPARISON BETWEEN SS TESTS AND UMPI TESTS

A detailed comparison between SS tests and UMPI tests is presented in Appendix I.
UMPI tests were introduced in Section 3.6.1, and are discussed in more detail in Chapter 5.

4.2.3. ARAIM BASELINE ARCHITECTURE

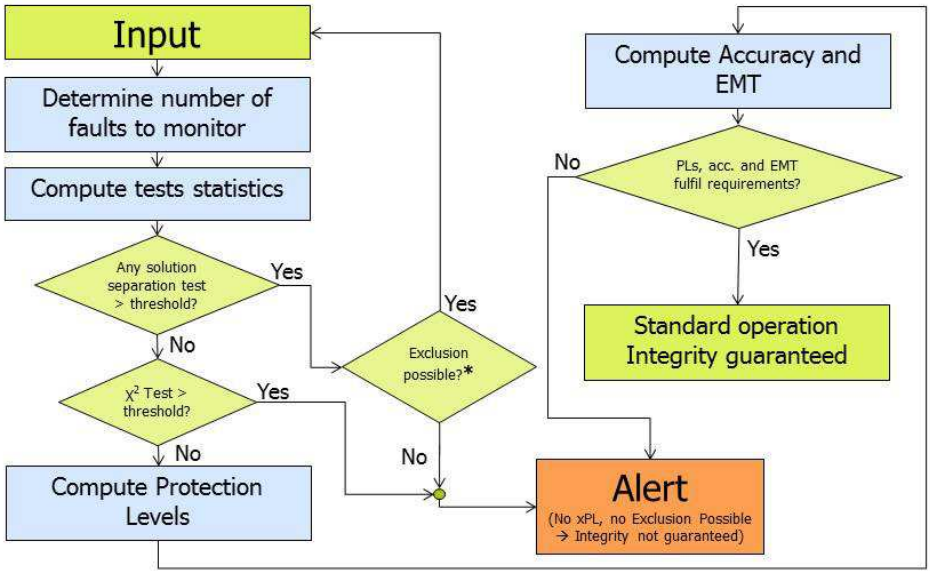


Figure 4.12: ARAIM baseline architecture: block diagram.

The ARAIM algorithm architecture is shown in Figure 4.12. The formulae employed at each stage are shown in next section. Each step is described in the following:

1. Input: main input to the algorithm are the observations y, the geometry matrix A, the variance matrix Q_{yy}, the nominal biases b, and the a-priori fault probabilities $P_{sat,j}$ Ű– via Integrity Support Message (ISM). The list of input extra over the standard one is reported in Table 4.4.

2. Determine the number of faults to monitor: on the basis of the prior probabilities $P_{sat,j}$, the maximum number of satellites simultaneously failing with probability larger than a pre-determined threshold is determined.

3. Compute test statistics: three tests statistics, for each dimension (North, East, Up) are computed for each failure mode, plus one χ^2 statistic for an overall sanity check (equivalent to the OMT). This sanity check is run to account for possible anomalies that cannot be detected by the other tests since these anomalies affect a larger number of measurements and/or are due to unexpected sources. In the standard DIA procedure, the OMT is run *before* the identification tests, cf. Chapter 5. We note that in ARAIM the $P_{FA_{\chi^2}}$ (False Alarm rate) allocated to the sanity check as well as its missed detection probability do not enter in the computation of the P_{HMI}. The $P_{FA_{\chi^2}}$ is considered in the total continuity allocation, but, as for the other tests, the correlation between this test statisitc and the others is neglected.

4. Tests (comparison of test statistics with their threshold): first the solution separation test statistics are compared to their threshold, and if none of the tests leads to rejection of the null hypothesis also the overall check is made. In case at least one solution separation test leads to rejection of the null hypothesis, the algorithm proceeds with adaptation of the model (exclusion of faulty satellites). In case instead the χ^2 test leads to rejection (after acceptance of the SS tests), an Alert would be given directly (unavailability of the navigation service). If no test leads to rejection of the null hypothesis, the algorithm proceeds to the Protection Levels computation step (step 6).

5. Exclusion: the suspected satellites corresponding to the alternative hypothesis reputed most likely are excluded, and afterwards the procedure is started again with the updated model (there are little modifications in the algorithm when it is run after a previous exclusion but these will be discussed later). To find the most likely alternative hypothesis a χ^2 statistic is computed for each alternative hypothesis and the one resulting with the minimum value is selected.

6. Compute Protection Levels (PL): the Multiple Hypothesis approach is applied to determine the PLs, in such a way that the sum of the Probabilities of Hazardous Misleading Information (P_{HMI}) under each different hypothesis is smaller than the total pre-set upper limit for the P_{HMI}. This is done considering the distribution of the position estimator being distributed as a sum of Gaussian distributions, weighted by the prior probabilities assigned to each hypothesis.

7. Check on PLs and on the other performance parameters: if the PLs are too large compared to the requirements for safe flight, an Alert is raised and the system is declared unavailable. If instead the requirements are fulfilled, standard operations can continue and integrity is guaranteed.

Note that after an exclusion has been performed, additional tests statistics are run to account for the eventuality of a Wrong Exclusion (WD), the exclusion of satellites that in fact are not faulty, which can finally affect also the computation of the PLs.

As further remark we add that the ARAIM procedure includes the computation of other two performance parameters, just before step 8: the accuracy a and the Effective Monitoring Threshold (EMT). Accuracy was already defined (in particular as accuracy parameters the 95% and the 99.99999% position error bounds in fault free condition are computed), whereas th e EMT is the 99.999% position error bound holding for any system state (even in case of fault). The EMT is fundamentally the same type of bound as the Protection Levels, but for a looser probability requirement: it serves as a protection from faults that are not large enough (in terms of biases in the measurements) to ensure detection. It makes sense to compute the error bound for a different probability confidence because the distribution of the position estimator is not a unimodal Gaussian distribution when all possible states of the system are considered. In this dissertation anyway the computation of the EMT will not be analyzed.

4.2.4. SUMMARY AND CONCLUSIONS

ARAIM constitutes the new generation RAIM and a huge step forward with respect to the Standard RAIM. On the other hand, as already mentioned, the ARAIM algorithm is still a work in progress, in its more advanced forms, and is currently in experimental phase. In

this dissertation we described only the baseline algorithm (mainly based on [11]). This algorithm is reckoned to be still sub-optimal from different points of view. We list here the main remarks to the baseline algorithm:

- We highlighted already in Section 4.2.1 that not all the possible iterations of the algorithm are taken into account in the computation of the P_{HMI}. Starting from Equation (4.37), the actual $P_{HMI|H_i}$ for a multiple steps algorithm should be computed:

$$\begin{gathered} P_{\text{HMI}|H_i} = P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} \cap \underline{T}^1_{SS} \in \Omega^1_T | H_i) + \\ \sum_{s=2}^{N_{steps,max}} P(\underline{\hat{x}}_s - x \notin \Omega_{\text{AL}} \cap \underline{T}^{s-1}_{SS} \notin \Omega^{s-1}_T \cap \underline{T}^s_{SS} \in \Omega^s_T | H_i) \end{gathered} \tag{4.44}$$

  where $\underline{\hat{x}}_s$, $\underline{T}^s_{SS}$ Ω^s_k are the position solution, the test statistics (all the tests, for any i) and the tests acceptance regions at the s^{th} step of the algorithm and $N_{steps,max}$ is the total number of steps that can be possibly run. In the new paper [14] on ARAIM exclusion the computation of the risk is slightly modified to take into account the exclusion tests that are run after the exclusions are performed. Still the successive iterations of the algorithm (after the first one) are not properly considered for the computation of the risk (see [51] for a possible solution of the problem).

- The RAIM requirement $P^{req}_{FA'}$ on the FA' rate is translated by ARAIM to a requirement on the P_{FA} of the first detection step of the algorithm, as if the algorithm was made of a single test/iteration (see Equation (3.31), when instead $P_{FA_{tot}}$ (for the full procedure) should be computed.

- The computation of the P_{HMI} through Equation (4.40), obtained by means of the approximation described in Section 4.2.1, may, as discussed, result quite conservative (see Appendix K for a comparison with a different approximation).

- The ARAIM employs SS tests instead of UMPI tests. Those tests are analyzed and compared in detail in Appendix I. As result of this analysis, both SS and UMPI tests are not optimal for the integrity problem — therefore the preference for the SS with respect to the UMPI test can be discussed. Furthermore, in case of one-dimensional anomalies the SS tests are redundant: in fact the same statistic is measured three times, one time for each position component. Finally in the ARAIM the SS tests are only a trigger for further tests that will constitute the identification step, so the SS steps are actually used only as detection step. Since different SS tests are run for each possible alternative hypothesis, one would expect that the tests that lead to rejection will automatically identify the possibly dangerous measurements to exclude, therefore the need to run extra identification tests should be justified.

- The ARAIM exclusion process, including tests to check for wrong exclusions, is in experimental phase. Due complexity of the underlying statistical problem, the formulae employed for such steps of the procedure are often subject to modifications. For instance the choice of the threshold in Equation (4.61) of the algorithm description is taken from [11] where it is not justified. In the more recent paper [14] the threshold is instead computed as:

$$\begin{gathered} k_{i,excl,l} = -\sigma_{\nabla \hat{x}_{excl,i},l} \cdot \\ \Phi^{-1}\left(\frac{1-\theta}{2N_{Ha}}\frac{p_0}{p_i}\left(1 - \frac{P_{sat,unmonitored} - P_{const,unmonitored}}{P^{req}_{\text{HMI}_{ver}} + P^{req}_{HMI_{hor}}}\right)\right) \end{gathered}$$

with $P_{sat-thresh}$, $P_{const-thresh}$ and θ are pre-determined input to the algorithm, see Table 4.4 and Section 4.2.5. In this new paper the computation of the risk is slightly modified to take into account the exclusion tests that are run after the exclusions are performed, and the thresholds for the exclusion tests are derived accordingly. As previously mentioned anyway the successive iterations of the algorithm (after the first detection and exclusion) are not considered for the computation of the risk and of the thresholds. Also the parameter θ was not present in the 2012 description [11], but only in the more recent paper [14]. It represents the integrity allocation between confirmed and not confirmed exclusions but it is not well explained how to choose the value for this parameter (should be comprised between 0 and 1).

Table 4.4: Specific input to the ARAIM algorithm (complement to Table 7.3).

Name	Description	Source
$P_{sat,j}$	Prior probability of fault in satellite j	ISM
$P_{const,q}$	Prior probability of a fault of constellation q (several satellites)	ISM
$P_{sat-thresh}$	Threshold for the integrity risk coming from unmonitored satellite faults	Requirement
$P_{const-thresh}$	Threshold for the integrity risk coming from unmonitored constellation faults	Requirement
$P^{req}_{FA'_{\chi^2}}$	Continuity budget allocated to False Alerts for the χ^2 test	Requirement

4.2.5. ARAIM ALGORITHM

The baseline input for the algorithm, common to all the other RAIM algorithms, is given in Tables 7.1, 7.3 and 7.4 (the values reported are the ones employed in the numerical simulations in Chapter 7).

1. Input (in addition to Tables 7.1, 7.3 and 7.4) in Table 4.4.
2. Check if the redundancy is sufficient:

$$\begin{cases} \text{Continue procedure if } N_{sat} > 3 + N_{const} \\ \text{Declare Unavailable if } N_{sat} \leq 3 + N_{const} \end{cases} \tag{4.45}$$

3. Compute the maximum number of satellites simultaneously failing, to be monitored, $N_{sat,max}$:

$$N_{sat,max} = \max(r) : \frac{(\sum_{k=1}^{N_{sat}} P_{sat,k})^r}{r!} > P_{sat-thresh} \quad \text{with } r < N_{sat} - 3 - N_{const} \tag{4.46}$$

The solution of this relation can be found iteratively. Equation (4.46) is presented in [11] (precisely in Appendix C of [11], a complete proof is absent though).

4. Compute the probability of occurrence of an unmonitored satellite fault, $P_{sat,unmonitored}$:

$$P_{sat,unmonitored} = \frac{(\sum_{k=1}^{N_{sat}} P_{sat,k})^{N_{sat,max}+1}}{(N_{sat,max}+1)!} \tag{4.47}$$

5. Compute the maximum number of constellations simultaneously failing monitored, $N_{const,max}$:

   (a) Compute:

$$P_{noconst} = \prod_{q=1}^{N_{const}} (1 - P_{const,q}) \tag{4.48}$$

   the probability of no constellation fault occurring.

   (b)

$$\begin{aligned} &\text{If:} && N_{const} \geq 3 \text{ and} \\ &&& 1 - P_{noconst} - P_{noconst} \sum_{q=1}^{N_{const}} \frac{P_{const,q}}{1-P_{const,q}} > P_{const-thresh} \\ &\text{then:} && N_{const,max} = 2 \\ &\text{else:} && N_{const,max} = 1 \end{aligned} \tag{4.49}$$

6. Compute the probability of occurrence of an unmonitored constellation fault, $P_{const,unmonitored}$:

$$\begin{aligned} &\text{If:} && N_{const,max} = 1 \\ &\text{then:} && P_{const,unmonitored} = 1 - P_{noconst} - P_{noconst} \sum_{q=1}^{N_{const}} \frac{P_{const,q}}{1-P_{const,q}} \\ &\text{else, if:} && N_{const,max} = 2 \\ &\text{then:} && P_{const,unmonitored} = 1 - P_{noconst} \left(1 + \sum_{q=1}^{N_{const}} \frac{P_{const,q}}{1-P_{const,q}}\right) - \\ &&& P_{noconst} \sum_{q_1<q_2} \frac{P_{const,q_1}}{1-P_{const,q_1}} \frac{P_{const,q_2}}{1-P_{const,q_2}} \end{aligned} \tag{4.50}$$

   where the last summation (over $q_1 < q_2$) is a summation over all the possible combinations of two constellations.

7. Compute the total number of fault modes to be considered in the algorithm (number of alternative hypotheses):

$$N_{H_a} = \sum_{r=1}^{N_{sat,max}} \binom{N_{sat}}{r} + \sum_{q=1}^{N_{const,max}} \binom{N_{const}}{q}$$

   where the first term represents the satellite fault modes and the second term the constellation fault modes.

8. Define subsets of satellites to be tested for faults (one subset for each alternative hypothesis):

   For $i = 1, 2, \ldots, N_{H_a}$ build a subset i of r satellites, indexed by an r-components vector idx_i, which contains the satellites reputed faulty in the corresponding

alternative hypothesis. The first $\sum_{r=1}^{N_{sat,max}} \binom{N_{sat}}{r}$ subsets (satellite faults) contain all the possible combinations of r satellites among the complete set, with $r = 1,2,\ldots,N_{sat,max}$. The remaining subsets (constellation faults) instead contain the satellites belonging to the different constellations reputed faulty in the different hypotheses, or to the different combinations of constellations. In this last case to each subset we can assign also a $cidx_i$, a vector containing the indexes of the constellations contained in the subset.

9. Compute the prior probability of occurrence of each fault mode:

$$p_i = \prod_{j \in idx_i} P_{sat,j} \quad \text{with } i \leq \sum_{r=1}^{N_{sat,max}} \binom{N_{sat}}{r}$$

for each satellite fault mode. For the constellation fault modes instead:

$$p_i = \prod_{q \in cidx_i} P_{const,q} \quad \text{with } i > \sum_{r=1}^{N_{sat,max}} \binom{N_{sat}}{r}$$

10. Detection and identification of faults:

(a) Compute SS test statistics:

$$\underline{T}_{SS_i} = \hat{\nabla}\underline{\hat{x}}_i = \underline{\hat{x}}_0 - \underline{\hat{x}}_i \quad i = 1,2,\ldots,N_{H_a} \tag{4.51}$$

with:

$$\begin{aligned} \underline{\hat{x}}_0 &= (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} \underline{y} = S\underline{y} \\ \underline{\hat{x}}_i &= (A^T Q_{yy_i}^{-1} A)^{-1} A^T Q_{yy_i}^{-1} \underline{y} = S_i \underline{y} \end{aligned} \tag{4.52}$$

where $Q_{yy_i}^{-1}$ is obtained from Q_{yy}^{-1} replacing the diagonal elements corresponding to the satellites reputed faulty in the i^{th} alternative hypothesis (subset i, satellites indexed by idx_i) with 0.

(b) Compute SS test thresholds:

A threshold is set for each of the three position components of x (which has $3 + N_{const}$ components), therefore three thresholds for each failure mode i. The thresholds $k_{i,l}$ (for the position components l) are set based on the availability requirement. They are defined by:

$$k_{i,l} = K_{fa,l}\sigma_{\hat{\nabla}\hat{x}_{i,l}} \quad \text{with } l = 1,2,3 \tag{4.53}$$

where:

$$\begin{aligned} K_{fa,1} = K_{fa,2} &= -\Phi^{-1}\left(\frac{P_{FA'_{hor}}^{req}}{4N_{H_a}}\right) \\ K_{fa,3} &= -\Phi^{-1}\left(\frac{P_{FA'_{vert}}^{req}}{2N_{H_a}}\right) \end{aligned} \tag{4.54}$$

with Φ being the left side of the cumulative distribution function of a zero mean unit Gaussian distribution. $\sigma_{\hat{\nabla}\hat{x}_{i,l}}$ is the standard deviation of the SS

test statistic, that is obtained as the square root of the l^{th} diagonal element of:

$$Q_{\hat{\nabla}\hat{x}_i} = (S - S_i) Q_{yy} (S - S_i)^T \tag{4.55}$$

(c) Perform SS tests: For $i = 1, 2, \ldots, N_{H_a}$:

$$\begin{cases} \text{Accept } H_0 \text{ if } |\hat{\nabla}\hat{x}_{i,l}| \leq k_{i,l} \ \ \forall l = 1,2,3 \\ \text{Reject } H_0 \text{ if } \exists\ l\ (l \in \{1,2,3\}) : |\hat{\nabla}\hat{x}_{i,l}| > k_{i,l} \end{cases} \tag{4.56}$$

where $\exists\ l$ means 'exists at least one l'. Note that $l = 1$ and $l = 2$ individuate the horizontal components, whereas $l = 3$ the vertical component. If H_0 is rejected, go to step 9. Otherwise, in case this is the first run of the algorithm (no satellite has been excluded yet) go to step 12, whereas in case an exclusion has already been performed go to step 11.

11. Check if the redundancy is sufficient to exclude satellites:

$$\begin{cases} \text{Continue to step 10 if } N_{sat} > 3 + N_{const} + 1 \\ \text{Declare Alert if } N_{sat} \leq 3 + N_{const} + 1 \end{cases} \tag{4.57}$$

12. Perform Fault exclusion:

    (a) First exclusion attempt:

    To select the subset of satellites most likely to be faulty, i_{ex}, a χ^2 statistic is computed for all the subsets i:

$$\chi^2_{ex,i} = \underline{y}^T (Q^{-1}_{yy_i} - Q^{-1}_{yy_i} A (A^T Q^{-1}_{yy_i} A)^{-1} A^T Q^{-1}_{yy_i}) \underline{y} \tag{4.58}$$

    where $Q^{-1}_{yy_i}$ is obtained from Q^{-1}_{yy} replacing the diagonal elements corresponding to the satellites reputed faulty in the i^{th} alternative hypothesis (indexed idx_i) with 0. For each size r of the subsets, the best candidate for exclusion is determined as:

$$i_{ex_r} = \arg\min_i \{\chi^2_{ex,i} | \text{size}(idx_i) = r\} \quad i = 1, 2, \ldots, N_{H_a} \tag{4.59}$$

    In this way, a candidate subset for exclusion i_{ex_r} is found for each size r of the monitored subsets. First the candidate with $r = 1$ is chosen, i_{ex_1}. In general, at the n^{th} exclusion attempt i_{ex_n} is chosen (the other candidate subsets with $r \neq 1$ are to be saved).

    The measurement corresponding to the faulty satellite of i_{ex_1} is removed from the model. A^{new}, Q^{new}_{yy} and y^{new} are the geometry matrix, the measurement covariance matrix and the observable re-constructed after exclusion: y^{new} does not include the measurements from the excluded satellites, A^{new} is obtained from A taking out the rows corresponding to the excluded satellites and Q^{new}_{yy} instead is obtained from Q_{yy} removing both rows and columns corresponding to the excluded satellites. N_{sat} is also updated accordingly ($N^{new}_{sat} = N_{sat} - 1$. With this new input we move back to step 2 and the algorithm is repeated.

(b) n^{th} $(1 < n \leq N_{sat,max})$ exclusion attempt:
In this case, we go back to the original A, Q_{yy}, y, as they were before any exclusion was performed, say A_0, Q_{yy_0}, y_0. With reference to the candidate subsets for exclusion i_{ex_r} found in the first exclusion attempt, i_{ex_n} is chosen this time. The measurements corresponding to the faulty satellites of i_{ex_n} are removed from the original model, constituted by A_0, Q_{yy_0}, y_0. A^{new}, Q_{yy}^{new} and y^{new} are the geometry matrix, the measurement variance matrix and the observable re-constructed after exclusion: y^{new} does not include the measurements from the excluded satellites, A^{new} is obtained from A_0 taking out the rows corresponding to the excluded satellites and Q_{yy}^{new} instead is obtained from Q_{yy_0} removing both rows and columns corresponding to the excluded satellites. N_{sat} is also updated accordingly. With this new input we move back to step 2 and the algorithm is repeated.

(c) n^{th} $(N_{sat,max} < n \leq N_{sat,max} + N_{const,max})$ exclusion attempt:
In this case, we go back to the original A, Q_{yy}, y, as they were before any exclusion was performed, say A_0, Q_{yy_0}, y_0. With reference to the candidate subsets for exclusion i_{ex_r} found in the first exclusion attempt, if this is step number $n = N_{sat,max} + 1$ the subset i_{ex_r} best candidate for exclusion corresponding to single constellation fault is selected, otherwise the subset i_{ex_r} best candidate for exclusion corresponding to double constellation fault is selected. Again the measurements corresponding to the selected subset are removed from the original model, constituted by A_0, Q_{yy_0}, y_0. A^{new}, Q_{yy}^{new} and y^{new} are the geometry matrix, the measurement variance matrix and the observable re-constructed after exclusion; N_{sat} is also updated accordingly. With this new input we move back to step 2 and the algorithm is repeated.

(d) n^{th} $(n > N_{sat,max} + N_{const,max})$ exclusion attempt:
Alert is declared because a number of satellites larger than allowed would need to be excluded.

13. Check for Wrong Exclusion (WE) (only if exclusion was performed)

(a) Compute WE test statistics
In this step we consider again the original A, Q_{yy}, y, as they were before any exclusion was performed, say A_0, Q_{yy_0}, y_0 (these are necessarily different from the current A, Q_{yy} and y, which are the result of previous exclusions). Compute, for any i:

$$\underline{\hat{x}}_{0i} = (A_0^T Q_{yy_{0i}}^{-1} A_0)^{-1} A_0^T Q_{yy_{0i}}^{-1} \underline{y} = S_{0i} \underline{y}$$

where $Q_{yy_{0i}}^{-1}$ is obtained from $Q_{yy_0}^{-1}$ replacing the diagonal elements corresponding to the satellites reputed faulty in the i^{th} alternative hypothesis (subset i) with 0.
Compute a test statistic for each position component l, and for each i:

$$\underline{\nabla \hat{x}}_{excl,i,l} = \underline{\hat{x}}_{i,l} - \underline{\hat{x}}_{0i,l} \tag{4.60}$$

(b) Compute WE test thresholds:
The threshold for each WE test is computed as:

$$k_{i,excl,l} = -\Phi^{-1}\left(\frac{p_i}{2}\right)\sigma_{\nabla\hat{x}_{excl,i},l} \tag{4.61}$$

with $\sigma_{\nabla\hat{x}_{excl,i},l}$ the square root of the l^{th} element of the diagonal of $Q_{\nabla\hat{x}_{excl,i}}$:

$$Q_{\nabla\hat{x}_{excl,i}} = (S_{0i} - S'_i)Q_{yy}(S_{0i} - S'_i)^T \tag{4.62}$$

where S'_i is obtained from S_i by adding extra columns of zeros in correspondence to the measurements excluded during the previous steps.

(c) Check Exclusion consistency:

$$\begin{cases} \text{Confirm Exclusion if } \underline{\nabla\hat{x}}_{excl,i,l} > k_{i,excl,l} \ \ \forall i,l \\ \text{Do not confirm Exclusion if } \exists\{i,l\} : \underline{\nabla\hat{x}}_{excl,i,l} \leq k_{i,excl,l} \end{cases} \tag{4.63}$$

The computation of the PLs and of the P_{HMI} changes depending on whether or not the exclusion is confirmed by these tests.

14. Check consistency of the measurements

(a) Compute $\underline{T}_{OMT}$ (χ^2 distributed) test statistic:
The $\underline{T}_{OMT}$ statistic is computed as:

$$\underline{T}_{OMT} = \underline{y}^T(Q_{yy}^{-1} - Q_{yy}^{-1}A(A^TQ_{yy}^{-1}A)^{-1}A^TQ_{yy}^{-1})\underline{y} \tag{4.64}$$

(b) Compute $\underline{T}_{OMT}$ test threshold:

$$k_{\chi^2} = \chi^{2inv}_{N_{sat}-3-N_{const}}(1 - P^{req}_{FA'_{\chi^2}}) \tag{4.65}$$

where the operator $\chi^{2inv}_{N_{sat}-3-N_{const}}$ is the inverse of the cdf of a central χ^2 distribution with $N_{sat} - 3 - N_{const}$ degrees of freedom.

(c) Determine overall consistency of measurements:

$$\begin{cases} \text{Accept } H_0 \text{ if } T_{OMT} \leq k_{\chi^2} \\ \text{Reject } H_0 \text{ if } T_{OMT} > k_{\chi^2} \end{cases} \tag{4.66}$$

If H_0 is rejected the sanity check is not passed and an Alert has to be declared.

15. Compute VPL and/or $P_{\text{HMI}_{ver}}$

(a) *Case No Exclusion performed* or *Exclusion performed and Confirmed*
The VPL is the solution of equation:

$$\begin{aligned} &2\Phi\left(\tfrac{b_{0,3}-VPL}{\sigma_{\hat{x}_0,3}}\right) + \textstyle\sum_{i=1}^{N_{Ha}} p_i\Phi\left(\tfrac{k_{i,3}+b_{i,3}-VPL}{\sigma_{\hat{x}_i,3}}\right) = \\ &P^{req}_{\text{HMI}_{ver}}\left(1 - \tfrac{P_{sat,unmonitored}-P_{const,unmonitored}}{P^{req}_{\text{HMI}_{ver}}+P^{req}_{\text{HMI}_{hor}}}\right) \end{aligned} \tag{4.67}$$

where $b_{0,l}$ and $b_{i,l}$ are the (maximum possible) biases on the l^{th} component of the position solution under null and alternative hypotheses due to the nominal biases $b_{nom,j}$ in the pseudoranges, and are computed as:

$$\begin{aligned} b_{0,l} &= \sum_{j=1}^{N_{sat}} |S_{(l,j)}| b_{nom,j} \\ b_{i,l} &= \sum_{j=1}^{N_{sat}} |S_{i,(l,j)}| b_{nom,j} \end{aligned} \tag{4.68}$$

As from Equation (4.67), the integrity risk computed from the algorithm, $P_{\mathrm{HMI}_{ver}}$, can be obtained substituting the VPL with VAL:

$$\begin{aligned} P_{\mathrm{HMI}_{ver}} = &\left(1 - \frac{P_{sat,unmonitored} - P_{const,unmonitored}}{P^{req}_{\mathrm{HMI}_{ver}} + P^{req}_{\mathrm{HMI}_{hor}}}\right)^{-1} \cdot \\ &\left[2\Phi\left(\frac{b_{0,3} - VAL}{\sigma_{\hat{x}_0,3}}\right) + \sum_{i=1}^{N_{Ha}} p_i \Phi\left(\frac{k_{i,3} + b_{i,3} - VAL}{\sigma_{\hat{x}_i,3}}\right)\right] \end{aligned} \tag{4.69}$$

Details on the computation of PL and P_{HMI} described above are in Section 4.2.1.

A method to solve the PL equation is shown below, Section 4.2.6.

(b) *Case Exclusion performed and Not Confirmed*

The Vertical PL (VPL) is the solution of equation:

$$\begin{aligned} &2\Phi\left(\frac{b_{0,3} - VPL}{\sigma_{\hat{x}_0,3}}\right) + \sum_{i=1}^{N_{Ha}} p_i \Phi\left(\frac{k_{i,3} + b_{i,3} - VPL}{\sigma_{\hat{x}_i,3}}\right) = \\ &p_{i_{ex}} P^{req}_{\mathrm{HMI}_{ver}} \left(1 - \frac{P_{sat,unmonitored} - P_{const,unmonitored}}{P^{req}_{\mathrm{HMI}_{ver}} + P^{req}_{\mathrm{HMI}_{hor}}}\right) \end{aligned} \tag{4.70}$$

where $b_{0,l}$ and $b_{i,l}$ are defined in Equation (4.68) and i_{ex} identifies the subset of satellites excluded.

As from Equation (4.70), the integrity risk computed from the algorithm, $P_{HMI_{ver}}$, can be obtained substituting the VPL with VAL:

$$\begin{aligned} P_{HMI_{ver}} = &\, p_{i_{ex}}^{-1} \left(1 - \frac{P_{sat,unmonitored} - P_{const,unmonitored}}{P^{req}_{HMI_{ver}} + P^{req}_{HMI_{hor}}}\right)^{-1} \cdot \\ &\left[2\Phi\left(\frac{b_{0,3} - VAL}{\sigma_{\hat{x}_0,3}}\right) + \sum_{i=1}^{N_{Ha}} p_i \Phi\left(\frac{k_{i,3} + b_{i,3} - VAL}{\sigma_{\hat{x}_i,3}}\right)\right] \end{aligned} \tag{4.71}$$

where i_{ex} identifies the subset of satellites excluded in the previous step.

16. Check if the integrity is guaranteed:

$$\begin{cases} \text{Continue standard operations if } VPL \leq VAL \\ \text{Declare Alert if } VPL > VAL \end{cases} \tag{4.72}$$

Or equivalently:

$$\begin{cases} \text{Continue standard operations if } P_{HMI_{ver}} \leq P^{req}_{HMI_{ver}} \\ \text{Declare Alert if } P_{HMI_{ver}} > P^{req}_{HMI_{ver}} \end{cases} \tag{4.73}$$

4.2.6. VPL NUMERICAL COMPUTATION (ARAIM)

We report here a method (from [11]) to compute numerically the ARAIM VPL. Equation (4.67) can be solved using a half interval search. With:

$$f(VPL) = 2\Phi\left(\frac{VPL - b_{0,3}}{\sigma_{\hat{x}_0,3}}\right) + \sum_{i=1}^{N_{Ha}} p_i \Phi\left(\frac{VPL - k_{i,3} - b_{i,3}}{\sigma_{\hat{x}_i,3}}\right)$$

the equation to solve is:

$$f(VPL) = P^{req}_{\mathrm{HMI}_{ver}}\left(1 - \frac{P_{sat,unmonitored} - P_{const,unmonitored}}{P^{req}_{\mathrm{HMI}_{ver}} + P^{req}_{\mathrm{HMI}_{hor}}}\right) = P^{req}_{\mathrm{HMI}_{ver,adj}}$$

This search can be started with the following lower and upper bounds:

$$VPL_{low,init} = \max\left\{\begin{array}{l} \Phi^{-1}\left(\frac{P^{req}_{\mathrm{HMI}_{ver,adj}}}{2}\right)\sigma_{\hat{x}_0,3} + b_{0,3}, \\ \max_i \Phi^{-1}\left(\frac{P^{req}_{\mathrm{HMI}_{ver,adj}}}{p_i}\right)\sigma_{\hat{x}_i,3} + k_{i,3} + b_{i,3} \end{array}\right\}$$

$$VPL_{up,init} = \max\left\{\begin{array}{l} \Phi^{-1}\left(\frac{P^{req}_{\mathrm{HMI}_{ver,adj}}}{2(N_{Ha}+1)}\right)\sigma_{\hat{x}_0,3} + b_{0,3}, \\ \max_i \Phi^{-1}\left(\frac{P^{req}_{\mathrm{HMI}_{ver,adj}}}{p_i(N_{Ha}+1)}\right)\sigma_{\hat{x}_i,3} + k_{i,3} + b_{i,3} \end{array}\right\}$$

The iterations stop when:

$$|VPL_{up} - VPL_{low}| \leq TOL_{PL}$$

where TOL_{PL} is the tolerance accepted for the PL.

5

DIA PROCEDURE

The Detection Identification and Adaptation (DIA) procedure is an FDE method originally developed by TU Delft for geodetic networks applications. The theory underlying the procedure can be found in [4], [5], [97], [99] and [101]. The DIA method can be readily applied to GNSS navigation applications. As a RAIM algorithm though, the DIA method lacks a straightforward way to compute the P_{HMI}; this computation is strictly required in a RAIM algorithm — mind the definition given in Chapter 2 of the RAIM problem.

5.1. BASIC FORMULATION

5.1.1. LINEAR MODEL

The DIA procedure is based on the statistical hypothesis testing theory for linear models introduced in Section 3.6.1, and presented in more detail for instance in [101]. In Section 3.6.1 we introduced the linear model (to which our GNSS model can be approximated) representing the system in fault-free state:

$$\underline{y} = Ax + \underline{e} \quad \text{cf. Equation (2.2)}$$

with $\underline{e} \sim N(0, Q_{yy})$, where $\underline{y}$ is the vector of measurements (m entries), x is the unknown position vector (n entries), A is the $m \times n$ geometry matrix and $\underline{e}$ the error vector (m entries), which is assumed to be normally distributed. We assigned this model to the null hypothesis H_0, and we introduced an alternative hypothesis H_a, describing the case an anomaly is affecting the system:

$$\begin{aligned} H_0 &: \quad \underline{y} = Ax + \underline{e} \\ H_a &: \quad \underline{y} = Ax + C_y\nabla + \underline{e} \end{aligned} \quad \text{cf. Equation (3.21)}$$

where C_y is a $m \times q$ matrix (the theory for the case C_y a vector was first developed by Baarda, and extended to the case of a matrix by Teunissen) which represents the 'signature' of the errors in the measurements and ∇ is a q sized vector that contains the sizes of the biases in each degree of freedom (q) of C_y.

In view of Detection and Identification of alternative hypotheses, i.e. testing the alternative hypothesis H_a with the error signature described throught the $m \times q$ matrix C_y against the null hypothesis H_0, the Delft school developed the UMPI test statistic, which reads:

$$\underline{T}_q = \underline{\hat{e}}_0^T Q_{yy}^{-1} C_y (C_y^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} C_y)^{-1} C_y^T Q_{yy}^{-1} \underline{\hat{e}}_0 \quad \text{cf. Equation (3.22)}$$

where $\underline{\hat{e}}_0 = \underline{y} - A\underline{\hat{x}}_0$ is the vector of residuals computed considering the null hypothesis holding true ($\underline{\hat{x}}_0$ being the position estimator under the null hypothesis). The vector of residuals is obtained through Best Linear Unbiased Estimation (BLUE), and $Q_{\hat{e}_0\hat{e}_0} = P_A^{\perp} Q_{yy} P_A^{\perp T}$. This is historically the first formulation of the test, presented in [5] for the one-dimensional case. This UMPI test can also be shown to follow from the Generalized Likelihood Ratio (GLR).

The statistic $\underline{T}_q$ can be also written in alternative formulations as shown in [99]:

$$\begin{aligned} \underline{T}_q &= \underline{\hat{e}}_0^T Q_{yy}^{-1} \underline{\hat{e}}_0 - \underline{\hat{e}}_a^T Q_{yy}^{-1} \underline{\hat{e}}_a = ||\underline{\hat{e}}_0||^2_{Q_{yy}^{-1}} - ||\underline{\hat{e}}_a||^2_{Q_{yy}^{-1}} \\ \underline{T}_q &= (\underline{\hat{y}}_0 - \underline{\hat{y}}_a)^T Q_{yy}^{-1} (\underline{\hat{y}}_0 - \underline{\hat{y}}_a) \\ \underline{T}_q &= \underline{\hat{\nabla}}^T C_y^T P_A^{\perp T} Q_{yy}^{-1} P_A^{\perp} C_y \underline{\hat{\nabla}} \\ \underline{T}_q &= \underline{\hat{\nabla}}^T Q_{\hat{\nabla}\hat{\nabla}}^{-1} \underline{\hat{\nabla}} \end{aligned} \tag{5.1}$$

where $\underline{\hat{e}}_a = \underline{y} - A\underline{\hat{x}}_a - C_y\underline{\hat{\nabla}}$ is the vector of residuals computed considering the alternative hypothesis holding true, with $\underline{\hat{\nabla}}$ the BLUE of the bias ∇, whereas $\underline{\hat{y}}_0 = A\underline{\hat{x}}_0$ and $\underline{\hat{y}}_a = A\underline{\hat{x}}_a + C_y\underline{\hat{\nabla}}$ are respectively the vectors of estimators for the means of the observables in null and alternative hypotheses, and $P_A^{\perp} = I - A(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$ is the projector onto the space $R(A)^{\perp}$, $P_A^{\perp} = I - A(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$. As evident in the first expression of Equations (5.1), the UMPI test statistic $\underline{T}_q$ is in fact the measurement-residual squared norm separation. The expression in Equation (3.22) can be particularly handy since it is solely based on the residual vector computed under the null hypothesis $\underline{\hat{e}}_0$.

This test statistic is χ^2 distributed:

$$H_0 : \underline{T}_q \sim \chi^2(q,0) \text{ and } \quad H_a : \underline{T}_q \sim \chi^2(q,\lambda) \quad \text{cf. Equation (3.23)}$$

with noncentrality parameter:

$$\lambda = \nabla^T Q_{\hat{\nabla}\hat{\nabla}}^{-1} \nabla \quad \text{cf. Equation (4.7)}$$

where $Q_{\hat{\nabla}\hat{\nabla}}^{-1} = C_y^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} C_y$.

For any $1 \leq q < m - n$ many different tests can be run, corresponding to alternative hypotheses characterized by different choices of C_y. Only in case $q = m - n$ all the possible choices for a C_y matrix (with rank $m - n$) lead to the same test statistic, as we will see later.

In case C_y is chosen to be a vector ($m \times 1$ matrix), the possible error is restricted to a single dimension, ∇ is a simple scalar and $q = 1$. In this case the corresponding test statistic $\underline{T}_1$ can be equivalently substituted by the w-test $\underline{w}$, which is normally distributed:

$$\underline{w} = \frac{\underline{\hat{\nabla}}}{\sigma_{\hat{\nabla}}} \tag{5.2}$$

and is:

$$H_0 : \underline{w} \sim N(0,1) \text{ and } \quad H_a : \underline{w} \sim N(\nabla w, 1) \tag{5.3}$$

with $\nabla w = \frac{\nabla}{\sigma_{\hat{\nabla}}}$, and the test is two sided, one sided in case $|\underline{w}|$ is considered. It is in fact $\underline{T}_1 = \underline{w}^2$. The test statistic in Equation (5.2) can be written also as:

$$\underline{w} = \frac{C_y^T Q_{yy}^{-1} \underline{\hat{e}}_0}{\sqrt{C_y^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} C_y}} \tag{5.4}$$

as function of the residual vector computed under the null hypothesis $\underline{\hat{e}}_0$, with C_y a vector.

When we assume that failures in one measurement at a time are the major threat, we can adapt the model to this kind of threat, restricting ourselves to take into consideration only those types of errors and testing against them. In this case we would choose each C_y as a canonical unit vector of the space R^m (for instance $C_{y_1} = [1\ 0\ \cdots\ 0]^T$ in case the first measurement y_1 is considered to be faulty).

Of course, to take into account the possibility of any of the measurements failing, m different alternative hypotheses are foreseen. For each of these m hypotheses the corresponding test statistic has to be computed. This procedure is called *data snooping*, and foresees the use of the corresponding w-tests. For each C_{y_i} a canonical unit vector for the i-th component, with $i = 1,2,\ldots,m$, the corresponding $\underline{w}_i$ is computed by Equation (5.2) with $C_y = C_{y_i}$, $\nabla = \nabla_i$ etc. In case Q_{yy} is diagonal, the w-tests for data snooping have the simple formula:

$$\underline{w}_i = \frac{\hat{e}_{0_i}}{\sigma_{\hat{e}_{0_i}}} \quad \text{for } i = 1,2,\ldots,m \tag{5.5}$$

Beside the single satellite fault threats, or some other possible faults with defined signature (q and C_y) that can be expected for any particular reason, there is always to consider the possibility that the system is affected by a general unanticipated inconsistency. Therefore it is standard practice to perform an Overall Model Test (OMT), to protect the system from non anticipated errors/faults of any possible type (except those for which $C_y \in R(A)$, which are never detectable). The OMT is performed allowing the error to assume the maximum possible dimensionality, so that $q = m - n$. In this case $C_{y_{\text{OMT}}}$ is a $m \times m - n$ matrix with rank $m - n$ and $[A \mid C_{y_{\text{OMT}}}]$ is a full rank $m \times m$ matrix, and the corresponding ∇_{OMT} is a $m - n$ components vector. The test statistic reads:

$$\underline{T}_{q=m-n} = \underline{\hat{e}}_0^T Q_{yy}^{-1} \underline{\hat{e}}_0 \tag{5.6}$$

It is $\underline{T}_{m-n} | H_0 \sim \chi^2(m-n, 0)$ and the test is one sided, accept H_0 if $T_{m-n} \leq k_{\text{OMT}}$.

The standard DIA procedure actually foresees the computation of the OMT as the first step (detection), followed then by the identification step in which particular types of error are tested for. This is explained in next Section.

5.1.2. DIA STEPS

The traditional DIA procedure from the Delft school foresees the following steps:

1. **Detection:** once the measurements are collected, the OMT is computed (Equation (5.6)) and compared to the respective threshold k_{OMT}. Note that it must be $m > n$ (the redundancy of measurements is necessary), otherwise the DIA cannot be implemented, and unavailability should be declared.

   (a) If the test is passed, i.e. $T_{m-n} \leq k_{\text{OMT}}$, the system can be considered available and no further action is required.

   (b) If the test is not passed, i.e. $T_{m-n} > k_{\text{OMT}}$, and $m \leq n+1$ (identification is not possible), an alert is declared (unless it is possible to re-do *all* the measurements, in which case the Detection step can be repeated).

(c) If the test is not passed, i.e. $T_{m-n} > k_{\text{OMT}}$, and $m > n+1$ (identification can be possible) we go to the second step.

The Detection step constitutes a safeguard against all possible types of errors (except, as stated above, the cases in which $C_y \in R(A)$); specific tests as w-tests and data snooping assume specifically formulated hypotheses, but are not effective against non anticipated errors.

2. **Identification:** In this step an attempt is made to determine whether the anomaly detected in the previous step can be attributed to some type of anticipated error. These anticipated errors should be characterized by a particular signature, which can be one- or multi-dimensional, i.e. their corresponding C_y in the model of Equation (3.21) should be a matrix $m \times q$ with $q < m-n$. Let N_{H_a} be the total number of specific alternative hypotheses anticipated. Let also $N_{H_{aq}}$ be the number of alternative hypotheses characterised by dimensionality q. For the case $q = 1$ the corresponding optimal detection test is the w-test, C_y in that case may represent a linearized signal refraction model for instance, or be a canonical unit vector for the data snooping case. A $\underline{T}_q$ is therefore computed based on the formula in Equation (5.1.1) for each anticipated alternative hypothesis (characterized by a different C_y), and compared to a threshold k_q. For each $\underline{T}_q$ with same dimensionality q the same threshold k_q is normally used, since all $\underline{T}_q$ have the same distribution under the null hypothesis. With this choice, the maximum of the $\underline{T}_q$ for each dimension q considered is compared to the threshold k_q. The maximum of the $\underline{T}_q$ is chosen since a larger value corresponds to a lower probability that the measurements can be considered to be a sample from the distribution under the null hypothesis. We can have two cases:

(a)
$$\max_i \underline{T}_{q_i} \leq k_q \quad \forall i : 1 \leq i \leq N_{H_{aq}} \quad \forall q \tag{5.7}$$

In this case, unless all measurements can be re-taken (in which case we can go back to step 1), an alert should be declared, since it is not possible to identify a particular type of fault and therefore perform an adaptation of the model, whereas in the previous step an unanticipated error has been detected.

(b)
$$\exists q : \max_i \underline{T}_{q_i} > k_q \quad \forall i : 1 \leq i \leq N_{H_{aq}} \tag{5.8}$$

In this case, we distinguish two cases:

- The threshold was exceeded only for one value of the dimension q: the corresponding $\max_i \underline{T}_{q_i}$ identifies the most likely alternative hypothesis/anomaly affecting the system.
- The threshold was exceeded for more than one value of the dimension q: it is necessary to choose between the corresponding alternative hypotheses. A possible criterion of choice is to pick the one which $\underline{T}_q$ realization is associated with the lowest p-value (see [33], from which especially [75] and [6]). An alternative is to compare the statistics $\underline{T}_q / k_q$, as proposed in [24].

Furthermore, after the anomaly has been identified, the following two cases can occur:

- It is possible (and opportune) to take re-measurements of the identified faulty observations: in this case, new measurements should be taken and we can go back to step 1 with the new set of measurements.
- Re-measurements are not foreseen (this is generally the case for GNSS model): in this case adaptation should be attempted, and we move to step 3.

Case of data snooping. It is very common to consider anomalies that affect only one measurement at a time as more likely and frequent. This is the case of the data snooping, that was described in the previous Section. For data snooping m w-tests are computed, in addition eventually to other possible $\underline{T}_q$'s relating to other specific alternative hypotheses.

Case of Q_{yy} diagonal. In case of Q_{yy} diagonal, the w-tests relative to the data snooping have the simplified expression given in Equation (5.5).

3. **Adaptation:** The model has to be adapted/updated, since it is not possible to have new measurements to replace the previously considered faulty. The most likely alternative state in which the system can be was identified in the previous step. In general, adaptation is made switching the model from the null to the more likely alternative one, identified in previous step. Therefore the model is updated in such a way that the alternative model corresponding to the identified hypothesis becomes the new null hypothesis model. This means:

$$H_{0_{new}} : \underline{y} = Ax + C_{y_{id}}\nabla + \underline{e}$$

where $C_{y_{id}}$ is the C_y associated to the previously identified alternative hypothesis. Once the choice for the new model is made we can move back to step 1 and the procedure is repeated with the new model for the new null hypothesis $H_{0_{new}}$ and with new alternative hypotheses. The new null hypothesis model has now a decreased redundancy compared with the original one, since now there are $n+q$ parameters to be estimated (∇ can be estimated now). The redundancy left is $m-n-q$. Note that with a smaller value for the redundancy the model has become weaker: the precision of the estimators and the reliability of the system have degraded. It is important to keep track of this weakening.

Case of data snooping. In this case switching to the alternative hypothesis is equivalent to excluding one measurement from the model (this equivalence is proved in Appendix C). The measurement corresponding to the maximum w-test must be excluded. Once the identified measurement is excluded, the measurements y and the geometry matrix A are updated, i.e. the alternative model is adopted. Therefore we will move back to step 1 with now $m = m-1$, and the procedure is repeated. As noted before with $m = m-1$ the model becomes weaker, so it is important to keep track of this weakening.

Fundamentally with hypothesis testing we compare the likelihood of the measurements being drawn under different hypotheses and next we make a *hard decision* on which hypothesis to use in the sequel — we adapt the model to an alternative one in case the original corresponding to the null hypothesis is no longer considered trustworthy.

Even when using only data snooping, with the described *iterative* procedure not only single failures but also multiple failures can be detected, since after a first adaptation another anomaly can be detected in the following iteration of the algorithm.

5.1.3. REMARK ON THE DETECTION OF MULTIPLE FAILURES

It has to be noted that when only data snooping is applied (only one-dimensional faults are anticipated), even if the algorithm can detect multiple failures through iteration, it will not be the most powerful in this circumstance, since a different C_y should be used in the model. In fact, in case q measurements are considered possibly faulty at the same time, the best way to test it (in the sense of maximum detection power, with a UMPI test) is to consider an alternative hypothesis as in the second line of Equation (2.2) with in this case C_y directly an $m \times q$ matrix (for instance $C_y = [C_{y_1} \;\; C_{y_2} \;\; C_{y_3}]$ in case the first three measurements are supposed faulty, where C_{y_i} is a canonical unit vector as before). To write the UMPI test, it is necessary to know which q measurements are faulty. When there is no knowledge about which q measurements are possibly faulty among the m available, and no knowledge about the number q of possibly simultaneously failing observations either, all possible combinations of q observations, for any possible value of q, should be considered, each of which setting a different alternative hypothesis. If we set a maximum number r of observations simultaneously failing to monitor (as it is practice in the ARAIM procedure), the total amount of alternative hypotheses to test against with specific tests will be the sum of the combinations of q observations out of m, for all the q considered:

$$N_{H_a} = \sum_{q=1}^{r} \binom{m}{q} \tag{5.9}$$

N_{H_a} can rapidly grow to very large values, providing heavy computational load to an eventual algorithm. For instance, with 30 satellites in view and $r = 3$ we would have $N_{H_a} = 4525$ alternative hypotheses to monitor.

5.1.4. SETTING THE THRESHOLDS

We have defined the test statistics to employ in Equations (5.1.1) and (5.1), and in Equations (5.2) and (L.2) for the special case $q = 1$, and we know their distributions under the null hypothesis:

$$H_0: \quad \underline{w} \sim N(0,1); \quad \underline{T}_q \sim \chi^2(q,0)$$

Under the alternative hypotheses, the distributions depend on the actual sizes of the biases ∇, which are unknown:

$$\begin{aligned} H_a: &\quad \underline{T}_q \sim \chi^2(q,\lambda) \\ H_a: &\quad \underline{w} \sim N(\nabla w, 1) \end{aligned}$$

where:

$$\begin{aligned} \lambda &= \nabla^T Q^{-1}_{\hat{\nabla}\hat{\nabla}} \nabla \\ \nabla w_i &= \frac{\nabla}{\sigma_{\hat{\nabla}}} \end{aligned} \tag{5.10}$$

The standard DIA procedure with reliability analysis foresees an a-priori setting of the parameters α and β (or α and γ) which are the error I and error II probabilities, for each test individually. In fact, once for a single test α is fixed the threshold is defined, and a choice

for β defines the reliability associated to the test. The probability of False Alarm α and the probability of Missed Detection β are defined, for general $\underline{T}_q$ and w-test (w subscript):

$$\begin{aligned} P_{FA} &= \alpha = P(\underline{T}_q > k|H_0) \\ P_{FA_w} &= \alpha_w = P(|\underline{w}| > k|H_0) \\ P_{MD} &= \beta = P(\underline{T}_q < k|H_a) \\ P_{MD_w} &= \beta_w = P(|\underline{w}| < k|H_a) \end{aligned} \tag{5.11}$$

The power γ is simply defined as $\gamma = 1-\beta$. Once β is fixed, it is possible to infer on the actual size of the measurements bias, and on the bias generated in the position solution that can be detected (or better, that is left undetected with probability β), that is on the reliability of the test. This is explained in the next section.

5.1.5. REMARKS ON α AND β CHOICE IN MODERN GNSS RAIM

Typical values chosen in the DIA implementation in geodetic applications were originally (in the '60s and '70s for instance) $\alpha = 0.05$, 0.01 or even 0.001, and $\beta = 0.2$, for each single w-test. This was due to the high cost related to carrying out a geodetic re-measurement in the field, hence high cost of a False Alarm. It has to be noted that with increasing availability of GNSS measurements, with new constellations being deployed in space, it will be possible to discard a large number of measurements at first inspection yet maintaining the availability of the positioning system. GNSS observations are becoming relatively 'cheap', and the high redundancy can allow one to choose values of α much larger than in the past (larger False Alarm rate), thereby gaining detection power.

5.1.6. INTERNAL AND EXTERNAL RELIABILITY

Reliability describes the ability of the system to check itself for modeling and measurement errors, and therefore describes the performance of the tests. Internal reliability refers specifically to the capability and effectiveness of the test to detect errors affecting the observation, and it is quantified by the *power* of the test given a determined size of the bias ∇, or conversely by the minimum size $|\nabla_{\text{MDB}}|$ of the error that can be detected with a pre-set power. Since it is standard practice to fix the value of the power $\gamma = 1-\beta$, we will refer mainly to the second definition.

With reference to Equations (5.11), once β is fixed, given the threshold k that was determined by the choice for α, we can retrieve the corresponding λ or ∇w for the w-tests. The reliability can be retrieved directly from ∇w for the w-tests with a scalar error size, whereas for the general $\underline{T}_q$ with $q > 1$ the possible ∇ vectors describe an ellipsoid in R^q once λ is fixed (see Equation (4.7)); also for the case $q > 1$ hence ∇ is constrained to lie on a surface (ellipsoid) and the influence of such bias in the position domain (external reliability) is constrained accordingly (the position bias must lie on an ellipsoid as well). This is described in detail in the following.

As from Equations (5.11), when fixing the value for β the distribution of the test under the alternative hypothesis will be consequently determined. In particular, the corresponding value of the non-centrality parameter λ can be retrieved through computing (numerically) the inverse of the cumulative (non-central) χ^2 or normal distribution function. The relation between λ and the bias size is:

$$\lambda = \nabla^T C_y^T Q_{yy}^{-1} Q_{\hat{e}_0 \hat{e}_0} Q_{yy}^{-1} C_y \nabla \tag{5.12}$$

In case of $q = 1$ the value of ∇ is uniquely determined (apart from its sign):

$$|\nabla| = |\nabla_{\text{MDB}}| = \sqrt{\frac{\lambda}{C_y^T Q_{yy}^{-1} Q_{\hat{e}_0 \hat{e}_0} Q_{yy}^{-1} C_y}} = \sqrt{\frac{\nabla w^2}{C_y^T Q_{yy}^{-1} Q_{\hat{e}_0 \hat{e}_0} Q_{yy}^{-1} C_y}} \tag{5.13}$$

In the case of data snooping and of Q_{yy} diagonal the last equation can be simplified as:

$$|\nabla_{\text{MDB}_i}| = \sigma_{y_i} \sqrt{\frac{\nabla w_i^2}{1 - \sigma_{\hat{y}_i}^2 / \sigma_{y_i}^2}} \tag{5.14}$$

where the subscript i refers to the i-th observation (as in Section 5.1.1). With these formulae (Equations (5.13) and (5.14)) the MDB can be easily retrieved from ∇w (the equivalent of λ for the w-test, $\lambda = \nabla w^2$).

External reliability instead refers to the impact an undetected error of size $|\nabla_{\text{MDB}}|$ has on the estimator of the unknown position x. Range bias and position bias are related by the simple linear relation:

$$\nabla \hat{x} = (A^T Q_{yy}^{-1} A)^{-1} A^T C_y \nabla \tag{5.15}$$

To find the incidence of the bias size $|\nabla_{\text{MDB}}|$ (when $q = 1$) on the position solution we can compute the corresponding $\nabla \hat{x}_{\text{MDB}}$:

$$\nabla \hat{x}_{\text{MDB}} = \pm (A^T Q_{yy}^{-1} A)^{-1} A^T C_y |\nabla_{\text{MDB}}| \tag{5.16}$$

This is the bias induced in the solution, once an error of size $|\nabla_{\text{MDB}}|$, detectable with power γ, remains undetected (with probability β). On the basis of this value, considering integrity applications, it is possible to decide whether the number of measurements available and the geometry A is adequate to guarantee integrity or the system should be declared unavailable instead. In general, as mentioned in Section 2.1, not the full vector $\nabla \hat{x}$ will be of interest for integrity check, but only some components of it (for instance the solution for the clock error is not typically of interest in GNSS), that we can represent as a linear combination (through matrix L) of it, let us say $\nabla \hat{x}_{\text{int}}$:

$$\nabla \hat{x}_{\text{int}} = L \nabla \hat{x} \tag{5.17}$$

Applying this transformation all the relations written for $\nabla \hat{x}$ can be transposed to $\nabla \hat{x}_{\text{int}}$ (and viceversa).

Once the impact that the measurements bias has on the position estimator is known, through the external reliability parameter $\nabla \hat{x}_{\text{MDB}}$, the full probability distribution of the estimator is known (conditioned on that specific reference bias size). Therefore, if a safety region Ω_x (Ω_{AL} in Chapter 2 is the equivalent region for RAIM application) around the true position x is defined, it is possible to compute the probability that the position solution will be inside this region, i.e. the probability $P(\underline{\hat{x}} \in \Omega_x | H_a)$. This probability, together with $\nabla \hat{x}_{\text{MDB}}$, represent 'general' purpose measures of the reliability of the position solution and of the DIA algorithm. These reliability considerations could naturally be extended to the RAIM problem, in which $P(\underline{\hat{x}} \in \Omega_x)$ is of main interest. The application of the DIA to the RAIM problem, based in fact on the reliability aspect, is dealt with in Chapter 6.

It is opportune to stress here that the DIA performance parameters discussed so far, i.e. False Alarm and Missed Detection probabilities (α and β) and the MDB just introduced, apply only to the *binary* situation of one alternative hypothesis tested against the null hypothesis and they do not describe instead the overall quality of the full DIA procedure. The overall performance of the DIA procedure is discussed in Section 5.3.

5.2. OMT + W-TESTS DIA

We focus here onwards on one of the most common DIA implementation, the one foreseeing the use of OMT and w-tests only — therefore restricting the identification only to one-dimensional threats (though possibly employing the procedure in an iterative way). In particular it is common to choose the w-tests for data snooping, but the following analysis applies to any type of w-tests.

5.2.1. BASIC DIA STRUCTURE AND NOTATION

Let $N_{H_{a_1}}$ be the number of one-dimensional alternative hypotheses monitored by our algorithm ($N_{H_{a_1}} = m$ in case of data snooping). We denote by subscript i, with $i = 1,\dots,N_{H_{a_1}}$, each variable related to a particular one-dimensional hypothesis, therefore to a particular w-test. Instead the subscript OMT is used for the OMT (as in Section 5.1.1).

The different hypotheses on the state of the system that can possibly hold are hereby defined:

$$\begin{array}{lll} H_0: & \underline{y} = Ax + \underline{e} & \\ H_i: & \underline{y} = Ax + C_{y_i}\nabla_i + \underline{e} & \dim C_{y_i} = m \times 1 \\ & \text{rank}([A\ C_{y_i}]) = n+1 & i = 1,\dots,N_{H_{a_1}} \\ H_{\text{OMT}}: & \underline{y} = Ax + C_{y_{\text{OMT}}}\nabla_{\text{OMT}} + \underline{e} & \dim C_{y_{\text{OMT}}} = m \times (m-n) \\ & \text{rank}([A\ C_{y_{\text{OMT}}}]) = m & \end{array} \tag{5.18}$$

with C_{y_i} possibly, though not necessarily, the canonical unit vectors for data snooping.

The w-tests and OMT will have the following distributions under the null hypothesis:

$$H_0: \quad \underline{w}_i \sim N(0,1); \quad \underline{T}_{m-n} \sim \chi^2(m-n,0)$$

Under the alternative hypotheses, the distributions depend on the actual sizes of the biases ∇_{OMT} and ∇_i, which are unknown:

$$\begin{array}{ll} H_{\text{OMT}}: & \underline{T}_{m-n} \sim \chi^2(m-n,\lambda_{\text{OMT}}) \\ H_i: & \underline{w}_i \sim N(\nabla w_i, 1) \end{array}$$

where with H_{OMT} we refer to the hypothesis of overall inconsistency tested for with the OMT, with H_i we consider the alternative hypothesis corresponding to the i^{th} w-test, and:

$$\begin{array}{c} \lambda_{\text{OMT}} = \nabla_{\text{OMT}}^T Q_{\hat{\nabla}_{\text{OMT}}}^{-1} \nabla_{\text{OMT}} \\ \nabla w_i = \frac{\nabla_i}{\sigma_{\hat{\nabla}_i}} \end{array} \tag{5.19}$$

Note that under a hypothesis $H_{k\neq i}$ the w-test $\underline{w}_i$ follows a yet different distribution, dependent on the correlation between the w-tests. More specifically, under H_k it holds $\underline{w}_i \sim N(\rho_{ik}\nabla w_k, 1)$, where ρ_{ik} is the correlation coefficient between the w-test statistics ($\rho_{ik} = \frac{C(\underline{w}_i,\underline{w}_k)}{\sigma_{\underline{w}_i}\sigma_{\underline{w}_k}} =$, with $C(\underline{w}_i,\underline{w}_k)$ the covariance between the two w-test statistics).

The OMT is run in the Detection step, as previously described, and reads:

$$\begin{array}{ll} T_{m-n} \leq k_{\text{OMT}} & \text{No Detection — Nominal Operations} \\ T_{m-n} > k_{\text{OMT}} & \text{Detection — Identification required} \end{array} \tag{5.20}$$

There is to note that for the case of the OMT, an infinite number of choices for the $C_{y_{\text{OMT}}}$ are possible, as long as $[A \mid C_{y_{\text{OMT}}}]$ is a full rank $m \times m$ matrix, as seen in Section 5.1.1. We

said in Section 5.1.1 that for the general $\underline{T}_q$ with $q > 1$ the possible ∇ vectors describe an ellipsoid in R^q (once the non-centrality parameter λ is fixed), therefore one may argue that for different choices of $C_{y_{\text{OMT}}}$, different ellipsoids are obtained for the corresponding ∇_{OMT}, and it is not possible to infer on the actual bias ∇ affecting the measurements. This is not true; in fact the probability of occurrence of $T_{m-n} < k_{\text{OMT}}$ is only (given m and n) dependent [99] on λ_{OMT} and:

$$\lambda_{\text{OMT}} = \|\nabla_{\text{OMT}}\|^2_{Q^{-1}_{\hat{\nabla}_{\text{OMT}}}}$$

This equation shapes an ellipsoid in R^{m-n} and for any different but proper choice of $C_{y_{\text{OMT}}}$ it will generate the same bound (ellipsoid) for the general bias $\nabla y = C_{y_{\text{OMT}}}\nabla$ in R^m that can affect the measurements. Different results from different $C_{y_{\text{OMT}}}$ are only re-parametrizations of the same ellipsoid. This is because the OMT monitors the residuals vector on all its degrees of freedom and is explained also later in Section 5.2.2[1].

With reference to the Identification step previously described, only the case $q = 1$ is considered, therefore Equations (5.7) and (5.8) simply reduce to:

$$\begin{array}{ll} \max_i |w_i| \le k_w \;\; \forall i: \; 1 \le i \le N_{H_{a_1}} & \text{No identification} \\ \max_i |w_i| > k_w \;\; \forall i: \; 1 \le i \le N_{H_{a_1}} & \text{Identification} \end{array} \tag{5.21}$$

with k_w the threshold for the w-tests. The coupling of the thresholds k_{OMT} and k_w between OMT and w-tests is dealt with in next section.

5.2.2. Overall Model Test and w-tests

As mentioned in Section 5.1.6, it is necessary to establish some criteria to define the thresholds for OMT and w-tests/data snooping, and a relation between them. This is a problem of simultaneous statistical inference, since at the same time it is necessary to choose among multiple possible hypotheses, and many different approaches have been developed. In the following we describe three of them, with reference to [5] and [68].

B-method

A possible approach (proposed by Baarda [5]) is the B-method, in which the main idea is to assure the same reliability for OMT and w-tests with regard to a certain error. Given α_w, β_w and λ_w the significance, the type II error probability and the non-centrality parameter of the w-tests (the same for all the w-tests), and α_{OMT}, β_{OMT} and λ_{OMT} again the performance parameters for the OMT, the B-method requires:

$$\begin{array}{l} \beta_{\text{OMT}} = \beta_w \;\; \text{for } \nabla y = \nabla y_{\text{MDB}} = C_{y_i}\nabla_{i_{\text{MDB}}} = C_{y_i}\nabla_{\text{OMT}_{\text{MDB}}} \\ \lambda_{\text{OMT}} = \lambda_w \;\; \text{for } \nabla y = \nabla y_{\text{MDB}} = C_{y_i}\nabla_{i_{\text{MDB}}} = C_{y_i}\nabla_{\text{OMT}_{\text{MDB}}} \end{array} \tag{5.22}$$

that means, depending on which alternative hypothesis holds true (from Equation (5.26)):

$$\begin{array}{ll} H_{\text{OMT}}: & \lambda_{\text{OMT}} = \nabla^T_{\text{OMT}_{\text{MDB}}} Q^{-1}_{\hat{\nabla}_{\text{OMT}}} \nabla_{\text{OMT}_{\text{MDB}}} = \lambda_w \\ H_i: & \lambda_w = \frac{\nabla^2_{i_{\text{MDB}}}}{\sigma^2_{\hat{\nabla}_i}} = \lambda_{\text{OMT}} \end{array} \tag{5.23}$$

[1] Note also that $\underline{T}_{m-n} = \underline{t}^T Q_t^{-1} \underline{t}$, where $\underline{t}$ is the misclosures vector (cf. [99]), so that an upperbound for the test fully bounds the misclosures vector.

With these choices, OMT and w-tests will have the same reliability. This means they will be able to detect biases of the same size with the same power, the OMT being used for detection purpose and thereafter the w-tests for identification purpose. To clarify the concept we report here an example.

Example. Given is $A = [1 \;\; 1 \;\; 1]^T$, $Q_{yy} = I_3$, therefore we have $m = 3$ measurements for $n = 1$ unknown parameter. We can define a w-test choosing $C_{y_w} = [1\ 0\ 0]^T$, which addresses the first measurement, whereas for the OMT a possible choice is:

$$C_{y_{\text{OMT}}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{5.24}$$

Suppose we defined a threshold for the w-test such that $\beta_w = 0.2$ for a $\nabla_{\text{MDB}} = 3$. This means that if a bias of size 3 is actually present in the first measurement, it will be detected with 80% probability with this specific w-test. The threshold set on the OMT applying the B-method is such that, when the same bias of size 3 (therefore the MDB) is present in the first measurement, it will be detected as well with 80% probability. In fact we have $\nabla_{\text{OMT}_{\text{MDB}}} = [3 \;\; 0]^T$ (same bias vector $\nabla y = [3\ 0\ 0]^T$ applied to the measurements) and for the variances:

$$\sigma^2_{\hat{\nabla}_i} = 1.5 \;\text{ and }\; Q_{\hat{\nabla}_{\text{OMT}}} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \tag{5.25}$$

Applying the corresponding formulas for the non-centrality parameters to the two cases (Equation (5.23)) leads to the same $\lambda_{\text{OMT}} = \lambda_w = 6$ for both w-test and OMT. But the B-method sets the threshold of the OMT statistic to give the test the same power (equivalently same β) for the same non-centrality parameter, therefore same power and β are obtained for the same bias (∇y) in the measurements. The idea of the B-method is that the same error should be detected with same probability by OMT and specific w-tests.

Note that $Q_{\hat{\nabla}_{\text{OMT}}}(1,1)$ and $\sigma_{\hat{\nabla}_w}$ are different, but still we obtain $\lambda_{\text{OMT}} = \lambda_w = 6$. This is because λ as determined in Equation (4.7) can be written also as (see [99]):

$$\lambda = \|P_A^{\perp} C_y \nabla\|^2_{Q_{yy}^{-1}} = \|P_A^{\perp} \nabla y\|^2_{Q_{yy}^{-1}} \tag{5.26}$$

so that the non-centrality parameter both for OMT and w-test is the same regardless the choice made for $C_{y_{\text{OMT}}}$. This is further explained in Figure 5.3, where the β behaviours are shown for the w-test and the OMT for the particular choice of $\nabla = \nabla_{\text{MDB}}$.

Let us go back to the general B-method. As a result of the choice made for setting the thresholds, we have $\alpha_{\text{OMT}} > \alpha_w$. This is naturally due to the fact that the w-test is the most powerful test in detecting an error with the particular signature it addresses, which means it shows the lowest significance level of all possible tests (so also lower than the OMT) for a chosen size of the MDB and a chosen power γ.

In Figure 5.1 (taken from [5]) the space of $R(P_A^{\perp})$ (as an example) is shown, i.e. perpendicular to the space of A with respect to the metric defined by Q_{yy}^{-1} ($P_A^{\perp}$ being the projector onto the space perpendicular to A with the metric defined by Q_{yy}^{-1}), on which we can represent all the w-tests performed. Remember that the value of the test statistic w equals the length of the projection of vector $\hat{e}_0$ onto vector $P_A^{\perp} C_y$ (and $\hat{e}_0 = P_A^{\perp} y$); in fact, from [101]:

$$P_A^{\perp} C_y \hat{\nabla} = \hat{e}_0 - \hat{e}_a$$

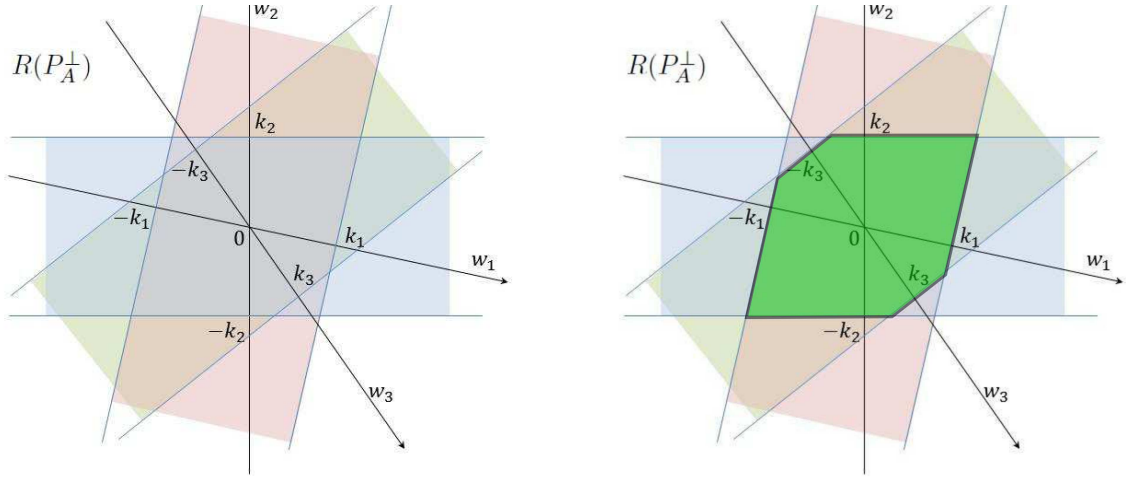


Figure 5.1: Space $R(P_A^\perp)$, acceptance regions of different w-tests, in this case for a Q_{yy} identity matrix. On the right highlighted the combined acceptance region, where no test is rejected.

5

and, pre-multiplying by $P_{P_A^\perp C_y}$:

$$P_A^\perp C_y \hat{\nabla} = P_{P_A^\perp C_y} \hat{e}_0$$

because $\hat{e}_a$ is perpendicular to the space $R(A\ C_y)$, from which:

$$w^2 = T_1 = \|P_A^\perp C_y \hat{\nabla}\|^2_{Q_{yy}^{-1}} = \|P_{P_A^\perp C_y} \hat{e}_0\|^2_{Q_{yy}^{-1}}$$

We can therefore represent the w-tests on axes $P_A^\perp C_y$ in the $R(P_A^\perp)$ plane. The acceptance region in terms of $\hat{e}_0$ of the i-th w-test is represented in this plane by the strip comprised between the two parallel rects at $\pm k$ perpendicular (in the metric defined by Q_{yy}^{-1}) to the corresponding $P_A^\perp C_{y_i}$. We can see how the intersection of the acceptance region of different w-tests can create a closed region around the origin.

The acceptance region for the OMT is instead represented by a circle (in the metric defined by Q_{yy}^{-1}). In the case of the B-method, the intersection of the perimeter of the circle with the w-tests axes will be beyond (at a distance larger than) the thresholds defined for each w-test. This is sketched in Figure 5.2. This causes acceptance of the OMT whereas the w-test(s) would be rejected.

The highlighted brown regions in the figure correspond to the cases in which the OMT gets rejected but the w-tests are (both) accepted, therefore the cases in which adaptation is not possible and which result in direct unavailability for our RAIM application.

The fact that the intersections of the perimeter of the circle with the w-tests axes are beyond the thresholds defined for each w-test follows directly from the choice of same β for defined λ for all the w-tests and the OMT. Choosing one of the w-tests for instance, the corresponding alternative hypothesis with defined λ_w will determine a distribution for the residuals $\underline{\hat{e}}_0$. The missed detection probability β is simply the integral of this distribution over the acceptance region, and in case the threshold for the OMT is chosen the same as for the w-test (Figure 5.4), the integral will always be smaller, the detection region for OMT being contained by the one for the w-test. Therefore with the B-method the OMT threshold is always larger than the squared w-test's threshold (the OMT threshold is the square of the

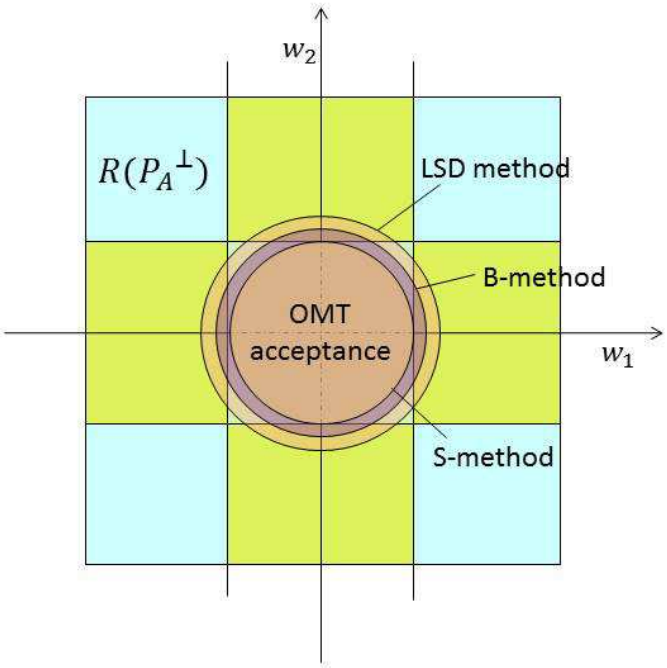


Figure 5.2: Space $R(P_A^{\perp})$, example of acceptance regions for w-tests and for OMT following S-method, B-method and LSD method, in case of Q_{yy} identity matrix (the axes for the w-tests are not necessarily perpendicular). The highlighted brown regions correspond to the cases in which the OMT gets rejected but the w-tests are accepted, therefore would cause direct unavailability in our RAIM application.

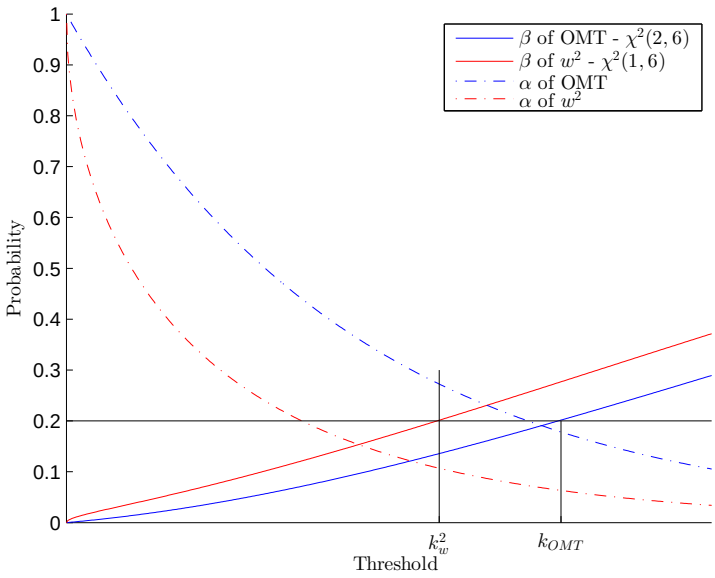
distance from the origin). Figure 5.3 shows β and α for the OMT and for a w-test for the example case in which $m-n=2$ and $\lambda_w = \nabla^2_{\mathrm{MDB}_w}/\sigma^2_{\hat{\nabla}_w} = 6$. It is clear as for same β, the threshold for the OMT will be larger than the squared threshold for the w-test, but at the same time $\alpha_{\mathrm{OMT}} > \alpha_w$.

One practical issue with the B-method is that with increased redundancy $m-n$ the significance of the test α_{OMT} tends to grow to very large values, with consequent large tendency to conservativeness. This behaviour is shown for instance in [99] (at page 136 in Figure C.1).

S-METHOD

The Scheffé method (S-method) [90] is a reknown method in the field of simultaneous statistical inference (see [68], [3], [66]). This method, as the B-method, combines the classicist approach to detect general divergence from the null hypothesis (with the F or χ^2 test) with the multiple comparisonist approach through which one wants to choose among specific alternative hypotheses. The following description is mainly based on [68].

With reference to Figure 5.1, if we imagine an infinite number of different w-tests for a corresponding infinite number of possible alternative hypotheses, we can imagine these to be represented by strips bounded by couples of parallel lines around the origin and extending to infinity perpendicularly away from the w-axis. As a matter of fact, each w-test tests for errors of a specific signature (vector C_y), and their values grow along specific directions away from the origin, the axes $P_A^{\perp}C_y$ in the space $R(P_A^{\perp})$. Any possible w-test can also be defined by a linear combination of $m-n$ other linearly independent w-tests. The strips of the w-tests acceptance regions would span all possible directions, so we can deduce that the



5

Figure 5.3: β and α for the OMT and for a w-test for different choices of threshold, case in which $m-n=2$ and $\lambda=6$. It is reported in particular the squared w-test because of easier comparison with the OMT case. The thresholds are chosen for same value of β, following the B-method. It is clear that for the same β the threshold for the OMT will be larger than the squared threshold for the w-test, but at the same time $\alpha_{\text{OMT}} > \alpha_w$.

intersection of all the acceptance regions will result in a circular area around the origin, or an hyper-sphere in the general case (always in the metric defined by Q_{yy}^{-1}). A circle is the shape of the acceptance region of the OMT, therefore the OMT constitutes a protection against all possible alternative hypotheses. Now for each w-test we set the threshold at the same distance from the origin, that is for the same value of residuals vector length. In this way the circular region is tangent to all the line segments defining the acceptance regions for the w-tests, or to say it in a different way, the thresholds for the w-tests are determined by the projection of the circle onto the w-tests axes. The threshold for the w-test will be computed therefore as:

$$\begin{gathered}\text{accept } H_0 \text{ whenever } |w_i| < k_w \\ \text{with } \; k_w : \int_0^{k_w^2} \chi(m-n,0) = \alpha_{\text{OMT}}\end{gathered} \tag{5.27}$$

that is, the threshold k_w for the w-tests is exactly the square root of the threshold used for the OMT.

Therefore the method is based on the following: first a threshold for the OMT is set based on a predefined significance α_{OMT}. Then the threshold for the w-tests is obtained by the intersection of the circle determined by the choice of the threshold with the w-tests axes. The S-method is sketched in Figure 5.4 (taken from [5]).

Comparing with the B-method, the OMT acceptance region of the S-method is smaller, as we previously mentioned, with $\alpha_{\text{OMT}} > \alpha_w$ even more pronounced than with the B-method. Differently from the B-method, the Scheffé method does not require an approach in two steps, since there are no cases in which a rejection from a w-test could have been instead accepted by the OMT. Similarly as for the B-method, if instead of starting by setting a threshold for the OMT we start from the threshold for the w-tests, with increased redundancy $m-n$ the significance of the test α_{OMT} tends to grow to very large values, with consequent large

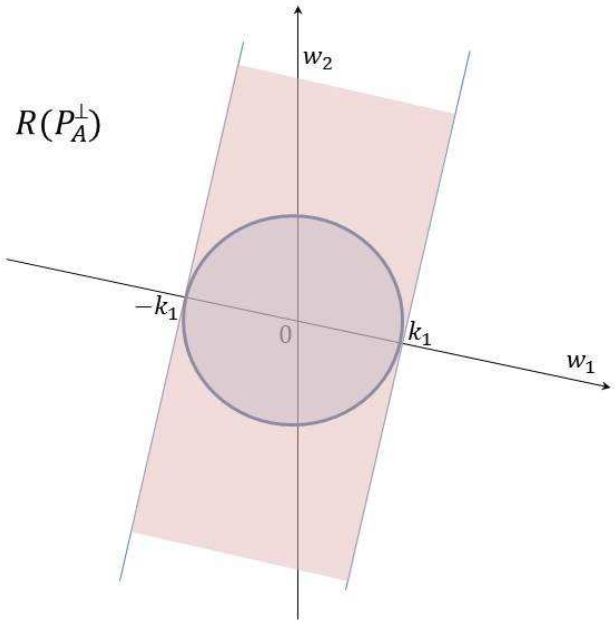


Figure 5.4: Space $R(P_A^{\perp})$, acceptance regions of OMT and a w-test for the S-method.

tendency to conservativeness. This is true to an even larger extent than with the B-method.

The Scheffé testing procedure can be summarized in the following steps (distinction between detection and identification steps is mainly for clarity purpose[2]):

1. The thresholds for OMT and w-tests are set: if we want to fix the total significance of the procedure, then first $\alpha_{\text{OMT}} = \alpha_{tot}$ is set and the threshold for the w-tests is decided accordingly, otherwise the threshold for the w-tests is first set for instance based on reliability reasons and the threshold for the OMT follows.

2. The OMT is run (detection step): if the null hypothesis is not rejected, then there is no need to go further in the procedure and a fault-free case is declared.

3. Once the OMT has led to rejection of the null hypothesis, the w-tests are run (identification step):

   - if any of the w-tests exceeds the threshold, then the maximum modulus w-test identifies the type of anomaly most likely to be present. Identification of the anomaly can possibly lead to adaptation of the model and re-application of the procedure.
   - if none of the w-tests exceeds its threshold, identification is not possible, and a general (non-anticipated) anomaly is declared.

The Scheffé method is not only restricted to the specific testing problem described but is of more general application in the context of simultaneous multiple comparisons. Given a

[2]As mentioned OMT and w-tests can be run at the same time because if a w-test gives rejection the OMT will reject as well (the tests are coherent).

linear model as in Equation (2.2), we can consider $\mathscr{L} = \{l = (l_1, \dots, l_n)\}$ a d-dimensional linear subspace of the n-dimensional space. The Scheffé technique gives confidence intervals on the linear combinations $l^T x$ for all $l \in \mathscr{L}$.

In fact if L is a $d \times n$ matrix whose rows form a basis for $\mathscr{L}$, then Lx would represent all possible linear combinations of x in $\mathscr{L}$. Given that the standard BLUE estimator $\underline{\hat{x}} = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} \underline{y}$, $L\underline{\hat{x}}$ has distribution:

$$L\underline{\hat{x}} \sim N(Lx, LQ_{\hat{x}\hat{x}}L^T)$$

A linear hypothesis on x has the form:

$$H_0 : \; Lx = \phi_0$$

and the likelihood ratio test (against an alternative $H_a : \; Lx \neq \phi_0$) is:

$$\underline{T}_d = (L\underline{\hat{x}} - \phi_0)^T (LQ_{\hat{x}\hat{x}}L^T)^{-1} (L\underline{\hat{x}} - \phi_0)$$

5

which has a $\chi^2(d, 0)$ distribution. Setting a threshold to the test based on a confidence α will complete the testing procedure. Graphically, this is equivalent to check whether the point ϕ_0 is contained in the $1 - \alpha$ confidence ellipsoid centered at $L\underline{\hat{x}}$.

To extend this concept to our specific case, we consider the same model as in Equation (2.2) in the general alternative hypothesis:

$$H_a : \quad \underline{y} = Ax + C_y \nabla + \underline{e} \quad \nabla \neq 0$$

with $C_y \nabla$ as general as possible, i.e. C_y an $m \times (m - n)$ matrix and ∇ an $m - n$ dimensional vector. The null hypothesis H_0 reads $\nabla = 0$, therefore the application of the just described Scheffé method results in a standard ellipsoidal confidence region around the origin for the vector ∇, with corresponding test statistic:

$$\underline{T}_{m-n} = \underline{\hat{\nabla}}^T Q_{\hat{\nabla}\hat{\nabla}}^{-1} \underline{\hat{\nabla}} \sim \chi^2(m - n, 0)$$

This is the standard OMT. Note that this problem formulation is equivalent to a parameter significance test [99] performed on the parameters $[x^T \; \nabla^T]^T$, namely on ∇. In fact in the alternative hypothesis the whole vector of parameters is considered and solved for, whereas in the null hypothesis a constraint is given to some linear combination of the parameters (in this case on the subvector ∇). This means testing whether the extra parameters ∇ are actually needed to describe the original observations model.

The OMT therefore accounts for all the possible linear combinations of all the possible redundant extra parameters that could describe the system, whereas the w-tests accounts for just one extra parameter.

FISHER'S LSD METHOD (LEAST SIGNIFICANT DIFFERENCE TEST)

The LSD test first proposed by Fisher is based on two steps as the B-method, first the OMT and following the multiple specific tests (comparisons), which are the w-tests in our case. In the LSD method, the thresholds are chosen in such a way that all the tests performed have the same significance level. That is:

$$\alpha_{\text{OMT}} = \alpha_w$$

The test results in a detection region similar to the one for the B-method (in Figure 5.2). With respect to the S-method and similarly to the B-method, this test requires distinction in two steps, since as it can be seen in figure the acceptance area of the OMT intersects with the rejection areas of the w-tests — this means some observations could lead to acceptance of H_0 using the OMT but rejection when testing with the w-tests, therefore the order of testing assumes importance.

The rationale behind this approach is simply that if we used a certain significance level to decide between acceptance or rejection of null hypothesis, why should we use a different significance level to decide whether we belong to one of the specific alternative hypotheses? This also eliminates the issue of the increase of significance of the OMT with the increase of the redundancy $m-n$. On the other hand, no consideration is made on the reliability of this test. Also the S-method lacks of any specific analysis of the reliability aspects.

Table 5.1 compares the three methods described, summarizing their main selection principles.

Table 5.1: Summary of the simultaneous inference methods discussed and their main underlying principles.

Method	Principle
S-method	$k_{\text{OMT}} = k_w^2$ (same threshold)
B-method	$\beta_{\text{OMT}} = \beta_w$ for $\nabla y = \nabla y_{\text{MDB}}$ (same power)
LSD method	$\alpha_{\text{OMT}} = \alpha_w$ (same significance)

5.2.3. REMARKS ON THE THREE MULTIPLE COMPARISON METHODS DESCRIBED

In the author's opinion, detection and exclusion of faults in a RAIM algorithm must be based on external reliability aspects. The rationale behind the choice between acceptance or rejection of the fault-free hypothesis should be related to the risk of occurrence of a Positioning Failure, i.e. an error in the position solution that exceeds a given safety requirement, the Alert Limit (AL). The only method that is strictly based on reliability reasoning is the B-method. Within this view, the S-method appears overly conservative whereas the LSD method is not conservative. The B-method assures same reliability only when considering the binary case of one alternative hypothesis to the null hypothesis, and only for a particular choice of β. For our integrity problem we would have to consider instead the overall procedure performance and the full range of β. In this view, if it is not feasible to monitor over different β choices, it would make sense to employ the S-method instead.

As a further remark, with the use of different prior probabilities, as proposed in the following, the OMT can be employed to detect only the faults that cannot be detected by the w-tests (therefore more unlikely). In this way the OMT would have a role complementary to the w-tests. This possible approach is explored in Section 6.6.

5.3. OVERALL PERFORMANCE OF THE DIA PROCEDURE

In the previous Sections, we considered a RAIM procedure based on an OMT and the w-tests. We analyzed in particular only the performance parameters α and β for each test by itself. Considering the procedure as a whole instead, it is important to compute the overall performance, that is taking into account all the tests together, that are run at the same time or in succession. In the following we focus on the most common case of data snooping,

therefore the case in which m w-tests are run ($N_{a_1} = m$). Furthermore it is assumed that Q_{yy} is diagonal.

5.3.1. DIA IDENTIFICATION PERFORMANCE

Let us leave out for the moment the OMT and consider only the w-tests. This means we consider only the Identification step. Remember that the rule for the step was defined in Equation (5.21), such that the general no identification case occurs when:

$$\max_i w_i \leq k_w \quad \forall i : \ 1 \leq i \leq N_{a_1} = m \tag{5.28}$$

FALSE ALARM

For each single w-test we defined α in Equation (5.11). In case a total $P_{FA_{tot}}$ is set instead, it is:

$$P_{FA_{tot}} = \alpha_{tot} = P(\underline{w}_M \notin \Omega_w(k_w)|H_0) \tag{5.29}$$

where $\underline{w}_M|H_0 \sim N_m(0, Q_{ww})$ is the vector of all m w-tests and $\Omega_w(k_w)$ is the hyper-rectangular acceptance region in R^m enclosed by the thresholds k_w of each w-test. Defining $D = \text{diag}(\sigma_{\hat{e}_i}^{-1})$ the diagonal matrix which diagonal elements are the $\sigma_{\hat{e}_i}^{-1}$, it is $\underline{w}_M = D\underline{\hat{e}} = DP_A^{\perp}\underline{y}$ (under assumption of data snooping and Q_{yy} diagonal) and $Q_{ww} = DP_A^{\perp}Q_{yy}P_A^{\perp T}D$. The probability in Equation (5.29) in fact corresponds to the intersection event constituted by all the w-tests being smaller than the threshold. This results in:

$$\alpha_{tot} = 1 - \int_{\Omega(k)} N_m(0, Q_{ww})\mathrm{d}w_M \tag{5.30}$$

We chose to make this region an hyper-cube with same threshold k_w for each w-test, since the same w-test value corresponds to the same probability that the related measurement has been drawn from the unbiased distribution. This choice can be discussed, the choice of different thresholds for the w-tests can optimize integrity, but requires much higher computational effort. Note that Q_{ww} is rank defect, since its rank is $m - n$ (but it is a square $m \times m$ matrix). Therefore the integration over an hypercube in R^m in Equation (5.29) is equivalent to an integration over a closed figure in the R^{m-n} sub-space in which the $\underline{w}$ distribution is lying.

An upper bound for the total α_{tot} can be obtained by the formula [102] (Šidák approximation):

$$\alpha_{tot} \leq 1 - \prod_{i=1}^{m}\left[2\Phi\left(\frac{k}{\sigma_{w_i}}\right) - 1\right] = 1 - \prod_{i=1}^{m}[2\Phi(k) - 1] \tag{5.31}$$

as $\sigma_{w_i} = 1$, where $\Phi(x)$ is the standard normal CDF. The equality holds when the w-tests are uncorrelated.

The total $P_{FA_{tot}}$ here considered is the actual probability of FA only in case the algorithm is made up of a single identification step (with m tests simultaneously or in parallel), or let us say, it is the $P_{FA_{tot}}$ within one identification step of the algorithm. In case multiple steps are foreseen (as for instance with iterative DIA), a FA can be followed (after adaptation) by another test, therefore the probability of FA of the successive steps has to be considered. These considerations are exposed in Section 6.5. Till then we will analyze the performance of the procedure as if it was constituted by a single step.

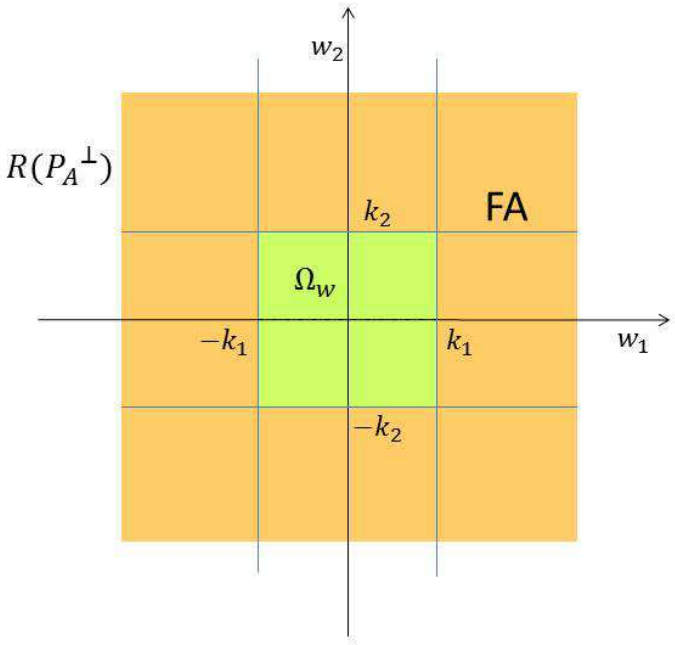


Figure 5.5: Example of acceptance/rejection regions in $R(A^{\perp})$. The region in green is the acceptance region of the w-tests ($\Omega_w(k_w)$), while the remaining area in orange is the rejection region and generates False Alarm in the fault-free case.

MISSED AND WRONG DETECTION

Another issue to consider is the occurrence of wrong detections/identifications. This is especially related to the reliability of the system. We consider again the distribution of each test under the corresponding alternative hypothesis:

$$H_i: \quad \underline{w}_i \sim N(\nabla w_i, 1) \quad \text{for } i = 1,2,\ldots,m \tag{5.32}$$

with $\nabla w_i = \frac{\nabla_i}{\sigma_{\hat{\nabla}_i}}$. If we fix a value for the probability of Missed Detection P_{MD}:

$$P_{MD_i} = \beta_i = P(|\underline{w}_i| \leq k_w | H_i) \quad \text{for } i = 1,2,\ldots,m \tag{5.33}$$

to a value $P_{MD_1} = \beta_1$ (same for all i) we can retrieve the so called Minimum Detectable Bias, MDB, that is the value of ∇_i that satisfies the previous equation.

With the previous equation we are considering each w-test separately, disregarding the possibility of wrong detections. A proper formulation should take into account that even if the correct w-test statistic exceeds the threshold, at the same time another one may be larger and a wrong detection may take place.

We must define then better the cases of Missed Detection (MD) and Wrong Detection (WD). For the MD we therefore consider only the case in which *none* of the w-tests is rejected, when in reality the corresponding (to the w-test considered) alternative hypothesis holds true:

$$P_{MD_i} = \beta_i = P(\underline{w}_M \in \Omega_w(k_w) | H_i) \tag{5.34}$$

Let us deal with the WD now. The WD, the case $|\underline{w}_i| \leq |\underline{w}_j| \cap |\underline{w}_j| \geq k_w |H_i$ for any j, is sometimes referred to in literature as a type III error (the type I refers to the FA, and the type II to the MD), see for instance [34]. In case the algorithm is made of a single identification and exclusion step, it is possible to treat this type III error as equivalent to the MD probability, but in case the algorithm foresees multiple steps it makes sense to distinguish between type II and type III errors. This because an MD causes the algorithm to stop and deliver a (likely) incorrect solution, whereas a WD leads to further identification steps with the chance of a subsequent correct detection. The type III error can be therefore defined:

Def. WD: rejection of H_0 and acceptance of $H_{j\neq i}$ $\forall\, j$ when in fact H_i is true.

Therefore, with reference also to Table 5.2, we distinguish the probability of MD previously stated in Equation (5.34) from the probability of Wrong Detection:

$$P_{WD_i} = \kappa_i = P(\bigcup_{j=1\ j\neq i}^{m} (|\underline{w}_j| > k_w \cap |\underline{w}_i| < |\underline{w}_j|)\,|H_i) \tag{5.35}$$

5

This means that a WD related to a fault in the i^{th} observation (hypothesis H_i) occurs when the maximum of the w-tests is not w_i, the one corresponding to the faulty observation, but w_j, and it does exceed the threshold. A WD would occur also if more than one 'wrong' w-test exceeds the threshold (this eventuality is taken into account by considering the union of the events in the formula). With respect to a FA (which could also be considered a sort of wrong detection), the difference is that a detection occurs when there *is* something to be detected. Just not the right one is detected.

Table 5.2: Different types of errors in a detection problem with multiple alternative hypotheses. The simple case of two alternative hypotheses ($m = 2$) is considered, and correspondingly two w-tests are run in parallel, in one step of identification. The probabilities on each row add up to a probability of one.

	Result of the test		
Unknown reality	H_0 $\lvert w_1\rvert \leq k_w, \lvert w_2\rvert \leq k_w$	H_1 $\lvert w_1\rvert > k_w, \lvert w_1\rvert > \lvert w_2\rvert$	H_2 $\lvert w_2\rvert > k_w, \lvert w_2\rvert > \lvert w_1\rvert$
H_0	OK	α_1 (type I error)	α_2 (type I error)
H_1	β_1 (type II error)	OK	κ_1 (type III error)
H_2	β_2 (type II error)	κ_2 (type III error)	OK

The probability in Equation (5.33) is easily computed through integration of a normal distribution (remembering the distribution of $\underline{w}_i|H_{a_i}$ in Equation (5.32)), whereas to find the probabilities in Equations (5.34) and (5.35) it is necessary to integrate the multivariate normal distribution of the w-tests $\underline{w}_M$ over the region $\Omega_w(k_w)$ in the first case, and over a region $\Gamma_i(k_w) = \{w_M \in R^m : \exists j \neq i,\ |w_j| > k_w \cap |w_i| \leq |w_j|\}$ in the WD case. That is:

$$\begin{aligned} \beta_i &= \textstyle\int_{\Omega_w(k_w)} N_m(\nabla w_{M_i}, Q_{ww})\mathrm{d}w_M \\ \kappa_i &= \textstyle\int_{\Gamma_i(k_w)} N_m(\nabla w_{M_i}, Q_{ww})\mathrm{d}w_M \end{aligned} \tag{5.36}$$

where $\nabla w_{M_i} = DP_A^{\perp}\nabla y_i$ with ∇y_i the bias vector $\nabla y_i = C_{y_i}\nabla_i$. An example of regions $\Gamma_i(k_w)$ and $\Omega_w(k_w)$ is shown in Figure 5.6 for a two-dimensional case. Together with Equation (5.30) the above Equation (5.36) determines the formulas for the errors probabilities for the identification step of the DIA procedure.

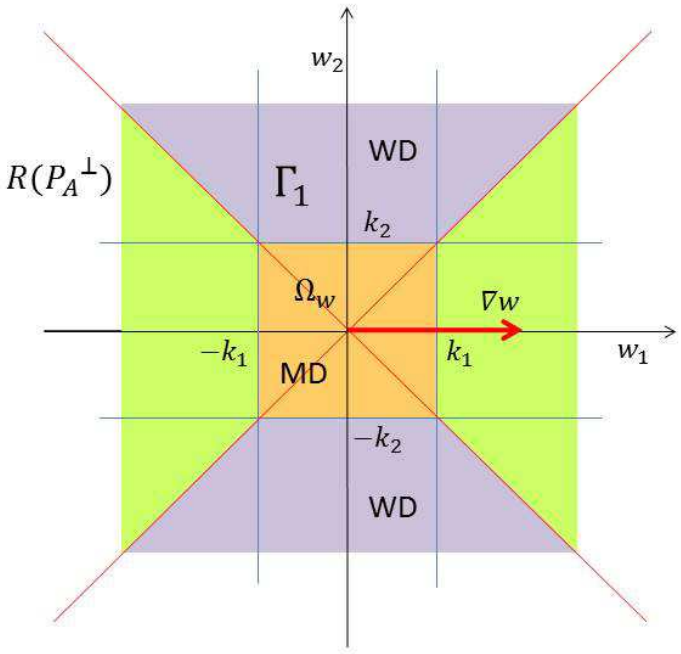


Figure 5.6: Example of acceptance/rejection regions in $R(A^{\perp})$. The region in orange denoted by MD ($\Omega(k_M)$) is the acceptance region of the w-tests, and generates Missed Detection in case a fault is present (presence of the bias ∇w). The region denoted by WD is $\Gamma_1(k_M)$, referring to the case of occurrence of H_1, an outlier in the first measurement. The diagonal lines are driven by selecting the *maximum* of the w-tests.

An upperbound for the MD probability in Equation (5.36) can be simply obtained using the formula for the binary case [100]:

$$\beta_i \leq \int_{-k_w}^{+k_w} N(\nabla w_i, 1)\mathrm{d}w_i = \Phi(k_w - \nabla w_i) - \Phi(-\nabla w_i - k_w) \tag{5.37}$$

where Φ indicates the cumulative distribution function of a zero mean unit Gaussian distribution. This equation represent an upperbound for the MD probability because, with comparison with Equation (5.36), the integration is now carried out over a strip delimited by the thresholds k_w which contains the acceptance region $\Omega_w(k_w)$, and the integrand is always positive.

Let us look at the MDB now. Fixing a value for $\beta_i = \beta_{0i}$, it is possible to know the size of the bias that we can detect with determined confidence. To do so it is necessary to invert Equations (5.36), that is possible numerically as long as the region $\Omega(k)$ is symmetric respect to the origin. Also the sign of ∇w_i is not important, $\pm\nabla w_i$ yielding the same β_i. The function $\beta_i(|\nabla w_i|)$ (as said the sign of ∇w_i is not important) is invertible because it is continuous and differentiable (it is the integral of a multinormal distribution) and is monotonously decreasing with $|\nabla w_i|$. This can be demonstrated by means of Anderson's theorem in [2]. As a result, considering the function $\beta_i(|\nabla w_i|)$ and fixing a confidence value for the missed detection $\beta_i = \beta_{0i}$, the MDB can be retrieved as:

$$|\nabla w_i|_{MDB} = \frac{|\nabla_i|_{MDB}}{\sigma_{\hat{\nabla}_i}} = \beta_i^{-1}(\beta_{0i}) \tag{5.38}$$

With regard to the WD probability instead, an upperbound can be obtained as:

$$\kappa_i \leq \min \begin{cases} \int_{\Lambda_i(k_w)} N_m(\nabla w_{M_i}, Q_{ww}) \mathrm{d}w_M \\ \int_{\Psi_i(k_w)} N_m(\nabla w_{M_i}, Q_{ww}) \mathrm{d}w_M \end{cases} \tag{5.39}$$

where the integration regions $\Lambda_i(k_w)$ and $\Psi_i(k_w)$ are defined as:

$$\begin{aligned} \Lambda_i(k_w) &= \{w_M \in R^m : \exists j \neq i, \ |w_j| > k_w\} \\ \Psi_i(k_w) &= \{w_M \in R^m : \exists j \neq i, \ |w_j| > |w_i|\} \end{aligned} \tag{5.40}$$

$\Lambda_i(k_w)$ describes the region in which any of the 'wrong' ($w_{j\neq i}$) w-test statistic exceeds the critical value (regardless of being larger or smaller than w_i), whereas $\Psi_i(k_w)$ describes the region in which any of the 'wrong' ($w_{j\neq i}$) w-test statistic is larger than w_i (regardless of exceeding or not the threshold). These regions ($\Lambda_i(k_w)$ and $\Psi_i(k_w)$) are shown in Figure 5.7. Fundamentally the two expressions represent two upperbounds to the WD probability, therefore we can choose the tightest of the two.

We first define the *specific* WD probabilities, κ_{ij}:

$$\kappa_{ij} = P(|\underline{w}_j| > k_w \cap |\underline{w}_i| < |\underline{w}_j| \ | H_i) \tag{5.41}$$

which is the probability that a WD occurs selecting specifically w_j (for a particular j) instead of w_i. Applying the Bonferroni inequality, we have:

$$\kappa_i \leq \sum_{j\neq i} \kappa_{ij} \tag{5.42}$$

Now we can find the two equivalent upperbounds of Equation (5.39) for each κ_{ij}:

$$\kappa_{ij} \leq \min \begin{cases} \int_{-\infty}^{-k} N(\nabla w_j | H_i, 1) \mathrm{d}w_j \\ \int_{0}^{+\infty} f_{|\underline{w}_j - \underline{w}_i|}(|w_j - w_i|) \end{cases} \tag{5.43}$$

The upperbounds in Equation (5.39) can thus be computed as:

$$\kappa_i \leq \min \begin{cases} \sum_{j\neq i} (1 - \Phi(k_w - \nabla w_i)) + \sum_{j\neq i} (1 - \Phi(k_w + \nabla w_i)) \\ \sum_{j\neq i} \left(\Phi\left(\frac{(\sigma_{w_i w_j} - 1)\nabla w_i}{2 - \sigma^2_{w_i w_j}} \right) + \Phi\left(\frac{(-\sigma_{w_i w_j} - 1)\nabla w_i}{2 - \sigma^2_{w_i w_j}} \right) \right) \end{cases} \tag{5.44}$$

The second equation can be obtained remembering that:

$$\underline{w}_j - \underline{w}_i \sim N((\sigma_{w_i w_j} - 1)\nabla w_i, 2 - \sigma^2_{w_i w_j})$$

The first upperbound is monotonously increasing with the bias size ∇ whereas the second one is monotonously decreasing. The upperbound given by the minimum of the two first increases to reach a maximum and then decreases monotonically with increasing bias size. This behaviour is visible in Figure 5.13, where the upperbound is applied to an actual geometry.

Until now we have considered only the identification step of the DIA, that is only the set of w-tests to be run.

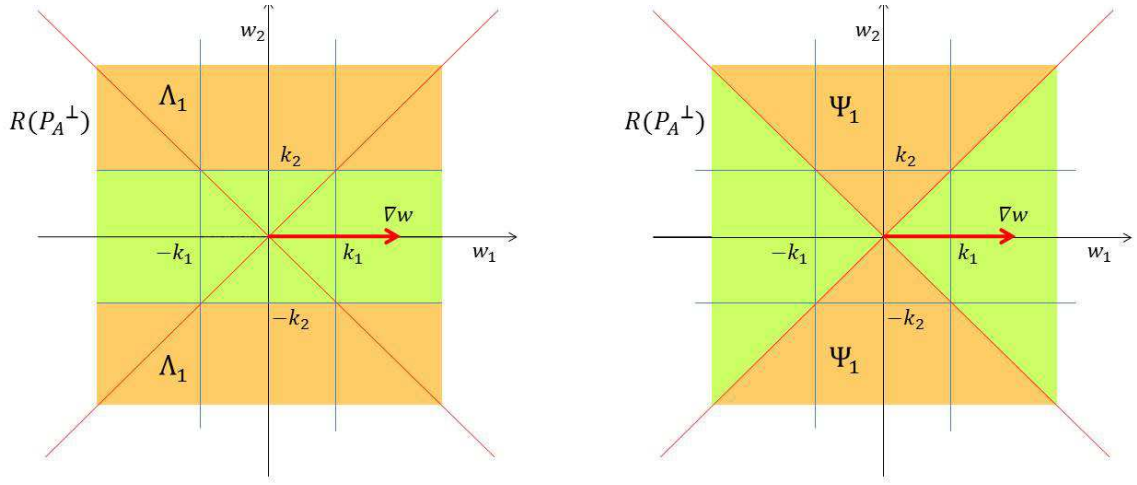


Figure 5.7: Example of regions $\Lambda_i(k_w)$ and $\Psi_i(k_w)$ employed to determine an upperbound to the WD probability. The regions Λ_1 and Ψ_1, relative to a failure in satellite 1, are in orange in the picture, which shows a two-dimensional $R(A)^{\perp}$ space. Both Λ_1 and Ψ_1 contain the actual Γ_1 region, previously defined (see Figure 5.6).

5

5.3.2. DIA DETECTION PERFORMANCE — ADDITION OF THE OMT

When detection is made by means of the OMT, significance and MD probability of an iteration of the DIA coincide with the α_{OMT} and β_{OMT}, performance of the OMT. Therefore, if originally a value for α_w was set, the α_{OMT} is derived employing the B-method (or another of the methods described in Section 5.2.2), and that corresponds to the significance of a full algorithm iteration. Alternatively, one could start from α_{OMT} and determine α_w using an inverse approach.

This means that when the OMT is run first, the w-tests only control the WD probability. Nevertheless, with the B-method, the MD probabilities of w-tests and OMT are strongly related — equal for a specific size of the bias, the MDB. In the GPS example of next Section, we show that with the adoption of the B-method the difference between MD probabilities of OMT and w-tests are very small for any size of the measurement biases. This result supports the adoption of the B-method for integrity monitoring purposes.

There is to note that in case of use of the B-method, there is a chance that the OMT detects an anomaly but none of the w-tests identifies it — occurrence resulting in an Alert. The probability of this type of occurrences cannot be computed analytically but it is generally quite small. In next Section this probability is estimated (by means of Monte Carlo integration) for a particular geometry.

It can be interesting to note that, when the w-tests are run following the OMT, they will follow a distribution that is different from the standard normal (since conditional on the detection). Knowledge of this distribution could be useful to determine the probability of (not) identifying after a detection (the event just discussed above), or actual (more precise) probabilities of WD and CD. This conditional distribution is analyzed in Appendix L.

5.3.3. AN EXAMPLE

We consider here an example GPS geometry and we compute for it the performance parameters α_{tot}, β_i, κ_i and $|\nabla w_i|_{MDB}$ using both correct formulae (Equations (5.30), (5.36) and (5.38)) and approximated ones (Equations (5.31), (5.37) and (5.44)). The skyplot of the geometry considered is shown in Figure 5.8.

The variance matrix Q_{yy} of the observables, employed in the computation, is diagonal and based on the elevation of the satellites. In fact for each observable we used the relation:

$$\sigma_{y_i} = \sigma_c \left(1 + 3\exp\left(\frac{-el_i}{el_0}\right)\right)$$

where $\sigma_c = 0.3m$ is the reference standard deviation, el_i is the elevation of the i^{th} satellite and $el_0 = 10°$ is the reference elevation.

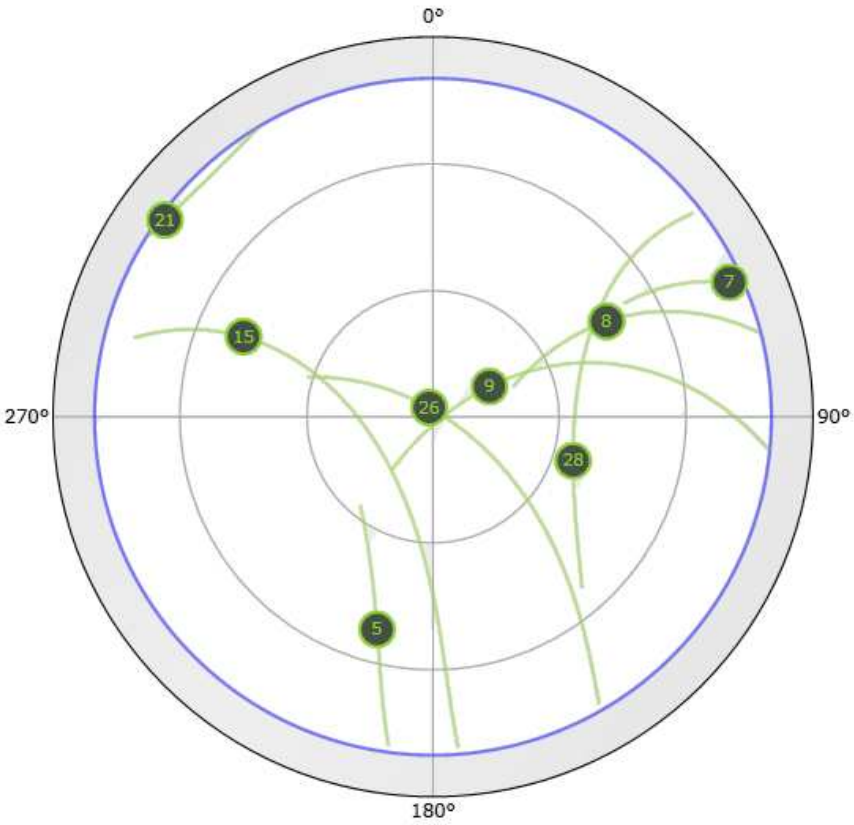


Figure 5.8: Skyplot with PRN numbers. Configuration with $m = 8$ satellites.

Figure 5.9 shows the results for α_{tot} obtained by MC numerical integration and by the approximation in Equation (5.31), for increasing values of the significance of each single w-test (α_i) and therefore decreasing value of the threshold k_w. The upperbound is quite tight, and exceeds the actual value by about 30%.

Figure 5.10 shows the computed β_i through MC integration of Equation (5.36) and through the approximated (upperbounded) Equation (5.37). The normalized difference of the two

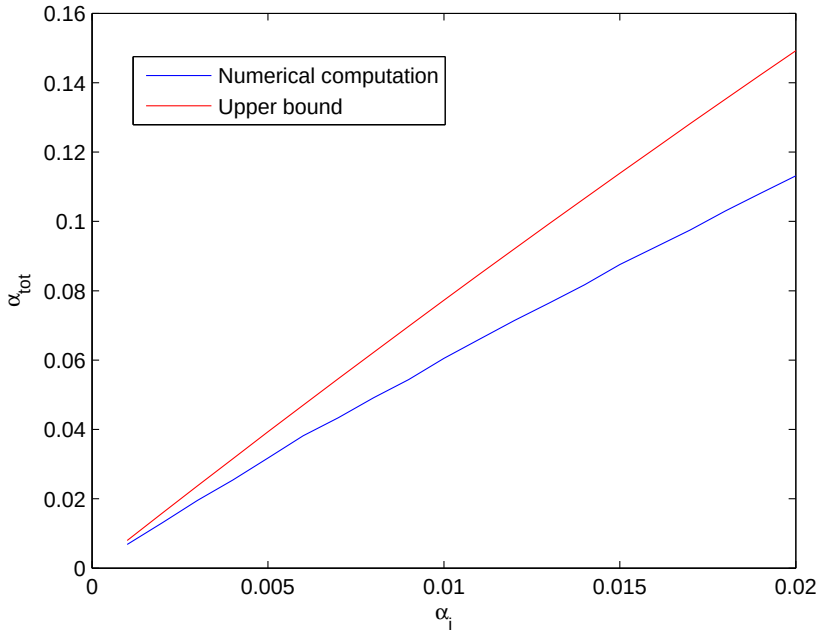


Figure 5.9: Total significance for the all m w-tests run together (first detection step), for increasing significance of each single w-test. Results of MC integration and of analytical approximation. For $\alpha_i = 0.001$ we have $\alpha_{tot} \approx 0.007$.

computations is shown in Figure 5.11. We can see that the relative error is generally smaller than 10%, and in almost all cases smaller than 20%. This means that, if for instance for $\nabla = 5$ we have $\beta = 10^{-3}$ with respect to a fault in a certain satellite, then the error $\beta_{approx} - \beta \approx 10^{-4}$. Note in Figure 5.10 that for a zero value of the bias size the curves for the β_i all start from the value $1 - \alpha_i = 0.999$ in the approximated (upperbounded) case, and from the value $1 - \alpha_{tot} \approx 0.993$ for the actual numerical computation.

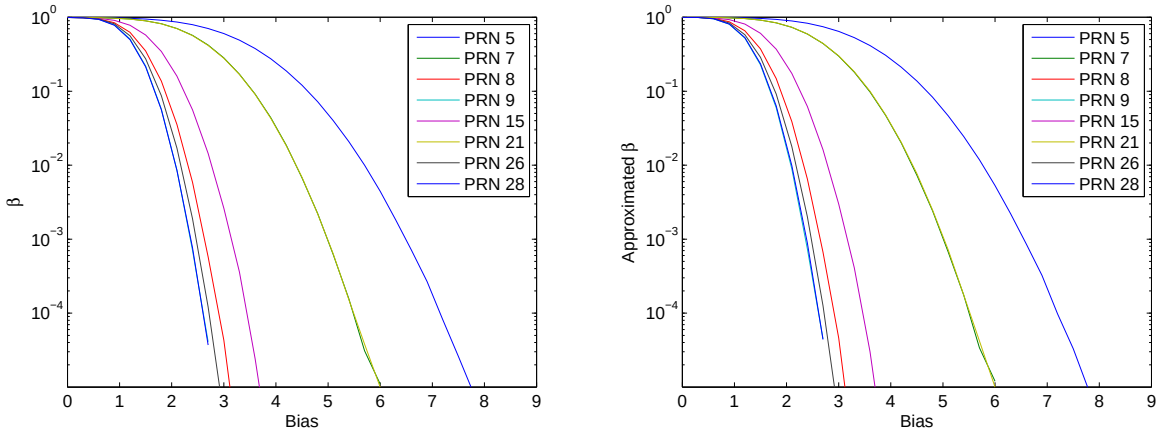


Figure 5.10: MD probabilities computed through MC integration of Equation (5.36) on the left and through the approximated Equation (5.37) on the right. The bias size ∇ on the horizontal axis is measured in meters, the threshold of the w-tests is set such that $\alpha_i = 0.001$.

Figure 5.12 shows the WD probabilities (κ_i) as functions of the size of the bias in the faulty

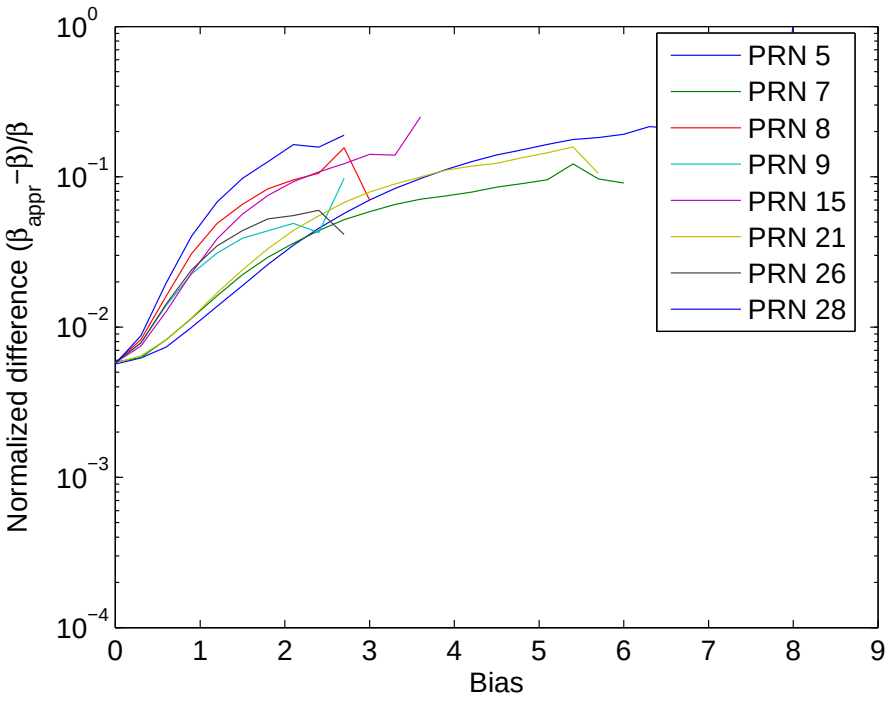


Figure 5.11: Normalized difference between the MD probabilities computed by the approximated formula in Equation (5.37) and by MC integration of the actual formula in Equation (5.36), shown separately in Figure 5.10: $(\beta_{i_{approx}} - \beta_{i_{MC}})/\beta_{i_{MC}}$.

satellite, given a fixed $\alpha_i = 10^{-3}$, obtained through MC integration. We can notice that all the curves start from the same value, for a zero value of the bias size. This is due to the fact that the w-tests distribution is symmetric with respect to all the bisectrix (or divisory planes) that separate the different detection regions shown in Figure 5.6. The detection region can in fact be divided in m different sub-regions, each relative to a different w-test. The probability of $\underline{w}_M$ falling in any of these sub-regions is just the same in case of zero bias size. In $m-1$ out of m cases $\underline{w}_M$ would fall in a wrong region (corresponding to WD), therefore all curves in figure start from the value $\kappa_0 = \frac{m-1}{m}(1-\alpha_{tot})$.

Figure 5.13 shows the approximated values as obtained by Equation (5.44) for the same case of Figure 5.12. As we can see, the upperbound is quite conservative, this is likely due to the high correlation present among the w-tests.

Finally, we consider the use of the OMT for detection, based on the B-method. Figure 5.14 shows α_{OMT} in comparison with α_{tot} as obtained from the multiplicity of w-tests (shown previosly in Figure 5.9), for different values of α_w. In the same Figure it is also shown the probability of detection but no identification (rejection by the OMT but acceptance of all the w-tests), indicated with the greek letter ζ. We can see that for small values of α_w, α_{OMT} is larger than both the actual and the upperbound (given by the Šidák approximation) of the α_{tot} of the combined w-tests, whereas for larger values of α_w it becomes sensibly smaller. This means that for larger values of α_w employing the OMT for Detection (with the B-method) results in a higher detection power even in case of single satellite faults. ζ results instead generally negligible compared to α_{tot} for any α_w. Figure 5.15 shows the difference $\beta_{OMT_i} - \beta_{i_{approx}}$, with the $\beta_{i_{approx}}$ computed with the approximating formula in Equa-

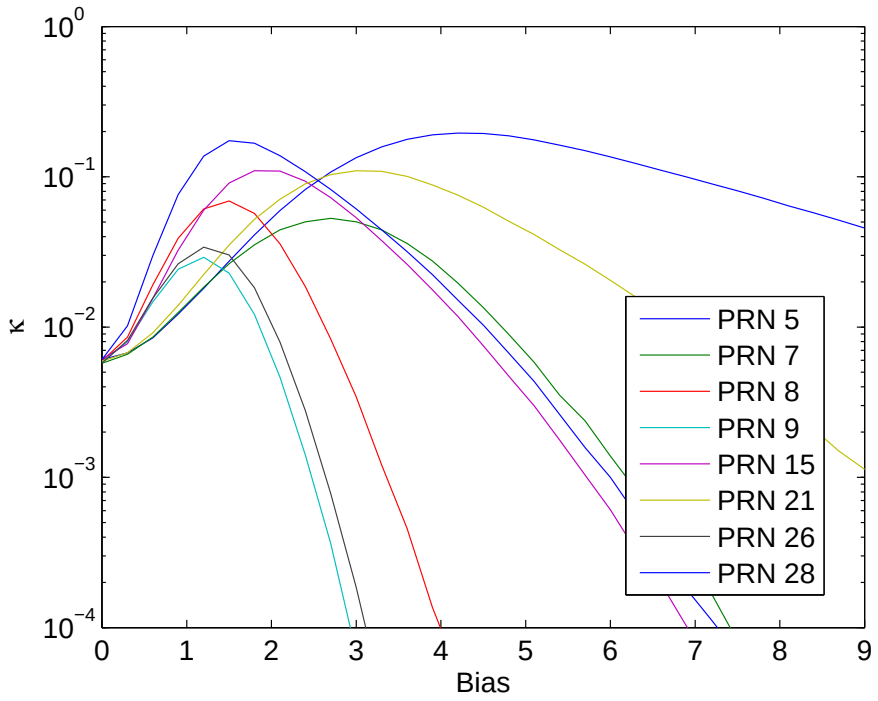


Figure 5.12: Probability of WD κ as a function of the size of the bias in the faulty satellite, with a fixed $\alpha = 10^{-3}$ and therefore fixed threshold for the w-tests, for the 8 satellites in view. Each curve exhibits a global extremum before decreasing monotonically and asymptotically to zero.

tion (5.37), i.e. the difference between the MD probability of OMT and w-tests (note that the difference is zero for the MDB, from the principle of the B-method). The graph shows that this difference is very contained.

5.4. SUMMARY AND CONCLUSIONS

The DIA procedure is a well established method for outlier detection in geodesy. It employs the most powerful test (UMPI) for the detection and identification of generic or (anticipated) specific anomalies and, with the B-method, defines a valid rationale for the choice of the thresholds of each single test.

Regarding the application of the DIA procedure to the RAIM problem, this is naturally possible, provided a direct way to compute the main RAIM parameters, in particular the P_{HMI}, is defined — the *reliability* monitoring provided by the DIA has to be extended to be able to cope specifically with the RAIM case. In a P_{HMI} monitoring set up the use of the B-method in not necessarily justified, and the choices of the input parameters for the method, e.g. α and β of each test can be discussed. An adaptation of the standard DIA algorithm for the RAIM case is proposed in Chapter 6.

In Chapter 7 the standard DIA is compared by means of simulations to the other RAIM algorithms discussed. To allow the comparison, the following choices have been made for the set up of the DIA algorithm:

- The maximum number of allowed iterations of the procedure is three. This choice is due to the assumption, employed also by the other algorithms tested in the simula-

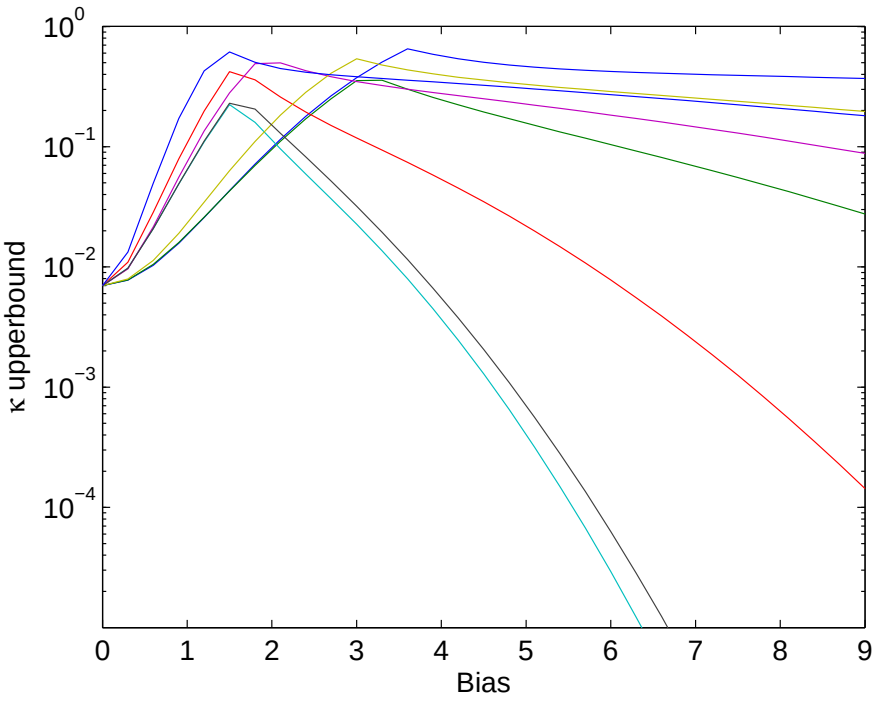


Figure 5.13: Approximated probability of WD κ, via Equation (5.44) as a function of the size of the bias in the faulty satellite, with a fixed $\alpha = 10^{-3}$ and therefore fixed threshold for the w-tests, for the 8 satellites in view. Each curve exhibits a global extremum before decreasing monotonically and asymptotically to zero.

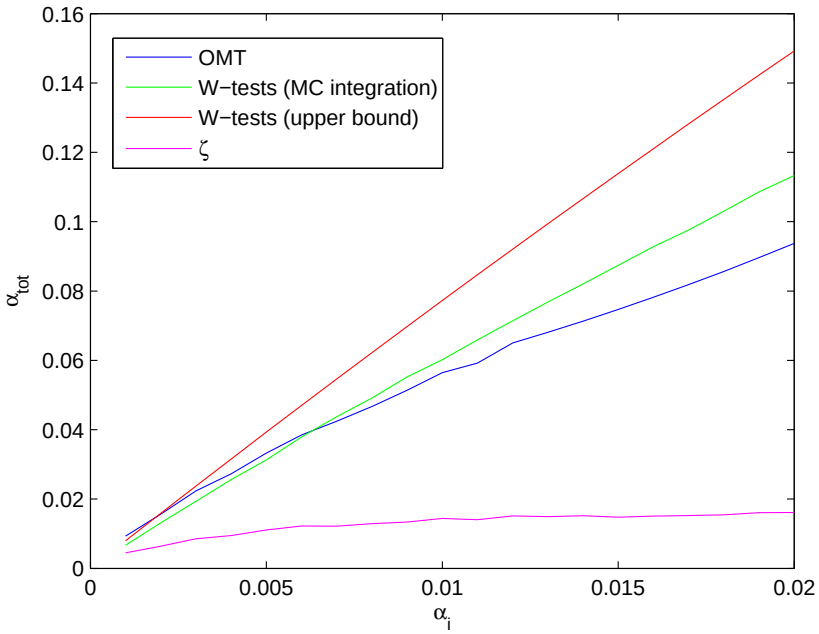


Figure 5.14: P_{FA} for the DIA Detection step (OMT), in comparison with the total P_{FA} in case of Detection based on the w-tests for data snooping, for varying significance of the individual w-tests, α_w (on the horizontal axis). Also shown the probability of rejection of the OMT but acceptance of all the w-tests (ζ).

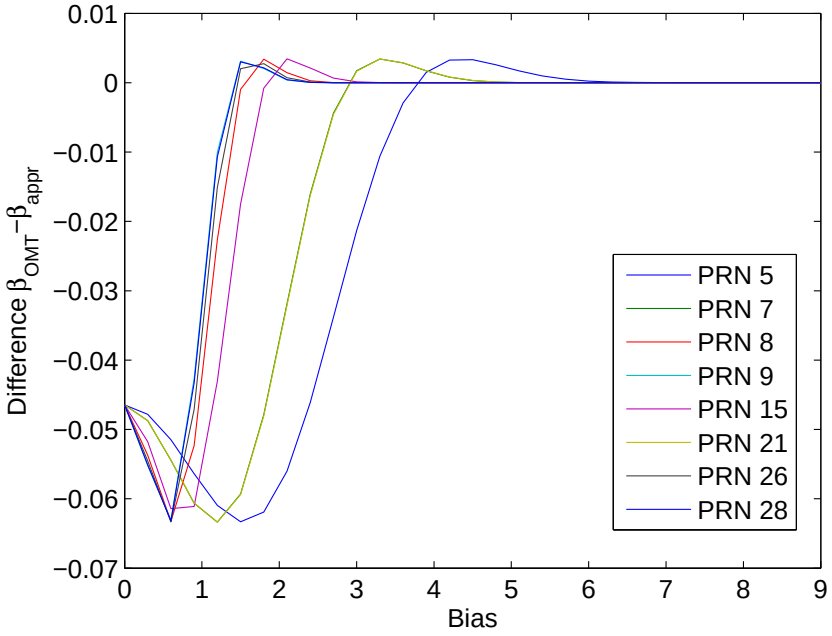


Figure 5.15: Difference between the MD probabilities of the OMT (β_{OMT_i}) and the (approximated) MD probabilities of the w-tests $\beta_{i_{approx}}$.

tions, that the occurrence of an anomaly affecting three or more satellites has probability of occurrence much lower than the requirement on the P_{HMI} (see also the formula employed in ARAIM, Equation (4.46)); this means that three (or more) satellites should never be considered faulty simultaneously, and therefore excluding more than two satellites is assumed unnecessary.

- The α_{OMT}, same for each iteration of the procedure, is set equal to the $P_{FA'}$ requirement given in Table 7.3 — in the standard DIA we would fix instead the α_i of the w-tests. This choice is made to assure that the requirements on $P_{FA'}$ is satisfied and to better compare the algorithm with the other ones, since the tests total significance is fixed. With this choice the possible multiple iterations are not taken into account (for the performance computation).

- The thresholds for the w-tests k_w are determined applying the B-method in *reverse* fashion, starting from the significance of the OMT. The details of this variant of the method are in the algorithm description, Section 5.5, highlighted in italic.

5.5. DIA ALGORITHM

In Baarda's DIA algorithm [4] [5] the OMT, run for detection purpose, precedes the w-tests, run for identification (iterative data-snooping). The baseline input for the algorithm, common to all the other RAIM algorithms, is given in Tables 7.1, 7.3 and 7.4 (the values reported are the ones employed in the numerical simulations in Chapter 7).

Table 5.3: Specific input to the Baarda's DIA algorithm.

Name	Description	Source
$\beta \equiv P_{MD}$	Maximum allowed probability of non-detecting a fault (usually set to 0.8)	A-priori
$\alpha_w \equiv P_{FA_w}$	False Alarm probability for each w-test, testing the null versus a single alternative hypothesis test	A-priori

1. Input (extra than Tables 7.1, 7.3 and 7.4) in Table 5.3.

2. Check if the redundancy is sufficient:

$$\begin{cases} \text{Continue procedure if } N_{sat} > 3 + N_{const} \\ \text{Declare Alert if } N_{sat} \leq 3 + N_{const} \end{cases} \tag{5.45}$$

3. Detection:

   (a) Compute $\underline{T}_{m-n}$ (χ^2 distributed) test statistic:
   The $\underline{T}_{m-n}$ statistic is computed as:

$$\underline{T}_{m-n} = \underline{y}^T (Q_{yy}^{-1} - Q_{yy}^{-1} A (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}) \underline{y} \tag{5.46}$$

   (b) Compute $\underline{T}_{OMT}$ test threshold:
   Compute first the w-tests thresholds:

$$k_w = \sqrt{\chi_1^{2inv}(1 - P_{FA_w})} \tag{5.47}$$

   where the operator χ_1^{2inv} is the inverse of the CDF of a central χ^2 distribution with 1 degree of freedom. Compute:

$$\lambda_0 : \chi_{(1,\lambda_0)}^{2cdf}(k_w^2) = P_{MD} \tag{5.48}$$

   where $\chi_{(1,\lambda))}^{2cdf}$ stands for the CDF of a non-central χ^2 distribution with 1 degree of freedom and non-centrality parameter λ. λ_0 can be found iteratively, solving for instance the following minimization:

$$\lambda_0 = \arg\min_{\lambda} : \chi_{(1,\lambda)}^{2cdf}(k_w^2) \leq P_{MD} \tag{5.49}$$

   Compute the $\underline{T}_{m-n}$ test threshold:

$$k_{\text{OMT}} = \chi_{(N_{sat}-3-N_{const},\lambda_0)}^{2inv}(P_{MD}) \tag{5.50}$$

   where $\chi_{(N_{sat}-3-N_{const},\lambda_0)}^{2inv}$ stands for the inverse CDF of a non-central χ^2 distribution with $N_{sat}-3-N_{const}$ degrees of freedom and non-centrality parameter

λ_0. The choice of the above described thresholds for OMT and w-tests follows the B-method [4] [5], as described in Section 5.2.2.

In the simulation implementation (Chapter 7), to better compare with the other algorithms, the total $P_{FA_{tot}}$ is fixed instead of P_{FA_w}. Two versions of DIA are implemented, for each of which the threshold for the OMT is computed as:

$$\begin{array}{ll} \text{DIA-}k_1 & k_{\text{OMT}} = \chi^{2inv}_{N_{sat}-3-N_{const}}(1-P_{FA'}) \\ \text{DIA} & k_{\text{OMT}} = \chi^{2inv}_{N_{sat}-3-N_{const}}(1-\sqrt[3]{P_{FA'}}) \end{array} \tag{5.51}$$

where the operator $\chi^{2inv}_{N_{sat}-3-N_{const}}$ is the inverse of the cdf of a central χ^2 distribution with $N_{sat}-3-N_{const}$ degrees of freedom. The choice for the DIA on the second line is due to the fact that a maximum of 3 iterations are performed in the simulation implementation (see Section 6.5.1 for more details on this approximated formula).

(c) Determine overall consistency of measurements:

$$\begin{cases} \text{Accept } H_0 \text{ if } T_{OMT} \leq k_{\text{OMT}} \\ \text{Reject } H_0 \text{ if } T_{OMT} > k_{\text{OMT}} \end{cases} \tag{5.52}$$

If H_0 is rejected an Identification has to be made, go to step 4. Otherwise standard operation is continued with solution estimate:

$$\hat{x} = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} y = Sy$$

.

4. Identification (only if H_0 was rejected by OMT):

(a) Check whether the redundancy is sufficient to exclude satellites:

$$\begin{cases} \text{Continue procedure if } N_{sat} > 3 + N_{const} + 1 \\ \text{Declare integrity unavailable if } N_{sat} \leq 3 + N_{const} + 1 \end{cases} \tag{5.53}$$

(b) Compute w-tests statistics: compute a $\underline{w}_i$ test statistic for each satellite i (data snooping):

$$\underline{w}_i = \frac{\underline{\hat{e}}_i}{\sigma_{\hat{e}_i}} \quad \text{for } i = 1,2,\ldots,N_{sat} \quad \text{cf. Equation (5.5)}$$

where $\underline{\hat{e}}_{0_i}$ is the i^{th} component of the residuals vector $\underline{\hat{e}}$ and $\sigma_{\hat{e}_i}$ is the i^{th} diagonal element of the residuals covariance matrix $Q_{\hat{e}\hat{e}}$ determined with:

$$\begin{aligned} \underline{\hat{e}} &= (I_{N_{sat}} - A(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}) y = P_A^{\perp} y \\ Q_{\hat{e}\hat{e}} &= Q_{yy} - A(A^T Q_{yy}^{-1} A)^{-1} A^T = P_A^{\perp} Q_{yy} P_A^{\perp T} \end{aligned} \tag{5.54}$$

Note that Q_{yy} is assumed diagonal to apply Equation (5.5).

(c) Recall the w-tests thresholds computed with Equation (5.47).

In the simulation implementation (Chapter 7) the following procedure is applied. Compute:

$$\lambda_0 = \arg\min_{\lambda} : \chi^{2cdf}_{(N_{sat}-3-N_{const},\lambda)}(k_{\text{OMT}}) \leq P_{MD} \tag{5.55}$$

where $\chi^{2cdf}_{(N_{sat}-3-N_{const},\lambda)}$ stands for the CDF of a non-central χ^2 distribution with $N_{sat}-3-N_{const}$ degrees of freedom and non-centrality parameter λ. λ_0 can be found iteratively. Compute:

$$k_w = \sqrt{\chi^{2inv}_{(1,\lambda_0)}(P_{MD})} \tag{5.56}$$

where $\chi^{2cdf}_{(1,\lambda_0))}$ stands for the CDF of a non-central χ^2 distribution with 1 degree of freedom and non-centrality parameter λ_0. The choice of the above described thresholds for OMT and w-tests is equivalent to the application of the B-method [4] [5] in reverse order, starting from the OMT threshold to obtain the w-tests thresholds.

(d) Identify faulty satellite:

If:

$$\max_i w_i \leq k_w \quad \forall i : \ 1 \leq i \leq N_{sat} \quad \text{cf. Equation (5.21)}$$

Declare Alert, since it is not possible to identify a particular type of fault and perform an adaptation of the model, whereas the OMT rejected H_0. Else, find:

(e)

$$i_{ex} = \arg\max_i w_i \quad \forall i : \ 1 \leq i \leq N_{sat} \tag{5.57}$$

Satellite i_{ex} has to to be excluded from the model. Go to step 5.

5. Adaptation:

The measurement corresponding to the faulty satellite i_{ex} is removed from the model. New geometry matrix, measurement variance matrix and observable vector, respectively A^{new}, Q^{new}_{yy} and $\underline{y}^{new}$, are re-constructed after exclusion: $\underline{y}^{new}$ does not include the measurement from the excluded satellite, A^{new} is obtained from A taking out the row corresponding to the excluded satellite and Q^{new}_{yy} instead is obtained from Q_{yy} removing both row and column corresponding to the excluded satellite. Also N_{sat} is correspondingly updated, $N^{new}_{sat} = N_{sat}-1$. With this new input we move back to step 2 and the algorithm is repeated until finally either standard operation is continued with a position estimate or Alert is declared.

6

DIA APPLIED TO RAIM

In the previous Chapter we introduced the DIA procedure, and discussed the basic principles of this FDE algorithm. Strengths and weaknesses of the procedure were highlighted. We also pointed out that the DIA procedure cannot directly fulfill the role of a RAIM algorithm due to lack of a direct computation of the P_{HMI}. In this Chapter a modification (and an extension) of the basic procedure is proposed, in order to comply with the requirements of a RAIM algorithm and to optimize the associated performance.

6.1. SPECIFIC APPLICATION NEEDS

With reference to Section 2.4, we have seen that a RAIM algorithm needs as input the parameters α_0 and P^{req}_{HMI}. To satisfy both those requirements, it has to be able to computate/estimate the P_{HMI} and α as functions of each other for each geometry: in such a way the availability of the geometry under consideration can be determined. The standard DIA as described in previous Chapter is based on a prior choice for the α and β of the tests that is mostly arbitrary, to be tuned to each specific application. In this sense, RAIM can represent a specific application, that requires monitoring of the parameter P_{HMI}.

The adaptation of the DIA procedure proposed in this Chapter consists in the development of a method to estimate/upperbound the P_{HMI} relative to an observation geometry in order to determine the availability of the geometry. This method also exploits the particular statistical model introduced in Section 2.7, foreseeing the assignment of prior probabilities to the alternative hypotheses. The assignment of prior probabilities is not required for the application of the standard DIA procedure, therefore the method proposed in this Chapter provides a way to exploit the extra information constituted by these prior probabilities.

Beside this integrity monitoring addition to the standard DIA, the standard role of the OMT is also discussed (Section 6.6). Alternatives to the B-method are taken into consideration to set the threshold of the OMT as well as the option to run the OMT *in parallel* with the w-tests. In the latter case both OMT and w-tests would assume the purpose of Detection (and Identification). This option would constitute a real modification to the standard DIA algorithm — we will refer to this algorithm as adapted DIA.

6.2. P_{HMI} COMPUTATION FOR THE DIA — INTRODUCING PRIOR INFORMATION

In previous Chapter we have described the DIA procedure and determined its performance in terms of α, β, κ and MDB. We have seen in Section 2.5 that other performance parameters are to be evaluated when monitoring the integrity of the system, in particular the P_{HMI}.

In this section a method to compute the P_{HMI} is proposed, based on the following concepts:

1. Prior probabilities of occurrence of each hypothesis are introduced: a prior probability for the occurrence of the fault-free hypothesis and a prior probability for the occurrence of each fault mode envisioned, following the model presented in Section 2.7.

2. An assumption of a worst-case bias is made: in case of an anomaly affecting the measurements (fault mode), the bias present in the measurements will be unknown. For computation of the P_{HMI} we conservatively assume that the bias is always of the size that maximizes the P_{HMI}.

In this approach, as described in section 2.7, the hypothesis H in which the system lies describes a state of nature, and is considered as a random variable, $\underline{H}$. This variable can assume the values $H_0, H_1, \ldots, H_{N_{H_a}}$ where N_{H_a} is the number of alternative hypotheses considered, and is characterized by a PMF. To each occurrence $\underline{H} = H_i$ is assigned therefore a probability mass p_i. The hypotheses are furthermore constructed in such a way to always represent mutually exclusive events. It holds $\sum_{i=0}^{N_{H_a}} p_i = 1$.

In Sections 6.3 and 6.4 an upperbound for the P_{HMI} is obtained for a single iteration of the DIA procedure. In Section 6.5 the case of multiple iterations is instead considered. The P_{HMI} is a function of the FDE procedure employed, in the same way as it is function of the position estimator in use. We will in fact see that the two different procedures, based on one or multiple iterations, lead to a different P_{HMI} — result that should steer the choice between the two algorithms.

6.3. PRIOR PROBABILITIES AND P_{HMI} CONTRIBUTIONS UNDER EACH HYPOTHESIS

Recall the definitions of P_{PF} and absolute P_{HMI} in Equations (2.24) and (2.26):

$$P_{PF} = P(\underline{\hat{x}} - x \notin \Omega_{\text{AL}})$$
$$P_{\text{HMI}} = P(PF \cap \text{No Alert})$$

Considering the first iteration of the algorithm, in case of a fault (alternative hypothesis H_a) no Alert will occur in case of Missed Detection (MD), and the $P_{\text{HMI}|H_a}$ can be computed as in Equation (3.29):

$$P_{\text{HMI}|H_a} = P(PF \cap \text{No alert}|H_a) = P_{PF|MD_{H_a}} \cdot \beta_{H_a} = P_{PF_{H_a}} \cdot \beta_{H_a} \tag{6.1}$$

where the last equivalence is true if test statistics and position error are independent (and this is the case when employing UMPI tests, as long as the input is normally distributed). Note that we employed here the subscript H_a to indicate the probabilities of MD and PF under (conditioned on) that hypothesis — β_{H_a} and $P_{PF_{H_a}}$ — and that $\beta_{H_a} = P(\text{No Alert}|H_a)$.

Assuming that the different hypotheses are mutually exclusive events (this is a quite common assumption for single satellite failures), the total P_{HMI} is obtained as a sum of the $P(HMI \cap H_i) = p_i P(HMI|H_i)$, the contributions of the different hypotheses being determined with Equation (6.1):

$$\begin{aligned} P_{\text{HMI}} &= \sum_{i=0}^{t} P(HMI \cap H_i) = \\ &= \sum_{i=0}^{t} p_i \cdot \beta_i \cdot P_{PF_i} \end{aligned} \tag{6.2}$$

where with the subscript i we mean that the quantity refers to (is conditioned on) the hypothesis H_i. Now we just have to find a way to compute β_i and P_{PF_i}, or an upperbound for them.

In the following we consider for simplicity the case x_{int} one-dimensional, say $x_{int} = x_3$, the third component of the full parameter vector x. This can for instance represent the case when only the integrity for one position dimension, the vertical, is monitored. We have for the conditional probability of PF:

$$P_{PF_i} = P(|\underline{\hat{x}}_3 - x_3| > \text{VAL}|H_i) \tag{6.3}$$

Suppose we use, as in the DIA procedure, the BLUE $\underline{\hat{x}} = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} \underline{y} = S\underline{y}$ under the null hypothesis as estimator $\underline{\hat{x}}$. Under each alternative hypothesis H_i the distribution of $\underline{\hat{x}}_3$ is known as function of the bias size vector ∇_i relative to that hypothesis. In fact under the H_i hypothesis $\underline{\hat{x}}_3$ has distribution:

$$\underline{\hat{x}}_3|H_i \sim N(x_3 + \nabla\hat{x}_{3_i}, \sigma^2_{\hat{x}_3}) \tag{6.4}$$

with $\nabla\hat{x}_{3_i} = S_{3j} C_{y_i} \nabla_i$ (as from Equation (5.15)), where S_{3j} is the third row of matrix $S = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$. Therefore P_{PF_i} can be computed as a function of ∇_i as the integral of a normal distribution:

$$P_{PF_i}(\nabla_i) = P(|\underline{\hat{x}}_3 - x_3| > \text{VAL}|H_i) = \Phi\left(\frac{S_{3j} C_{y_i} \nabla_i - VAL}{\sigma_{\hat{x}_3}}\right) + \Phi\left(\frac{-S_{3j} C_{y_i} \nabla_i - VAL}{\sigma_{\hat{x}_3}}\right) \tag{6.5}$$

with Φ being the cumulative distribution function of a standard normal distribution. β_i, as function of ∇_i, was already approximated in Equation (5.37) as:

$$\beta_i \leq \Phi(k - \nabla w_i) - \Phi(-\nabla w_i - k) \tag{6.6}$$

Putting together Equations (6.5) and (6.6) to determine the $P_{HMI|H_a}$ defined in Equation (6.1) as function of ∇_i, we have:

$$\begin{aligned} P_{\text{HMI}|H_i}(\nabla_i) \leq [\Phi(k - \nabla w_i) - \Phi(-\nabla w_i - k)] \cdot \\ \left[\Phi\left(\frac{S_{3j} C_{y_i} \nabla_i - VAL}{\sigma_{\hat{x}_3}}\right) + \Phi\left(\frac{-S_{3j} C_{y_i} \nabla_i - VAL}{\sigma_{\hat{x}_3}}\right)\right] \end{aligned} \tag{6.7}$$

Note that in case of multi-dimensional x_{int}, the reasoning described so far still holds true, and the only difference is that the P_{PF} cannot be computed as a simple integral of a standard normal distribution but requires the integration of a multinormal distribution outside an integrity region Ω_{AL} (as in Equation (2.24)).

6.4. WORST-CASE BIAS

The $P_{\mathrm{HMI}|H_i}$ has been expressed in Equation (6.7) as a function of the bias size ∇_i. At this point we apply the worst-case bias approach, and look for the bias size that maximizes the P_{HMI}. We have to distinguish here between single satellite faults and multiple satellite faults. In the first case only the size of the bias is unknown, whereas in the second case both bias direction and size are unknown — therefore it is necessary to find a worst-case bias direction and a worst-case bias size.

6.4.1. SINGLE SATELLITE FAULTS

In case of single satellite fault, we maximize directly each of the terms $P_{\mathrm{HMI}|H_i}(\nabla_i)$ over ∇_i and perform a weighted sum as in Equation (6.2), where the weights are the a-priori probabilities of occurrence of each hypothesis. This maximization can be made directly employing a line-search algorithm. This method has been already employed in many other RAIM works, for instance [60] and [52]. As a final result we have, for the P_{HMI} relative to the single component x_3:

$$P_{\mathrm{HMI}} \leq \sum_{i=0}^{N_{Ha}} p_i \cdot \max_{\nabla_i} \left\{ [\Phi(k-\nabla w_i) - \Phi(-\nabla w_i - k)] \cdot \left[\Phi\left(\frac{S_{3j} C_{y_i} \nabla_i - VAL}{\sigma_{\hat{x}_3}} \right) + \Phi\left(\frac{-S_{3j} C_{y_i} \nabla_i - VAL}{\sigma_{\hat{x}_3}} \right) \right] \right\} \tag{6.8}$$

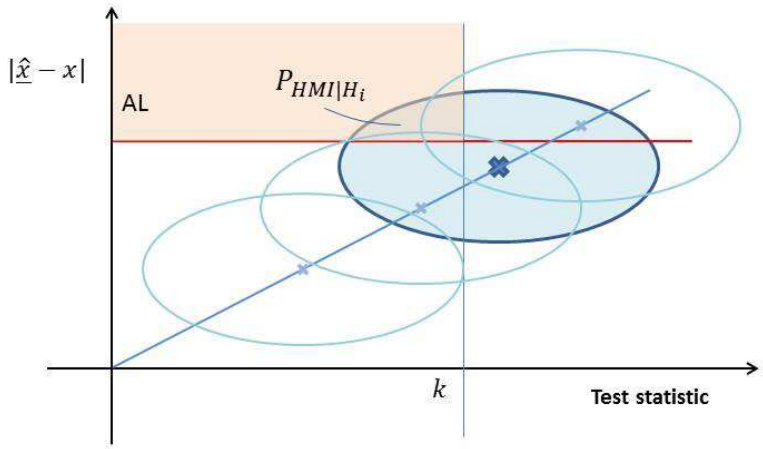


Figure 6.1: Position error and w-test distributions. On the horixontal axis is the w-test statistic for a faulty measurement, on the vertical axis the position error. The ellipse represents their joint distribution for a determined size of the bias affecting the measurement. The center of the ellipse moves along a straight line (in figure) when the bias size changes, as can be seen from Equations (5.3) and (5.15). The integral of the joint distribution over the area labeled as $P_{\mathrm{HMI}|H_i}$ corresponds to Equation (6.7), the P_{HMI} conditioned on the occurrence of that special fault with that particular size of the bias. $\underline{w}$ and $\underline{\hat{x}}$ are assumed uncorrelated in the diagram.

To explain this P_{HMI} computation we refer to Figure 6.1. The ellipse represents the joint distribution of w-test and position error, which is a simple bivariate normal distribution,

supposed a certain bias affects the measurement corresponding to the w-test considered (note that in case x is multi-dimensional the joint distribution will be a multivariate normal distribution with the corresponding extra degrees of freedom, whereas if we consider also the full distribution of the $\underline{w}_M$ vector we would have a multivariate normal distribution with $m-1$ extra degrees of freedom). With a different bias, the ellipse would be translated to a different position, but its shape, size and orientation would be the same. The Alert Limit and the threshold are pre-set, therefore the conditional $P_{\mathrm{HMI}|H_i}$ is determined computing the integral of the joint distribution over the area correspondingly labeled $P_{\mathrm{HMI}|H_i}$, which is the integrity risk associated with this particular choice of fault and bias size. Since the bias size is not known, to be on the safe side a worst-case scenario is considered, that means the bias size that maximizes the integrity risk is foreseen, and the $P_{\mathrm{HMI}|H_i}$ in Equation (6.7) is computed for all possible bias sizes. This computation is shown in Figure 6.2.

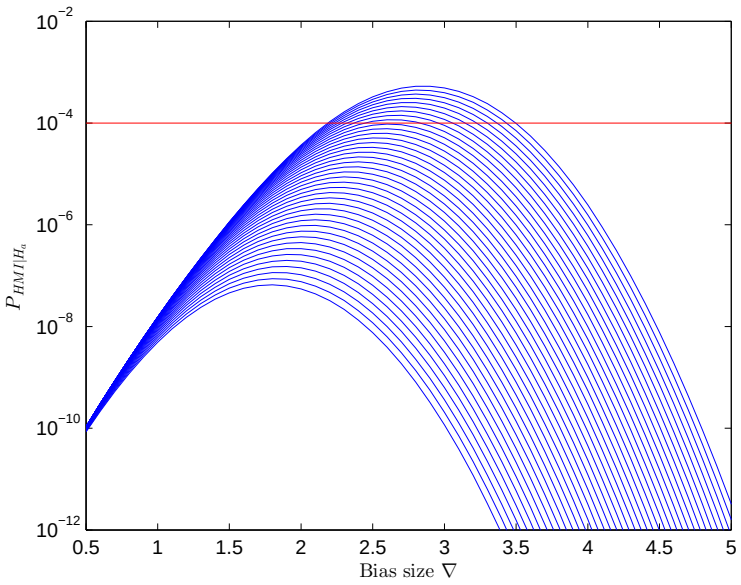


Figure 6.2: $P_{\mathrm{HMI}|H_a}$, as computed in Equation (6.1), in dependence of bias size for different choices of the threshold k. On the horizontal axis is the size of the bias ∇_i, on the vertical axis the $P_{\mathrm{HMI}|H_a}$, in case of fault in satellite PRN26 from the skyplot in Figure 5.8. Each parabolic blue line represents a different threshold k, in this example ranging from $\alpha = 0.0357$ to $\alpha = 2 \times 10^{-9}$.

This graph shows how, for different values of the threshold k, the integrity risk varies with the size of the bias. For many different values of the threshold, a blue line is traced, which shows a sort of parabolic shape: with increasing bias size, the risk increases up to a certain maximum and then decreases again. The maximum of each curve corresponds to the *worst-case bias* for each choice of the threshold. Each curve cannot be asymptotically increasing because the P_{HMI} is the product of $P_{PF} = P(|\underline{\hat{x}}_3 - x_3| > AL)$ and β_i, of which the first term is usually increasing with the bias size but converging to a maximum of 1 (since it is a probability, bounded in $[0,1]$), and the second term is instead decreasing with the bias size (for the usual unimodal type of distribution, see [99] for the case of simple approximation in Equation (5.37)), tending to zero in the limit $\nabla \to 0$. Therefore the product is either only decreasing or has a parabolic trend, finally decreasing once the convergence of the first term to 1 has been reached. The monotonously decreasing behavior of the second term for the

actual case of β_i as in Equation (5.36) can be demonstrated again by means of Anderson's theorem in [2]. This proves that a worst-case bias that maximixes Equation (6.8) exists. The worst-case bias size ∇_{WC_i} for each satellite i can be expressed as:

$$\begin{gathered}\nabla_{WC_i} = \arg\max_{\nabla_i}(\beta_i \cdot P_{PF_i}) = \\ \arg\max_{\nabla_i}\left\{[\Phi(k-\nabla w_i)-\Phi(-\nabla w_i-k)]\cdot\left[\Phi\left(\frac{S_{3j}C_{y_i}\nabla_i - AL}{\sigma_{\hat{x}_3}}\right)+\Phi\left(\frac{-S_{3j}C_{y_i}\nabla_i - AL}{\sigma_{\hat{x}_3}}\right)\right]\right\}\end{gathered} \tag{6.9}$$

In Appendix M results for the computation of the P_{HMI} and related quantities for the geometry in Figure 5.8 are reported.

6.4.2. MULTIPLE SATELLITE FAULTS

In case of multiple satellite fault, to determine the worst-case bias it is necessary to first determine the worst-case bias direction and only then find the worst-case bias size, by means of a line-search algorithm as in the single satellite case. The worst-case bias direction can be found analytically as explained in the following.

WORST-CASE BIAS DIRECTION

We want to find the worst-case direction for the bias ∇y, given C_y (which means the worst case direction of ∇) from the integrity point of view, i.e. the bias direction that maximizes $\|\nabla\hat{x}\|$ (or some other function of $\nabla\hat{x}$, e.g. $|\nabla\hat{x}_3|$) and at the same time minimizes the detection power of the test. To minimize the detection power we can find the bias direction that minimizes the non-centrality parameter λ. To maximize $\|\nabla\hat{x}\|_{Q_{\hat{x}\hat{x}}}$, we can maximize the length of the projection onto $R(A)$ of ∇y, $\|P_A C_y \nabla\|_{Q_{yy}^{-1}}$. Since λ is the squared length of the projection of $C_y\nabla$ onto $R(A)^\perp$, in this case the problem is equivalent to finding the vector lying on $R(C_y)$ that forms the largest angle with $R(A)^\perp$.

The tangent of this angle, that we call φ (see also Appendix F, and [99]), can be written as:

$$\tan\varphi = \frac{\|P_A C_y \nabla\|_{Q_{yy}^{-1}}}{\|P_A^\perp C_y \nabla\|_{Q_{yy}^{-1}}} \tag{6.10}$$

The problem consists in finding the elements of vector ∇ that maximizes this tangent, keeping the length of the vector constant/fixed $\|\nabla\| = 1$ (changing only its direction):

$$\nabla_{WCD} = \arg\max_{\nabla} \tan\varphi \quad \text{with } \|\nabla\| = 1 \tag{6.11}$$

By Linear Algebra, the maximum of $\tan\varphi$ is given (see also [99]) by the largest eigenvalue of:

$$(C_y^T P_A^{\perp T} Q_{yy}^{-1} P_A^\perp C_y)^{-1} C_y^T P_A^T Q_{yy}^{-1} P_A C_y$$

From a geometrical point of view, the worst-case direction of ∇ solution of the problem is given by any vector lying in the space $R(P_{C_y}A)$, obtained from the projection of A onto $R(C_y)$. Multiple directions can be solutions of the problem, depending on the dimension of $R(A)$.

In case we intend to monitor only one dimension of the position solution at a time, for instance $|\nabla\hat{x}_3|$, then there is only one direction (disregarding its sign) solution of the problem. The worst-case direction for this particular case was derived also in [51]. The equivalent of $\tan\varphi$ to be maximized in this case is:

$$s = \frac{|\nabla\hat{x}_3|}{\|P_A^\perp C_y \nabla\|_{Q_{yy}^{-1}}} \tag{6.12}$$

The problem is equivalent to finding the maximum of:

$$s^2 = \frac{|\nabla \hat{x}_3|^2}{\|P_A^{\perp} C_y \nabla\|^2_{Q_{yy}^{-1}}} = \frac{(LSC_y\nabla)^2}{\nabla^T C_y^T P_A^{\perp T} Q_{yy}^{-1} P_A^{\perp} C_y \nabla} \tag{6.13}$$

where:

$$L = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ and } S = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$$

This is the same type of problem as in Equation (6.11). The maximum s^2 coincides with the largest eigenvalue of:

$$(C_y^T P_A^{\perp T} Q_{yy}^{-1} P_A^{\perp} C_y)^{-1} C_y^T S^T L^T L S C_y$$

In this case, this matrix is derived from an outer product (since $C_y^T S^T L^T$ is a $q \times 1$ vector), therefore it has only one non-zero eigenvalue (corresponding to the maximum s^2), which value is:

$$s^2_{max} = LSC_y (C_y^T P_A^{\perp T} Q_{yy}^{-1} P_A^{\perp} C_y)^{-1} C_y^T S^T L^T \tag{6.14}$$

Substituting this value for s^2 in the expression in Equation (6.13), we find by inspection that the worst-case direction for ∇ is:

$$\nabla_{WCD} = (C_y^T P_A^{\perp T} Q_{yy}^{-1} P_A^{\perp} C_y)^{-1} C_y^T S^T L^T \tag{6.15}$$

WORST-CASE BIAS SIZE

Once the worst-case bias direction is known (Equation (6.15)), the worst-case bias size can be found following the same procedure as for single satellite fault case. That is, maximizing Equation (6.8), with a line-search algorithm.

The DIA algorithms implemented in the simulations of Chapter 7 do not take into account multiple satellite faults in the P_{HMI} computation (their contribution is assumed negligible).

6.5. SUCCESSIVE ITERATIONS

In Sections 5.1.4 and 5.1.6 we analyzed the performance of the OMT and of the w-tests in the simple binary context H_0 vs H_a. In Section 5.3 we considered the multiplicity of the w-tests and analyzed the performance of the full Identification step, in which the w-tests are run at the same time. Later in Section 5.3, we coupled the Identification step with the Detection step (with the OMT), and analyzed the performance of the two combined steps, Detection and Identification, which together complete a full algorithm iteration[1].All quantities computed (α, β, κ) up to previous Chapter therefore refer only to one iteration of the RAIM algorithm, the first one, i.e. the first round of Detection, Identification and Adaptation. This can be considered as the first round of the whole iterative procedure, which can end immediately if no adaptation is required, but can also entail further iterations. Therefore, while

[1]By iteration we mean any round of the procedure, including the first one.

the computed α, β, κ measure the performance of the first step, it is the performance of the whole algorithm, i.e. the total performance resulting from the succession of the DIA steps, that has to be eventually monitored.

At each iteration of the DIA algorithm, depending on which hypothesis holds true, the following events can occur:

H_0 : False Alarm *or*
No Alarm (Nominal Operations)
H_i : Missed Detection *or*
Detection and Correct Identification *or*
Detection and Wrong Identification (Wrong Detection)

Note that in case we are considering an iteration of the algorithm different from the first, H_0 refers to an already adapted model, resulting from an earlier adaptation step (from the previous iteration).

Suppose as an example that only two full iterations of DIA are attempted, that is, the algorithm foresees a new detection step after a first exclusion is performed, but if the new detection step leads to a further rejection the algorithm stops and an Alert is declared. Then, the possibly occurring events are the ones shown in Tables 6.1 and Tables 6.2, respectively in the cases of H_0 and H_i holding true. The events of False Alarm, Missed Detection and Wrong Detection are indicated with respectively the notation FA, MD and WD.

Table 6.1: Scheme of the possible occurring events in case of a RAIM algorithm made up of two successive DIA iterations, case of null hypothesis H_0 holding true. The subscripts 1 and 2 refer to the corresponding iteration of the procedure.

	H_0		
Step 1	No Detection (Nominal Operations)	Detection (FA_1) & Exclusion	
Step 2 (New tests)	(No step 2)	No Detection (Nominal Operations)	Detection → FA ($FA_2 \vert FA_1$)

Table 6.2: Scheme of the possible occurring events in case of a RAIM algorithm made up of two successive DIA iterations, case of alternative hypothesis H_i holding true. The subscripts 1 and 2 refer to the corresponding iteration of the procedure.

	H_i					
Step 1	MD_1	Correct Detection & Exclusion		WD_1 & Exclusion		
Step 2 (New tests)	(No step 2)	No Detection (Nominal Operations)	Detection $FA_2 \vert CD_1$	Correct Detection (Alarm)	$WD_2 \vert WD_1$ (Alarm)	$MD_2 \vert WD_1$

6.5.1. P_{HMI} COMPUTATION FOR MULTIPLE ITERATIONS

As from Equation (6.1), the $P_{\text{HMI}|H_a}$ for an algorithm made up of a single iteration is computed:

$$\begin{gathered} P_{\text{HMI}|H_a} = P(PF \cap \text{No alert}|H_a) = \\ P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} \cap \underline{T} \in \Omega_T|H_a) \end{gathered} \tag{6.16}$$

where $\underline{\hat{x}}_0$ is the position estimator under the null hypothesis (for a single iteration algorithm), $\underline{T}$ stands for all the test statistics computed and Ω_T is the whole acceptance region for the tests. When a multiple step algorithm is foreseen, this equation becomes:

$$\begin{gathered} P_{\text{HMI}|H_a} = P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} \cap \underline{T}_1 \in \Omega_{T_1}|H_a) + \\ \sum_{s=2}^{N_{iter,max}} P\left(\underline{\hat{x}}_{s-1} - x \notin \Omega_{\text{AL}} \cap \underline{T}_s \in \Omega_{T_s} \cap \bigcap_{j<s} \underline{T}_j \notin \Omega_{T_j}|H_a\right) \end{gathered} \tag{6.17}$$

where $\underline{\hat{x}}_s$, $\underline{T}_s$ Ω_{T_s} are the position solution, the test statistics (all the tests run for the specific step s) and the tests acceptance regions at the s^{th} step of the procedure and $N_{iter,max}$ is the total number of iterations that can be possibly run in the procedure. In particular $\underline{\hat{x}}_s$ is the position estimator that is chosen after the tests $\underline{T}_s$ led to rejection of the null hypothesis and the model was accordingly adapted (at that stage).

The formula for the total $P_{\text{HMI}|H_a}$ in Equation (6.17) presents extra terms over the $P_{\text{HMI}|H_a}$ computed in case of single step (Equation (6.16)), thus it may seem that running multiple iterations causes an increase of the P_{HMI}. In fact, while the P_{HMI} increases, with multiple iterations the total probability of False Alarm consistently decreases. The total probability of False Alarm under the null hypothesis H_0, which can approximate the probability of False Alert for the whole algorithm, is computed as:

$$P_{FA_{tot}} = P(\underline{T}_1 \notin \Omega_{T_1}) \cdot \prod_{s=2}^{N_{iter,max}} P\left(\underline{T}_s \notin \Omega_{T_s} \middle| \bigcap_{j<s} \underline{T}_j \notin \Omega_{T_j}\right) \tag{6.18}$$

In case of a correct Detection and Exclusion, the subsequent iteration of the FDE procedure could lead anyway to a FA, e.g. the case of FA_2 in Table 6.2, a case which is not included in the formula in Equation (6.18), since we are fundamentally interested only in the P_{FA} under the fault-free hypothesis (as discussed in Sections 2.5 and 3.7). Therefore, if we want to keep constant the $P_{FA_{tot}}$, the acceptance regions Ω can be made smaller, and as final result the total P_{HMI} will eventually decrease. The $P_{FA_{tot}}$ as computed in Equation (6.18) can be upperbounded by:

$$P_{FA_{tot}} \leq (P_{FA_s})^{N_{iter,max}} \tag{6.19}$$

where P_{FA_s} is the FA probability (assumed constant) of each detection step. The equality holds in case successive rejections of the null hypothesis are independent events (but they are not). The smaller or equal relation holds because there is a negative correlation between the occurrences of FAs, i.e. the occurrence of a FA in a first algorithm iteration reduces the probability of occurrence of an FA in the following iteration. Therefore the P_{FA_s} of each iteration can be set at the value $P_{FA_s} = (P_{FA_{tot}})^{\frac{1}{N_{iter,max}}}$.

Let us go back now to Equation (6.17). Note that the event $\underline{T}_j \notin \Omega_{T_j}$ corresponds to either WD or CD:

$$\underline{T}_j \notin \Omega_{T_j} \equiv WD \text{ or } CD$$

The probability of occurrence of such events depends on the size of the bias affecting the measurements. It is possible to find a worst-case bias which maximizes the probability of

occurrence of these events (in particular of the WD), in a similar way as considered in Section 6.4.

For easier computation of Equation (6.17), the probability of the event $\underline{T}_j \notin \Omega_{T_j}$ can be approximated with the P_{WD} (κ), since in case of CD the P_{PF} can be considered negligible. The P_{WD} (κ) can be approximated with the upperbound in Equation (5.44). Therefore a simpler expression for Equation (6.17) can be:

$$\begin{gathered} P_{\mathrm{HMI}|H_a} = P^1_{PF_{H_a}} \beta^1_{H_a} + \\ \sum_{s=2}^{N_{iter,max}} P\left(PF_s | \bigcap_{j<s} WD_j, H_a\right) P\left(MD_s | \bigcap_{j\leq s} WD_j, H_a\right) P\left(\bigcap_{j<s} WD_j, H_a\right) \end{gathered} \tag{6.20}$$

where the subscript of each quantity (superscript in case of P_{PF} and β_{H_a}) stands for the corresponding iteration of the algorithm. As a result, the upperbound for the total P_{HMI} can be written as:

$$\begin{gathered} P_{\mathrm{HMI}} \leq \sum_{i=0}^{N_{H_a}} p_i \max_{\nabla_i} \left\{ P^1_{PF_i} \beta^1_i + \sum_{s=2}^{N_{iter,max}} P\left(PF_s | \bigcap_{j<s} WD_j, H_i\right) \cdot \right. \\ \left. P\left(MD_s | \bigcap_{j\leq s} WD_j, H_i\right) P\left(\bigcap_{j<s} WD_j, H_i\right) \right\} \end{gathered} \tag{6.21}$$

In Equation (6.20), whilst $P^1_{PF_{H_a}}$ can be computed with Equation (6.5) (for a single dimension) and $\beta^1_{H_a}$ can be easily approximated, all the other quantities conditioned on previous WDs are not easy to estimate since the conditional distribution of the remaining measurements after a WD is not anymore Gaussian. Because of this it may be necessary to upperbound the conditional P_{MD} with 1.

$$P_{\mathrm{HMI}|H_a} = P^1_{PF_{H_a}} \beta^1_{H_a} + \sum_{s=2}^{N_{iter,max}} P\left(PF_s | \bigcap_{j<s} WD_j, H_a\right) P\left(\bigcap_{j<s} WD_j, H_a\right) \tag{6.22}$$

Now it is possible to approximate $P\left(PF_s | \bigcap_{j<s} WD_j\right)$ with the corresponding unconditional probability. This approximation is reasonable, but is *not* conservative. The alternative would be to upperbound also $P\left(PF_s | \bigcap_{j<s} WD_j\right)$ with 1. As a result we can write the upperbound for the total P_{HMI} as:

$$\begin{gathered} P_{\mathrm{HMI}} \leq \sum_{i=0}^{N_{H_a}} p_i \max_{\nabla_i} \left\{ [\Phi(k - \nabla w_i) - \Phi(-\nabla w_i - k)] \cdot \left[\Phi\left(\frac{S_{3j} C_{y_i} \nabla_i - VAL}{\sigma_{\hat{x}_3}} \right) + \right. \right. \\ \left. \Phi\left(\frac{-S_{3j} C_{y_i} \nabla_i - VAL}{\sigma_{\hat{x}_3}} \right) \right] + \sum_{s=2}^{N_{iter,max}} \left\{ \prod_{j<s} \kappa^s_i \left[\Phi\left(\frac{S^s_{3j} C^s_{y_i} \nabla_i - VAL}{\sigma^s_{\hat{x}_3}} \right) + \Phi\left(\frac{-S^s_{3j} C^s_{y_i} \nabla_i - VAL}{\sigma^s_{\hat{x}_3}} \right) \right] \right\} \Bigg\} \end{gathered} \tag{6.23}$$

where the superscript s indicates the iteration to which each quantity refers and κ stands for the probability of WD that can upperbounded with Equation (5.44).

In the simulation of Chapter 7, the P_{HMI} is computed in the multiple iterations DIA algorithm employing Equation 6.21, via numerical approximation (Monte Carlo integration).

6.5.2. IMPACT OF MULTIPLE ITERATIONS ON INTEGRITY

As we have mentioned, the P_{HMI} relative to any geometry depends on the FDE procedure employed. In this chapter we have analyzed a single iteration of the DIA procedure, for which we have obtained an upperbound of the P_{HMI} in Equation (6.8), and a multiple iterations DIA procedure, for which an upperbound was given in Equations (6.21) and (6.23). Now the simple question arise: does allowing multiple iterations of the DIA decrease or increase the P_{HMI} for any geometry? Or similarly, is the upperbound in Equation (6.8) larger or smaller

than the one in Equation (6.21)? The first question is of more theoretical interest, whereas the second is more practical. Results from the simulations (discussed in Chapter 7) show that multiple iterations of the DIA procedure result in both larger P_{HMI}, and larger upperbounds of it, for most of the geometries analyzed.

This result may come as quite counter-intuitive. In fact, we would expect that the capability of excluding faults and consequently adapting the observation model should allow a substantial increase of the significance of the detection tests, with consequent increase of power and reduction of P_{MD}. In reality instead this effect, which corresponds to a shrinking of the term $P^1_{PF_{H_a}}\beta^1_{H_a}$ in Equation (6.20), is overruled by the addition of the new terms connected to WDs in the equation. There is to note that, even though the P_{HMI} increases when multiple iterations are allowed, so does the detection power of the procedure, and to a much larger extent. As a result the procedure guarantees a significantly larger continuity *in case of faults* (see also the simulation results in Chapter 7). In view of this consideration, we maintain a preference for the multiple iteration algorithm, as long as the increase of risk is contained and still within the navigation requirements.

The reason why WDs are particular detrimental to integrity is that in case of a WD, the successive iteration of the algorithm becomes much less effective than expected — it becomes extremely more difficult to find the actual faulty satellite once a WD has occurred. Furthermore the position estimation is strongly jeopardized in case of a WD, both because of the weakening of the geometry after exclusion, and because of the upraising of an extra bias conditioned on the WD. The mathematical root-cause of the problem, that makes extremely more difficult to monitor the risk after a WD, is that after the first iteration of the procedure we have to deal with a *conditional* distribution of the remaining measurements. Theoretical performance parameters that can be computed for the first iteration of the procedure do not constitute good approximations for the successive iterations — $P_{MD|WD}$ cannot be approximated by the theoretical unconditional P_{MD} in Equation (6.20), as well as $P_{PF|WD}$ should not be approximated by the theoretical unconditional P_{PF}. This is especially true in case of WD, but in principle holds also in case of CDs (though experience shows that the effects are generally small in this case).

Figure 6.3 can be helpful to understand why $P_{MD|WD}$ is sensibly larger than the theoretical unconditional P_{MD} in the second iteration of the DIA algorithm. The detection space ($R(P_A^{\perp})$) in the simple case of two correlated w-tests is shown. The dark green area in the centre is the region of acceptance of both w-tests, and, in case an anomaly is affecting the system, a projection of the measurement falling in that area would represent an MD. Light green and yellow areas correspond to single detection of a fault by w-test (during the full iterative procedure): light green areas are characterized by single w-test statistic exceeding the threshold in the first iteration of the DIA, whereas yellow areas correspond to both w-test statistic exceeding their threshold. The pink areas represent the cases for which a double detection would occur, employing the iterative DIA procedure. Let us consider a different procedure in which instead of excluding a satellite at a time we instead exclude directly at the first iteration all the satellites which w-test statistic exceeds the threshold: in this case all the areas in yellow in Figure (as well as the pink areas) would correspond to a double exclusion. This indicates that in an iterative procedure the correlation between the w-tests can make the detection after the first iteration much more difficult. Note furthermore that in case of no correlation between the w-tests, the yellow areas and pink areas in Figure would coincide (Figure 5.5).

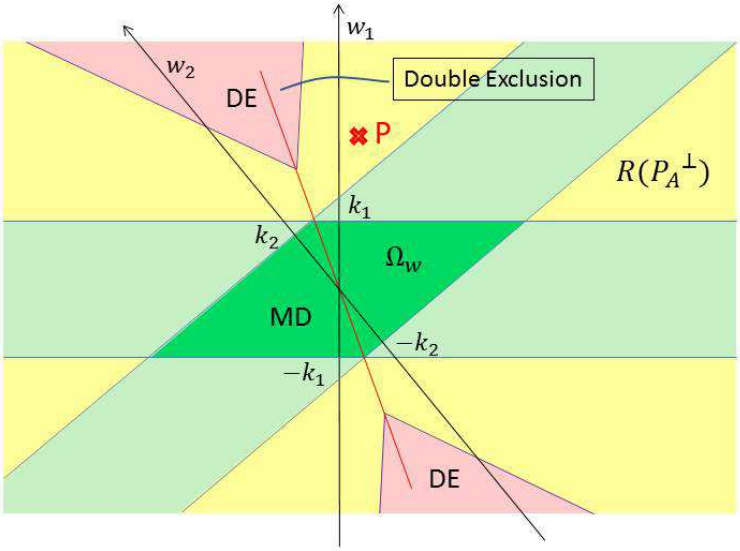


Figure 6.3: Detection/identification areas in the space $R(P_A^{\perp})$ for the DIA (iterative) procedure, in the simple case of two correlated w-tests. The dark green area correspond to no-detection (MD), the light green and yellow areas correspond to single satellite identification and exclusion and the pink areas correspond to double satellite identification and exclusion (through two iterations of the algorithm). As an example, computed w_1 and w_2 statistics corresponding to point P in the graph imply *only one* exclusion, because with the first DIA iteration satellite 1 is excluded ($w_1 > w_2 > k$ at the first iteration), and at the second iteration $w_2 < k$, due to the (large) correlation between w-tests, and no further satellite is excluded.

This issue leads to the conclusion that the standard exclusion mechanism of the DIA is not optimal, and, in general, any iterative exclusion method based on a forward selection approach neither. Even though not fully based on forward selection, the ARAIM exclusion method is equally affected by this problem, since exclusion of a single satellite is similarly prioritized with respect to simultaneous multiple exclusions (before excluding two or more satellites, the algorithm first attempts to exclude a single satellite).

6.6. ROLE OF THE OMT

In the standard DIA procedure the OMT constitutes the Detection step of the procedure, whereas the w-tests are used for Identification. This means that the OMT is run first and only if it leads to rejection, the w-tests are consequently run. In this Section we discuss upon the possible roles of the OMT, as a Detection test as in the DIA procedure or alternatively as an Identification test for special types of fault, at the same level of the w-tests.

6.6.1. THE OMT AS DETECTION STEP

The OMT is run beforehand, and if no detection is made there is no need to run the w-tests. The threshold for the OMT can be set based on the requirement on the total $P_{FA_{tot}}$ of the algorithm. The following questions arise: how to set subsequently the thresholds for the w-tests and how to compute the P_{HMI}? We consider here a few options:

- The Scheffé method: this method assures that any failure that would lead to a rejection of a w-test is also detected by the OMT. Therefore the computation of the P_{HMI} can be

made safely using Equation (6.2) (and therefore Equation (6.8)). The problem of this method is that it is very conservative, and would result in high loss of testing power and a overly-conservative computed P_{HMI}. Furthermore, with this method it is possible that the OMT rejects the null hypothesis whereas all the w-tests accept it — this would lead to Alert with no clear reason.

- The B-method: to apply the method a choice should be made on the β or MDB values to employ. For instance we could choose as MDB the smallest of the worst-case biases found for each w-test using the method in Section 6.4. There is no warranty though that this worst-case bias is the worst-case bias when using the OMT as a detection test, since these bias sizes were obtained considering the use of the w-tests as detection tests.

- P_{HMI} computation based on OMT detection: instead of computing the P_{HMI} considering the w-tests as detection tests as in Section 6.2, the computation could follow a similar procedure but assuming that the detection comes from the OMT. A worst-case bias can be found for each satellite, and a P_{HMI} as well (likely different from those with the w-tests), with formula:

$$P_{\text{HMI}} = \sum_{i=0}^{t} P(H_i) \cdot \beta_{OMT_i} \cdot P_{PF_i} \tag{6.24}$$

  as opposed to Equation (6.2), where β_{OMT_i} is the MD probability when using the OMT for Detection. In this way the w-tests are used only for Identification purpose, and are not used to determine the P_{HMI}. Therefore no threshold for the w-tests is needed. Identification is just based on finding the maximum value of the w-tests. In this way all OMT detections lead to an Adaptation of the model (exclusion), and there is no chance of an Alert being raised directly after Detection without an attempt of Adaptation.

  The reason why, in Multiple Comparisons, methods like Scheffé's are employed, is that maximum care is given to the control of the FA rate — it is very important to legitimate the identification of a specific anomaly, therefore it is important to assure an upper-bound to the FA rate of a single identification step. This is not the case for GNSS based navigation, because an FA in a single identification step means an exclusion of a satellite, and generally few satellites can be excluded from the model with no problem and without leading to any disservice.

In any case, using the OMT as a Detection step results in an increase of the computed P_{HMI}, given a fixed $P_{FA_{tot}}$ of the procedure. This is due to the lower power of the OMT compared to the w-tests in detecting the one-dimesional faults (see Figure 5.3), based on which the P_{HMI} is computed. Among the Multiple Comparisons techniques, the Scheffé method is employed only when it is required to test for un-planned (non-anticipated) comparisons, because otherwise the method has a sensibly lower power than other methods (Tukey, Bonferroni etc.).

The use of the OMT as a Detection test somewhat implies that the same weight is given to any type of anomaly, because it has the same power in detecting any type of anomaly — a single satellite fault, or an anomaly affecting two or more satellites etc. This seems appropriate when the different types of faults have the same a priori probability of occurrence, but not in case different type of faults are assigned different a priori probabilities.

6.6.2. THE OMT AS IDENTIFICATION TEST

The OMT could be employed in parallel with the w-tests to detect and identify types of faults different from those addressed by the w-tests. The w-tests can be employed to detect anomalies with one degree of freedom, whereas the OMT can be employed to detect the anomalies characterized by all other signatures.

The OMT has more power than the w-tests in detecting a multi-dimensional anomaly. The idea behind the use of the OMT only for detection of multi-dimensional anomalies is that the prior probabilities of occurrence of the (anticipated) one-dimensional anomalies are generally much higher than the prior probabilities of occurrence of the multi-dimensional ones. This difference in a-priori probabilities should be taken into account when setting the thresholds for the detection of different faults. If the a-priori probability of occurrence of a hypothesis is smaller, we can likely tolerate a larger β for it (assuming similar associate probabilities of PF). We elaborate on this in the following.

We can consider the anomalies addressed by the w-tests as a subset of the general set of all anomalies possibly affecting the system, detectable by the OMT (i.e. with the exception of the anomalies that can never be detected, with $C_y \in R(A)$). The type of anomaly affecting the system is fully characterized by the matrix C_y in Equation (3.21); we define $\mathscr{C}$ the set of all the C_y of all the possible anomalies that can affect the system, and $\mathscr{C}_{w_1}$ the set of anomalies detectable by data-snooping, i.e. characterized by canonical unit vectors as C_y. We assume that the iterative DIA procedure, in which up to $N_{steps,max}$ steps of DIA are run, is able to deal properly with anomalies constituted by combinations of up to $N_{steps,max}$ canonical unit vectors; therefore we can add these anomalies to set $\mathscr{C}_{w_1}$ to obtain the set $\mathscr{C}_w$, which contains all the anomalies detectable by w-tests iterative procedure. The set $\mathscr{C}_w$ is a subset of $\mathscr{C}$, $\mathscr{C}_w \subset \mathscr{C}$. The OMT has to be run to provide a protection against the anomalies that belong to $\mathscr{C} \setminus \mathscr{C}_w$, that are not detectable by the w-tests, or at least not with sufficient power.

If we assign a priori probabilities to the hypotheses of occurrence of faults, say $P(H_{w1})$ to single satellite faults and $P(H_{OMT})$ to multi-dimensional faults, and we recall the total P_{HMI} as sum of the contributions coming from each hypothesis as in Equation (6.2), we can write each contribution as:

$$\begin{aligned} P_{\text{HMI},H_{w1}} &= P(H_{w1})\beta_{H_{w1}}P_{PF_{H_{w1}}} \\ P_{\text{HMI},H_{OMT}} &= P(H_{OMT})\beta_{H_{OMT}}P_{PF_{H_{OMT}}} \end{aligned} \tag{6.25}$$

From these equations we can see that if $P(H_{w1}) >> P(H_{OMT})$ and $P_{PF_{H_{w1}}} \approx P_{PF_{H_{OMT}}}$ (same order of magnitude), then we can accept to have a lower power in detecting multi-dimensional faults ($\beta_{H_{OMT}} > \beta_{H_{w1}}$) to have similar contributions to the total P_{HMI} from each fault. This assumption is quite realistic, since non-anticipated faults are extremely unlikely.

6.6.3. NEED OF RUNNING THE OMT

We can consider to first choose a threshold k_{OMT} loose enough to *not* require an OMT test to be actually run, i.e. such that any OMT rejection would always coincide with a rejection of some w-tests. We explain the concept in the following. The m w-tests run in the DIA procedure describe by themselves a closed acceptance region in $R(A)^{\perp}$ ($m-n$ dimensional). The acceptance region of the OMT is a hyper-ellipsoid in $R(A)^{\perp}$. It is always possible to circumscribe the hyper-ellipsoid acceptance region of the OMT around the acceptance region of the w-tests: in this way the whole rejection region of the OMT is always inside the rejection

region of the w-tests. The acceptance region of the w-tests Ω_w is characterized by:

$$\Omega_w : \ \underline{\hat{e}}_i \leq \sigma_{\hat{e}_i} k_w \ , \ \ i = 1,2,\ldots,m$$

where k_w is the threshold for each w-test. We can write this acceptance region in terms of the misclosures vector $\underline{t} = B^T \underline{\hat{e}}$ (with B any $m \times (m-n)$ matrix such that $B^T A = 0$, see [98]):

$$\Omega_w \equiv \Omega_t : \ \underline{t}_j \leq k_{t_j} \ , \ \ j = 1,2,\ldots,m-n$$

with

$$k_t = B^T \begin{bmatrix} \sigma_{\hat{e}_1} k_w \\ \sigma_{\hat{e}_2} k_w \\ \vdots \\ \sigma_{\hat{e}_m} k_w \end{bmatrix}$$

Now since for the OMT statistic it holds $\underline{T}_{m-n} = \underline{t}^T Q_{tt}^{-1} \underline{t}$, we can find the threshold for the OMT that would circumscribe the OMT rejection region outside the w-tests rejection region as:

$$k_{OMT} = \max_{t \in \Omega} (t^T Q_{tt}^{-1} t) \tag{6.26}$$

With respect to the methods to set the OMT threshold considered in Section 5.2.2, i.e. B-method, S-method and LSD method, this approach would yield the largest threshold for the OMT — the OMT acceptance region would completely include the w-tests acceptance region, whereas in the other methods it intersects or is completely inside the w-tests acceptance region, as can be seen in Figure 5.2.

With the choice of such threshold for the OMT, or of a larger threshold, the OMT is in fact not needed, since any anomaly that it could detect would anyway be detected by at least one of the w-tests employed in the identification step. Therefore, if the threshold for the OMT is set on the basis of a criterion different than B-method, S-method or LSD, for instance with a method inspired by Section 6.6.2 or on the basis of an a priori allocation of the P_{FA} (as in ARAIM), it can be useful to compare it with the limiting value computed via Equation (6.26). If the threshold exceeds this value, there is actually no need to run the OMT (as long as w-tests are run for detection/identification).

6.7. SUMMARY AND CONCLUSIONS

In this chapter we discussed the application of the DIA algorithm to the RAIM problem. The following additions/modifications have been proposed with respect to the standard DIA algorithm:

- A method to estimate the P_{HMI} for a satellite geometry as a function of the thresholds choice for the tests has been proposed (Section 6.2), based on the worst-case bias approach (Equation (6.9)). This allow to set the thresholds for the tests in order to satisfy the requirement P_{HMI}^{req}, so that the integrity is guaranteed for the geometry under consideration. In standard DIA the choice of the α_i for the w-test is somewhat arbitrary and depends on the specific application: this means that this P_{HMI} estimation falls within the need of a *specialization* of the DIA to the application at hand. A method to determine the worst-case bias in case of multi-dimensional fault was furthermore proposed (Section 6.4.2).

- A method to take into account the possible multiple iterations of the DIA algorithm has been proposed (Section 6.5.1), in order to compute the total P_{HMI} and α_{tot} of the algorithm (Equations (6.19), (6.20), (6.22)). The standard performance evaluation for the DIA procedure does not analyze in detail the possibility of multiple iterations of the algorithm.

- Possible different roles (other than the Detector one) for the OMT have been envisioned (Section 6.6). In particular, beside using the OMT beforehand as detector, and renouncing to run the w-tests in case of OMT acceptance, it is proposed to run OMT and w-tests in parallel, with responsibility of both Detection and Identification. This proposal is linked to the fact that, for low values of α_i and large number of satellites in view, the significance of the OMT (employing the B-method) tends to be sensibly larger than the significance of the multiple w-tests run (as total multiple testing). The change of the role of the OMT would constitute a consistent modification for the DIA algorithm; regarding the use of the OMT for Detection instead, methods different from the B-method were also discussed.

In the simulation implementation (Chapter 7), two different versions of adapted DIA are run. In the first version only one iteration of DIA is considered, whereas in the second multiple iterations (a maximum of three) are taken into account. The computation of the P_{HMI} for each geometry is made by Monte Carlo (MC) integration. In both versions, the following choices have been made:

- The OMT is run after the w-tests (parallel detection/identification)

- The threshold for the OMT is set based on a previous allocation of the total $P_{FA'}^{req}$ among anticipated faults and not-anticipated, using the same approach as in ARAIM: only a small part of the total $P_{FA'}^{req}$ is allocated to general non-anticipated faults, and the threshold for the OMT is chosen to guarantee that significance of the test.

Table 6.3: Specific input to the adapted DIA algorithm.

Name	Description	Source
$P_{sat,i}$	Prior probability of fault in satellite i	ISM
$P_{FA_{\chi^2}}$	Continuity budget allocated to False Alerts for the OMT (χ^2) test	Requirement

6.8. ADAPTED DIA ALGORITHM (DIA-W)

In this adapted DIA algorithm the w-tests precede the OMT, and they are run for both detection and identification purposes (as iterative data-snooping). The OMT is run afterwards, to detect extra anomalies undetectable by w-tests. The baseline input for the algorithm, common to all the other RAIM algorithms, is given in Tables 7.1, 7.3 and 7.4 (the values reported are the ones employed in the numerical simulations in Chapter 7).

1. Input (extra than Tables 7.1, 7.3 and 7.4) in Table 6.3.
2. Check if the redundancy is sufficient:

$$\begin{cases} \text{Continue procedure if } N_{sat} > 3 + N_{const} \\ \text{Declare Alert if } N_{sat} \leq 3 + N_{const} \end{cases} \tag{6.27}$$

3. Determine the maximum number of satellites simultaneously failing, to be monitored, $N_{sat,max}$. This can be set a-priori or computed through Equation (4.46), as in ARAIM.
4. Determine geometry integrity, for both single iteration and multiple iteration cases:
   (a) Determine the FA rate allocated to each iteration of the algorithm (in case of multiple iterations):

$$P_{FA}^{iter} = (P_{FA'_{ver}}^{req})^{\frac{1}{N_{sat,max}}} \tag{6.28}$$

   (b) Compute w-tests thresholds for a single iteration:

$$k_w^{si} = \Phi^{-1}\left[\frac{(1 - P_{FA'_{ver}}^{req})^{\frac{1}{m}} + 1}{2}\right] \tag{6.29}$$

   with Φ being the left side of the cumulative distribution function of a zero mean unit Gaussian distribution. This formula is obtained inverting Equation (5.31), which is obtained from the Šidák approximation as described in Section 5.3.
   (c) Compute w-tests thresholds for multiple iterations:

$$k_w^{mi} = \Phi^{-1}\left[\frac{(1 - P_{FA_{iter}}^{req})^{\frac{1}{m}} + 1}{2}\right] \tag{6.30}$$

   (d) Compute $P_{\mathrm{HMI}_{ver}}^{si}$ for single iteration:

$$P_{\mathrm{HMI}}^{si} \leq \sum_{i=0}^{N_{sat}} P(H_i) \cdot \max_{\nabla_i} \left\{\left[\Phi(k_w^{si} - \nabla w_i) - \Phi(-\nabla w_i - k_w^{si})\right] \cdot \left[\Phi\left(\frac{S_{3j} C_{y_i} \nabla_i - VAL}{\sigma_{\hat{x}_3}}\right) + \Phi\left(\frac{-S_{3j} C_{y_i} \nabla_i - VAL}{\sigma_{\hat{x}_3}}\right)\right]\right\}$$

cf. Equation (6.8)

   with $P_A^{\perp} = I_{N_{sat}} - A(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$, $S = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$, $\sigma_{\hat{x}_3}$ the third diagonal element of $Q_{\hat{x}\hat{x}} = (A^T Q_{yy}^{-1} A)^{-1}$ and $\sigma_{\hat{e}_i}$ the i^{th} diagonal element of the residuals variance matrix $Q_{\hat{e}\hat{e}}$ determined with:

$$Q_{\hat{e}\hat{e}} = Q_{yy} - A(A^T Q_{yy}^{-1} A)^{-1} A^T = P_A^{\perp} Q_{yy} P_A^{\perp T}$$

   With this formula for each satellite we search for a *worst-case* bias, the bias that maximizes the product $P_{PF} P_{MD}$. For the computation of the P_{MD} the one-dimensional approximation is employed. The above formula was derived in Section 6.2. This maximization can be solved by iterative search.

(e) Compute $P^{mi}_{\mathrm{HMI}_{ver}}$ for multiple iterations:

$$P^{mi}_{\mathrm{HMI}_{ver}} \leq \sum_{i=0}^{N_{sat}} p_i \max_{\nabla_i} \left\{ \left[\Phi(k_w^{mi} - \nabla w_i) - \Phi(-\nabla w_i - k_w^{mi}) \right] \cdot \left[\Phi\left(\frac{S_{3j} C_{y_i} \nabla_i - VAL}{\sigma_{\hat{x}_3}} \right) + \Phi\left(\frac{-S_{3j} C_{y_i} \nabla_i - VAL}{\sigma_{\hat{x}_3}} \right) \right] + \sum_{s=2}^{N_{sat,max}+1} \left\{ \prod_{j<s} \kappa_i^s \left[\Phi\left(\frac{S^s_{3j} C^s_{y_i} \nabla_i - VAL}{\sigma^s_{\hat{x}_3}} \right) + \Phi\left(\frac{-S^s_{3j} C^s_{y_i} \nabla_i - VAL}{\sigma^s_{\hat{x}_3}} \right) \right] \right\} \right\}$$

cf. Equation (6.23)

with values for κ obtainable from Equation (5.44), or employ the more general Equation (6.21), with Monte Carlo integration (the latter method is employed for the simulations in Chapter 7). Also this maximization can be solved by iterative search.

(f) Check if the integrity is guaranteed:

$$\begin{cases} \text{Continue procedure allowing multiple iterations (adaptations) if} \\ P^{mi}_{\mathrm{HMI}_{ver}} \leq P^{req}_{HMI_{ver}} \\ \text{Continue procedure but do not allow multiple iterations (adaptations) if} \\ P^{mi}_{\mathrm{HMI}_{ver}} > P^{req}_{HMI_{ver}} \text{ but } P^{si}_{HMI_{ver}} \leq P^{req}_{HMI_{ver}} \\ \text{Declare Alert if} \\ P^{mi}_{\mathrm{HMI}_{ver}} > P^{req}_{HMI_{ver}} \text{ and } P^{si}_{HMI_{ver}} > P^{req}_{HMI_{ver}} \end{cases} \tag{6.31}$$

(g) Choose the w-test threshold k_w: if multiple iterations are allowed, $k_w = k_w^{mi}$, otherwise $k_w = k_w^{si}$.

5. Detection and Identification of single satellite faults:

(a) Compute w-tests statistics:

$$\underline{w}_i = \frac{\hat{\underline{e}}_i}{\sigma_{\hat{e}_i}} \quad \text{for } i = 1, 2, \ldots, N_{sat} \quad \text{cf. Equation (5.5)}$$

where $\hat{\underline{e}}_{0_i}$ is the i^{th} component of the residuals vector $\hat{\underline{e}}$ determined with:

$$\hat{\underline{e}} = (I_{N_{sat}} - A(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}) \underline{y} = P_A^{\perp} \underline{y} \tag{6.32}$$

Note that Q_{yy} is assumed diagonal.

(b) Detect single satellite faults:
if

$$\max_i w_i \leq k_w \quad \forall i : \ 1 \leq i \leq N_{sat} \quad \text{cf. Equation (5.21)}$$

No single satellite fault is identified, go to step 6. Else, if no multiple iterations of the DIA are allowed, declare Alert; otherwise, continue to next step.

(c) Check whether redundancy is sufficient to exclude satellites:

$$\begin{cases} \text{Continue procedure if } N_{sat} > 3 + N_{const} + 1 \\ \text{Declare Alert if } N_{sat} \leq 3 + N_{const} + 1 \end{cases} \tag{6.33}$$

(d) Identify faulty satellite:

$$i_{ex} = \arg\max_i w_i \quad \forall i: \ 1 \le i \le N_{sat} \quad \text{cf. Equation (5.57)}$$

Satellite i_{ex} has to to be excluded from the model. Go to step 5.

6. Adaptation:

The measurement corresponding to the faulty satellite i_{ex} is removed from the model. New geometry matrix, measurement variance matrix and observable vector, respectively A^{new}, Q_{yy}^{new} and $\underline{y}^{new}$, are re-constructed after exclusion: $\underline{y}^{new}$ does not include the measurement from the excluded satellite, A^{new} is obtained from A taking out the row corresponding to the excluded satellite and Q_{yy}^{new} instead is obtained from Q_{yy} removing both row and column corresponding to the excluded satellite. Also N_{sat} is correspondingly updated, $N_{sat}^{new} = N_{sat} - 1$. With this new input we move back to step 2 and the algorithm is repeated until finally either standard operation is reached with a position estimate, or an Alert is declared.

7. Check for other generic anomalies:

(a) Compute $\underline{T}_{m-n}$ (χ^2 distributed) test statistic:
The $\underline{T}_{m-n}$ statistic is computed as:

$$\underline{T}_{m-n} = \underline{y}^T (Q_{yy}^{-1} - Q_{yy}^{-1} A (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}) \underline{y} \tag{6.34}$$

(b) Compute $\underline{T}_{m-n}$ test threshold:

$$k_{\text{OMT}} = \chi^{2inv}_{N_{sat}-3-N_{const}} (1 - P^{req}_{FA'_{\chi^2}}) \tag{6.35}$$

where the operator $\chi^{2inv}_{N_{sat}-3-N_{const}}$ is the inverse of the CDF of a central χ^2 distribution with $N_{sat} - 3 - N_{const}$ degrees of freedom.

(c) Determine overall consistency of measurements:

$$\begin{cases} \text{Accept } H_0 \text{ if } T_{m-n} \le k_{\text{OMT}} \\ \text{Reject } H_0 \text{ if } T_{m-n} > k_{\text{OMT}} \end{cases} \tag{6.36}$$

If H_0 is rejected an Alert has to be declared. Otherwise standard operation is continued with solution estimate:

$$\hat{x} = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} y = Sy$$

.

7

SIMULATIONS AND RESULTS

In this Chapter the RAIM algorithms discussed so far, i.e. Standard RAIM, ARAIM and DIA, are tested and compared by means of numerical simulations. Different geometries are analyzed to assess the availability of each algorithm and observations from GPS and Galileo constellations are simulated in different scenarios, fault-free and faulty. The performance of the full procedures are evaluated, in terms of Missed Detections, Hazardous Misleading Informations, False Alert rates, compliance to the requirements, and quality of upperbounds and approximations are assessed.

7.1. GNSS MODEL

The GNSS model employed for the simulations was presented in Section 2.8. The parameters employed to build the stochastic model adopted for the simulations are given in Table 7.1. The GNSS functional and stochastic model is part of the input required from the RAIM algorithms, as described in Section 2.5, together with the navigation requirements (see next Section). A 24 satellite GPS satellite constellation (based on [65]) and a 27 satellite Galileo constellation (constructed as a 27/3/1 plus constellation as described for instance in [89] with modified altitude after 2004 as described in [43]) are assumed (the nominal constellations)[1].

7.2. NAVIGATION REQUIREMENTS

The simulations are based on the CAT-I approach navigation requirements, that are reported in Table 7.3. These correspond to the integrity requirements P_{HMI}^{req}, a_0 and Ω_{AL} introduced in Section 2.5, input to the RAIM algorithm ($a_0 \approx P_{FA'}$ as previously discussed). The simulations and their analysis focus on monitoring of the vertical position coordinate, for which the aircraft approach navigation requirements are often most difficult to fulfil. For this reason the total P_{HMI}^{req} and $P_{FA'}$ are split into vertical and horizontal components, and only the vertical share is used as input for the RAIM algorithms. Parameters that result from the simulations of geometry and observations, and serve as input to the RAIM algorithms, are listed in Table 7.4.

[1] Note anyway that the baseline Galileo constellation was recently reduced from 27 to 24 satellites.

Table 7.1: Simulation parameters. ξ stands for the elevation of the satellite, in degrees.

Parameter	Description	Value
σ^{GPS}_{URA}	GPS SV clock and orbit error used for integrity (URA)	0.75 m
σ^{Gal}_{URA}	Galileo SV clock and orbit error used for integrity (URA)	0.957 m
$\sigma_{tropo,i}$	Residual tropospheric error, for satellite i	$0.12\frac{1.001}{(0.002001+\sin^2\xi)^{1/2}}$ m
$\sigma^{GPS}_{MP,i}$	GPS smoothed code multipath	$0.13+0.53e^{-\xi/10}$ m
$\sigma_{noise,i}$	Smoothed code receiver noise	$0.15+0.43e^{-\xi/6.9}$ m
$\sigma^{GPS}_{user,i}$	GPS smoothed code total noise generated at user level	$\sqrt{\frac{f_{L1}^4+f_{L5}^4}{(f_{L1}^2-f_{L5}^2)^2}}\sqrt{\sigma^2_{MP}+\sigma^2_{noise}}$ m
$\sigma^{Gal}_{user,i}$	Galileo smoothed code total noise generated at user level	See Table 7.2
σ^{GPS}_{i}	GPS smoothed code total standard deviation used for integrity	$\sqrt{\sigma^{GPS^2}_{URA}+\sigma^2_{tropo}+\sigma^{GPS^2}_{user}}$ m
σ^{Gal}_{i}	Galileo smoothed code total standard deviation used for integrity	$\sqrt{\sigma^{Gal^2}_{URA}+\sigma^2_{tropo}+\sigma^{Gal^2}_{user}}$ m

Table 7.2: Galileo Elevation Dependent SIS user error.

Elevation	σ^{Gal}_{user}	Elevation	σ^{Gal}_{user}
5°	0.4529 m	50°	0.2359 m
10°	0.3553 m	55°	0.2339 m
15°	0.3063 m	60°	0.2302 m
20°	0.2638 m	65°	0.2295 m
25°	0.2593 m	70°	0.2278 m
30°	0.2555 m	75°	0.2297 m
35°	0.2504 m	80°	0.2310 m
40°	0.2438 m	85°	0.2274 m
45°	0.2396 m	90°	0.2277 m

7.3. ALGORITHMS TESTED

The algorithms tested are:

- Standard Slope RAIM — algorithm described in Section 4.1.4.
- ARAIM — algorithm described in Section 4.2.5.
- DIA procedure (maximum 3 iterations) with thresholds based on a single iteration (only for study purpose) — algorithm described in Section 5.5, with threshold k_{OMT} computed with the specific formula for the DIA-k_1 (Equation (5.51)). The algorithm is in fact referred to in the following as DIA-k_1.

Table 7.3: Navigation requirements for CAT-I approach.

Parameter	Description	Value
P^{req}_{HMI}	Total integrity budget	10^{-7}
$P_{\text{HMI}_{ver}}$	Integrity budget for the vertical component	9.8×10^{-8}
$P^{req}_{\text{HMI}_{hor}}$	Integrity budget for the horizontal component	2×10^{-9}
$P^{req}_{FA'}$	Continuity budget allocated to False Alerts (the total continuity budget is 8×10^{-6} per approach[1] [46])	4×10^{-6}
$P^{req}_{FA'_{ver}}$	Continuity budget allocated to False Alerts for the vertical mode	3.9×10^{-6}
$P^{req}_{FA'_{hor}}$	Continuity budget allocated to False Alerts for the horizontal mode	9×10^{-8}
VAL	Vertical Alert Limit	10 m
HAL	Horizontal Alert Limit	40 m

Table 7.4: Simulated parameters (geometry and measurements obtained during the simulation to be input to the RAIM algorithms).

Name	Description	Source
N_{const}	Number of GNSS constellations	Receiver
N_{sat}	Number of satellites in view	Receiver
A	$N_{sat} \times (3 + N_{const})$ satellites geometry matrix (one clock parameter for each constellation)	Receiver
y_i	Pseudorange of satellite i after dual frequency correction, tropospheric correction, satellite clock error correction and smoothing are performed	Receiver

- DIA procedure (with maximum 3 iterations) — algorithm described in Section 5.5.
- Adapted DIA procedure (with maximum 3 iterations), with OMT following the w-tests — algorithm described in Section 6.8. This algorithm is referred to in the following as DIA-w.

Each algorithm is tuned in such a way that the total False Alarm probability ($P_{FA_{tot}}$), as computed by the algorithm itself, is within the requirement defined in Table 7.3 on the False Alert probability ($P^{req}_{FA'}$) — as discussed in Section 3.7, the False Alert probability is not easily computed/estimated, therefore it is always approximated with the False Alarm probability.

We have distinguished three implementations of the DIA procedure, indicated with DIA-k_1, DIA and DIA-w. Both DIA-k_1 and DIA are characterized by the standard DIA procedure, but in DIA-k_1 the $P_{FA'}$ computation is made as if no multiple iterations were allowed, whereas in DIA it is based on the multiple iterations capability. In fact, likely for safety reasons, in ARAIM the $P_{FA'}$ is computed as if no adaptation can be successful; for comparison purposes, also the DIA-k_1 is implemented making the same approximation, i.e. the OMT threshols is set as if a Detection would result immediately in an Alert. For both DIA-k_1 and

DIA a maximum of three iterations are run. This means that a maximum of two satellites can be excluded, and after two exclusions a further Detection step is run again: then, if a Detection occurs also this last time, Alert is declared. The third DIA algorithm tested, DIA-w, is an adapted DIA algorithm that runs the OMT *after* the w-tests, as proposed in Section 6.6.

7.4. SIMULATIONS PROCEDURE

With reference to Figure 2.1, a RAIM algorithm first processes the satellite geometry, to determine the availability of the integrity for such geometry (this is the reliability analysis in the case of the DIA, Sections 5.1.6 and 6.3), and next checks the consistency of the measurements within the assumed model. Therefore it is considered opportune to split the simulations into two parts, geometry simulations and observation simulations:

1. Geometry simulations: once a location for the receiver is chosen, different satellite geometries are simulated, to compute the *predicted* performance of the different RAIM algorithms. These simulations yield as a result the P_{HMI} for each geometry and therefore the integrity availability (the estimated P_{HMI} can be either smaller or larger than the requirement). Actual observations are not simulated in this first step.

2. Observations simulations, to verify the *actual* performance of the algorithms. They are organized as follows:

   - Two reference satellite geometries are chosen — these geometries must be such that integrity is available for all the algorithms tested (a-priori, ahead of measurements).
   - For the geometries chosen, GPS and Galileo observations are simulated in different scenarios:
     - Fault free case. This scenario allows to verify the total FA (and FA') probability — most of the RAIM algorithms tested use in fact approximate ways to compute the FA rate, since closed form expressions do not exist for their testing procedure, therefore the quality of the approximations can be checked.
     - 1 pre-determined biased measurement (pre-determined single fault) — a single satellite is considered faulty, the bias size to apply is chosen such that its (un-detected) impact is not negligible, yet not too easy or obvious to detect (see Section 7.6, the introduced biases are generally in the range of 5 to 10 meters). Two different cases for each geometry (two different satellites assumed faulty, once at a time) are analyzed.
     - 1 random biased measurement (random single fault) — a single satellite is considered faulty, satellite and bias size are determined randomly
     - 2 pre-determined biased measurements (pre-determined double fault) — two satellites are considered faulty simultaneously, the bias sizes are chosen again to have a sensible impact to the positioning without being too easy to detect.

[1] The total continuity budget accounts also for disruptions coming from the ground segment (non-autonomous integrity monitoring). The requirement per approach is considered equivalent to a requirement per epoch since the duration of the approach is 150 seconds, but high correlation is present between successive observations and smoothing is performed over a 100 seconds window.

  - 2 random biased measurements (random double fault) — two satellites are considered faulty simultaneously, satellites and bias sizes are determined randomly.
- For each condition (state) of the system, apply the same RAIM algorithms, described in Section 7.3.

Once the simulations are run, the performance of the methods are compared: from the geometry simulations it is possible to compare the integrity risk that each method gives as result (given a fixed continuity requirement of the navigation system); from the actual observation simulations it is possible to verify the performance predicted in the previous analysis, to compare the power of each method in detecting faults, both hazardous and not hazardous, and of course to finally check whether each method satisfies the integrity requirements.

7.4.1. SIMULATIONS OUTPUT

The output of a RAIM algorithm were discussed in Section 2.5. In particular, as results from the geometry simulations we can compute the P_{HMI} (for each algorithm), given the requirement on $P_{FA'}$ being satisfied, whereas from the observation simulations we can count (for each algorithm):

- *Fault-free case*:
  1. N. of Alerts — from which we compute the actual $P_{FA_{tot}}$
  2. N. of False Alerts (for completeness) — Alerts given when there is no Positioning Failure (a PF might occur accidentally even when there are no faults), from which we compute the actual $P_{FA'}$
  3. N. of Positioning Failures — positioning errors larger than the AL, from which we compute the actual P_{PF}
  4. N. of HMI events — from which we compute the actual P_{HMI}
  5. N. of (wrong) Detections — this differs from the number of False Alerts because detections can be followed by successful adaptation and *not* result in an Alert
- *Faulty case*:
  1. N. of Alerts — from which we can possibly determine the cases in which the system was available and no HMI occurred (N. of Epochs – Alerts – HMIs)
  2. N. of False Alerts — Alerts given when there is no Positioning Failure
  3. N. of Positioning Failures — positioning errors larger than the AL, from which we compute the actual P_{PF}
  4. N. of HMI events — from which we compute the actual P_{HMI}
  5. N. of MD — at least one faulty satellite remaining undetected, and no healthy satellite detected as faulty, from which we compute the actual P_{MD}
  6. N. of WD (Wrong Detections) — at least one healty satellite detected as faulty and at least one faulty satellite remaining undetected, from which we compute the actual P_{WD}

7. N. of OD (Over Detections) — detections and identification of a larger number of faulty satellites than the actual one (but the faulty satellites are all correctly identified), from which we compute the actual P_{OD}
8. N. of CD (Correct Detections) — correct detections and identifications, from which we compute the actual P_{CD}

Note that the first four performance parameters of the faulty case list are actually integrity related parameters, whereas the others are more related to the general hypothesis testing problem. For the last four parameters it holds furthermore:

$$P_{MD} + P_{WD} + P_{OD} + P_{CD} = 1 \tag{7.1}$$

except for the Standard RAIM, for which WD, OD and CD are not really defined (because no exclusion is performed).

7.4.2. SIGNIFICANCE OF THE RESULTS

As these results are in fact output of a Monte Carlo simulation, their accuracy can be computed employing the formula [101]:

$$\sigma_{N_p} = \sqrt{Np(1-p)} \tag{7.2}$$

where p is the actual probability of an event (for instance the actual P_{HMI}), N the number of samples simulated, N_p the number of occurrences of that event and σ_{N_p} the standard deviation of N_p. If we estimate a probability p from the empirical rate of occurrence of the event:

$$\hat{p} = \frac{N_p}{N} \tag{7.3}$$

the standard deviation of the corresponding estimator can be computed as:

$$\sigma_{\hat{p}} = \frac{\sqrt{N\hat{p}(1-\hat{p})}}{N} \tag{7.4}$$

For example, if we simulate $N = 10^6$ epochs and we count $N_{MD} = 2000$ MDs, we can estimate the P_{MD} with Equation (7.3):

$$\hat{P}_{MD} = 2000/10^6 = 2 \cdot 10^{-3}$$

and the standard deviation of this estimator is (from Equation (7.4)):

$$\sigma_{\hat{P}_{MD}} = \frac{\sqrt{10^6 \cdot 2 \cdot 10^{-3} \cdot (1 - 2 \cdot 10^{-3})}}{10^6} \cong 4.47 \cdot 10^{-5}$$

7.5. GEOMETRY SIMULATION

First of all the Standard RAIM, the ARAIM and the DIA algorithms have been tested simulating different geometries (but no actual measurements) to determine the availability that they can assure. Given a fixed receiver position, at our GNSS observatory in Delft, the satellite geometries over a 24 hour time stretch (the GPS constellation repeat period) have been considered to compute the geometry integrity availability (the P_{HMI} each geometry can guarantee). We believe that Delft location, at a mid-latitude, is representative for performance in

Western Europe and US. The graph in Figure 7.2 shows the P_{HMI} with respect to the vertical dimension as computed by the three algorithms, for a full day. In Figure 7.1 it is shown the skyplot of the period considered. Two computations are made for the DIA algorithm, a first one based on a single iteration of the DIA, assuming thereofore that no exclusion is attempted (this algorithm is not implemented in the next observation simulations), and a second one based on 3 iterations DIA, for which up to two exclusions are allowed.

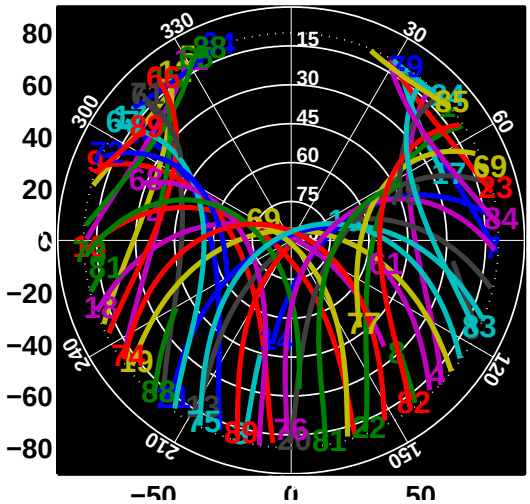


Figure 7.1: Skyplot at TU Delft, during a full day.

7.6. OBSERVATIONS SIMULATION

The geometries shown in Figure 7.3 and Figure 7.4, that we will refer to as geometry A and geometry B respectively, have been chosen to simulate 10^6 observations for the fault-free case, for the cases of pre-determined single and double satellite fault, and for random single and double satellite fault. 10^6 epochs are considered a sufficient amount to determine the performance of the algorithms in a reliable way because probabilities of events in the order of 10^{-4} or 10^{-5} can be estimated with reasonable accuracy. In fact, if we apply Equation (7.4) to estimated probabilities $\hat{p}$ of 10^{-4} and 10^{-5}, we obtain respectively $\sigma_{\hat{p}} \cong 10^{-5}$ and $\sigma_{\hat{p}} \cong 3 \times 10^{-5}$, which make the results reasonably significant. To check whether the requirement on integrity is satisfied, probabilities of much lower order of magnitude have to be estimated (at least of the order of 10^{-7} or 10^{-8}), but if we assume that significant contributions to the P_{HMI} occur in practice only under the hypothesis that an anomaly is affecting the system (i.e. the integrity risk under nominal operations is negligible), and that anomalies occur with a known a-priori probability of the order 10^{-3}-10^{-4}, we can conclude that being able to estimate in those cases probabilities of the order 10^{-4}-10^{-5}, becomes sufficient to estimate correctly the P_{HMI}.

The algorithms are compared setting the same $P_{FA'}$ for each of them. The

$$P_{FA'} \leq P_{FA'}^{req}$$

is computed for each algorithm following its own method. Furthermore each algorithm computed a P_{HMI} (an upperbound of it). The simulation shows whether these upperbounds

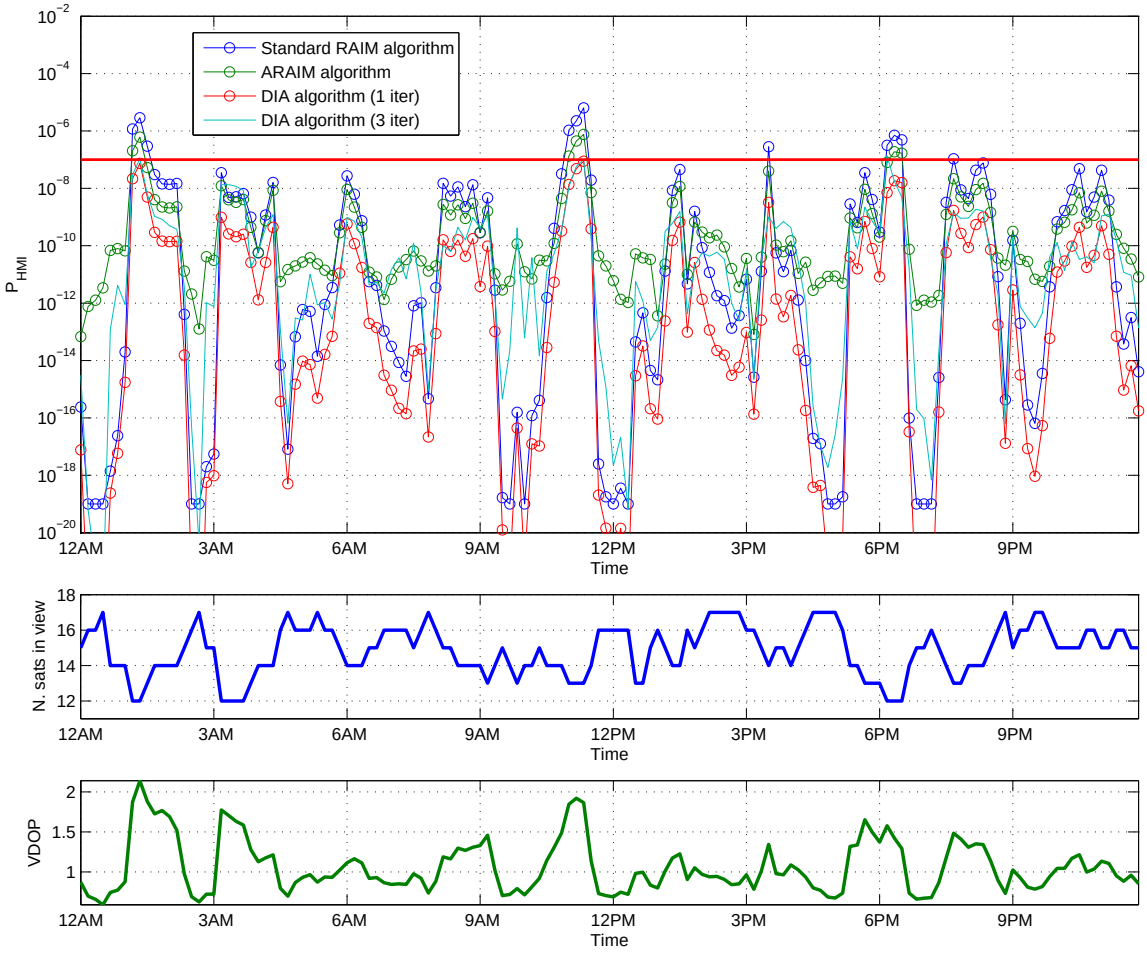


Figure 7.2: Probability of Hazardous Misleading Information (P_{HMI}) relative to the height estimation computed through Standard RAIM (Equation (4.31)), ARAIM (Equation (4.69)) and DIA algorithms (Equations (6.8) and (6.20)) during a full day. At the bottom, the number of satellites in view and the VDOP is reported for each geometry. We can notice that the P_{HMI} follows a similar behaviour when computed by any of the algorithms, and, as expected, is highly positively correlated with the VDOP, and negatively correlated with the number of satellites in view. We can notice furthermore that the ARAIM algorithm distinguishes itself from the others in almost all the geometries in which the P_{HMI} drops under about 10^{-10}, i.e. the strongest geometries. This distinct behaviour is mainly due to the fact that ARAIM computation explicitly takes into account multiple satellite faults and constellation faults: the contribution of these faults is in fact generally negligible for P_{HMI}s larger than 10^{-9} but it becomes the main contribution in strong geometries, with larger number of satellites in view.

are correct. The upperbounds for the P_{HMI} computed by each algorithm are the following:

$$\text{Geometry A}\begin{cases}\text{Standard RAIM:} & P_{\text{HMI}} \leq 1.1\times10^{-8}\\ \text{ARAIM:} & P_{\text{HMI}} \leq 2.7\times10^{-9}\\ \text{DIA (3 iterations):} & P_{\text{HMI}} \leq 4.4\times10^{-10}\end{cases}$$

$$\text{Geometry B}\begin{cases}\text{Standard RAIM:} & P_{\text{HMI}} \leq 2.5\times10^{-7}\\ \text{ARAIM:} & P_{\text{HMI}} \leq 8\times10^{-8}\\ \text{DIA (3 iterations):} & P_{\text{HMI}} \leq 1.7\times10^{-7}\end{cases}$$

We can see that the above upperbounds are (generally) close to the requirement $P^{req}_{\text{HMI},ver} = 9.8 \times 10^{-8}$. This is in fact the reason why these particular geometries have been chosen for the simulations: the relative weakness of the geometries can allow us to witness a number of HMIs sufficient to guarantee the significance of the results. Note that for geometry B the Standard RAIM and DIA upperbounds exceed the original requirement $P^{req}_{\text{HMI},ver} = 9.8 \times 10^{-8}$, and unavailability would be declared before taking the measurements; therefore the $P^{req}_{\text{HMI},ver}$ is set in this case at 10^{-6}. This geometry was chosen because it is one of the few in which the estimated P_{HMI} was larger for the DIA than for the ARAIM and because it is a reasonably weak geometry, prone to show a significant number of HMIs.

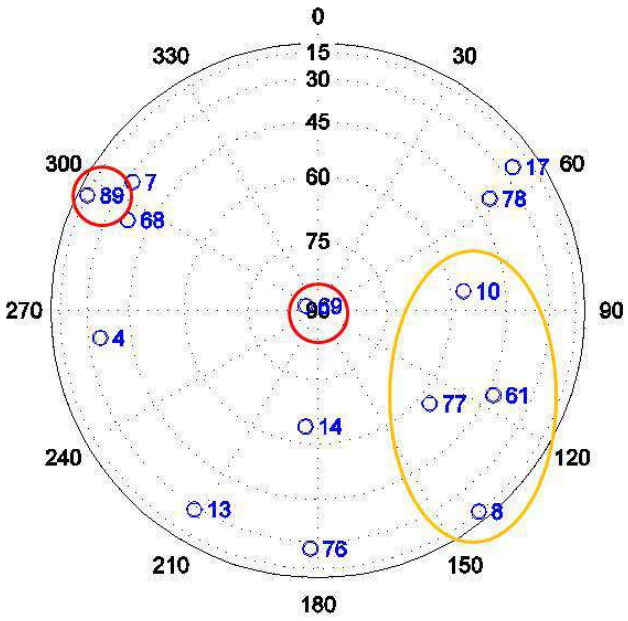


Figure 7.3: Skyplot of geometry A employed for the observations simulation. PRN numbers from 1 to 24 stand for the GPS satellites, from 61 to 90 for the Galileo satellites. In the circles are the satellites assumed faulty during the simulations.

7.6.1. FAULT-FREE

Tables 7.5 and 7.6 show the results of 10^6 epochs of simulations for the fault-free case, for geometries A and B respectively. Total number of Alerts, False Alerts, HMI events, Missed Detections, Wrong Detections and Over Detections are reported for each of the algorithm implemented. For this case (fault-free) it has been chosen:

$$P^{req}_{FA'} = 10^{-4}$$

instead of $P^{req}_{FA'} = 3.9 \times 10^{-6}$ that was given in Table 7.3 and is employed in the next simulations, because in this way we can be sure to witness a sufficient number of False Alerts (and

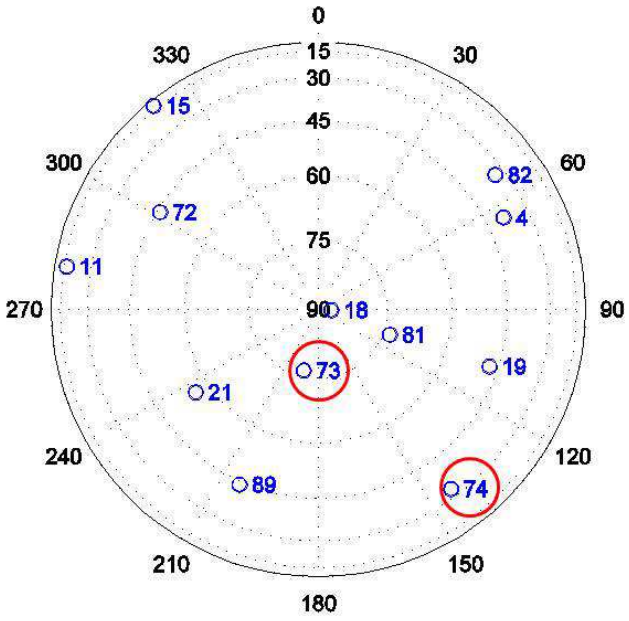


Figure 7.4: Skyplot of geometry B employed for the observations simulation. PRN numbers from 1 to 24 stand for the GPS satellites, from 61 to 90 for the Galileo satellites. In the circles are the satellites assumed faulty during the simulations.

check if the $P_{FA'}$ is correctly approximated by the algorithms). Analysis and conclusions are presented in Section 7.7.

Table 7.5: Results of the 10^6 epochs of simulation for the fault-free case, geometry A. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	WDs
Standard RAIM	1.04×10^{-4}	1.04×10^{-4}	0	0	1.04×10^{-4}
ARAIM	5.5×10^{-5}	5.5×10^{-5}	0	0	5.5×10^{-5}
DIA-k_1	0	0	0	0	1.04×10^{-4}
DIA	8.3×10^{-5}	8.3×10^{-5}	0	0	4.6×10^{-2}
DIA-w	1.2×10^{-5}	1.2×10^{-5}	0	0	4.4×10^{-2}

7.6.2. SINGLE FAULT

For each geometry, two pre-determined single satellite fault cases have been simulated.

Geometry A. Table 7.7 shows the results of 10^6 epochs of simulation for the case of satellite PRN69 failing with a bias size of 7m in geometry A. Table 7.8 instead shows the case of satellite PRN89 failing with a bias size of 9m. Total number of Alerts, False Alerts, HMI

Table 7.6: Results of the 10^6 epochs of simulation for the fault-free case, geometry B. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	WDs
Standard RAIM	1.23×10^{-4}	1.23×10^{-4}	0	0	1.23×10^{-4}
ARAIM	6.5×10^{-5}	6.5×10^{-5}	0	0	6.5×10^{-5}
DIA-k_1	0	0	0	0	1.23×10^{-4}
DIA	2.3×10^{-5}	2.3×10^{-5}	0	0	4.6×10^{-2}
DIA-w	1×10^{-6}	1×10^{-6}	0	0	4.2×10^{-2}

events, Missed Detections, Wrong Detections and Over Detections are reported for each of the algorithms implemented.

Table 7.7: Results of the 10^6 epochs of simulation for the case of single satellite fault, in geometry A. Measurements from satellite PRN69 have been biased by 7 meters. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	MDs	WDs	ODs	CDs
Standard RAIM	0.8658	0.8657	4.0×10^{-5}	4×10^{-6}	0.1342	-	-	-
ARAIM	0.9426	0.9426	8×10^{-6}	3×10^{-6}	5.74×10^{-2}	5.8×10^{-5}	0	0.9426
DIA-k_1	0	0	9×10^{-6}	9×10^{-6}	0.1342	7.3×10^{-5}	5×10^{-6}	0.8657
DIA	3.47×10^{-4}	3.47×10^{-4}	5×10^{-6}	5×10^{-6}	7.91×10^{-4}	1.20×10^{-4}	1.58×10^{-2}	0.9833
DIA-w	8.5×10^{-5}	8.5×10^{-5}	5×10^{-6}	5×10^{-6}	1.06×10^{-4}	9.9×10^{-5}	1.49×10^{-2}	0.9849

Geometry B. Table 7.9 shows the results of 10^6 epochs of simulation for the case of satellite PRN74 failing with bias size of 8m in geometry B. Table 7.10 instead shows the case of satellite PRN73 failing with bias size of 5m. Total number of Alerts, False Alerts, HMI events, Missed Detections, Wrong Detections and Over Detections are reported for each of the algorithms implemented.

We report here the standard DIA performance parameters for geometry B. For the case of satellite PRN74 failing with a bias size of 8m ($\nabla w = 6.3406$), the DIA-k_1 (employing the B-method) has the following theoretical performance parameters:

$$\beta_{\text{OMT}} = 0.2398$$
$$\beta_w = 0.2343$$
$$\alpha_w = 9.8 \times 10^{-9}$$

where α_w and β_w are the significance and the MD probability of each single w-test. The DIA

Table 7.8: Results of the 10^6 epochs of simulation for the case of single satellite fault, in geometry A. Measurements from satellite PRN89 have been biased by 9 meters. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	MDs	WDs	ODs	CDs
Standard RAIM	0.9264	0.9264	0	0	7.36×10^{-2}	-	-	-
ARAIM	0.9735	0.9735	1×10^{-6}	0	2.65×10^{-2}	2.21×10^{-4}	0	0.9733
DIA-k_1	0	0	1×10^{-6}	1×10^{-6}	7.36×10^{-2}	2.05×10^{-4}	5×10^{-6}	0.9262
DIA	3.20×10^{-4}	3.20×10^{-4}	0	0	2.16×10^{-4}	2.63×10^{-4}	1.58×10^{-2}	0.9838
DIA-w	9.7×10^{-5}	9.7×10^{-5}	0	0	2.5×10^{-5}	2.52×10^{-4}	1.51×10^{-2}	0.9846

has instead the following performance parameters:

$$\beta_{\text{OMT}} = 0.0028$$
$$\beta_w = 0.0018$$
$$\alpha_w = 2.9\times 10^{-4}$$

7

We can see that the simulation results, with an MD rate of about 0.24 for the DIA-k_1 and 2.8×10^{-3} for the DIA (in Table 7.9, MDs column), are in line with these theoretical values. The DIA-k_1 and the standard RAIM have the same MD rate, since detection is made through the OMT (with threshold based on a single iteration).

Figure 7.5 provides a representation of the DIA integrity monitoring, similar to the SLOPE representation in Figure 4.9 (employed by the Standard RAIM), for the specific geometry and fault conditions described above. In the top graph the vertical positioning error (on the vertical axis) is plotted against the w-test for PRN74 (on the horizontal axis), the test statistic most influenced by the anomaly. In the middle graph instead the vertical positioning error is plotted against the w-test for PRN89, which is the w-test most correlated with the first one. The bottom graph shows instead the projection of the (simulated) measurements onto the plane $R(P_A^{\perp}C_y)$, with $C_y = [C_{y_1}\ C_{y_2}]$ and C_{y_1}, C_{y_2} the anomaly signature vectors for the faults in satellites PRN74 and PRN89. In this plane (which lies in the detection space $R(A)^{\perp}$) the areas corresponding to MD, WD and CD can be distinguished (WD and CD areas are separated by the purple line in the middle). 10^4 observations (red dots) are reported in the graphs. It can be seen that, due to the high correlation between the two w-test statistics, the probability of WD is very high. The WD event, as previously mentioned, is almost as dangerous as an MD event, because the chances of correctly detecting the anomaly *after* a WD are quite scarce (the WD next 'masks' the actual anomaly).

To interpret the results in Table 7.9 (and generally of all Tables 7.7 to 7.10), we have to remember that the occurrence of an anomaly in one of the satellites has a probability of only $p_i = 10^{-4}$ (occurrence of an alternative hypothesis) therefore all the probability results should be multiplied by that factor (as in the simulation *all* samples were affected by the

Table 7.9: Results of the 10^6 epochs of simulation for the case of single satellite fault, in geometry B. Measurements from satellite PRN74 have been biased by 8 meters. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	MDs	WDs	ODs	CDs
Standard RAIM	0.7597	0.7594	4.44×10^{-4}	1.19×10^{-4}	0.2403	-	-	-
ARAIM	0.8494	0.8489	5.43×10^{-4}	7.9×10^{-5}	0.1506	1.06×10^{-2}	2×10^{-6}	0.8388
DIA-k_1	0	0	5.49×10^{-4}	5.49×10^{-4}	0.2403	9.7×10^{-3}	1×10^{-6}	0.7500
DIA	1.41×10^{-4}	1.41×10^{-4}	5.98×10^{-4}	5.98×10^{-4}	2.8×10^{-3}	1.61×10^{-2}	1.36×10^{-2}	0.9674
DIA-w	4.5×10^{-5}	4.5×10^{-5}	6.00×10^{-4}	6.00×10^{-4}	8.55×10^{-4}	1.62×10^{-2}	1.24×10^{-2}	0.9705

anomaly). For instance, the $P_{\mathrm{HMI}|H_a}$ reported in Table 7.9 for all the DIA algorithms, of about 6×10^{-4}, actually means a prior probability (P_{HMI}) of 6×10^{-8}. Since the type of anomaly chosen is practically the most dangerous among single satellite anomalies, this result appears in line to satisfy the original prediction by the algorithm (the DIA with 3 iterations gave an upperbound of 1.7×10^{-7} to the P_{HMI}, in Section 7.6).

To determine the significance of the results we apply Equation (7.4), with $N = 10^6$. For the estimated $\hat{P}_{\mathrm{HMI}|H_a} = \hat{p} = 6\times 10^{-4}$, for instance, we have $\sigma_{\hat{p}} \cong 2.45\times 10^{-5}$. This means that the estimate is quite accurate, and the actual $P_{\mathrm{HMI}|H_a}$ is comprised between 5×10^{-4} and 7×10^{-4} with probability larger than 99%.

For the case of satellite PRN73 failing with a bias size of 5m ($\nabla w = 5.4476$, Table 7.10), the DIA-k_1 has the following performance parameters:

$$\beta_{\mathrm{OMT}} = 0.5589$$
$$\beta_w = 0.5668$$
$$\alpha_w = 9.8\times 10^{-9}$$

The DIA has instead the following performance parameters:

$$\beta_{\mathrm{OMT}} = 0.0271$$
$$\beta_w = 0.0221$$
$$\alpha_w = 2.9\times 10^{-4}$$

Also in this case the simulation results are in line with the theoretical values.

Finally random single satellite faults have been simulated, for both geometries A and B. The results are shown in Tables 7.11 and 7.12. 10^6 epochs have been simulated, each of which presenting a bias of random size (uniformly random between 0 and 15 meters) in the observation from a single satellite chosen randomly as well among the full set. Note that the prior probability of occurrence of *any* of the single satellite faults is assumed to be $p_i \cong 1.4\times 10^{-3}$ in geometry A and $p_i \cong 1.2\times 10^{-3}$ in geometry B (because there are respectively respectively $m = 14$ and $m = 12$ satellites in view, and satellite faults are assumed uncorrelated).

Table 7.10: Results of the 10^6 epochs of simulation for the case of single satellite fault, in geometry B. Measurements from satellite PRN73 have been biased by 5 meters. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	MDs	WDs	ODs	CDs
Standard RAIM	0.4419	0.4419	0	0	0.5581	-	-	-
ARAIM	0.5396	0.5396	0	0	0.4604	3.38×10^{-4}	1×10^{-6}	0.5393
DIA-k_1	0	0	0	0	0.5581	5.04×10^{-4}	5×10^{-6}	0.4414
DIA	2.03×10^{-4}	2.03×10^{-4}	4×10^{-6}	4×10^{-6}	2.71×10^{-2}	3.2×10^{-3}	1.52×10^{-2}	0.9545
DIA-w	5.7×10^{-5}	5.7×10^{-5}	4×10^{-6}	4×10^{-6}	1.22×10^{-2}	2.8×10^{-3}	1.34×10^{-2}	0.9715

7.6.3. DOUBLE FAULT

A case of pre-determined double satellite fault has been simulated in both geometries A and B. Table 7.13 shows the results of 10^6 epochs of simulation for the case of satellites PRN69 and PRN89 failing in geometry A with bias size of respectively 6m and 8m. Table 7.14 shows the results of 10^6 epochs simulations for the case of satellites PRN73 and PRN74 failing in geometry B with bias size of respectively 4m and 7m. Total number of Alerts, False Alerts, HMI events, Missed Detections, Wrong Detections and Over Detections are reported for each of the algorithm implemented. The prior probability of the double satellite fault is assumed to be $p_i = 10^{-8}$.

Finally random double satellite faults have been simulated. The results are shown in Tables 7.15 and 7.16 for geometry A and B respectively. 10^6 epochs have been simulated, each of which presenting biases of random size (uniformly random between 0 and 15 meters) in the observations from a couple of satellites chosen randomly among the full set. The prior probability of *any* double satellite fault is assumed to be $p_i \cong 9.1\times 10^{-7}$ in geometry A and $p_i \cong 6.6\times 10^{-7}$ in geometry B (because 91 and 66 are the number of combinations of two satellites out of the $m = 14$ and $m = 12$ in view respectively). For a discussion of the results we refer to Section 7.7.3.

7.6.4. GENERAL ANOMALY

Beside single and double satellite faults, a more generic anomaly has been simulated and the results are shown in this section. The case of a bias affecting all the satellites of azimuth between 70° and 150° in geometry A, i.e. PRN8, PRN10, PRN61, and PRN77 (see Figure 7.3), has been simulated. This type of anomaly could exemplify the occurrence of a ionospheric front in that specific sky region, in a single frequency observations context. The same bias of $5m$ has been introduced in the four measurements at each simulated epoch. Table 7.17 shows the results of 10^5 epochs simulations. Total number of Alerts, False Alerts, HMI events and Missed Detections are reported for each of the algorithm implemented. In this case an MD means that no anomaly at all is detected and no satellites are excluded. If we would use

Table 7.11: Results of the 10^6 epochs of simulation for the case of random single satellite fault, in geometry A. The simulated bias size is a random value between 0 and 15 meters. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	MDs	WDs	ODs	CDs
Standard RAIM	0.5588	0.5577	1.11×10^{-2}	0	0.4312	-	-	-
ARAIM	0.5983	0.5983	0	0	0.4017	8.6×10^{-5}	0	0.5982
DIA-k_1	0	0	0	0	0.4312	1.19×10^{-4}	0	0.5687
DIA	2.50×10^{-4}	2.50×10^{-4}	0	0	0.2575	1.00×10^{-2}	1.16×10^{-2}	0.7208
DIA-w	5.7×10^{-5}	5.7×10^{-5}	0	0	0.2371	7.80×10^{-3}	1.10×10^{-2}	0.7441

the same definition of MD as before, the MD rate would simply be the complement to one of the Alert rate, since no algorithm can correctly detect a 4 satellites fault.

We can observe from the results that, as expected, the algorithms that employ the OMT for detection perform better (are more likely to detect an anomaly) than those that employ at first stage more specialized tests. In particular, among the algorithms that set their test thresholds assuming a single iteration of the FDE mechanism, i.e. Standard RAIM (which does not do exclusion though), ARAIM and DIA-k_1, we can see that Standard RAIM and DIA-k_1 perform better than ARAIM (show fewer MDs), as they employ the OMT for detection. Between the algorithms that assume multiple iterations instead, we observe a better performance of the DIA with respect to the DIA-w, with fewer number of MDs, again thanks to the use of the OMT for detection instead of the w-tests. We note furthermore that these last two algorithms tend to (wrongly) adapt the model more frequently than the others, giving rise to more numerous PFs.

7.7. ANALYSIS AND CONCLUSIONS

Going through the simulation results reported in Figure 7.2 and in Tables 7.5 to 7.16, we can identify recurring patterns among the different scenarios. In the following we discuss first the geometry simulations results (reliability analysis) and then the actual observation simulations results, considering separately the two main aspects of the RAIM algorithms, the integrity monitoring and the fault detection and exclusion (hypothesis testing).

7.7.1. GEOMETRY SIMULATIONS RESULTS

From Figure 7.2 we can notice that the P_{HMI} follows a similar behaviour when computed by any of the algorithms, and, as expected, is highly positively correlated with the VDOP, and negatively correlated with the number of satellites in view. We can notice furthermore that the ARAIM algorithm distinguishes itself from the others in almost all the geometries in which the P_{HMI} drops under about 10^{-10}, i.e. the strongest geometries. This distinct

Table 7.12: Results of the 10^6 epochs of simulation for the case of random single satellite fault, in geometry B. The simulated bias size is a random value between 0 and 15 meters. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	MDs	WDs	ODs	CDs
Standard RAIM	0.5568	0.5451	1.18×10^{-2}	1×10^{-6}	0.4432	-	-	-
ARAIM	0.5749	0.5749	4.9×10^{-5}	0	0.4251	6.77×10^{-4}	7×10^{-6}	0.5742
DIA-k_1	0	0	5.0×10^{-5}	5.0×10^{-5}	0.4432	6.82×10^{-4}	2×10^{-6}	0.5561
DIA	1.11×10^{-4}	1.11×10^{-4}	5.1×10^{-5}	5.1×10^{-5}	0.2620	1.47×10^{-2}	1.12×10^{-2}	0.7122
DIA-w	3.5×10^{-5}	3.5×10^{-5}	5.0×10^{-5}	5.0×10^{-5}	0.2476	1.23×10^{-2}	1.0×10^{-2}	0.7301

behaviour is mainly due to the fact that ARAIM computation explicitly takes into account multiple satellite faults and constellation faults: the contribution of these faults is generally negligible for P_{HMI}s larger than 10^{-9} but it becomes the main contribution in the strong geometries, with larger number of satellites in view.

Figure 7.2 also shows that the upperbound of the P_{HMI} computed by the DIA algorithm based on single iteration (therefore without foreseeing any exclusion) is always smaller than the upperbounds computed by Standard RAIM and ARAIM. The upperbound computed by ARAIM tends to be smaller than Standard RAIM in weak geometries but larger in stronger geometries (this is due to the fact that ARAIM takes into account double satellite faults in the computation of the integrity risk). The DIA based on three iterations, on the other hand, always (or almost) estimates a larger P_{HMI} than the DIA based on one iteration; this estimate is generally smaller than the ARAIM's, except for few geometries (e.g. between 15:40 and 16:10).

These results signify that, as long as the upperbounds are correct, the DIA algorithms (especially the one based on single iteration) guarantee higher availability than ARAIM and Standard RAIM. This result is a consequence of the different methods used by DIA and ARAIM to upperbound the P_{HMI}, as described in Sections 4.2.1 and 6.4. The two methods to upperbound the P_{HMI} are directly compared in Appendix K, where it is shown that for a simple model with one unknown parameter and four observations the upperbound determined by ARAIM is about one order of magnitude larger than the DIA's one. Note that the worst-case bias method employed by the DIA algorithm is more computational expensive than the ARAIM approximation.

As we mentioned, the DIA based on three iterations always estimates a larger P_{HMI} than the DIA based on one iteration. This means that, in weak geometries, it is possible that the single iteration DIA can guarantee integrity, when the multiple iterations DIA cannot: renouncing to attempting any exclusion, and simply declaring Alert in case of detection from the OMT, can reduce the integrity risk enough to make the geometry available. The higher risk associated with multiple iterations DIA means that performing exclusions, with the DIA

Table 7.13: Results of the 10^6 epochs of simulation for the case of double satellite fault, in geometry A. Measurements from satellites PRN69 and PRN89 have been biased by 6 and 8 meters respectively. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	MDs	WDs	ODs	CDs
Standard RAIM	0.9994	0.9004	0	0	5.87×10^{-4}	-	-	-
ARAIM	0.9986	0.9985	1.01×10^{-4}	0	0.9754	1.93×10^{-2}	0	5.3×10^{-3}
DIA-k_1	6.2×10^{-5}	6.2×10^{-5}	6.6×10^{-5}	6.6×10^{-5}	0.5381	1.93×10^{-2}	0	0.4422
DIA	2.69×10^{-2}	2.69×10^{-2}	6.0×10^{-5}	4.3×10^{-5}	1.34×10^{-2}	2.14×10^{-2}	0	0.9652
DIA-w	2.80×10^{-2}	2.80×10^{-2}	5.8×10^{-5}	3.8×10^{-5}	3.6×10^{-3}	2.14×10^{-2}	0	0.9750

method, introduces more risk (due to a larger WD rate) than what is saved by the reduction of the MD rate (due to larger significance of the OMT, and therefore higher detection power: the significance of the OMT can be increased in a multiple iteration procedure applying Equation (6.19)). As we previously mentioned, WDs are practically as dangerous as MDs, since they may mask the actual anomaly, making it more difficult to detect.

Note again that the DIA method with a single iteration was not tested specifically in the observation simulations because its performance would result identical to the Standard RAIM.

7.7.2. INTEGRITY MONITORING — FAS AND HMIS

The fault-free simulations (Tables 7.5 and 7.6) show that all the algorithms analyzed respect the requirement on $P_{FA'}^{req}$, though all the DIA algorithms appear generally conservative (the rate of FAs is significantly smaller than expected). This shows that the approximation for the $P_{FA_{tot}}$ computation in Equation (6.19) is quite conservative. We also register a larger rate of WDs for DIA and DIA-w, this is due to the choice of smaller thresholds for the detection tests (Equation (5.51)), and therefore a higher number of adaptations of the model (that anyway do not result in Alerts).

In any scenario including failing satellites, we notice that for Standard RAIM and ARAIM all the detections (WDs, ODs, CDs) lead to Alerts (and almost always False Alerts, since the positioning failures occur at a much lower rate). This is expected for Standard RAIM because no exclusion is foreseen in this algorithm, but it occurs also in ARAIM because the geometry selected is too weak to allow for exclusion. In fact ARAIM recomputes the P_{HMI} (or the PLs) after an exclusion is performed, and checks if this still complies with the requirements. The geometries chosen for the simulations are quite weak and are just within the requirements, for each of the algorithms, therefore the P_{HMI} computation update in ARAIM always results in an Alert. On the other hand, successful exclusions are possible employing the DIA algorithms, since the P_{HMI} computation (before taking the measurements) includes wrong exclusions occurrences (DIA computation based on 3 iterations, Equation 6.21). As a result

Table 7.14: Results of the 10^6 epochs of simulation for the case of double satellite fault, in geometry B. Measurements from satellites PRN73 and PRN74 have been biased by 4 and 7 meters respectively. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	MDs	WDs	ODs	CDs
Standard RAIM	0.9037	0.9037	0	0	9.63×10^{-2}	-	-	-
ARAIM	0.8147	0.8146	6.1×10^{-5}	0	0.6814	0.3184	0	1.20×10^{-4}
DIA-k_1	9×10^{-6}	7×10^{-6}	1.74×10^{-4}	1.74×10^{-4}	0.6037	0.3595	0	3.69×10^{-2}
DIA	6.39×10^{-2}	6.37×10^{-2}	4.79×10^{-4}	2.98×10^{-4}	0.1345	0.4102	0	0.4552
DIA-w	5.52×10^{-2}	5.50×10^{-2}	4.37×10^{-4}	2.64×10^{-4}	9.89×10^{-2}	0.4078	0	0.4933

all the DIA algorithms have by far a much smaller number of FAs. This lower FA' rate in faulty conditions can be a very desirable performance, in particular when, with a large number of satellites in view, the likelihood of occurrence of an anomaly is not negligible. The results of the random single satellite fault cases (Tables 7.11 and 7.12) suggest that, in case of biases smaller than 15 meters affecting single satellite measurements, the DIA algorithms can guarantee a rate of FA' several order of magnitude smaller than ARAIM.

With regard to the HMI rate, we observe that all the algorithms analyzed respect their original prediction/upperbound (in Section 7.6). The lowest number of HMIs is observed with the ARAIM algorithm, which results as the safest algorithm, though this is also due to the fact that no exclusion was actually successful in ARAIM (all detections resulted in an Alert). All algorithms performed safely in the scenarios analyzed.

7.7.3. FAULT DETECTION AND EXCLUSION — MDS, WDS AND ODS

The last four columns of the Tables 7.5 to 7.16 determine the detection and exclusion performance of the different algorithms. DIA and DIA-w are more powerful than the other algorithms: in fact a larger rate of CDs and a smaller rate of MDs is observed. This is due to the fact that the possibility of performing additional iterations allows to sensibly increase the significance (P_{FA}) of the detection tests. We can see from Equation (5.51) how the significance of the tests is increased in the multiple iteration algorithm: to increased significance corresponds increased power of the test. On the other hand this increased power leads also to a larger number of WDs and ODs.

We note furthermore that Standard RAIM and DIA-k_1 show the same MD rate in case of single satellite faults, and that this MD rate is larger than with the ARAIM algorithm. This is due to the fact that both Standard RAIM and DIA employ the OMT (χ^2 test) for detection, and this test is less powerful than the specific tests (SS tests or w-tests), with Bonferroni adjustment, in detecting single satellite faults. Note that the B-method in the DIA procedure guarantees equal power of OMT and w-tests, but not equal P_{FA}: if we set the OMT to have the same significance as the combination of w-tests (or SS tests), the final power will be different

Table 7.15: Results of the 10^6 epochs of simulation for the case of random double satellite fault, in geometry A. The simulated bias sizes are random values between 0 and 15 meters. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	MDs	WDs	ODs	CDs
Standard RAIM	0.8387	0.8085	3.02×10^{-2}	6×10^{-6}	0.1613	-	-	-
ARAIM	0.8412	0.8405	7.86×10^{-4}	2×10^{-6}	0.7804	7.6×10^{-3}	0	0.2120
DIA-k_1	4.2×10^{-3}	3.2×10^{-3}	1.60×10^{-3}	6.35×10^{-4}	0.6745	1.20×10^{-2}	0	0.3135
DIA	1.81×10^{-2}	1.64×10^{-2}	1.80×10^{-3}	1.23×10^{-4}	0.4402	3.77×10^{-2}	0	0.5221
DIA-w	1.52×10^{-2}	1.38×10^{-2}	1.70×10^{-3}	3.15×10^{-4}	0.4194	3.12×10^{-2}	0	0.5494

(lower for the OMT in case of single satellite faults). The use of the w-tests for detection, as in the adapted DIA, is more powerful even in a single iteration implementation.

7.7.4. SUMMARY AND CONCLUSIONS

In view of the simulation results reported above, we can gather the following conclusions, classified under different performance parameters:

- *Integrity requirements and safety*

  The results from the simulation show that the upperbounds on the P_{HMI} of each algorithm considered are satisfied, strongly suggesting that **all algorithms are safe**. Judging from the number of HMI events, Standard RAIM and ARAIM are **the most conservative** and safest algorithms, but this performance comes at the price of **much a lower continuity in case of a fault**, due to a much lower capability to adapt the model to the anomaly. We stress that if with the DIA-k_1 algorithm we renounce to perform any adaptation and we restrict to the first iteration, it would produce the same number of HMIs as the Standard RAIM — the increased risk is due to the choice made of performing adaptation and increase the continuity also in case of fault. The results furthermore suggest that in ARAIM the P_{HMI} **is** significantly **overestimated**, whereas the DIA algorithm is quite precise in risk estimation (the number of HMIs is in line with the prediction), and can afford to attempt adaptations of the model. Nevertheless the increased P_{HMI} that comes with the adaptation of the model suggests that the current DIA's adaptation method is not yet optimal (see below).

- *Integrity availability*

  The geometry simulations results, in Figure 7.2, show that the DIA algorithm's upperbound to the P_{HMI} based on single iteration is at least one order of magnitude tighter than the ARAIM's (which in turn is about one order of magnitude tighter than Standard RAIM). This means that DIA can guarantee **higher integrity availability** — in

Table 7.16: Results of the 10^6 epochs of simulation for the case of random double satellite fault, in geometry B. The simulated bias sizes are random values between 0 and 15 meters. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	MDs	WDs	ODs	CDs
Standard RAIM	0.8225	0.7711	5.2×10^{-2}	5×10^{-6}	0.1775	-	-	-
ARAIM	0.8190	0.8098	9.8×10^{-3}	5.11×10^{-4}	0.7863	2.98×10^{-2}	0	0.1838
DIA-k_1	1.04×10^{-2}	3.7×10^{-3}	1.85×10^{-2}	1.18×10^{-2}	0.6878	4.52×10^{-2}	0	0.2670
DIA	3.66×10^{-2}	2.20×10^{-2}	1.90×10^{-2}	4.5×10^{-3}	0.4467	8.56×10^{-2}	0	0.4677
DIA-w	3.24×10^{-2}	1.88×10^{-2}	1.90×10^{-2}	5.4×10^{-3}	0.4377	7.81×10^{-2}	0	0.4842

Table 7.17: Results of the 10^5 epochs of simulation for a generic anomaly case, in geometry A. The algorithms tested are Standard RAIM, ARAIM, DIA (maximum 3 iterations) with conservative threshold k_1 (indicated with DIA-k_1), DIA (maximum 3 iterations) and adapted DIA (indicated with DIA-w).

	Alerts	False Alerts	PFs	HMIs	MDs
Standard RAIM	0.9527	0.9527	0	0	0.0473
ARAIM	0.8618	0.8618	2×10^{-5}	0	0.1382
DIA-k_1	0.9329	0.9329	2×10^{-5}	0	0.0473
DIA	0.7438	0.7434	4.1×10^{-4}	0	5×10^{-5}
DIA-w	0.5021	0.5017	4.1×10^{-4}	4×10^{-5}	0.0149

the location considered with CAT I requirements the DIA (single iteration) is available 99.3% of the time whereas ARAIM only 95.8% of the time (Figure 7.2). The observation simulations support the correctness of the DIA's upperbounds and the conservative character of the ARAIM's one. The DIA based on three iterations (DIA) shows instead similar availability as ARAIM (Figure 7.2) — but a much higher continuity in case of occurrence of an anomaly (Tables 7.7 to 7.17). Note that in fact renouncing to exclude satellites and adapting the observation model reduces the risk of HMI but comes to the price of lower continuity in faulty cases: this is acceptable because the prior probability of occurrence of faults is assumed to be very small, but can become an issue to take into account when the number of satellites in view is large.

- *Testing effectiveness and detection power*

  The results from the simulation show that the DIA algorithm (based on multiple iterations: DIA) **is the most powerful** method in detecting and identifying anomalies (see for instance Tables 7.11, 7.12, 7.15, 7.16, the CDs column). The higher power is linked to the possibility of **sensibly increasing the significance** of the detection tests, under the assumption that successive iterations of the algorithm are to be run. Furthermore

the DIA-w (employing w-tests for detection instead of the OMT) is more powerful in detecting single and double anomalies than the standard DIA, in the same way as the ARAIM is more powerful than DIA-k_1 and Standard RAIM. The OMT results to be more powerful than specific tests in detecting generic anomalies with more than 2 corrupted measurements (Table 7.17). A preference between the two methods of detection (OMT vs specific tests) can be expressed on the basis of the weight/probability that is given to the occurrence of generic anomalies, compared to the one given to single satellite faults.

A large detection power is a desirable property, but we can see from the results that in the DIA algorithms **a larger detection power** corresponds also to **larger WD and OD rates** (e.g. see Tables 7.7 to 7.17). While a larger detection power contributes to a reduction of the P_{HMI}, an associated larger WD (and OD) rate acts in the opposite direction. The increase of WD is most likely the main issue in the DIA procedure, and it is linked to the exclusion method employed (and most exclusion methods in general), further discussed below.

- *Exclusion mechanisms and position estimation*

  The observation simulations results provide us information on the performance of the DIA exclusion (adaptation) mechanism (and in general of any iterative exclusion method based on a forward selection approach). The ARAIM exclusion mechanism performance cannot really be assessed because no adaptation was performed successfully by ARAIM in the scenarios analyzed — ARAIM requires much stronger geometry or looser integrity requirements to allow exclusion to be performed (in this sense it is a fairly conservative algorithm).

  Regarding the DIA algorithm, we observe that **multiple iterations do not result in an improvement in the positioning/reduction of the risk** (due to the increased rate of Wrong Detections). Both geometry simulations (Figure 7.2) and observations simulations (see Tables 7.7 to 7.17: HMI rate is higher in the DIA algorithms than in Standard RAIM and ARAIM) show that the P_{HMI} increases when multiple iterations of the DIA are foreseen. The reason why WDs are particular detrimental to integrity is that in case of WD, the successive iteration of the algorithm becomes much less effective than expected — it becomes extremely more difficult to find the actual faulty satellite once a WD has occurred (see also Section 6.5.2). Furthermore the position estimation is strongly jeopardized in case of a WD, both because of the weakening of the geometry after exclusion and because of the upraising of an extra bias in the position conditioned on the WD. This leads to the conclusion that the standard exclusion mechanism of the DIA is not optimal. Improvement could be made by introducing checks on the WD probability, (similarly as to ARAIM, or employing Equation (5.44)) and making the exclusion mechanism more selective, or considering exclusion of multiple satellites at a time (with a different Subset Selection method).

  The (mathematical) root-cause of the problem, that makes extremely more difficult to monitor the risk after a WD, is that after the first iteration of the procedure we have to deal with a conditional distribution of the remaining measurements. Theorical performance parameters that can be computed for the first iteration of the procedure do not constitute good approximations for the successive iterations. This is especially true in

case of WD, but holds in principle also in case of CDs.

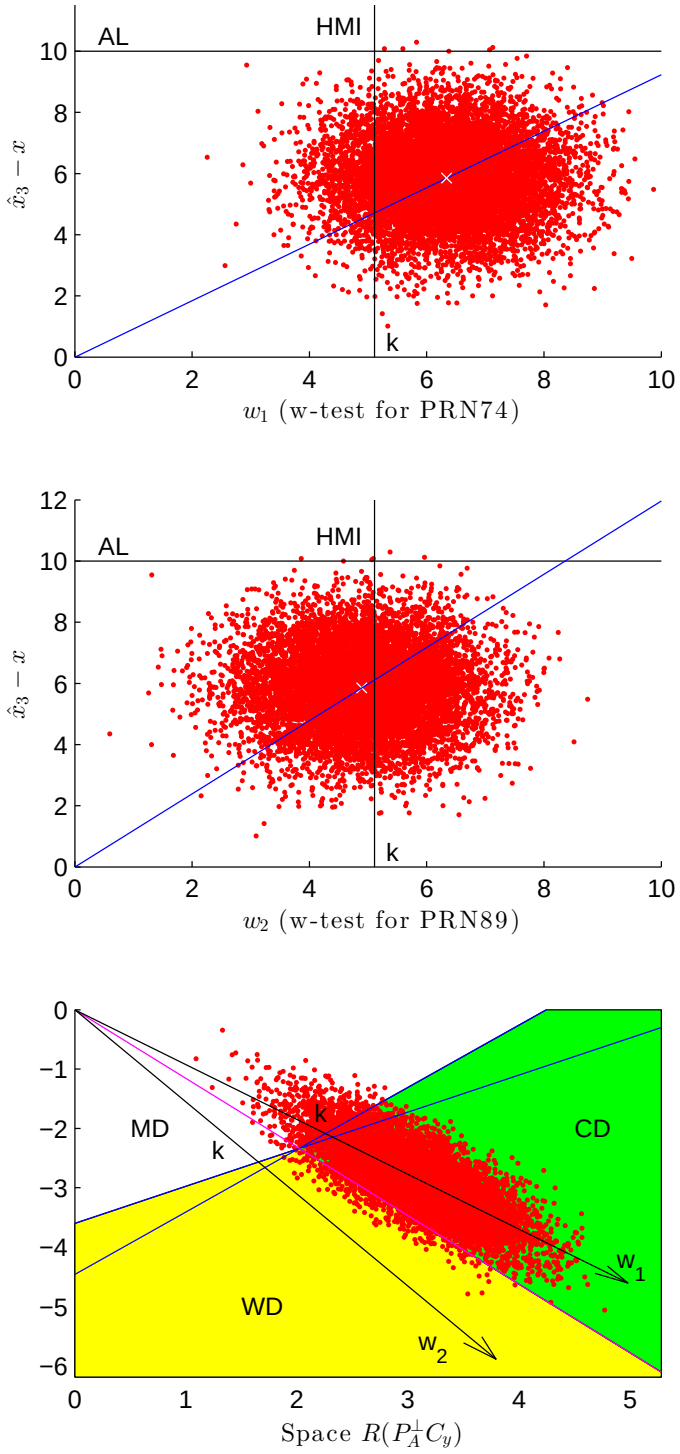


Figure 7.5: Distribution of w-tests statistics and vertical positioning error, together with a detection space representation, in case of observations from satellite PRN74 being biased by 8 meters (10^4 samples represented). Top and middle graph plot on the horizontal axis the two w-test statistics of main interest (w_1 for PRN74 and w_2 being most correlated to w_1, most likely to wrongly detect the fault) against the positioning error on the vertical axis. AL and w-test threshold are reported, with the corresponding HMI area. The bottom graph shows instead the space $R(P_A^{\perp} C_y)$, the detection space, with the distribution of the projected measurements, where the areas of MD, WD and CD can be identified.

8

CONCLUSIONS

8.1. SUMMARY AND RESULTS

In this thesis the RAIM problem has been introduced and given a mathematical formulation. On the basis of such formulation the statistical approaches available in literature to confront the problem have been reviewed and discussed, as well as the most popular RAIM algorithms developed by the scientific community so far — Least Square Residuals RAIM (Standard RAIM) and ARAIM. Next, the DIA procedure, the FDE method developed by TU Delft, has been presented and its application to the RAIM problem proposed. The three RAIM algorithms, Standard RAIM, ARAIM and DIA, have been implemented on software and tested by means of simulations. The results showed that all the algorithms are generally able to monitor the integrity effectively and their performance can be measured and compared. In the simulated scenarios the DIA algorithm based on single iteration delivered higher availability than the other RAIM algorithms, whilst the multiple-iterations implementation delivered higher detection power and continuity. The results showed furthermore that both the Exclusion methods employed by ARAIM and DIA are not optimal from the point of view of maximizing the integrity — this led to the recommendation, for the development of future RAIM concepts, of avoiding any prioritization of single exclusion with respect to simultaneous multiple exclusions.

We now go through the dissertation in detail and highlight the main results of each chapter.

- After the first introductory chapter, in Chapter 2 we introduced the concept of integrity in navigation and of RAIM, and we provided a mathematical definition of the RAIM problem (Equations (2.11)-(2.15)). We outlined the functions that a RAIM algorithm should fulfil, and defined its input, output and performance parameters. We hope to have clarified the many concepts and terms of which earlier definitions have led to confusion or misunderstandings.

- In Chapter 3 we gave an overview of the most relevant tools available in statistics to deal with the RAIM problem. Most RAIM algorithms currently in use and under development make use of FDE procedures, which are based on Multiple Comparisons and Subset Selection theories presented in this Chapter. The currently in use FDE methods can find their place among the other different methods, can be compared and

discussed, and ideas for alternatives or modifications can arise. This is particularly true when we consider that the Exclusion mechanism of current RAIM algorithms (in particular ARAIM) is still a work in progress.

- In Chapter 4 the two RAIM algorithms for use in aircraft navigation most popular in literature were reviewed: the classic Weighted RAIM by Walter and Enge (referred to as Standard RAIM) [107] and the modern ARAIM developed by the Stanford group [11]. For both the procedures the full algorithm is provided. The Standard RAIM is a simple and sound algorithm, which does not foresee exclusion of satellites and adaptation of the observation model, and is not designed to cope with multiple satellite faults: the main shortcomings were pointed out and possible solutions have been proposed. The ARAIM algorithm is still a work in progress, in its more advanced forms, and is currently in experimental phase. We described only the baseline algorithm (mainly based on [11]). This algorithm is reckoned to be still sub-optimal from different points of view: most of the approximations employed have been discussed and ideas for improvement have been proposed. In particular:
  - Not all the possible iterations of the algorithm are taken into account in the computation of the P_{HMI} (problem already addressed in [51]).
  - The RAIM requirement $P_{FA'}^{req}$ on the FA' rate is translated by ARAIM to a requirement on the P_{FA} of the first detection step of the algorithm, as if the algorithm was made of a single test/iteration , when instead $P_{FA_{tot}}$ (for the full procedure) should be computed.
  - The computation of the P_{HMI} through Equation (4.40), obtained by means of the approximation described in Section 4.2.1, may result quite conservative (see Appendix K for a comparison with a different approximation).
  - The ARAIM employs SS tests instead of UMPI tests, this choice can be discussed (see Appendix I).
  - The ARAIM exclusion process, including tests to check for wrong exclusions, is in experimental phase. Some formulae from [11] are not justified and/or subject to modifications.

- In Chapter 5 the DIA procedure was reviewed. The DIA procedure is a well established method for gross error detection in geodesy developed by TU Delft. The application of the DIA procedure to the RAIM problem for use in aircraft navigation is naturally possible, provided a direct way to compute the main RAIM parameters, in particular the P_{HMI}, is defined — the *reliability* monitoring provided by the DIA has to be extended to be able to cope specifically with the RAIM case.

- In Chapter 6 we discussed the application of the DIA algorithm to the RAIM problem. The following additions/modifications have been proposed with respect to the standard DIA algorithm:
  - A method to upperbound the P_{HMI} of a satellite geometry as a function of the thresholds choice for the tests has been proposed (Section 6.2), based on the worst-case bias approach (Equation (6.9)). A method to determine the worst-case bias in case of multi-dimensional fault was furthermore proposed (Section 6.4.2).

  - A method to take into account the possible multiple iterations of the DIA algorithm has been proposed (Section 6.5.1), to quantify/approximate the FA rate and P_{HMI} (Equations (6.19), (6.20), (6.22)).
  - Possible different roles (other than the Detector one) for the OMT have been envisioned (Section 6.6).

- In Chapter 7 numerical simulations were run to compare Standard RAIM, ARAIM and DIA procedure. After scrutiny of the results, the following conclusions on the performance of the different algorithms were gathered:

  - *Integrity requirements and safety*

    The results from the simulation show that the upperbounds on the P_{HMI} of each algorithm considered are satisfied, strongly suggesting that all algorithms are safe. Judging from the number of HMI events, Standard RAIM and ARAIM are the most conservative and safest algorithms, but this performance comes at the price of a much lower continuity in case of a fault, due to a much lower capability to adapt the model to the anomaly. The results furthermore suggest that in ARAIM the P_{HMI} is significantly overestimated, whereas the DIA algorithm is quite precise in its risk estimation.

  - *Integrity availability*

    The DIA algorithm's upperbound to the P_{HMI} based on a single iteration is at least one order of magnitude tighter than the ARAIM's (which in turn is about one order of magnitude tighter than Standard RAIM). DIA can therefore guarantee a higher integrity availability — in the location considered with CAT I requirements the DIA is available 99.3% of the time, whereas ARAIM only 95.8% of the time. The observation simulations support the correctness of the DIA's upperbounds and the conservative character of the ARAIM's one. The DIA based on three iterations shows instead a similar availability as ARAIM — but a much higher continuity in case of occurrence of an anomaly.

  - *Testing effectiveness and detection power*

    The results from the simulation show that the DIA algorithm (based on multiple iterations) is the most powerful method in detecting and identifying anomalies. A large detection power is a desirable property, but we can see from the results that in the DIA algorithms a larger detection power corresponds also to larger WD and OD rates. While a larger detection power contributes to reduce the P_{HMI}, a larger WD (and OD) rate acts in the opposite direction. The increase of WD is most likely the main issue in the DIA procedure, and it is linked to the exclusion method employed (and exclusion methods in general), further discussed below.

  - *Exclusion mechanisms and position estimation*

    The ARAIM exclusion mechanism performance cannot really be assessed because no adaptation was performed successfully by ARAIM in the scenarios analyzed — ARAIM requires a much stronger geometry or looser integrity requirements to allow exclusion to be performed (in this sense it is a fairly conservative algorithm).

    Regarding the DIA algorithm, we observe that multiple iterations do not result in an improvement in the positioning/reduction of the risk (due to the increased

rate of Wrong Detections). Both geometry simulations (Figure 7.2) and observations simulations show that the P_{HMI} increases when multiple iterations of the DIA are foreseen. The reason why WDs are particular detrimental to integrity is that in case of a WD, the successive iteration of the algorithm becomes much less effective than expected — it becomes extremely more difficult to find the actual faulty satellite once a WD has occurred. Furthermore the position estimation is strongly jeopardized in case of a WD, both because of the weakening of the geometry after exclusion and because of the upraising of an extra bias conditioned on the WD. This leads to the conclusion that the standard exclusion mechanism of the DIA is not optimal, and in general any iterative exclusion method based on a forward selection approach. The (mathematical) root-cause of the problem, that makes extremely more difficult to monitor the risk after a WD, is that after the first iteration of the procedure we have to deal with a conditional distribution of the remaining measurements. Theorical performance parameters that can be computed for the first iteration of the procedure do not constitute good approximations for the successive iterations. This is especially true in case of WD, but holds also in case of CDs.

8.2. RECOMMENDATIONS AND FUTURE WORK

As it is apparent from the methods review and simulation results, all RAIM algorithms discussed have margin of improvement. Even though ARAIM performs safely, some of the approximations employed seem rather conservative and the exclusion mechanism seems not particularly effective. The DIA algorithm is also safe, but its exclusion mechanism does not appear to perform optimally.

RELIABILITY — P_{HMI} COMPUTATION

From a reliability analysis point of view — i.e. prior computation of the P_{HMI} based on the satellite geometry alone — the DIA seems to perform better than ARAIM. For instance, the worst-case bias approximation employed in the DIA to upperbound the P_{HMI} is to be preferred over the corresponding ARAIM approximation. Also the method, proposed in this dissertation, to compute the P_{HMI} for multiple DIA iterations, performs well and allows to monitor the risk associated with wrong exclusions. The method relies on Monte Carlo integration, but since the computation is based on the satellite geometry alone, it can be performed in advance with respect to the epoch of observation. On the other hand, from the exclusion/adaptation mechanism point of view, the simple method employed by the DIA procedure appears to possibly incur into high risk of wrong exclusions and associated higher risk of HMI: under this perspective the ARAIM algorithm, with its monitoring against wrong exclusion, is acknowledging the problem and starting to propose possible solutions.

FDE PROCEDURE

One of the main weaknesses of the RAIM algorithms analyzed is the Exclusion (or Adaptation) mechanism. Both ARAIM and DIA procedures reckognize that in many geometries — given a requirement on the continuity — attempting exclusion introduces more risk than just declaring an Alert (in ARAIM the risk is recomputed after a detection, and in the available geometries considered in the simulations it always exceeded the requirement). This suggests

that the Exclusion methods employed are sub-optimal. We would like to have, ideally, an Exclusion mechanism for which the possibility of excluding some satellites results in a lower or at most equal P_{HMI} for any geometry under consideration. If it is not possible to conceive an exclusion method with such performance, then it may be better renouncing to any exclusion/adaptation and performing only detection, or resorting to robust estimation methods. Improvement could be made in the DIA by introducing checks on the WD probability (similarly as to ARAIM, or employing Equation (5.44)) and making the exclusion mechanism more selective, or considering exclusion of multiple satellites at a time (with a different Subset Selection method, see Section 3.6.3). For instance, it could be possible to restrict the exclusion of a single satellite to the cases in which only one w-test exceeds the threshold (excluding two satellites or declaring Alert if multiple w-tests exceed their threshold), or/and introduce a check on the correlation between w-tests and avoid to exclude a single satellite when its w-test is highly correlated with another. Also methods as the p-value or $\underline{C}_p$ methods could be employed to possibly exclude multiple satellites at a time. Improvement to the ARAIM Exclusion method could also be made, by modifying the exclusion confirmation tests or finding less conservative ways to compute the risk after exclusion.

A

INTEGRITY — BAYESIAN VIEW

This Appendix constitutes an extension to Section 2.4, in which the RAIM problem was defined. It provides an interpretation of the RAIM problem from a Bayesian point of view. Assuming a Bayesian perspective, the probability distribution of the observable $\underline{y}$ can be considered to be fully known.

Let the observable $\underline{y}$ depend upon some parameter x we want to infer about. Suppose that the probability distribution function of the observable $\underline{y}$, $f_{\underline{y}}(y)$, is fully known. This means that either:

1. the parameter x is deterministically known or
2. the PDF of $\underline{x}$ is fully known.

Only the latter case is of interest for our problem.

The natural way of inferring about the actual realization of the parameter x, on the basis of a measurement y, is to determine the posterior probability $f_{\underline{x}|y}(x|y)$. Given everything is known by hypothesis — i.e. prior distribution of state of nature $\underline{x}$ and distribution of the observable conditioned on $\underline{x}$, $f_{\underline{y}|\underline{x}}$ — the posterior probability can be obtained for any y.

We have:

$$\begin{aligned} f_{\underline{x},\underline{y}} &= f_{\underline{y}|\underline{x}} \cdot f_{\underline{x}} \\ f_{\underline{y}} &= \int_{R^n} f_{\underline{x},\underline{y}} \mathrm{d}x \\ f_{\underline{x}|\underline{y}} &= \frac{f_{\underline{x},\underline{y}}}{f_{\underline{y}}} \end{aligned} \tag{A.1}$$

If we write a function of the observable $\underline{\hat{x}} = F(\underline{y})$ (the estimation function), we can define the following probability Probability of Positioning Failure:

$$P_{PF} = P(\underline{\hat{x}} - \underline{x} \notin \Omega_{AL}) \tag{A.2}$$

From $f_{\underline{y}}$ and $F(\underline{y})$ the distribution of $\underline{\hat{x}}$ can be obtained by transformation rule.

The probability in Equation (A.2) can anyway be decomposed in two parts to highlight the posterior probability as:

$$P_{PF} = \int_{R^m} P(\hat{x}(y) - \underline{x} \notin \Omega_{AL}|y) f_{\underline{y}}(y) \mathrm{d}y \tag{A.3}$$

The probability $P(\hat{x}(y) - \underline{x} \notin \Omega_{AL}|y)$ can be computed knowing $f_{\underline{x}|y}(x|y)$. We can find the function $\underline{\hat{x}} = F(\underline{y})$ that minimizes the above probability, but also we can define a smaller domain of integration, instead of the full y domain R^m:

$$P'_{PF} = \int_{\Omega} P(\hat{x}(y) - \underline{x} \notin \Omega_{AL}|y) f_{\underline{y}}(y) \mathrm{d}y \tag{A.4}$$

This means selecting only some measurements, and computing the risk of having a positioning error bigger than the AL only for them. This probability is always equal or smaller than the probability in Equation (A.3). Actually, by properly choosing Ω, this probability can be made small as desired. Therefore if minimizing the probability in Equation (A.3) is not sufficient to guarantee the integrity (the minimum obtained is not small enough), it is possible to resort to the selection of measurements employed in Equation (A.4), that would allow to reduce further the risk of having a positioning failure.

Example. Let:

$f_{\underline{x}}(x) = U(0,10)$	Uniform distribution between 0 and 10
$f_{\underline{y}|\underline{x}}(y|x) = \frac{9}{10}N(x,1) + \frac{1}{10}N(x+5,1)$	Sum of two normal distributions

With the above conditional distribution $f_{\underline{y}|\underline{x}}(y|x)$, the measurements can be biased by the amount 5 with respect the true position x with probability 1/10. We have:

$$\begin{aligned}
&f_{\underline{x},\underline{y}}(x,y) = \tfrac{1}{10}(\tfrac{9}{10}N_y(x,1) + \tfrac{1}{10}N_y(x+5,1)) \;\; \forall x \in [0,10],\ 0 \text{ otherwise} \\
&f_{\underline{y}}(y) = \textstyle\int_0^{10} f_{\underline{x},\underline{y}}(x,y)\mathrm{d}x = \tfrac{1}{10}(\tfrac{9}{10}\Phi_x|_0^{10}(y,1) + \tfrac{1}{10}\Phi_x|_0^{10}(y-5,1)) = \\
&\quad = \tfrac{1}{10}(\tfrac{9}{10}\Phi_x|_0^{10}(y,1) + \tfrac{1}{10}\Phi_x|_5^{15}(y,1)) \\
&f_{\underline{x}|\underline{y}} = \frac{f_{\underline{x},\underline{y}}}{f_{\underline{y}}}
\end{aligned}$$

where the notation $\Phi_z|_a^b(\mu,\sigma^2)$ stands for the cumulative distribution between a and b of a normal distribution with mean μ and variance σ^2, for the random variable z. Figure A.1 shows the corresponding distribution $f_{\underline{y}}(y)$, whereas Figure A.2 shows the conditional distributions $f_{\underline{x}|\underline{y}}(x|y)$ for three different realizations of y, $y = 2, 8, 12$.

Now we choose an estimator, $\underline{\hat{x}} = \underline{y}$ (for simplicity). We set furthermore the ALs at a distance 2 from the true position (arbitrarily).

Figure A.3 shows the quantity $P(\hat{x}(y) - \underline{x} \notin \Omega_{AL}|y) f_{\underline{y}}(y)$ (function of y)[1]. As in Equations (A.3) and (A.4), the$P(\hat{x}(y) - \underline{x} \notin \Omega_{AL}|y) f_{\underline{y}}(y)$ has to be integrated over its full domain or over a region Ω of selected measurements. We can see that measurements between about 5 and 13 contains the highest risk density, therefore excluding measurements in this area can be beneficial for integrity, in order not to let them contribute to the P'_{PF}.

[1] This quantity has been computed numerically as a sum over the x domain, discretized in 1000 steps ($x = 0 : 0.01 : 10$).

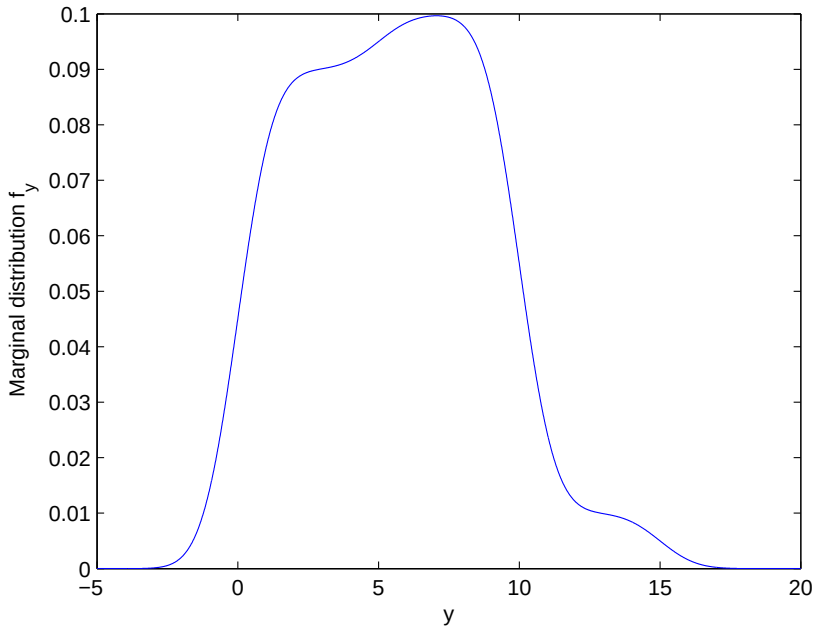


Figure A.1: Marginal distribution of the observable.

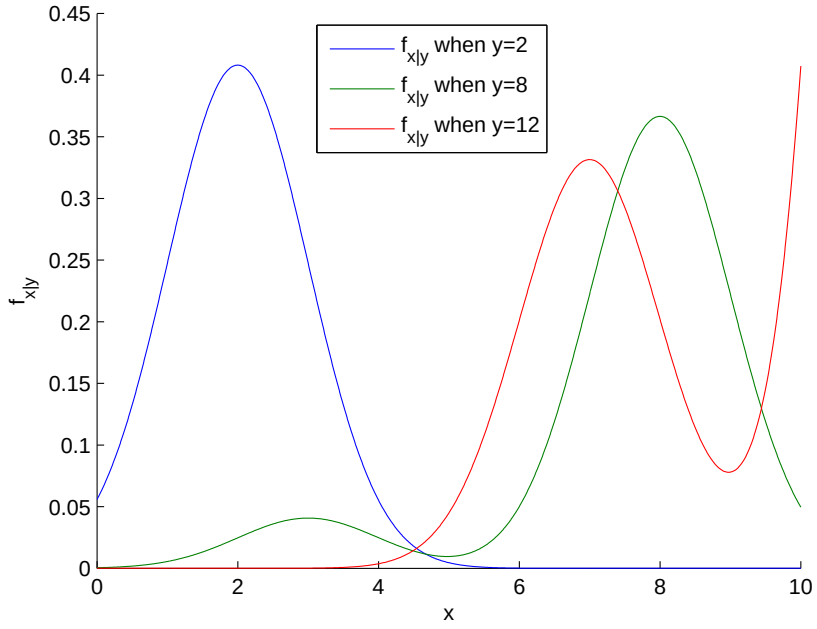


Figure A.2: Posterior distributions of the true position $\underline{x}|y$, given the observation $y = 2, 8, 12$.

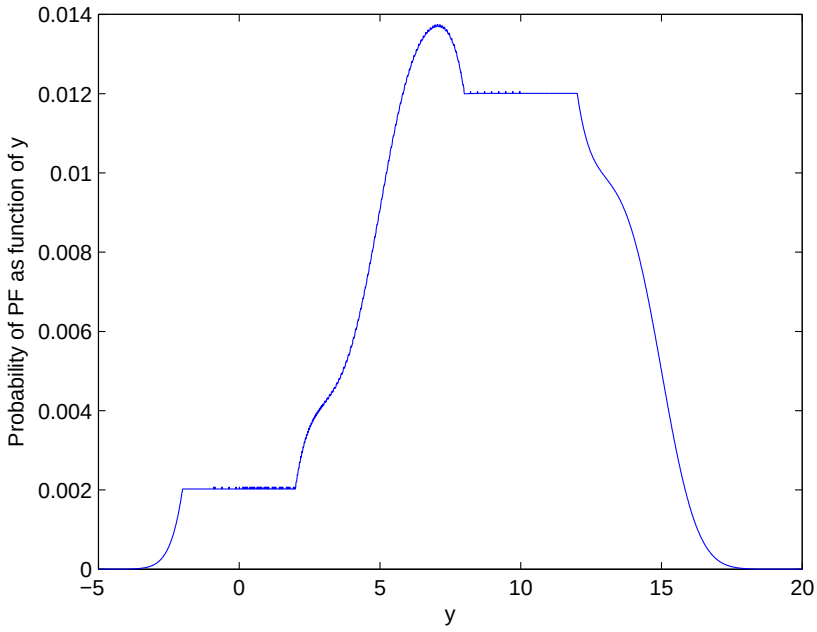


Figure A.3: Probability density function for the Positioning Failure event, choosing arbitrarily $AL = 2$ and $\underline{\hat{x}} = \underline{y}$.

B

FORMULATION BASED ON PROTECTION LEVELS

This Appendix constitutes an extension to Section 2.4, and provides an alternative definition of the RAIM problem.

An alternative formulation of the RAIM problem can be made introducing the Protection Levels (PLs), which make the problem more complex but are said to provide more information to the user. The problem can be defined as: find a region Ω *and* an estimation/detection rule that to the observable $\underline{y} \in \Omega$ assigns a position estimator $\underline{\hat{x}}$ and a random region bounded by the PLs, $\underline{\Omega}_{\text{PL}}$:

$$\underline{y} \in \Omega \rightarrow [\underline{\hat{x}}, \underline{\Omega}_{\text{PL}}] \tag{B.1}$$

such that:

$$P(\underline{\Omega}_{\text{PL}} \in \Omega_{\text{AL}} \cap \underline{y} \in \Omega) \geq 1 - P_{FA'}^{req} \quad \forall x \tag{B.2}$$

and

$$P(\underline{\hat{x}} - x \notin \Omega_{\text{AL}} \cap \underline{\Omega}_{\text{PL}} \in \Omega_{\text{AL}} \cap \underline{y} \in \Omega) \leq P_{\text{HMI}_0} \quad \forall x \tag{B.3}$$

$\underline{\Omega}_{\text{PL}}$, most likely of the same shape as Ω_{AL} (cilindrical for instance), therefore would just outbound (circumscript) the confidence region for the estimator associated to the extremely tight probability requirement on the P_{HMI}.

Also in this definition we keep specifying the event $\underline{y} \in \Omega$ since it may be not possible to determine a finite $\underline{\Omega}_{\text{PL}}$ for any $\underline{y} \in R^m$. In case $\underline{\Omega}_{\text{PL}}$ can instead be chosen also of infinite size, then would be possible to define $\underline{\hat{x}}$ and $\underline{\Omega}_{\text{PL}}$ for any $\underline{y}$ and we can formulate the problem as: find an an estimation/detection rule that to the observable $\underline{y}$ assigns a position estimator $\underline{\hat{x}}$ and a random region bounded by the Protection Levels $\underline{\Omega}_{\text{PL}}$:

$$\underline{y} \rightarrow [\underline{\hat{x}}, \underline{\Omega}_{\text{PL}}] \tag{B.4}$$

such that:

$$P(\underline{\Omega}_{\text{PL}} \in \Omega_{\text{AL}}) \geq 1 - P_{FA'}^{req} \quad \forall x \tag{B.5}$$

and

$$P(\underline{\hat{x}} - x \notin \Omega_{\text{AL}} \cap \underline{\Omega}_{\text{PL}} \in \Omega_{\text{AL}}) \leq P_{\text{HMI}_0} \quad \forall x \tag{B.6}$$

The idea of this formulation with Protection Levels is to provide the user with the further position information connected with the PLs, $\underline{\Omega}_{\text{PL}}$, which can be of course tighter than Ω_{AL}.

The PLs are typically employed in the so-called Stanford diagram, which provides a way to represent graphically the integrity related events defined in Section 2.4, such as PFs and HMIs. The Stanford diagram can be drawn only when the true position of the receiver (aircraft) is known, therefore it is a tool employed for research/experimentation, not in real life monitoring. It can be used to analyze a set of position solutions coupled with a relative PL, obtained from simulated or real data. An example Stanford diagram is shown in Figure B.1, for the horizontal position. On the horizontal axis is the horizontal positioning error, whereas on the vertical axis is the (H)PL (given from the WAAS). The bisectrix in the middle splits the diagram in two parts: one containing the cases in which the PL overbounds the positioning error and the other containing the cases in which the PL does *not* overbounds the true error. The cases in which the PL does not overbound the position error constitutes the occurrences of Misleading Information (MI), i.e. the RAIM monitoring did not provide a correct information (bound to the position error). MIs are not necessarily hazardous event (integrity failures): in fact to have a HMI the positioning error must be larger than the AL (set at 30m in this example), while the PL stays smaller than the AL (rectangular area on the lower-right angle). The cases in which the PL exceeds the AL represent the events of Alert, in which the system becomes unavailable.

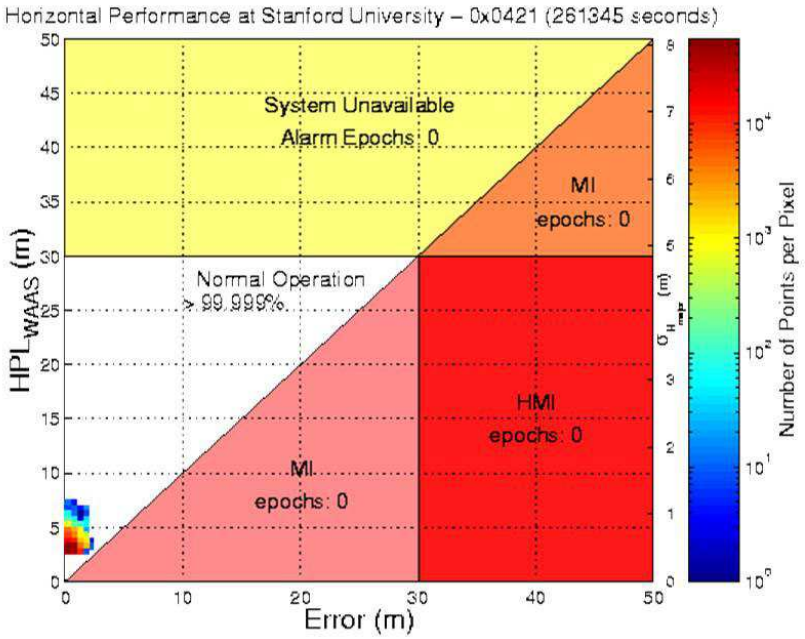


Figure B.1: Example of Stanford Diagram, for the horizontal positioning. On the horizontal axis is the horizontal positioning error, whereas on the vertical axis is the HPL (given from the WAAS). A position solution obtained at a certain epoch, together with its coupled PL, can be represented as a point on the graph. If the PL exceeds the AL (set to 30m), an Alert is raised (System Unavailable), which can be anyway justified or unjustified depending on the actual error. In case instead the PL is smaller than the AL, we have occurrence of NO (eventually with Misleading Information) in case the actual error is smaller than the AL, and occurrence of HMI in case the actual error is larger than the AL. Figure taken from [45].

C

EQUIVALENCE BETWEEN ADDITION OF EXTRA UNKNOWN PARAMETERS AND EXCLUSION OF OBSERVABLES FROM THE MODEL

In this Appendix we demonstrate the equivalence of Equations (3.11) and (3.12) presented in Section 3.3, with regards to the computation of $\underline{\hat{x}}_i$. Given the null hypothesis:

$$H_0: E(\underline{y}) = Ax \quad \text{cf. Equation (2.2)}$$

we focus here on the following alternative hypothesis:

$$H_a: E(\underline{y}) = Ax + C_y\nabla \quad \text{cf. Equation (3.21)}$$

with C_y the i^{th} canonical unit vector of R^m or a $m \times q$ matrix made up of different canonical unit vectors of R^m (say indexed by vector i_q), e.g.:

$$C_y = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ (case } i = 1\text{) or } C_y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \text{ (case } i_2 = [1\ 2]\text{)}$$

We demonstrate that this alternative hypothesis is equivalent to the following:

$$H_a: E(\underline{\tilde{y}}) = \tilde{A}x$$

where $\underline{\tilde{y}}$ is obtained from $\underline{y}$ elimimating the i^{th} component or the components with index i_q and $\tilde{A}$ is obtained from A eliminating the i^{th} row or the rows with index i_q. The two alternative hypotheses are equivalent in the sense that the BLUE of x (with its distribution) under both the alternative models, $\underline{\hat{x}}_a$, is the same, and so are consequently all the related parameters, as residuals $\underline{\hat{e}}_a$ and estimated measurements $\underline{\hat{y}}_a$. To say in different words, we

show here that adding extra parameters to the model with the above peculiar choice of C_y is equivalent to excluding/ignoring some of the available measurements (components of the full vector y).

Proof. Starting from the alternative model in (3.21), we write the *normal equations* [101] [98] for this model:

$$\begin{bmatrix} A^T Q_{yy}^{-1} A & A^T Q_{yy}^{-1} C_y \\ C_y^T Q_{yy}^{-1} A & C_y^T Q_{yy}^{-1} C_y \end{bmatrix} \begin{bmatrix} \hat{x}_a \\ \hat{\nabla} \end{bmatrix} = \begin{bmatrix} A^T Q_{yy}^{-1} y \\ C_y^T Q_{yy}^{-1} y \end{bmatrix} \tag{C.1}$$

we perform a Gaussian elimination pre-multiplying both sides of the equation by matrix:

$$\begin{bmatrix} I_n & -A^T Q_{yy}^{-1} C_y (C_y^T Q_{yy}^{-1} C_y)^{-1} \\ 0 & I_q \end{bmatrix}$$

to obtain the system:

$$\begin{bmatrix} A^T Q_{yy}^{-1} (I_m - C_y (C_y^T Q_{yy}^{-1} C_y)^{-1} C_y^T Q_{yy}^{-1}) A & 0 \\ C_y^T Q_{yy}^{-1} A & C_y^T Q_{yy}^{-1} C_y \end{bmatrix} \begin{bmatrix} \hat{x}_a \\ \hat{\nabla} \end{bmatrix} = \begin{bmatrix} A^T Q_{yy}^{-1} (I_m - C_y (C_y^T Q_{yy}^{-1} C_y)^{-1} C_y^T Q_{yy}^{-1}) y \\ C_y Q_{yy}^{-1} y \end{bmatrix}$$

We can now look just at the first row of the above system, and remembering that $I_m - C_y \cdot (C_y^T Q_{yy}^{-1} C_y)^{-1} C_y^T Q_{yy}^{-1} = P_{C_y}^{\perp}$, the projector onto the space orthogonal to $R(C_y)$, we have:

$$A^T Q_{yy}^{-1} P_{C_y}^{\perp} A \hat{x}_a = A^T Q_{yy}^{-1} P_{C_y}^{\perp} y$$

Applying the properties of the orthogonal projectors, $Q_{yy}^{-1} P_{C_y}^{\perp} = P_{C_y}^{\perp T} Q_{yy}^{-1} P_{C_y}^{\perp} = P_{C_y}^{\perp T} Q_{yy}^{-1}$, and defining $A_a \equiv P_{C_y}^{\perp} A$, we can write the above equation as:

$$A_a^T Q_{yy}^{-1} A_a \hat{x}_a = A_a^T Q_{yy}^{-1} y \tag{C.2}$$

This is the standard normal equation for an observation model of the type $E(y) = A_a x_a$. Substituting the specific type of C_y adopted in our alternative hypothesis into the formula for $P_{C_y}^{\perp}$, we can realize that this matrix has the form of an identity matrix (I_m) except that the i^{th} row (or the rows indexed by i_q), which is all zeros. As a result $A_a = P_{C_y}^{\perp} A$, the projection of A onto the space orthogonal to $R(C_y)$, is identical to A except for the i^{th} row (i_q rows), which is made of all zeros. This means that the i^{th} measurement (i_q^{th} measurements) is ignored from the original vector y, therefore H_a reads: $E(\tilde{y}) = \tilde{A}x$. This concludes the proof. □

We note here furthermore that computing $\hat{x}_a$ with:

$$\hat{x}_a = (A_a^T Q_{yy}^{-1} A_a)^{-1} A_a^T Q_{yy}^{-1} y$$

as from the normal equation (C.2) is equivalent to employing the following formula (presented in Section 4.2), adopted in the ARAIM algorithm [11]:

$$\hat{x}_a = (A^T Q_{yy_a}^{-1} A)^{-1} A^T Q_{yy_a}^{-1} y$$

with $Q_{yy_a}^{-1}$ obtained from Q_{yy}^{-1} by substituting the i^{th} (i_q^{th}) diagonal element with zero. This in fact means assigning infinite standard deviation to the i^{th} observation, and effectively completely de-weight/ignore such observation.

D

BAYESIAN APPROACH

In this Appendix we try to develop some basic ideas on the Bayesian approach to RAIM, as mentioned in Section 3.5. The Bayesian approach assumes that the PDF of the observable $\underline{y}$ is fully known, generally adopting *non-informative* prior distribution for the unknown parameters for which any a priori guess cannot be made. As reference Bayesian literature we mention [7], [55] [62] [96]. A Bayesian RAIM has been proposed also in [82].

As described in Appendix A, given an unknown set of parameters x and a conditional distribution for the observable $\underline{y}$, $f_{\underline{y}|x}$, the basic idea is to determine the posterior distribution of the parameters x, $f_{\underline{x}|y}(x|y)$. A prior distribution has to be assigned to $\underline{x}$ and the posterior distribution $f_{\underline{x}|y}(x|y)$ can be found employing the formulae in Equation (A.1).

Here we apply this approach to the GNSS model given in Section 2.7 (most often adopted throughout the dissertation). In this model the unknown parameters are x, H and ∇_i, and for each of them a prior distribution has to be assumed. We analyze in the following different choices for the prior distributions, and for each choice obtain the posterior distributions for the occurrence of the hypotheses, $f_{\underline{H}|y}$, and for the position x, $f_{\underline{x}|y}$.

D.1. UNIFORM PRIORS

Given are the hypotheses:

$$H_i : E(\underline{y}) = A\underline{x} + C_i\underline{\nabla}_i \ , \ i = 1,2,\dots,k \tag{D.1}$$

and the observable conditional distribution:

$$\underline{y}|H_i \sim N(E(\underline{y})|H_i, Q_{yy}) \tag{D.2}$$

Note that we shortened the notation from C_{y_i} to C_i and we did not explicitly consider the null hypothesis H_0 as a different hypothesis per se, since it can be readily included to the model as a particular H_i with zero ∇_i parameters, $q_i = 0$. We take a Bayesian approach and assume the following priors for the unknown parameters:

$$\begin{aligned}
&\underline{H} \sim P(H = H_i) = p_i \ , \ i = 1,2,\dots,k \\
&\underline{x}_j \sim c_x \ \text{ for } -\tfrac{1}{2c_x} \le \underline{x}_j \le +\tfrac{1}{2c_x} \ , \ j = 1,2,\dots,n \\
&\underline{\nabla}_{i_z}|H \sim \begin{cases} c_\nabla \ \text{ for } -\tfrac{1}{2c_\nabla} \le \underline{\nabla}_{i_z} \le +\tfrac{1}{2c_\nabla} \ , \ z = 1,2,\dots,q_i \ , \ \underline{H} = H_i \\ 0 \ \text{ for } \underline{H} \ne H_i \end{cases}
\end{aligned} \tag{D.3}$$

with c_x and c_∇ *small* constant scalars (the value given to c_x turns out not to be important, since this constant always cancels out in the next derivations, whereas this is not the case for c_∇), say $c_x \to 0$ and $c_\nabla \to 0$. For $\underline{x}_j$ it is also possible to use a standard *improper* non-informative prior of the type:

$$\underline{x}_j \sim c_x \text{ or } \underline{x}_j \sim 1 \text{ for } -\infty \leq \underline{x}_j \leq +\infty\,,\ j = 1,2,\ldots,n \tag{D.4}$$

with c_x any constant scalar, which is not a proper PDF because its integral from $-\infty$ to $+\infty$ does not result in the unity, but such type of prior cannot be used for $\underline{\nabla}_i$, unless all the hypotheses considered are characterized by the same q_i (see for instance [7]).

We apply the Bayes rule to determine the posterior probability for the occurrence of the hypotheses, $\underline{H}|y$:

$$f_{\underline{H}|y}(H|y) = \frac{f_{\underline{H},\underline{y}}}{f_{\underline{y}}} = \frac{\int_{x,\nabla} f_{\underline{x},\underline{\nabla}_i,\underline{H},\underline{y}} \mathrm{d}x\mathrm{d}\nabla}{\sum_H \int_{x,\nabla} f_{\underline{x},\underline{\nabla}_i,\underline{H},\underline{y}} \mathrm{d}x\mathrm{d}\nabla} \tag{D.5}$$

where the integration over parameters x and ∇_i is made between $-\frac{1}{2c_x}$ and $+\frac{1}{2c_x}$ for each component of x, and between $-\frac{1}{2c_\nabla}$ and $+\frac{1}{2c_\nabla}$ for each component of ∇_i. The joint distribution, computed as $f_{\underline{x},\underline{\nabla}_i,\underline{H},\underline{y}} = f_{\underline{H}} \prod_j f_{\underline{x}_j} \prod_z f_{\underline{\nabla}_{i_z}|H} f_{\underline{y}|H}$, writes:

$$\begin{aligned} &f_{\underline{x},\underline{\nabla}_i,\underline{H},\underline{y}} = p_i c_x^n c_\nabla^{q_i} \frac{1}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}\|y - Ax - C_i\nabla_i\|^2_{Q_{yy}^{-1}}\right), \\ &\forall i = 1,2,\ldots,k \end{aligned} \tag{D.6}$$

The integral at the numerator of Equation D.5 can be developed as ($\forall i = 1,2,\ldots,k$):

$$f_{\underline{H},\underline{y}} = \int_{x,\nabla} f_{\underline{x},\underline{\nabla}_i,\underline{H},\underline{y}} \mathrm{d}x\mathrm{d}\nabla =$$

$$\begin{aligned} &p_i c_x^n c_\nabla^{q_i} \int_{x,\nabla_i} \frac{1}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}\|y - Ax - C_i\nabla_i\|^2_{Q_{yy}^{-1}}\right) \mathrm{d}x\mathrm{d}\nabla = \\ &p_i c_x^n c_\nabla^{q_i} \frac{1}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}}} \int_{x,\nabla_i} \exp\left(-\tfrac{1}{2}\|y - \hat{y}_i\|^2_{Q_{yy}^{-1}} - \tfrac{1}{2}\|\hat{y}_i - Ax - C_i\nabla_i\|^2_{Q_{yy}^{-1}}\right) \mathrm{d}x\mathrm{d}\nabla = \\ &p_i c_x^n c_\nabla^{q_i} \frac{1}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}\|y - \hat{y}_i\|^2_{Q_{yy}^{-1}}\right) \int_{x,\nabla_i} \exp\left(-\tfrac{1}{2}\|\hat{y}_i - Ax - C_i\nabla_i\|^2_{Q_{yy}^{-1}}\right) \mathrm{d}x\mathrm{d}\nabla = \\ &p_i c_x^n c_\nabla^{q_i} \frac{1}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}\|y - \hat{y}_i\|^2_{Q_{yy}^{-1}}\right) \cdot (2\pi)^{\frac{n+q_i}{2}} |Q_{\hat{x}'_i,\hat{x}'_i}|^{\frac{1}{2}} = \end{aligned}$$

$$p_i c_x^n c_\nabla^{q_i} (2\pi)^{\frac{n+q_i-m}{2}} \left(\frac{|Q_{\hat{x}'_i,\hat{x}'_i}|}{|Q_{yy}|}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\|y - \hat{y}_i\|^2_{Q_{yy}^{-1}}\right) \tag{D.7}$$

where $\hat{y}_i$ is the estimated measurement vector under hypothesis H_i and $\hat{x}'_i = \begin{bmatrix} \hat{x} \\ \hat{\nabla}_i \end{bmatrix}$. Note that we employed the fact $c_x, c_\nabla \to 0$ when computing the integral between $-\frac{1}{2c_x}$ and $+\frac{1}{2c_x}$ for each component of x, and between $-\frac{1}{2c_\nabla}$ and $+\frac{1}{2c_\nabla}$ for each component of ∇_i:

$$\int_{-\frac{1}{2c_\nabla}}^{+\frac{1}{2c_\nabla}} \int_{-\frac{1}{2c_x}}^{+\frac{1}{2c_x}} \exp\left(-\frac{1}{2}\|\hat{y}_i - Ax - C_i\nabla_i\|^2_{Q_{yy}^{-1}}\right) \mathrm{d}x\mathrm{d}\nabla \to (2\pi)^{\frac{n+q_i}{2}} |Q_{\hat{x}'_i,\hat{x}'_i}|^{\frac{1}{2}}$$

In case of adoption of the prior in Equation (D.4), there is no need to make any approximation for the parameter x, since the integration is from $-\infty$ to $+\infty$. The denominator of

Equation D.5 is simply the sum over the different i of the terms at the numerator above derived. Therefore Equation D.5 can be written:

$$f_{\underline{H}|y}(H|y) = \frac{p_i c_\nabla^{q_i} (2\pi)^{\frac{q_i}{2}} |Q_{\hat{x}'_i,\hat{x}'_i}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\|y-\hat{y}_i\|^2_{Q_{yy}^{-1}}\right)}{\sum_{t=1}^{k} p_t c_\nabla^{q_t} (2\pi)^{\frac{q_t}{2}} |Q_{\hat{x}'_t,\hat{x}'_t}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\|y-\hat{y}_t\|^2_{Q_{yy}^{-1}}\right)}, \; i = 1,2,\ldots,k \tag{D.8}$$

Furthermore we have:

$$Q^{-1}_{\hat{x}'_i,\hat{x}'_i} = \begin{bmatrix} A^T \\ C_i^T \end{bmatrix} Q_{yy}^{-1} [A \; C_i] = \begin{bmatrix} A^T Q_{yy}^{-1} A & A^T Q_{yy}^{-1} C_i \\ C_i^T Q_{yy}^{-1} A & C_i^T Q_{yy}^{-1} C_i \end{bmatrix} \tag{D.9}$$

from which:

$$|Q_{\hat{x}'_i,\hat{x}'_i}|^{-1} = |A^T Q_{yy}^{-1} A| \cdot |C_i^T Q_{yy}^{-1} C_i - C_i^T Q_{yy}^{-1} A (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} C_i| \tag{D.10}$$

so that in the fraction in Equation D.8 the term $|A^T Q_{yy}^{-1} A|$ can be put in evidence and simplified between numerator and denominator, and the remaining term can be reckognized as the determinant of:

$$C_i^T Q_{yy}^{-1} C_i - C_i^T Q_{yy}^{-1} A (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} C_i = \bar{C}_i^T Q_{yy}^{-1} \bar{C}_i = Q^{-1}_{\hat{\nabla}_i \hat{\nabla}_i} \tag{D.11}$$

with $\bar{C}_i = P_A^{\perp} C_i$. On the other hand also the term $\|y-\hat{y}_i\|^2_{Q_{yy}^{-1}}$, squared norm of the projection of y over $R(A \; C_i)^{\perp}$, can be rewritten to highlight the effect of the extension of $R(A)$ by means of C_i:

$$\|y-\hat{y}_i\|^2_{Q_{yy}^{-1}} = \|y-\hat{y}_0\|^2_{Q_{yy}^{-1}} - \|\hat{y}_i-\hat{y}_0\|^2_{Q_{yy}^{-1}} \tag{D.12}$$

where $\hat{y}_0$ is the estimated measurement vector with the model $E(\underline{y}) = Ax$ (that is $\hat{y}_0 = P_A y$). The term $\|y-\hat{y}_0\|^2_{Q_{yy}^{-1}}$ can be simplified between numerator and denominator of Equation D.8, whereas the term $\|\hat{y}_i-\hat{y}_0\|^2_{Q_{yy}^{-1}}$ can be reckognized as the standard UMPI test statistic, and can be written in the equivalent form $\|\hat{\nabla}_i\|^2_{Q_{\hat{\nabla}_i\hat{\nabla}_i}}$.

Finally Equation D.8 can be rewritten in the form:

$$f_{\underline{H}|y}(H|y) = \frac{p_i w_i}{\sum_{t=1}^{k} p_t w_t} \tag{D.13}$$

with:

$$w_i = \left(\frac{1}{c_\nabla^{q_i} (2\pi)^{\frac{q_i}{2}} |Q_{\hat{\nabla}_i\hat{\nabla}_i}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\|\hat{\nabla}_i\|^2_{Q_{\hat{\nabla}_i\hat{\nabla}_i}}\right) \right)^{-1} \tag{D.14}$$

which in case of hypotheses characterized by equal q_i simplifies to:

$$w_i = \left(\frac{1}{|Q_{\hat{\nabla}_i\hat{\nabla}_i}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\|\hat{\nabla}_i\|^2_{Q_{\hat{\nabla}_i\hat{\nabla}_i}}\right) \right)^{-1} \tag{D.15}$$

D.2. GAUSSIAN PRIORS

Let us assume alternatively the following priors for the unknown parameters:

$$\begin{array}{l} \underline{H} \sim P(H = H_i) = p_i \ , \ i = 1,2,\ldots,k \\ \underline{x} \sim N(x_0, Q_{xx}) \\ \underline{\nabla}_i | H \sim \left\{ \begin{array}{l} N(\nabla_{0,i}, Q_{\nabla_i \nabla_i}) \ \text{ for } \underline{H} = H_i \\ 0 \ \text{ for } \underline{H} \neq H_i \end{array} \right. \end{array} \tag{D.16}$$

The joint distribution then writes:

$$\begin{array}{l} f_{\underline{x},\underline{\nabla}_i,\underline{H},\underline{y}} = p_i \frac{1}{(2\pi)^{\frac{m+n+q_i}{2}} |Q_{yy}|^{\frac{1}{2}} |Q_{xx}|^{\frac{1}{2}} |Q_{\nabla_i \nabla_i}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\|y - Ax - C_i\nabla_i\|^2_{Q^{-1}_{yy}}\right) \cdot \\ \exp\left(-\frac{1}{2}\|x - x_0\|^2_{Q^{-1}_{xx}}\right) \exp\left(-\frac{1}{2}\|\nabla_i - \nabla_{0,i}\|^2_{Q^{-1}_{\nabla_i \nabla_i}}\right) , \forall i = 1,2,\ldots,k \end{array} \tag{D.17}$$

Define:

$$x'_i = \left[\begin{array}{c} x \\ \nabla_i \end{array} \right] \ , \ x'_{i0} = \left[\begin{array}{c} x_0 \\ \nabla_{i0} \end{array} \right] \ , \ A'_i = [A \ C_i]$$

The joint distribution can thus be written:

$$\begin{array}{c} f_{\underline{x},\underline{\nabla}_i,\underline{H},\underline{y}} = f_{\underline{x}'_i,\underline{H},\underline{y}} = p_i \frac{1}{(2\pi)^{\frac{m+n+q_i}{2}} |Q_{yy}|^{\frac{1}{2}} |Q_{x'_i x'_i}|^{\frac{1}{2}}} \cdot \\ \exp(-\frac{1}{2}((y - A'_i x'_i)^T Q^{-1}_{yy}(y - A'_i x'_i) + (x'_i - x'_{i0})^T Q^{-1}_{x'_i x'_i}(x'_i - x'_{i0}))) \ , \ \forall i = 1,2,\ldots,k \end{array} \tag{D.18}$$

If we define:

$$\mu_i = (A'^T_i Q^{-1}_{yy} A'_i + Q^{-1}_{x'_i x'_i})^{-1} (A'^T_i Q^{-1}_{yy} y + Q^{-1}_{x'_i x'_i} x'_{i0}) \tag{D.19}$$

then the argument of the exponential in the above joint distribution can be developed as:

$$\begin{array}{c} -\frac{1}{2}(y^T Q^{-1}_{yy} y + x'^T_{i0} Q^{-1}_{x'_i x'_i} x'_{i0} - 2x'^T_i (A'^T_i Q^{-1}_{yy} y + Q^{-1}_{x'_i x'_i} x'_{i0}) + x'^T_i (A'^T_i Q^{-1}_{yy} A'_i + Q^{-1}_{x'_i x'_i}) x'_i) = \\ -\frac{1}{2}(y^T Q^{-1}_{yy} y + x'^T_{i0} Q^{-1}_{x'_i x'_i} x'_{i0} - \mu^T_i (A'^T_i Q^{-1}_{yy} A'_i + Q^{-1}_{x'_i x'_i}) \mu_i + \\ (x'_i - \mu_i)^T (A'^T_i Q^{-1}_{yy} A'_i + Q^{-1}_{x'_i x'_i})(x'_i - \mu_i)) \end{array}$$

The numerator of the posterior distribution in Equation D.5 is obtained taking the integral over x'_i of the joint distribution in Equation D.18. As a result we have:

$$\begin{array}{c} f_{\underline{H},\underline{y}} = p_i \frac{1}{(2\pi)^{\frac{m+n+q_i}{2}} |Q_{yy}|^{\frac{1}{2}} |Q_{x'_i x'_i}|^{\frac{1}{2}}} \cdot \\ \exp(-\frac{1}{2}(y^T Q^{-1}_{yy} y + x'^T_{i0} Q^{-1}_{x'_i x'_i} x'_{i0} - \mu^T_i (A'^T_i Q^{-1}_{yy} A'_i + Q^{-1}_{x'_i x'_i}) \mu_i)) \cdot \\ \int_{x'_i} \exp(-\frac{1}{2}(x'_i - \mu_i)^T (A'^T_i Q^{-1}_{yy} A'_i + Q^{-1}_{x'_i x'_i})(x'_i - \mu_i)) \mathrm{d}x'_i = \\ p_i \frac{|A'^T_i Q^{-1}_{yy} A'_i + Q^{-1}_{x'_i x'_i}|^{-\frac{1}{2}}}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}} |Q_{x'_i x'_i}|^{\frac{1}{2}}} \cdot \\ \exp(-\frac{1}{2}(y^T Q^{-1}_{yy} y + x'^T_{i0} Q^{-1}_{x'_i x'_i} x'_{i0} - \mu^T_i (A'^T_i Q^{-1}_{yy} A'_i + Q^{-1}_{x'_i x'_i}) \mu_i)) \end{array} \tag{D.20}$$

where we employed the fact that the integral at the third line is equal to $(2\pi)^{\frac{n+q_i}{2}} |A_i'^T Q_{yy}^{-1} A_i' + Q_{x_i' x_i'}^{-1}|^{-\frac{1}{2}}$. For the posterior $f_{\underline{H}|y}(H|y) = \frac{f_{\underline{H},\underline{y}}}{f_{\underline{y}}}$ we thus have:

$$f_{\underline{H}|y} = \frac{p_i \frac{|A_i'^T Q_{yy}^{-1} A_i' + Q_{x_i' x_i'}^{-1}|^{-\frac{1}{2}}}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}} |Q_{x_i' x_i'}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(y^T Q_{yy}^{-1} y + x_{i0}'^T Q_{x_i' x_i'}^{-1} x_{i0}' - \mu_i^T (A_i'^T Q_{yy}^{-1} A_i' + Q_{x_i' x_i'}^{-1})\mu_i))}{\sum_{t=1}^{k} p_t \frac{|A_t'^T Q_{yy}^{-1} A_t' + Q_{x_t' x_t'}^{-1}|^{-\frac{1}{2}}}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}} |Q_{x_t' x_t'}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(y^T Q_{yy}^{-1} y + x_{t0}'^T Q_{x_t' x_t'}^{-1} x_{t0}' - \mu_t^T (A_t'^T Q_{yy}^{-1} A_t' + Q_{x_t' x_t'}^{-1})\mu_t))}$$

This expression can be simplified in case each hypothesis is characterized by equal $Q_{x_i' x_i'}$ (which means equal $Q_{\nabla_i \nabla_i}$, since $Q_{x_i' x_i'}$ is a block-diagonal matrix). This assumption can be reasonable in case the different hypotheses are characterized by the same number of unknown parameters, q_i. In fact, if we suppose that each hypothesis is characterized by equal $Q_{x_i' x_i'}$, the term $|Q_{x_i' x_i'}|^{\frac{1}{2}}$ at the denominator of $f_{\underline{H},\underline{y}}$ simplifies between numerator and denominator of the expression above:

$$f_{\underline{H}|y} = \frac{p_i \frac{|A_i'^T Q_{yy}^{-1} A_i' + Q_{x_i' x_i'}^{-1}|^{-\frac{1}{2}}}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(y^T Q_{yy}^{-1} y + x_{i0}'^T Q_{x_i' x_i'}^{-1} x_{i0}' - \mu_i^T (A_i'^T Q_{yy}^{-1} A_i' + Q_{x_i' x_i'}^{-1})\mu_i))}{\sum_{t=1}^{k} p_t \frac{|A_t'^T Q_{yy}^{-1} A_t' + Q_{x_t' x_t'}^{-1}|^{-\frac{1}{2}}}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(y^T Q_{yy}^{-1} y + x_{t0}'^T Q_{x_t' x_t'}^{-1} x_{t0}' - \mu_t^T (A_t'^T Q_{yy}^{-1} A_t' + Q_{x_t' x_t'}^{-1})\mu_t))}$$

Taking now the limit $Q_{x_i' x_i'} \to +\infty$ we have:

$$\begin{aligned}
&A_i'^T Q_{yy}^{-1} A_i' + Q_{x_i' x_i'}^{-1} \to A_i'^T Q_{yy}^{-1} A_i' = Q_{\hat{x}_i' \hat{x}_i'}^{-1} \\
&\mu_i \to (A_i'^T Q_{yy}^{-1} A_i')^{-1} A_i'^T Q_{yy}^{-1} y = \hat{x}_i' \\
&x_{i0}'^T Q_{x_i' x_i'}^{-1} x_{i0}' \to 0
\end{aligned}$$

As a result the term at the numerator (and each term of the sum at the denominator) becomes:

$$\begin{aligned}
&p_i \frac{|Q_{\hat{x}_i' \hat{x}_i'}|^{\frac{1}{2}}}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(y^T Q_{yy}^{-1} y - \hat{x}_i'^T (A_i'^T Q_{yy}^{-1} A_i')\hat{x}_i')) = \\
&p_i \frac{1}{(2\pi)^{\frac{m}{2}}} \left(\frac{|Q_{\hat{x}_i' \hat{x}_i'}|}{|Q_{yy}|}\right)^{\frac{1}{2}} \exp(-\frac{1}{2}(y^T Q_{yy}^{-1} y - \hat{y}_i^T Q_{yy}^{-1} \hat{y}_i))
\end{aligned}$$

But this is the same as Equation D.7 except for the multiplicative constants $(2\pi)^{\frac{n+q_i}{2}}$ and $c_x^n c_\nabla^{q_i}$. Thus the posterior $f_{\underline{H}|y}(H|y)$ becomes:

$$f_{\underline{H}|y}(H|y) = \frac{p_i w_i}{\sum_{t=1}^{k} p_t w_t} \tag{D.21}$$

with:

$$w_i = \left(\frac{1}{|Q_{\hat{\nabla}_i \hat{\nabla}_i}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\|\hat{\nabla}_i\|^2_{Q_{\hat{\nabla}_i \hat{\nabla}_i}}\right)\right)^{-1} \tag{D.22}$$

which coincides with Equation D.15.

In case instead each hypothesis is *not* characterized by equal $Q_{\nabla_i \nabla_i}$, the expression for $f_{\underline{H}|y}(H|y)$ does not simplify. If we assume for instance $Q_{\nabla_i \nabla_i} = \sigma^2 I_{q_i}$, i.e. equal variance for

each component of ∇_i, then $|Q_{\nabla_i\nabla_i}| = \sigma^{2q_i}$. When we take the limit $\sigma \to +\infty$ in the posterior $f_{\underline{H}|y}(H|y)$, we will have that the posterior probability for hypotheses characterized by q_i larger than others tends to zero. This result suggests that the procedure of taking the limit $Q_{\nabla_i\nabla_i} \to +\infty$ does not work properly when comparing hypotheses with different number of degrees of freedom q_i. This is an expected result, because when we take the limit $\sigma \to +\infty$, the prior distribution for the parameter $\underline{\nabla}_i$ becomes an *improper* prior. The use of improper priors is not allowed when testing hypotheses with different number of degrees of freedom, as this is equivalent to testing a point null hypothesis, and improper priors are not allowed in such context (see for instance [7]).

D.3. SINGLE POINT PRIOR

Let us assume alternatively the following priors for the unknown parameters:

$$\begin{aligned} &\underline{H} \sim P(H = H_i) = p_i \ , \ i = 1,2,\dots,k \\ &\underline{x} \sim 1 \text{ (uniform non-informative prior)} \\ &\underline{\nabla}_i|H \sim \begin{cases} 1 & \text{for } \nabla_i = \hat{\nabla}_i \text{ and } \underline{H} = H_i \\ 0 & \text{for } \nabla_i \neq \hat{\nabla}_i \text{ or } \underline{H} \neq H_i \end{cases} \end{aligned} \tag{D.23}$$

where $\hat{\nabla}_i$ is the ML estimate of ∇_i under the alternative hypothesis H_i. The prior for each $\underline{\nabla}_i$ is a single point priors, with all the probability mass concentrated in a specific value $\hat{\nabla}_i$. Furthermore this prior depends on the actual observation y, since $\hat{\nabla}_i$ is a function of the observation y. In (D.23) the single point prior are expressed as a PMF, an equivalent notation as PDF can be:

$$\underline{\nabla}_i|H \sim \begin{cases} \delta(\nabla_i - \hat{\nabla}_i) \text{ for } H = H_i \\ 0 \text{ for } H \neq H_i \end{cases} \tag{D.24}$$

where δ stands for the Dirac function. The joint distribution then writes:

$$\begin{aligned} &f_{\underline{x},\underline{\nabla}_i,\underline{H},\underline{y}} = p_i \frac{1}{(2\pi)^{\frac{m}{2}}|Q_{yy}|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}\|y - Ax - C_i\nabla_i\|^2_{Q_{yy}^{-1}}\right)\delta(\nabla_i - \hat{\nabla}_i) \\ &\forall i = 1,2,\dots,k \end{aligned} \tag{D.25}$$

Equation (D.5) is employed to compute the posterior probability for the occurrence of the hypotheses, $\underline{H}|y$. The numerator can be developed as ($\forall i = 1,2,\dots,k$):

$$f_{\underline{H},\underline{y}} = \int_{x,\nabla} f_{\underline{x},\underline{\nabla}_i,\underline{H},\underline{y}} \mathrm{d}x\mathrm{d}\nabla =$$

$$\begin{aligned} &p_i \int_{x,\nabla} \frac{1}{(2\pi)^{\frac{m}{2}}|Q_{yy}|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}\|y - Ax - C_i\nabla_i\|^2_{Q_{yy}^{-1}}\right)\delta(\nabla_i - \hat{\nabla}_i)\mathrm{d}x\mathrm{d}\nabla = \\ &p_i \int_{x} \frac{1}{(2\pi)^{\frac{m}{2}}|Q_{yy}|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}\|y - Ax - C_i\hat{\nabla}_i\|^2_{Q_{yy}^{-1}}\right)\mathrm{d}x = \\ &p_i \frac{1}{(2\pi)^{\frac{m}{2}}|Q_{yy}|^{\frac{1}{2}}} \int_{x} \exp\left(-\tfrac{1}{2}\|y - \hat{y}_i\|^2_{Q_{yy}^{-1}} - \tfrac{1}{2}\|\hat{y}_i - C_i\hat{\nabla}_i - Ax\|^2_{Q_{yy}^{-1}}\right)\mathrm{d}x = \\ &p_i \frac{1}{(2\pi)^{\frac{m}{2}}|Q_{yy}|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}\|y - \hat{y}_i\|^2_{Q_{yy}^{-1}}\right) \int_{x} \exp\left(-\tfrac{1}{2}\|\hat{y}_i - C_i\hat{\nabla}_i - Ax\|^2_{Q_{yy}^{-1}}\right)\mathrm{d}x = \\ &p_i \frac{1}{(2\pi)^{\frac{m}{2}}|Q_{yy}|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}\|y - \hat{y}_i\|^2_{Q_{yy}^{-1}}\right) \cdot (2\pi)^{\frac{n}{2}}|Q_{\hat{x}_0,\hat{x}_0}|^{\frac{1}{2}} = \end{aligned}$$

$$p_i(2\pi)^{\frac{n-m}{2}} \left(\frac{|Q_{\hat{x}_0,\hat{x}_0}|}{|Q_{yy}|}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\|y - \hat{y}_i\|^2_{Q_{yy}^{-1}}\right) \tag{D.26}$$

where $\hat{y}_i$ is the (ML) estimation of the observation vector y under hypothesis H_i and $Q_{\hat{x}_0,\hat{x}_0} = (A^T Q_{yy}^{-1} A)^{-1}$. Therefore Equation (D.5) can be written:

$$f_{\underline{H}|y}(H|y) = \frac{p_i \exp\left(-\frac{1}{2}\|y-\hat{y}_i\|^2_{Q_{yy}^{-1}}\right)}{\sum_{t=1}^{k} p_t \exp\left(-\frac{1}{2}\|y-\hat{y}_t\|^2_{Q_{yy}^{-1}}\right)} \tag{D.27}$$

that can be rewritten in the form:

$$f_{\underline{H}|y}(H|y) = \frac{p_i w_i}{\sum_{t=1}^{k} p_t w_t} \tag{D.28}$$

with:

$$w_i = \exp\left(\frac{1}{2}\|\hat{\nabla}_i\|^2_{Q_{\hat{\nabla}_i\hat{\nabla}_i}}\right) \tag{D.29}$$

This result is valid also comparing hypotheses with different q_i. We can compute as well the posterior distribution of the variable $\underline{x}$ from:

$$f_{\underline{x}|y}(x|y) = \frac{f_{\underline{x},\underline{y}}}{f_{\underline{y}}} = \frac{\sum_H \int_\nabla f_{\underline{x},\underline{\nabla}_i,\underline{H},\underline{y}} \mathrm{d}\nabla}{\sum_H \int_{x,\nabla} f_{\underline{x},\underline{\nabla}_i,\underline{H},\underline{y}} \mathrm{d}x\mathrm{d}\nabla} \tag{D.30}$$

The numerator of Equation (D.30) can be computed as:

$$\sum_H \int_{\nabla_i} f_{\underline{x},\underline{\nabla}_i,\underline{H},\underline{y}} \mathrm{d}\nabla_i = \sum_{i=1}^{k} p_i \frac{1}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\|y-Ax-C_i\hat{\nabla}_i\|^2_{Q_{yy}^{-1}}\right) \tag{D.31}$$

and, given that the denominator is a summation over H of the terms in Equation (D.26), the posterior can be obtained:

$$f_{\underline{x}|y}(x|y) = \frac{\sum_{i=1}^{k} p_i \frac{1}{(2\pi)^{\frac{m}{2}} |Q_{yy}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\|y-Ax-C_i\hat{\nabla}_i\|^2_{Q_{yy}^{-1}}\right)}{\sum_{i=1}^{k} p_i (2\pi)^{\frac{n-m}{2}} \left(\frac{|Q_{\hat{x},\hat{x}}|}{|Q_{yy}|}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2}\|y-\hat{y}_i\|^2_{Q_{yy}^{-1}}\right)} = \frac{\sum_{i=1}^{k} p_i \exp\left(-\frac{1}{2}\|y-\hat{y}_i\|^2_{Q_{yy}^{-1}}\right) \exp\left(-\frac{1}{2}\|\hat{y}_i-Ax-C_i\hat{\nabla}_i\|^2_{Q_{yy}^{-1}}\right)}{\sum_{i=1}^{k} p_i (2\pi)^{\frac{n}{2}} |Q_{\hat{x},\hat{x}}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\|y-\hat{y}_i\|^2_{Q_{yy}^{-1}}\right)} \tag{D.32}$$

Noticing that the terms

$$\frac{1}{(2\pi)^{\frac{n}{2}} |Q_{\hat{x},\hat{x}}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\|\hat{y}_i - Ax - C_i\hat{\nabla}_i\|^2_{Q_{yy}^{-1}}\right) = \frac{1}{(2\pi)^{\frac{n}{2}} |Q_{\hat{x},\hat{x}}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\|A\hat{x}_i - Ax\|^2_{Q_{yy}^{-1}}\right)$$

are Gaussian distributions for the variable $\underline{x}$, of the form $N(\hat{x}_i, Q_{\hat{x}_0,\hat{x}_0})$, with $\hat{x}_i = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}(y - C_i\hat{\nabla}_i)$, the posterior distribution can be written as:

$$f_{\underline{x}|y}(x|y) = \sum_{i=1}^{k} \frac{p_i \exp\left(\frac{1}{2}\|\hat{\nabla}_i\|^2_{Q_{\hat{\nabla}_i\hat{\nabla}_i}}\right)}{\sum_{t=1}^{k} p_t \exp\left(\frac{1}{2}\|\hat{\nabla}_t\|^2_{Q_{\hat{\nabla}_t\hat{\nabla}_t}}\right)} N(\hat{x}_i, Q_{\hat{x}_0,\hat{x}_0}) \tag{D.33}$$

that is a multimodal Gaussian distribution, in which each single Gaussian is weighted by the term:

$$f(H_i|y) = \frac{p_i \exp\left(\frac{1}{2}\|\hat{\nabla}_i\|^2_{Q_{\hat{\nabla}_i\hat{\nabla}_i}}\right)}{\sum_{t=1}^{k} p_t \exp\left(\frac{1}{2}\|\hat{\nabla}_t\|^2_{Q_{\hat{\nabla}_t\hat{\nabla}_t}}\right)}$$

i.e. the posterior probability of occurrence of hypothesis H_i as in Equation (D.28). This means:

$$f_{\underline{x}|y}(x|y) = \sum_{i=1}^{k} f(H_i|y) N(\hat{x}_i, Q_{\hat{x}_0,\hat{x}_0}) \tag{D.34}$$

The use of the type of priors proposed in this section (single point priors) tends to favor the hypotheses with largest q_i, since the full prior probability is assigned to the ML estimate of ∇_i, which is the most likely value to occur given the measurements. With reference to Equation (D.28) and (D.29), if two nested hypotheses with different q_i have the same prior probability p_i, the hypothesis with larger q_i will always have larger posterior probability of occurrence. I considered adopting this type of prior in RAIM because it maximizes the posterior probability of occurrence of an outlier (for instance, in case of testing the point null hypothesis H_0 against any H_a). In [7], this prior is adopted to maximize the posterior probability of occurrence of the alternative hypothesis in the testing of point null hypothesis problem.

D.4. CONSIDERATIONS ON THE PRIOR DISTRIBUTIONS

The result of previous Section may leave us a little puzzled, wondering why the posterior distribution of $\underline{x}|y$ is a weighted sum of the Gaussian distributions $N(\hat{x}_i, Q_{\hat{x}_0,\hat{x}_0})$ and not $N(\hat{x}_i, Q_{\hat{x}_i,\hat{x}_i})$. This is due to the use of the single point prior, that does not leave room to the uncertainty on ∇_i. Intuitively, we would like to have as result a posterior distribution for the occurrence of the hypotheses as the one obtained in previous section, Equations (D.28) and (D.29), but a posterior distribution for $\underline{x}|y$ of the form:

$$f_{\underline{x}|y}(x|y) = \sum_{i=1}^{k} f(H_i|y) N(\hat{x}_i, Q_{\hat{x}_i,\hat{x}_i}) \tag{D.35}$$

instead of the one in Equation (D.34).

This posterior distribution can in fact be obtained adopting as prior for $\underline{\nabla}_i$ the following:

$$\underline{\nabla}_i \sim \frac{1}{(2\pi)^{\frac{q}{2}} |Q_{\hat{\nabla}_i\hat{\nabla}_i}|^{\frac{1}{2}}} = const \tag{D.36}$$

which is a non-informative prior, since it is just a constant. This result can be readily verified substituting this expression to $c_{\nabla}^{q_i}$ in Equation D.14. The use of such prior therefore results in preserving the unknown nature of the bias size ∇_i and at the same time maximizing the posterior probability of occurrence of higher dimensional alternative hypotheses, as much as the use of the single point prior presented in previous section (since the posterior $f_{\underline{H}|y}(H|y)$ is the same).

E

SUBSET SELECTION METHODS

This Appendix provides a coincise description of the most popular Subset Selection methods employed in linear regression. The concept of Subset Selection was introduced in Section 3.6.3, and the main methods were mentioned. Subset Selection in fact deals with the problem of choosing a subset of parameters (from an original pool of candidates) that is able to describe 'thoroughly' a certain phenomenon/observation. This subset should include the parameters that are actually involved in the physical process behind the observation taking, parameters that should be included in the model that describes the physical process. Consider the model in Equation (2.2) (or the model in Equation (3.27)): the problem of choosing how many and which parameters to include in vector x (or ∇) is a Subset Selection model. Choosing between the two hypothesis H_0 and H_a in Equation (3.21) can be treated as a Subset Selection problem, since there is to choose between two different subsets of regression parameters.

E.1. SPJØTVOLL METHOD

Let us consider for the moment the simplified problem with $Q_{yy} = \sigma^2 I_m$. Suppose we want to compare two subsets of parameters, composed of respectively q_1 and q_2 numbers of parameters, to which correspond the geometry matrices A_1 and A_2. The model to consider is the one in Equation (3.27), A_1 and A_2 being two competitive C_y among which to choose. Define instead A_t the actual (unknown) $m \times q_t$ geometry matrix linking measurements to parameters (the actual C_y), to be found among different alternatives (of which A_1 and A_2 are two possible choices). For the expectation of the measurements (that we will denote with η) it holds:

$$E(\underline{y}) = \eta = A_t x_t$$

where A_t and x_t are the *actual* unknown geometry matrix and regression parameters. If $\underline{\hat{x}}_t$ is the estimator related to the actual geometry matrix, $\underline{\hat{x}}_t = (A_t^T A_t)^{-1} A_t^T \underline{y}$, we have:

$$\frac{(\underline{\hat{x}}_t - x_t)^T A_t^T A_t (\underline{\hat{x}}_t - x_t)}{q_t \underline{\hat{\sigma}}^2} \sim F(q_t, m - q_t) \tag{E.1}$$

for the case of unknown σ (case described in [94]), and:

$$\frac{1}{\sigma^2}(\underline{\hat{x}}_t - x_t)^T A_t^T A_t (\underline{\hat{x}}_t - x_t) \sim \chi^2(q_t, 0)$$

in case of σ known. When instead of using A_t we use the most comprehensive geometry matrix that we can consider in our problem of finding the best fitting regression function for $\underline{y}$ among a pool of candidates, A (in the special case considered it can be $A = [A_1\ A_2]$), with a corresponding $\underline{\hat{x}}$ estimator, it can be proven (see [90]) that:

$$P(\frac{(\underline{\hat{x}} - x)^T A^T A(\underline{\hat{x}} - x)}{q\underline{\hat{\sigma}}^2} \leq F_\alpha(q, m-q)) \geq 1 - \alpha \tag{E.2}$$

where $x = E(\underline{\hat{x}})$ and $F_\alpha(q, m-q)$ is the value of the F distribution corresponding to the significance α. That is, the error distribution when using a set of parameters larger than the actual one, but still including the actual one, is overbound by an F distribution with all q degrees of freedom. For the case σ known we have similarly:

$$P(\frac{(\underline{\hat{x}} - x)^T A^T A(\underline{\hat{x}} - x)}{\sigma^2} \leq \chi^2_\alpha(q, 0)) \geq 1 - \alpha$$

When we employ the regression matrix A_1, we come up with a measurement estimator $\underline{\hat{y}} = A_1\underline{\hat{x}}_1$, which expectation is $\eta_1 = E(\underline{\hat{y}})$. The Spjøtvoll method measures the goodness of fit of a regression function (say $A_1 x_1$) computing the quantity:

$$(\eta - \eta_1)^T(\eta - \eta_1)$$

i.e. the squared length of the difference between η and η_1. Since $\eta_1 = A_1(A_1^T A_1)^{-1}A_1\eta$, the quantity in the expression above can be computed as:

$$\eta^T A_1^T (A_1^T A_1)^{-1} A_1 \eta = \|P_{A_1}\eta\|^2$$

which is the squared length of the projection of η on $R(A_1)$ (the subspace spanned by matrix A_1). The larger the value the better the fit. The difference of goodness between two regression functions (e.g. described by $A_1 x_1$ and $A_2 x_2$) can be therefore measured by the quantity:

$$\begin{gathered}\eta^T[A_1(A_1^T A_1)^{-1}A_1^T - A_2(A_2^T A_2)^{-1}A_2^T]\eta = \\ x^T A^T[A_1(A_1^T A_1)^{-1}A_1^T - A_2(A_2^T A_2)^{-1}A_2^T]Ax\end{gathered} \tag{E.3}$$

If this quantity is bigger than zero the subset A_1 performs better than subset A_2, therefore our aim is to infer on its value. Since $\eta = A_t x_t$ is unknown, also this quantity is unknown: to infer whether it is larger or smaller than zero, we can estimate a confidence region for it. Having a confidence region means having a lower and upper bound inside which the unknown is expected to lie with a determined confidence (α): in our case, if both bounds are positive, we can infer with the relative confidence α that the quantity is positive as well.

A confidence region for the quantity in Equation (E.3) can be defined exploiting the relation in Equation (E.2). Spjøtvoll derived this confidence region employing the Scheffé multiple comparison method. If both the bounds of this confidence region are bigger than zero, then we can declare with the corresponding significance level that the quantity in Equation (E.3) is bigger than zero and therefore the subset A_1 is significantly better than A_2. The bounds of the confidence region are derived in [67] and [94], here is shown a part of the proof, to be complemented by the results obtained in [32].

As we have said the quantity $(\underline{\hat{x}} - x)^T A^T A(\underline{\hat{x}} - x)$ is bounded by an F-distribution, as in Equation (E.2). Now, given this bound, we want to derive the corresponding bound holding

for the expression in Equation (E.3). Define $C \equiv [A_1(A_1^T A_1)^{-1}A_1^T - A_2(A_2^T A_2)^{-1}A_2^T]$. The idea is to find confidence bounds for all the quadratic functions $x^T Cx$, where C varies over all symmetric matrices with real elements, while Equation (E.2) holds. This means for any C we want to find the set of all possible values of $x^T Cx$ when x is contained in the ellipsoid $(\underline{\hat{x}} - x)^T A^T A(\underline{\hat{x}} - x) \leq q\underline{\hat{\sigma}}^2 F_\alpha(q, m-q)$. Any given symmetric C matrix can be diagonalized by a $q \times q$ matrix P such that $P^T R_{xx}^{-1} P = I$ (with $R_{xx}^{-1} = A^T A$) and $P^T CP = D$, with D diagonal with diagonal elements $d_1, d_2, \ldots, d_q$. If we define the vector $\underline{\hat{\gamma}} = P\underline{\hat{x}}$, our problem is equivalent to finding the set of values $\gamma^T D\gamma$ can possibly assume when γ is in the sphere $(\gamma - \underline{\hat{\gamma}})^T(\gamma - \underline{\hat{\gamma}}) \leq q\hat{\sigma}^2 F_\alpha$. Given $\underline{\lambda}_{min}$ and $\underline{\lambda}_{max}$ the minimum and maximum roots, respectively, of

$$\sum_{i=1}^{q} d_i^2 \underline{\hat{\gamma}}_i^2 / (d_i - \underline{\lambda})^2 = q\underline{\hat{\sigma}}^2 F_\alpha(q, m-q)$$

and let $\underline{a} = -\min(\min_i d_i, \underline{\lambda}_{min})$ and $\underline{b} = \max(\max_i d_i, \underline{\lambda}_{max})$, the final result is that $x^T Cx$ is bound with probability at least $1 - \alpha$ by:

$$\underline{K}_1 \leq x^T Cx \leq \underline{K}_2 \tag{E.4}$$

where

$$\underline{K}_1 = \underline{a} \sum_{i=1}^{q} d_i \underline{\hat{\gamma}}_i^2 / (\underline{a} + d_i) - q\underline{\hat{\sigma}}^2 F_\alpha(q, m-q) \tag{E.5}$$

and

$$\underline{K}_2 = \underline{b} \sum_{i=1}^{q} d_i \underline{\hat{\gamma}}_i^2 / (\underline{b}b - d_i) + q\underline{\hat{\sigma}}^2 F_\alpha(q, m-q) \tag{E.6}$$

with exception of the following cases:

$$\underline{K}_1 = 0 \text{ if all } d_i \geq 0 \text{ and } \sum_{i=1}^{q} \underline{\hat{\gamma}}_i^2 \leq q\underline{\hat{\sigma}}^2 F_\alpha(q, m-q), \tag{E.7}$$

and

$$\underline{K}_2 = 0 \text{ if all } d_i \leq 0 \text{ and } \sum_{i=1}^{q} \underline{\hat{\gamma}}_i^2 \leq q\underline{\hat{\sigma}}^2 F_\alpha(q, m-q) \tag{E.8}$$

If both computed K_1 and K_2 are larger than zero, then subset A_1 performs significantly better than A_2. Therefore the Spjøtvoll method procedure reads:

- Compute K_1 and K_2 with Equations (E.5), (E.6), (E.7) and (E.8)
- If $K_1 > 0$ and $K_2 > 0$ select A_1, otherwise stay with A_2.

Proof. The problem is determining the maximum and minimum of

$$T(\gamma) = \sum_{i=1}^{q} d_i \gamma_i^2$$

subject to $(\gamma - \hat{\gamma})^T(\gamma - \hat{\gamma}) \leq c^2$ where $c^2 = q\hat{\sigma}^2 F_\alpha$. We consider the case of finding the minimum, the problem of finding the maximum being analogous.

First consider the case when all $d_i \geq 0$ and $\sum_{d_i>0} \hat{\gamma}_i^2 \leq c^2$. In this case since the d_i are non-negative $T(\gamma)$ is always ≥ 0. Then the value 0 is obtained by setting $\gamma_i = 0$ for the indexes i with $d_i \geq 0$.

Next consider the case when all $d_i \geq 0$ and $\sum_{d_i>0} \hat{\gamma}_i^2 > c^2$. Then also $\sum_{i=1}^{q} \hat{\gamma}_i^2 > c^2$ and the sphere $(\gamma-\hat{\gamma})^T(\gamma-\hat{\gamma}) \leq c^2$ does not contain the origin. The ellipsoid $T(\gamma) = \sum_{i=1}^{q} d_i\gamma_i^2$ is centered at the origin. The minimum of $T(\gamma)$ when γ is in the sphere is obtained by expanding the ellipsoid until it touches the sphere. The minimum is therefore obtained at some point satisfying $(\gamma-\hat{\gamma})^T(\gamma-\hat{\gamma}) = c^2$.

Finally consider the case when not all $d_i \geq 0$. We can write:

$$T(\gamma) = \sum_{d_i<0} d_i(\gamma_i - \hat{\gamma}_i + \hat{\gamma}_i)^2 + \sum_{d_i>0} d_i\gamma_i^2$$

Any value of $T(\gamma)$ with $(\gamma-\hat{\gamma})^T(\gamma-\hat{\gamma}) \leq c^2$ can be decreased by increasing $|\gamma_i - \hat{\gamma}_i|$ for some γ_i with $d_i < 0$ and letting $\gamma_i - \hat{\gamma}_i$ having the same sign as γ_i. Hence the minimum takes place for a γ on the sphere $(\gamma-\hat{\gamma})^T(\gamma-\hat{\gamma}) = c^2$.

Similarly it can be shown that, apart from the exception in the theorem, the maximum of $T(\gamma)$ is also attained for $(\gamma-\hat{\gamma})^T(\gamma-\hat{\gamma}) = c^2$.

The solution to the problem of finding the minimum and maximum of $\gamma^T D\gamma$ subject to $(\gamma-\hat{\gamma})^T(\gamma-\hat{\gamma}) = c^2$ can be found from the results in [32]. There the expressions of K_1 and K_2 are derived. This completes the proof. □

Since the acceptance/rejection intervals for the hypothesis testing of this approach are obtained through the Scheffé multiple comparison method, they are tailored to comparisons with any possible subsets of size smaller or equal q. This means that the significance α refers to testing against all possible combinations of subsets of size smaller or equal q. This is the reason why the Spjøtvoll method is considered to have a certain degree of conservativeness (from a MC point of view), meaning that tends to prefer the null hypothesis, and rejects it only when there is very strong evidence against it[1].

Note that for our application certain parameters to be included in the candidate subsets are already defined (the n parameters for unknown position and clock error), so that a variation of the method described must be applied, also described in [94]. In our case in fact the regression functions have the form $Ax + C_y\nabla$, where the variables x are to be included in every regression function. Furthermore we can consider a known variance matrix Q_{yy}. For this problem instead of using Equation (E.2), we can start from:

$$P((\underline{\hat{\nabla}} - \nabla)^T Q_{\hat{\nabla}\hat{\nabla}}^{-1} (\underline{\hat{\nabla}} - \nabla) \leq \chi_\alpha^2(q,0)) \geq 1-\alpha \tag{E.9}$$

where q is the number of columns of the most comprehensive C_y considered, which should include the true C_{y_t}. The method applies in the same way, with the correct substitution to be made (A^TA with $Q_{\hat{\nabla}\hat{\nabla}}^{-1}$ and $q\underline{\hat{\sigma}}^2 F_\alpha(q, m-q)$ with $\chi_\alpha^2(q,0)$).

E.2. FORWARD AND BACKWARD SELECTION METHODS — UMPI TESTS

In case we want to compare two subsets such that one is fully contained in the other (*nested* models), the standard UMPI test can be defined and the corresponding test statistic has a

[1] In view of RAIM applications instead, this would mean that the method tends to have low power in detecting *anticipated* alternative hypotheses, since it is equally powerful in detecting any type of alternative.

known distribution, which is not always the case for the general test statistic derived by Spjøtvoll. Defining H_0 the hypothesis of anomaly in q_1 satellites (subset 0), characterized by design matrix $A_0 = [A\ C_{y_0}]$, and H_a the hypothesis of anomaly in $q_2 > q_1$ satellites (subset a), characterized by design matrix $A_a = [A\ C_{y_a}]$, the UMPI test statistic is:

$$\underline{T}_{q_2-q_1} = \underline{\hat{e}}_0^T Q_{yy}^{-1} \underline{\hat{e}}_0 - \underline{\hat{e}}_a^T Q_{yy}^{-1} \underline{\hat{e}}_a \tag{E.10}$$

where $\hat{e}_0 = P_{A_0}^{\perp} \underline{y}$ and $\hat{e}_2 = P_{A_a}^{\perp} \underline{y}$. For this test statistic it holds:

$$\begin{aligned} H_0: &\quad \underline{T}_{q_2-q_1} \sim \chi^2(q_2-q_1, 0) \\ H_a: &\quad \underline{T}_{q_2-q_1} \sim \chi^2(q_2-q_1, \lambda) \end{aligned} \tag{E.11}$$

with:

$$\lambda = \nabla^T Q_{\hat{\nabla}\hat{\nabla}}^{-1} \nabla \quad \text{cf. Equation (4.7)}$$

where ∇ is the bias sizes vector for the eventual anomaly present in the $q_2 - q_1$ satellites that complement the set 0 of faulty satellites to the set a (hypothesis H_a). It can be shown furthermore that the statistic in Equation (E.10) is actually the *parameter significance* test, developed by the TU Delft school, on the $q_2 - q_1$ additional faults. The parameter significance test is presented in [101]. The derivation of the equivalence is in the following.

Proof. H_a can be defined as:

$$H_a: \quad E(\underline{y}) = A_2 x_2$$

with x_2 the unknown $n+q_2$-components vector including q_2 components for the bias vector and A_2 a $m \times n + q_2$ geometry matrix including the 'bias geometry' in the last q_2 columns ($A_2 \equiv A_a$). The null hypothesis would read:

$$H_0: \quad E(\underline{y}) = A_2 x_2 \quad C^T x_2 = 0$$

with:

$$C^T = \begin{bmatrix} 0 & I_{q_2-q_1} \end{bmatrix}$$

stating that the extra $q_2 - q_1$ components of x are not significant and can be set to zero. The PDF of the observable $\underline{y}$ under the two hypotheses reads:

$$\begin{aligned} H_0: &\quad f_{\underline{y}}(y|x_2) = \frac{1}{(2\pi)^{m/2} \det(Q_{yy})^{1/2}} e^{-\frac{1}{2}(y-A_2x_2)^T Q_{yy}^{-1}(y-A_2x_2)} \ \text{ with } \ C^T x_2 = 0 \\ H_a: &\quad f_{\underline{y}}(y|x_2) = \frac{1}{(2\pi)^{m/2} \det(Q_{yy})^{1/2}} e^{-\frac{1}{2}(y-A_2x_2)^T Q_{yy}^{-1}(y-A_2x_2)} \end{aligned}$$

In order to apply the GLR criterion to obtain our test statistic, we have to maximize the likelihood under the two hypotheses. Under the alternative hypothesis the maximum is attained for $\hat{x}_{2_a} = (A_2^T Q_{yy}^{-1} A_2)^{-1} A_2^T Q_{yy}^{-1} y$ [101], the ML estimate, whereas under H_0 the ML estimate has to be searched within the null space of C^T, $\hat{x}_{2_0} \in R(D)$ with $C^T D = 0$. One form of D is:

$$D = \begin{bmatrix} I_{n+q_1} & 0 \\ 0 & 0 \end{bmatrix}$$

and the maximum likelihood under the null hypothesis is obtained with:

$$\hat{x}_{2_0} = D((A_2 D)^T Q_{yy}^{-1} (A_2 D))^{-1} (A_2 D)^T Q_{yy}^{-1} y$$

E

But since A_2D corresponds to just the first $n+q_1$ columns of A_2, containing only the 'bias geometry' for the first q_1 extra parameters, and therefore this estimate corresponds to the ML estimate for the model including only the q_1 extra parameters, say $\hat{x}_1$. The GLR reads then:

$$e^{-\frac{1}{2}(\hat{e}_{2_0}^T Q_{yy}^{-1}\hat{e}_{2_0}-\hat{e}_{2_a}^T Q_{yy}^{-1}\hat{e}_{2_a})} = e^{-\frac{1}{2}(\hat{e}_1^T Q_{yy}^{-1}\hat{e}_1-\hat{e}_2^T Q_{yy}^{-1}\hat{e}_2)}$$

since $\hat{e}_{2_0}=\hat{e}_1$ is the residual vector for the model with q_1 extra parameters only and $\hat{e}_{2_a}=\hat{e}_2$ is the standard residual vector for the second model with q_2 extra parameters. Taking the natural logarithm, the test statistic becomes just the difference between the weighted sum of squared residuals computed for the two models, which is just the difference of the OMT statistics for the two models:

$$\hat{e}_1^T Q_{yy}^{-1}\hat{e}_1 - \hat{e}_2^T Q_{yy}^{-1}\hat{e}_2$$

This difference also equals $\hat{\nabla}^T Q_{\hat{\nabla}\hat{\nabla}}^{-1}\hat{\nabla}$ in previous notation (Equations (E.11), (4.7)), the normalized squared length of the estimate of the biases in the extra q_2-q_1 parameters. □

Therefore setting a threshold on this test statistic in Equation (E.11) can be made setting a predetermined P_{FA} — where a FA here would mean excluding more satellites than what is actually needed. The P_{FA} for a single test is easily obtained but computing the total P_{FA} from the comparison of one set with all the different candidates is more complex. In fact, although each single test statistic has a known χ^2 distribution, we would have to monitor only the maximum of the test statistics, which distribution is not known, or alternatively monitor the full multivariate distribution of the test statistics and compute its integral over a convex volume, which is not an easy task (this is a problem of MC, considered in Section 3.6.2). Tables or numerical approximation can possibly be used to compute this value, or more simply Bonferroni or Šidák approximations.

E.3. MINIMIZING MALLOWS'S $\underline{C}_p$ (OR THE MSEP) AND OTHER METHODS

E.3.1. MALLOWS'S $\underline{C}_p$

As mentioned in Section 3.6.3, the idea behind this method is to select the dimension of the subset of parameters for which a balance is reached between:

1. the error we make by leaving out parameters that are actually significant in the model description (omission bias), decreasing when more parameters are included, and
2. the random error naturally propagated in the estimation process, that instead increases when more parameters are to be estimated.

Let us start from the model in Equation (3.28), with $Q_{yy}=\sigma^2 I$ known. Considering each component $\underline{y}_i$ of the observable $\underline{y}$, y_i the mean of each observation (unknown) and $\underline{\hat{y}}_i$ its estimator, the corresponding Mean Squared Error of Prediction (MSEP) [67] is:

$$E[(\underline{\hat{y}}_i - y_i)^2] \tag{E.12}$$

As mentioned a subset selection criterion is minimizing the average MSEP over all $\underline{y}$ components (or the sum of the MSEPs). Note that in the TU Delft terminology this is just the Mean Squared Error of the Estimation of the observable, y estimated by $\underline{\hat{y}}$.

Define A_t $(m \times n)$ the actual geometry matrix, unknown, that relates the actual unknown parameters x to the expected measurements y. If we partition A_t in two submatrix:

$$A_t = [A_A \; A_B]$$

and we use only a submatrix, say A_A of dimensions $m \times q$, for estimating the corresponding parameter vector x_A (of dimension q smaller than the true x, $q < n$), we have:

$$\underline{\hat{x}}_A = (A_A^T Q_{yy}^{-1} A_A)^{-1} A_A^T Q_{yy}^{-1} \underline{y}$$

and this estimator is biased by:

$$E(\underline{\hat{x}}_A) - x_A = (A_A^T Q_{yy}^{-1} A_A)^{-1} A_A^T Q_{yy}^{-1} A_B x_B \tag{E.13}$$

where x_B are the regression parameters left out by the use of the submatrix A_A only. This bias is usually referred to as *omission* bias, because is due to the omission of some regression parameters[2]. We have, for each measurement i:

$$\underline{\hat{y}}_i - y_i = A_{A_i} \underline{\hat{x}} - [A_{A_i} \; A_{B_i}] \begin{bmatrix} x_A \\ x_B \end{bmatrix} = A_{A_i} (\underline{\hat{x}}_A - E(\underline{\hat{x}}_A) + E(\underline{\hat{x}}_A) - x_A) - A_{B_i} x_B$$

where A_{A_i} is the i-th row of matrix A_A. If we define $b_i = A_{A_i}(E(\underline{\hat{x}}_A) - x_A) - A_{B_i} x_B = (A_{A_i} \cdot (A_A^T Q_{yy}^{-1} A_A)^{-1} A_A^T Q_{yy}^{-1} A_B - A_{B_i}) x_B$ (with Equation (E.13) the omission bias in the expected measurement, we can rewrite the previous espression as:

$$\underline{\hat{y}}_i - y_i = A_{A_i} (\underline{\hat{x}}_A - E(\underline{\hat{x}}_A)) + b_i$$

Following, the expected squared error of prediction (estimation) is:

$$E[(\underline{\hat{y}}_i - y_i)^2] = A_{A_i} Q_{\hat{x}_A \hat{x}_A} A_{A_i}^T + b_i^2 \tag{E.14}$$

because the cross terms that results from taking the square $(\underline{\hat{y}}_i - y_i)^2$ present the quantity $\underline{\hat{x}}_A - E(\underline{\hat{x}}_A)$, which has expectation zero, and therefore cancel out.

We furthermore have for the expectation of the Sum of Squared Errors (SSE):

$$E\left[\sum_{i=1}^{m} (\underline{y}_i - \underline{\hat{y}}_i)^2\right] = \sum_{i=1}^{m} b_i^2 + \text{trace}(Q_{\hat{e}_A \hat{e}_A}) = \sum_{i=1}^{m} b_i^2 + (m-q)\sigma^2 \tag{E.15}$$

where $\underline{\hat{e}}_A = \underline{y} - A_A \underline{\hat{x}}_A$, and $\text{trace}(Q_{\hat{e}_A \hat{e}_A}) = (m-q)\sigma^2$ because $m-q$ is the dimension of the projection space $R(A_A)$. Note that this is not the sum of the quantities in Equation (E.14), since we have no longer the (unknown) true y_i but the observable $\underline{y}_i$. From the above Equation we can obtain an estimate of the sum of the omission biases by replacing the left hand side of this equation by the sample sum of squared errors $SSE = \underline{\hat{e}}_A^T \underline{\hat{e}}_A$. If we sum the mean squared error of prediction in Equation (E.14) over all the observations, we obtain the Mean Sum of Squared Errors of Prediction:

$$MSSEP = E\left[\sum_{i=1}^{m} ((\underline{\hat{y}}_i - y_i)^2\right] = \text{trace}(A_A Q_{\hat{x}_A \hat{x}_A} A_A^T) + \sum_{i=1}^{m} b_i^2 = q\sigma^2 + \sum_{i=1}^{m} b_i^2 \tag{E.16}$$

[2]Another type of bias can possibly be present, the *selection* bias, when the model has been selected from the same data used to estimate the parameters. This bias is generally difficult to compute and we refer to [67] and [21] for further details.

and estimating the sum of biases using Equation (E.15) and replacing the expected SSE by the sample SSE as mentioned we obtain:

$$MSSEP \approx q\sigma^2 + \underline{\hat{e}}_A^T \underline{\hat{e}}_A - (m-q)\sigma^2 = \underline{\hat{e}}_A^T \underline{\hat{e}}_A - (m-2q)\sigma^2 \tag{E.17}$$

If we divide through by σ^2 we obtain Mallows's $\underline{C}_p$ statistic[3] [64] :

$$\underline{C}_p = \frac{\underline{\hat{e}}_A^T \underline{\hat{e}}_A}{\sigma^2} - (m-2q) \tag{E.18}$$

The subset to be chosen is the one for which the $\underline{C}_p$ assumes the lowest value, therefore for which the MSEP is lowest (the MSEP is obtained dividing the MSSEP by the number of observations m). Other methods exist which employ modified $\underline{C}_p$ values but the rationale behind the criteria is about the same. We note that the method does not foresee further checking on significance — just selection among a set of similar candidate models is made.

In case of generic Q_{yy} a similar approach can be followed but instead of estimating the MSSE as in Equation (E.16) we would estimate the Mean Weighted Sum of Squared Errors of Prediction:

$$MWSSEP = E[(\underline{\hat{y}} - \underline{y})^T Q_{yy}^{-1} (\underline{\hat{y}} - \underline{y})] \tag{E.19}$$

This leads to the following expression for the Generalized $\underline{C}_p$:

$$\underline{GC}_p = \underline{\hat{e}}_A^T Q_{yy}^{-1} \underline{\hat{e}}_A - (m-2q) \tag{E.20}$$

Again the subset that minimizes the $\underline{GC}_p$ should be chosen as best candidate model.

E.3.2. OTHER METHODS

Similarly to Mallows's C_p, the $\underline{R}^2_{adj}$ statistic can be used to determine the fit of a model to the data. Let us consider the standard ANOVA model in Equation (3.28). The $\underline{R}^2$ statistic, known as coefficient of multiple determination, is defined as:

$$\underline{R}^2 = 1 - \frac{\underline{\hat{e}}_M^T \underline{\hat{e}}_M}{\underline{\hat{e}}_0^T \underline{\hat{e}}_0} = 1 - \frac{\underline{SS}_E}{\underline{SS}_T} \tag{E.21}$$

where $\underline{\hat{e}}_M$ is the residual vector computed using the model M (subset of parameters) under consideration (in fact is the $\underline{\hat{e}}_a$) and $\underline{\hat{e}}_0$ is instead the residual under the standard null hypothesis $E(\underline{y}) = Ax_0$ with $A = [1,1,\ldots,1]^T$ and x_0 a single parameter (the mean $\bar{y}$) as previously described. $\underline{SS}_T$ stands for Total Sum of Squares and $\underline{SS}_E$ stands for Error Sum of Squares. The larger is the value assumed by the $\underline{R}^2$, the better is the fit of the model M.

The quantities SS_E and SS_T are represented geometrically in Figure E.1 with a comparison with the relative parameters from TU Delft theory presented in Chapter 5 (the residuals under null hypothesis and alternative hypothesis, $\hat{e}_0$ and $\hat{e}_a$, and the estimated measurements under null and alternative hypothesis $\hat{y}_0$ and $\hat{y}_a$).

The adjusted $\underline{R}^2_{adj}$ statistic was developed to compensate for the issue that the $\underline{R}^2$ always increases when extra regression parameters are added to the model. The $\underline{R}^2_{adj}$ statistic is defined as:

$$\underline{R}^2_{adj} = 1 - \frac{\underline{\hat{e}}_M^T \underline{\hat{e}}_M/(m-q)}{\underline{\hat{e}}_0^T \underline{\hat{e}}_0/(m-1)} = 1 - \frac{\underline{SS}_E/(m-q)}{\underline{SS}_T/(m-1)} \tag{E.22}$$

[3]The subscript p is used because the number of parameters q is sometimes referred to with the letter p.

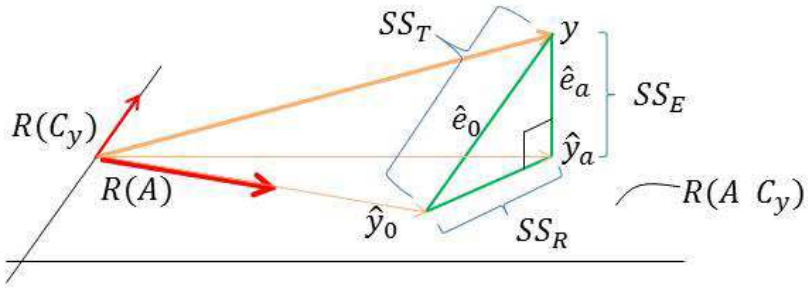


Figure E.1: Geometric visualization of the ANOVA parameters SS_E, SS_T and SS_R (Sum of Squares of the Regression) and their correspondence to the equivalent of the TU Delft theory: $SS_E \equiv \hat{e}_a^T \hat{e}_a$, $SS_T \equiv \hat{e}_0^T \hat{e}_0$, $SS_R \equiv \|\hat{y}_0 - \hat{y}_a\|^2$.

where q is the number of parameters employed in model M. $\underline{SS}_E/(m-q)$ is the residual mean square and since $\underline{SS}_T/(m-1)$ stays constant when comparing different models, the $\underline{R}^2_{adj}$ will only increase when a parameter is added if the new parameter reduces the residual mean square.

Both $\underline{R}^2$ and $\underline{R}^2_{adj}$ do not have a standard known distribution. For our case of generic variance matrix Q_{yy} the $\underline{R}^2_{adj}$ assumes the form:

$$\underline{R}^2_{adj} = 1 - \frac{\underline{\hat{e}}_M^T Q_{yy}^{-1} \underline{\hat{e}}_M/(m-n-q)}{\underline{\hat{e}}_0^T Q_{yy}^{-1} \underline{\hat{e}}_0/(m-n)} \tag{E.23}$$

and also in this case the $\underline{R}^2_{adj}$ increases when the residual mean square $\underline{\hat{e}}_M^T Q_{yy}^{-1} \underline{\hat{e}}_M/(m-n-q)$ decreases. This quantity is the same as the OMT statistic for the model M divided by its degrees of freedom.

The subset selection criterion of choosing the subset with maximum $\underline{R}^2_{adj}$ tends to prefer subsets with larger number of parameters than the $\underline{C}_p$ method, and generally the $\underline{C}_p$ method is preferred. Recalling the $\underline{C}_p$ formula:

$$\underline{C}_p = \frac{\underline{\hat{e}}_A^T \underline{\hat{e}}_A}{\sigma^2} - (m-2q)$$

the quantity $2q$ can be considered as a form of *penalty* for adding an extra parameter in the regression model. It is shown in [67] that the $\underline{R}^2_{adj}$ method is approximately equivalent to the Mallows's method in which instead of using $2q$ as penalty measure, a lower value, just q, is adopted.

Note that both $\underline{R}^2_{adj}$ and $\underline{C}_p$ methods are fundamentally equivalent to employing UMPI tests (GLR criterium) to select between one or another subset, with a particular choice for the significance of the test — a quite large significance compared to the standard 0.05 or

0.1 commonly employed in hypothesis testing. In case of the $\underline{R}^2_{adj}$ criterium, when nested models are compared this is in fact equivalent to running UMPI test with significance α set to 50%.

Another subset selection method is the Akaike's Information Criterion (AIC), which is the Mallows's equivalent developed in information theory (for the same model of Equation (3.27)). The AIC results in a formula identical to the $\underline{C}_p$, obtained though through different reasoning. In fact the AIC is obtained looking for the subset that minimizes the expectation of the Kullback-Leibler mean information. Details on this method can be found in [1].

The p-values, briefly introduced in Section 3.6.2, can also be employed as discriminants to select among multiple candidate subsets. To employ this Subset Selection criterium, a null hypothesis has to be chosen and a statistical test has to be run for each alternative model. A p-value is computed for each alternative model and, in case of rejection of the null hypothesis, the alternative hypothesis for which the smallest p-value was computed is chosen as a new model. The p-value was defined for the realization T_0 of a test statistic $\underline{T}$, in case of test with right-sided rejection region, as:

$$p(T_0) = \int_{T_0}^{+\infty} f_{\underline{T}|H}(T|H)\mathrm{d}T \quad \text{cf. Equation (3.26)}$$

In case of a symmetric two-sided rejection region (of the type $K: \ T < -k \cup T > k$, as it is the case for the w-tests), the p-value for T_0 can be defined as:

$$p(T_0) = 2\int_{T_0}^{+\infty} f_{\underline{T}|H}(T|H)\mathrm{d}T \tag{E.24}$$

The p-value method reads therefore:

1. Choose a null hypothesis H_0 (reference model/subset of parameters) and define the alternative hypotheses H_i (alternative models/subsets)
2. Run a UMPI test for each alternative hypothesis
3. If, applying an MC criterium (for instance if none of the comparison tests run is rejected)[4], the null hypothesis is not rejected, accept the null hypothesis model
4. If the null hypothesis is rejected, compute the p-values for each of the test statistic computed
5. Choose the alternative hypothesis/model corresponding to the test statistic realization characterized by the smaller p-value

We refer to [33] for further discussion on the p-values.

[4] An MC method that foresees a specific test for detection as the OMT can also be applied.

F

V_{slope} AND RELIABILITY

In Section 4.1.2 the V_{slope} was introduced as a parameter to monitor the effect of a measurement bias on the position domain. In this Appendix we explain this parameter making use of the reliability concepts developed by the TU Delft school and we provide a geometrical interpretation for it. More on the TU Delft testing theory can be found in Section 3.6.1, in Chapter 5 and in [101]. We defined already in Section 3.6.1 the quantity:

$$\lambda = \nabla^T Q^{-1}_{\hat{\nabla}\hat{\nabla}} \nabla \text{ cf. Equation (4.7)}$$

With reference to [101], the following quantities can also be defined:

$$\lambda_{\hat{x}} = \nabla \hat{x}^T Q^{-1}_{\hat{x}_0 \hat{x}_0} \nabla \hat{x} \tag{F.1}$$

$$\lambda_y = \nabla y^T Q^{-1}_{yy} \nabla y \tag{F.2}$$

These quantities are visualized in Figure F.1. They allow writing in simple way the relation between internal and external reliability. λ measures the incidence of the error on the test statistic (internal reliability, i.e. detectability of the fault), while $\lambda_{\hat{x}}$ measures the effect on the positioning error (external reliability, i.e. effect of undetected fault).

As a general relation it holds:

$$\lambda_y = \lambda_{\hat{x}} + \lambda \tag{F.3}$$

In particular:

$$\sqrt{\lambda_{\hat{x}}} = \| A \nabla \hat{x} \|_{Q^{-1}_{yy}} \tag{F.4}$$

With reference to Figure F.1:

$$\sqrt{\lambda_{\hat{x}}} = \| P_A \nabla y \|_{Q^{-1}_{yy}} \tag{F.5}$$

where $P_A = A(A^T Q^{-1}_{yy} A)^{-1} A^T Q^{-1}_{yy}$ is the orthogonal projector onto $R(A)$ in the metric defined by Q^{-1}_{yy}. Considering the vectors ∇y and $P^{\perp}_A \nabla y$ and naming φ the angle between the two, it is possible to write:

$$\tan\varphi = \frac{\| P_A \nabla y \|_{Q^{-1}_{yy}}}{\| P^{\perp}_A \nabla y \|_{Q^{-1}_{yy}}} = \frac{\sqrt{\lambda_{\hat{x}}}}{\sqrt{\lambda}} \tag{F.6}$$

where $P_A = I - A(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$ is the orthogonal projector onto $R(A)^\perp$ in the metric defined by Q_{yy}^{-1}. Therefore the V_{slope} is directly linked to this angle, with the difference that only the vertical component of x is taken into account:

$$\tan\varphi' = \frac{\|A_{j3}\nabla\hat{x}\|_{Q_{yy}^{-1}}}{\|P_A^\perp \nabla y\|_{Q_{yy}^{-1}}} = \sqrt{A_{j3}^T Q_{yy}^{-1} A_{j3}} \cdot \frac{|(\nabla\hat{x})_3|}{\sqrt{\lambda}} = \frac{1}{\sigma_{\hat{x}_3}} V_{slope} \tag{F.7}$$

where $\sigma_{\hat{x}_3}^{-1} = \sqrt{A_{j3}^T Q_{yy}^{-1} A_{j3}}$ (used in the last equivalence) accounts for the transformation from the solution space $R^4 \equiv R(A)$ to the measurement space R^m through the matrix A. In the last passage we made use of:

$$V_{slope} = \frac{|(\nabla\hat{x})_3|}{\sqrt{\lambda}} \tag{F.8}$$

with $\nabla\hat{x} = E(\underline{\hat{x}} - x)$. As a matter of fact, under the assumption of the presence of an outlier in the measurements from a single satellite (C_y is a vector of all zeros except a 1 at the i-th component/satellite and ∇ is just a scalar):

$$\nabla y = C_y \nabla = \begin{pmatrix} 0 \\ \vdots \\ \nabla \\ \vdots \\ 0 \end{pmatrix} \rightarrow \quad i^{th}\text{component}$$

and of Q_{yy} being a diagonal matrix, it is possible to obtain the original Equation (4.4):

$$\frac{|(\nabla\hat{x})_3|}{\sqrt{\lambda}} = \frac{|\sum_j K_{3j}\nabla y_j|}{\sqrt{\nabla y^T Q_{yy}^{-1}(I - P_A)\nabla y}} = \frac{|K_{3i}|\nabla}{\sqrt{\nabla^2 (I - P_A)_{ii}/\sigma_i^2}} = \frac{|K_{3i}|\sigma_i}{\sqrt{(1 - P_{A_{ii}})}}$$

In Equation (F.7) we obtained again the same expression of Equation (4.19), a well known result of the Delft school.

As concluding remark, the V_{slope} represents the ratio between the expectation of the positioning error (in the vertical dimension) and the projection, on the space orthogonal to the position space, of the expectation of the range measurement error.

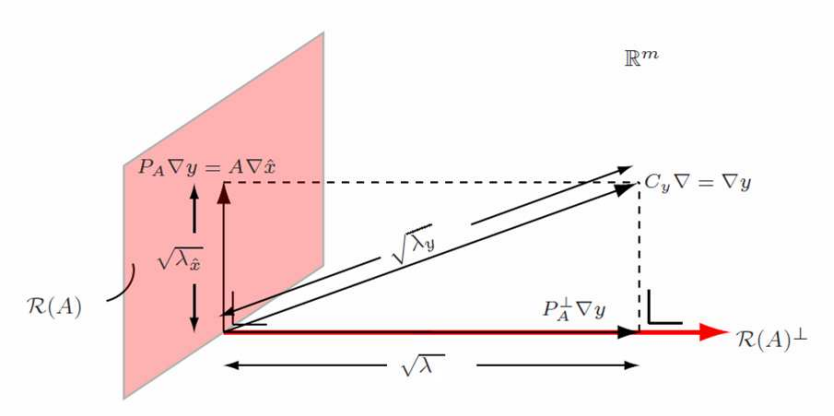


Figure F.1: Geometrical visualization of λ, $\lambda_{\hat{x}}$ and λ_y, from [101].

G

OPTIMALITY OF W-TEST FOR TESTING A SPECIFIC ERROR

In Section 4.1.2 we discussed the increase in detection power (in case of single satellite faults) obtainable employing the w-tests instead of the OMT. In this Appendix we provide a demonstration of the optimality of the w-test in detecting the specific type of faults that it addresses, with respect to any other test built around different fault models.

The optimality of the w-test in detecting an error in one single measurement, when the w-test targets that specific measurement, can be demonstrated by assuming the measurements being biased by a different generic error vector. It is shown that the highest power is obtained when the w-test targets the specific error (has same direction, therefore affects the same measurement).

Proof. The w-test is constructed with alternative hypothesis:

$$E(\underline{y}) = Ax + C_y\nabla$$

with C_y an m-dimensional vector and ∇ scalar. Instead the measurement is considered to be biased in reality as:

$$E(\underline{y}) = Ax + \tilde{C}_y\tilde{\nabla}$$

Also $\tilde{C}_y$ is an m-dimensional vector and $\tilde{\nabla}$ a scalar. This approach allows considering even failures that affect multiple measurements, but with a particular "pattern", i.e. with only one degree of freedom.

The next step is to determine the distribution of the test statistic under the alternative hypothesis considering this particular error. The idea is to maximize the power of the test, and this is equivalent to maximize the non-centrality parameter of the test distribution λ, since the power is a monotonously increasing function of λ.

The non-centrality parameter of the test distribution can be obtained computing:

$$E(\underline{\hat{e}_0}) = P_A^{\perp}(Ax + \tilde{C}_y\tilde{\nabla}) = P_A^{\perp}\tilde{C}_y\tilde{\nabla} = Q_{\hat{e}_0\hat{e}_0}Q_{yy}^{-1}\tilde{C}_y\tilde{\nabla}$$

$$E(\underline{\hat{\nabla}}) = (C_y^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} C_y)^{-1} C_y^T Q_{yy}^{-1} E(\underline{\hat{e}_0})$$

$$= (C_y^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} C_y)^{-1} C_y^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} \tilde{C}_y\tilde{\nabla}$$

$$\lambda = E(\hat{\underline{\nabla}})^T Q^{-1}_{\hat{\nabla}\hat{\nabla}} E(\hat{\underline{\nabla}})$$

$$Q_{\hat{\nabla}\hat{\nabla}} = (C_y^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} C_y)^{-1}$$

From this an expression for λ can be obtained:

$$\lambda = \tilde{\nabla} \tilde{C_y}^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} C_y (C_y^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} C_y)^{-1} C_y^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} \tilde{C_y} \tilde{\nabla}$$

This is an expression of the form $a^2\sigma^2_{\hat{\nabla}\hat{\nabla}}$, and can be maximized by maximizing a, that is maximizing:

$$a = \tilde{\nabla} \tilde{C_y}^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} C_y \tag{G.1}$$

We can take out $\tilde{\nabla}$ from the expression because it is just a weight, and quantifies the size of the error. We are interested in maximizing over C_y. From the last expression we can write:

$$\tilde{C_y}^T Q_{yy}^{-1} Q_{\hat{e}_0\hat{e}_0} Q_{yy}^{-1} C_y = \tilde{C_y}^T P_A^{\perp T} Q_{yy}^{-1} P_A^{\perp} C_y$$

This represents a scalar product in the metric defined by Q_{yy}^{-1}. This means it can be written as:

$$(P_A^{\perp}\tilde{C_y})^T Q_{yy}^{-1} (P_A^{\perp} C_y) = \|P_A^{\perp}\tilde{C_y}\|_{Q_{yy}^{-1}} \|P_A^{\perp} C_y\|_{Q_{yy}^{-1}} \cos\phi$$

where ϕ is the angle between the two vectors $P_A^{\perp} C_y$ and $P_A^{\perp}\tilde{C_y}$. Therefore the expression reaches its maximum for $\phi = 0$ and for $\phi = \pi$ (considering unity/fixed length of $\tilde{C_y}$), when the two projections onto the space perpendicular to $R(A)$ are aligned (the direction of the error is not important indeed). At the end what really counts is the component perpendicular to the space spanned by A, $R(A)$, since no test can detect an error lying in the space $R(A)$, and in $R(A)^{\perp}$, C_y has to be aligned with $\tilde{C_y}$.

In the general case, when C_y is not a vector but a matrix $m \times q$, i.e. we want to use a test that can detect errors in more than one dimension, the expression that was before $a^2\sigma^2_{\hat{\nabla}\hat{\nabla}}$ becomes now:

$$a^T Q_{\hat{\nabla}\hat{\nabla}} a$$

equivalent to the squared length of vector a in the metric of $Q_{\hat{\nabla}\hat{\nabla}}$. This time each component of a is a scalar product as before between the projection on $R(A^{\perp})$ of the corresponding column of C_y and the projection of $\tilde{C_y}$. Therefore to maximize the length of vector a it is necessary to maximize all the components of a, this is the case when the projections of all the columns of C_y on $R(A^{\perp})$ are the same as the projection of $\tilde{C_y}$. This means the best test could use a matrix C_y made of columns equal to each other — this would yield a non centrality parameter λ_q proportional to ql^2, where l is the length of each vector of C_y. But using a C_y matrix with columns equal to each other is equivalent to use a single vector C_y, since no extra degree of freedom is allowed for the bias.

As a conclusion the highest power is reached using a w-test, when knowing exactly the signature of the error, rather than using a more general test with $q > 1$. □

H

VPL COMPUTATION IN THE STANDARD RAIM

In this Appendix we provide an interpretation for the Equation (4.25) given in Section 4.1.2 (based on [107]). With reference to Equation (4.29), we have:

$$P_{\text{HMI}|H_a} = P(\underline{T} < k \cap |\underline{\hat{x}}_3 - x_3| > \text{AL}|H_a) = \\ P(|\underline{\hat{x}}_3 - x_3| > \text{AL}|H_a) \cdot P(\underline{T} < k||\underline{\hat{x}}_3 - x_3| > \text{AL}|H_a)$$

The two probabilities on the second line are uncorrelated, since $\underline{T}$ is a function of $\underline{\hat{e}}$ and $\underline{\hat{e}}$ is uncorrelated with $\underline{\hat{x}}_3$ (under the common assumption that the observables are normally distributed), therefore it holds:

$$P_{\text{HMI}|H_a} = P(|\underline{\hat{x}}_3 - x_3| > \text{AL}|H_a) \cdot P(\underline{T} < k|H_a) \tag{H.1}$$

The vertical position estimator will be distributed as $\underline{\hat{x}}_3|H_a \sim N(x_3 + \nabla\hat{x}_3, \sigma^2_{\hat{x}_3})$, where $\nabla\hat{x} = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} \nabla y$ is the bias in the position estimate directly proportional to the bias in the measurement. Therefore:

$$P_{\text{HMI}|H_a} = \Phi\left(\frac{-AL + \nabla\hat{x}_3}{\sigma_{\hat{x}_3}}\right) \cdot P(\underline{T} < k|H_a)$$

with Φ the cumulative distribution function of a standard normal distribution. If we approximate $P_{HMI} \approx P_{HMI|H_a} \cdot P(H_a)$, and consider $\nabla\hat{x}_3 = \max[V_{slope}]k$, i.e. we suppose $E(\underline{T}) = k$, we have:

$$\Phi\left(\frac{-AL + \max[V_{slope}]k}{\sigma_{\hat{x}_3}}\right) = \frac{P_{\text{HMI}}}{P(H_a)P(\underline{T} < k|H_a)}$$

If we now impose a required P_{HMI} and invert the above Equation, the equivalent of the AL would be in this case the PL and we have:

$$\max[V_{slope}]k - PL = \sigma_{\hat{x}_3}\Phi^{-1}\left(\frac{P_{\text{HMI}}}{P(H_a)P(\underline{T}<k|H_a)}\right) \Rightarrow \\ PL = \max[V_{slope}]k - \sigma_{\hat{x}_3}\Phi^{-1}\left(\frac{P_{\text{HMI}}}{P(H_a)P(\underline{T}<k|H_a)}\right) \Rightarrow \\ PL = \max[V_{slope}]k + \sigma_{\hat{x}_3}k_{MD}$$

which coincides with Equation (4.25) as long as $P_{MD} = \frac{P_{\text{HMI}}}{P(H_a)P(\underline{T}<k|H_a)}$. We suppose the P_{MD} is meant to be set at $\frac{P_{\text{HMI}}}{P(H_a)} \leq \frac{P_{\text{HMI}}}{P(H_a)P(\underline{T}<k|H_a)}$, when the use of a prior probability for H_a is foreseen.

I

SOLUTION SEPARATION AND UMPI TESTS

Here we derive the relations existing between the UMPI test presented in the Section 3.6.1 and the SS test presented in Section 4.2.1. First we show the main relationships holding between the test statistics and between the actual quantities in observation and position domains. Following we analyze in more detail these relationships treating separately the one-dimensional threat and the multi-dimensional threat cases.

I.1. TEST STATISTICS IN OBSERVATION AND SOLUTION DOMAINS

As in Equation (5.1) a general formulation of $\underline{T}_q$ is:

$$\underline{T}_q = \|P_A^{\perp} C_y \underline{\hat{\nabla}}\|^2_{Q_{yy}^{-1}} = \|P_A^{\perp} \underline{\hat{\nabla} y}\|^2_{Q_{yy}^{-1}} \tag{I.1}$$

On the other hand we know from for instance [97] that:

$$\underline{\hat{x}}_0 - \underline{\hat{x}}_a = \underline{\hat{\nabla}\hat{x}} = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} C_y \underline{\hat{\nabla}} \tag{I.2}$$

from which we can write:

$$A\underline{\hat{\nabla}\hat{x}} = P_A C_y \underline{\hat{\nabla}} = P_A \underline{\hat{\nabla} y} \tag{I.3}$$

where $\underline{\hat{x}}_0$ is the solution under the null hypothesis while $\underline{\hat{x}}_a$ is the solution computed under the alternative hypothesis.

Therefore we can notice that the realization of the test statistic T_q is the norm (in the metric defined by matrix Q_{yy}^{-1}) of the projection of the estimator of the bias vector $\hat{\nabla} y = C_y \hat{\nabla}$ in the space perpendicular to $R(A)$, whereas the SS is directly related to the projection of the $\hat{\nabla} y$ in the space $R(A)$. This means the two test statistics are just the two orthogonal components of the same vector $\hat{\nabla} y$.

From the previous equation we can further derive:

$$\|\underline{\hat{\nabla}\hat{x}}\|^2_{Q^{-1}_{\hat{x}_0\hat{x}_0}} = \|P_A C_y \underline{\hat{\nabla}}\|^2_{Q_{yy}^{-1}} \tag{I.4}$$

Therefore, the Pythagoras relation holds, and we can write the following relation between the UMPI test statistic and the norm of the SS:

$$\|C_y \underline{\hat{\nabla}}\|^2_{Q_{yy}^{-1}} = \|P_A^{\perp} C_y \underline{\hat{\nabla}}\|^2_{Q_{yy}^{-1}} + \|P_A C_y \underline{\hat{\nabla}}\|^2_{Q_{yy}^{-1}} = \underline{T}_q + \|\underline{\hat{\nabla}\hat{x}}\|^2_{Q^{-1}_{\hat{x}_0\hat{x}_0}} \tag{I.5}$$

Note that this represents the central relation holding between the estimators of the biases in observation and position domains, and can be visualized as in Figure I.1, though note that in figure only the representation (through matrix A) of the position domain in the R^m space is shown ($A\underline{\hat{\nabla}\hat{x}}$ and $A\underline{\hat{x}}$ for instance). $\underline{\hat{\nabla}y}$ is obtained by projecting $\underline{y}$ onto $R(A\ C_y)$ to get $\underline{\hat{y}}_a$ and then decomposing $\underline{\hat{y}}_a$ in $\underline{\hat{\nabla}y} + A\underline{\hat{x}}_a$, therefore the procedure to obtain the test statistics is: project $\underline{y}$ onto $R(A\ C_y)$ to get $\underline{\hat{\nabla}y} + A\underline{\hat{x}}_a$, then project $\underline{\hat{\nabla}y}$ onto $R(A)$ to obtain $A\underline{\hat{\nabla}\hat{x}}$ (representation through matrix A of the SS statistic $\underline{\hat{\nabla}\hat{x}}$) and onto $R(A)^\perp$ to obtain $P_A^\perp\underline{\hat{\nabla}y}$ whose norm is the $\underline{T}_q$ test statistic.

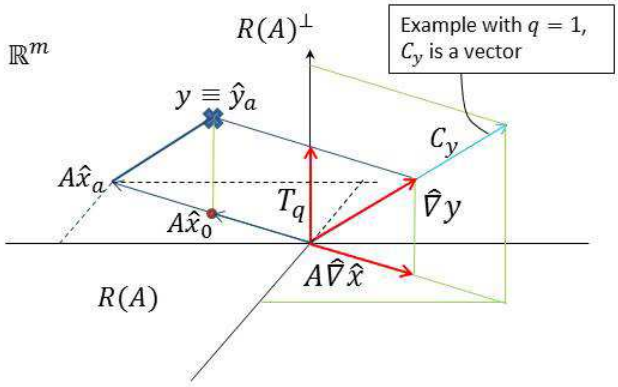


Figure I.1: Comparison between UMPI and SS test statistics. Observation space R^m with $m = 3$, $n = 2$ and $q = 1$. The error vector $\nabla y = C_y\nabla$ has only one degree of freedom. Given a measurement y, this is projected onto $R(A\ C_y)$ to get $\hat{\nabla}y$ (and $A\hat{x}_a$), then $\hat{\nabla}y$ is projected onto $R(A)$ to obtain the SS statistic (more precisely a linear combination of it), $A\hat{\nabla}\hat{x}$, and onto $R(A)^\perp$ to obtain $P_A^\perp\hat{\nabla}y$ which norm is the T_q test statistic realization.

The relation described by Equation (I.5) is not the only one relating the two test statistics: in fact $\underline{\hat{\nabla}y}$ is constrained to lie on $R(C_y)$; if $R(C_y)$ spans any of the base vectors of $R(A)^\perp$, or in case $q > n$, to the same $\underline{\hat{\nabla}\hat{x}}$ may correspond multiple values of $\underline{T}_q$, otherwise to $\underline{\hat{\nabla}\hat{x}}$ corresponds only one value of $\underline{T}_q$.

This second relation can be visualized as well in Figure I.1. In this case $R(A\ C_y)$ is the full R^m space, so $y \equiv P_{(A\ C_y)}y = \hat{y}_a$, but $\hat{\nabla}y$ lies on $R(C_y)$ so that $A\hat{\nabla}\hat{x}$ and T_q are fully determined as the two orthogonal components of it.

I.2. ACTUAL BIASES IN OBSERVATION AND SOLUTION DOMAINS

The same relation as Equation (I.5) holds between the *true* unknown biases in observations and position domains, and was developed by the Delft school when defining the concepts of internal and external reliability of the test [101]. Internal reliability relates to the power of the test to detect a bias in the observations whereas the external reliability relates to the effect an undetected bias has on the position solution. In particular, given a certain anomaly is present, causing a bias $C_y\nabla$ in the measurements, the corresponding effect on the position solution will be, as in Equation (I.2), $\nabla\hat{x} = (A^TQ_{yy}^{-1}A)^{-1}A^TQ_{yy}^{-1}C_y\nabla$ (note the notation now with a single hat).

We already defined λ in Equation (4.7). With reference to [101], the following quantities can be also defined:

$$\lambda_{\hat{x}} = \nabla\hat{x}^T Q^{-1}_{\hat{x}_0\hat{x}_0} \nabla\hat{x} = \|A\nabla\hat{x}\|^2_{Q^{-1}_{yy}} = \|P_A\nabla y\|^2_{Q^{-1}_{yy}} \tag{I.6}$$

$$\lambda_y = \nabla y^T Q^{-1}_{yy} \nabla y = \|\nabla y\|^2_{Q^{-1}_{yy}} \tag{I.7}$$

These quantities are visualized in Figure F.1. They allow writing in simple way the relation between internal and external reliability. λ is in fact the non-centrality parameter of the distribution of the test $\underline{T}_q$ under the alternative hypothesis (Equation (4.7)) and measures the incidence of the error on the test statistic (i.e. detectability of the fault), λ_y relates to the size of the detectable bias in the observation domain (internal reliability) while $\lambda_{\hat{x}}$ measures the effect on the position solution (external reliability, i.e. effect of undetected fault). A similar visualization can be made for the quantities in Equation (I.5). As a general relation, the equivalent of Equation (I.5), it holds:

$$\lambda_y = \lambda + \lambda_{\hat{x}} \tag{I.8}$$

This represents the main relation holding between internal and external reliability (λ_y and $\lambda_{\hat{x}}$).

I.3. CASE $q = 1$ (W-TEST)

In case $q = 1$ (that is the case of the w-test) $\hat{\underline{\nabla}}$ is a scalar and combining Equations (I.1) and (I.4):

$$\frac{\underline{T}_1}{\|\underline{\hat{x}}_0 - \underline{\hat{x}}_a\|^2_{Q^{-1}_{\hat{x}_0\hat{x}_0}}} = \frac{\|P_A^{\perp} C_y \hat{\underline{\nabla}}\|^2_{Q^{-1}_{yy}}}{\|P_A C_y \hat{\underline{\nabla}}\|^2_{Q^{-1}_{yy}}} = \frac{\|P_A^{\perp} C_y\|^2_{Q^{-1}_{yy}}}{\|P_A C_y\|^2_{Q^{-1}_{yy}}}$$

Therefore in this case, as long as the denominator $\|P_A C_y\|^2_{Q^{-1}_{yy}} \neq 0$, $\underline{T}_q$ and the norm of the SS are directly proportional, the proportionality constant depending from the geometry matrices (the dependency from the bias size has been taken out). First let us consider the standard case $\|P_A C_y\|^2_{Q^{-1}_{yy}} \neq 0$.

Since the direction of $\underline{\hat{x}}_0 - \underline{\hat{x}}_a = \hat{\underline{\nabla}}\hat{x}$ is fully determined by C_y as seen in Equation (I.2), there is a univocal correspondence between a value of the statistic $\underline{T}_q$ and the SS $\hat{\underline{\nabla}}\hat{x}$, except for the sign ($\pm\hat{\underline{\nabla}}\hat{x}$ yield the same $\underline{T}_1$).

This can be seen easily in Figure I.2. The space R^m of the observations is shown. The space $R(C_y)$ on which the estimate of the bias vector $\hat{\nabla}y$ has to lie is a simple straight line and a value (for instance a threshold) for the $\underline{T}_q$ defines a specific unique vector (also in the projection on $R(A)^{\perp}$), to which corresponds a unique projection on $R(A)$.

If in the SS approach a threshold is set for each component of $\hat{\underline{\nabla}}\hat{x}$ (this is represented as a parallelogram constraint in the space $R(A)$ with dashed lines in figure), it is evident that only one of the n thresholds ($n = 2$ in the Figure) actually constrains the SS, since the direction of this vector is already known.

Let us consider the limiting case of $\|P_A C_y\|^2_{Q^{-1}_{yy}} = 0$. This means that C_y lies on $R(A)^{\perp}$, and $\hat{\underline{\nabla}}\hat{x}$ would always be zero for any value of the bias ∇. In this case the anomaly considered has

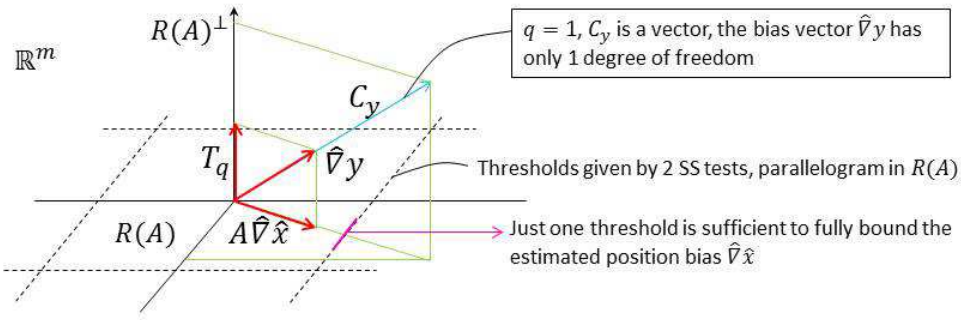


Figure I.2: Comparison between w-test and SS test statistic. Observation space R^m with $m = 3$, $n = 2$ and $q = 1$. The error vector $\nabla y = C_y \nabla$ has only one degree of freedom, therefore only one threshold on the statistic $\underline{T}_q$ is sufficient to fully constrain it. Vector $\hat{\nabla} y$ is given by $\hat{\nabla} y = C_y \hat{\nabla}$.

no effect at all on the position domain, therefore the SS test never leads to rejection, whereas the UMPI test can still lead either to rejection or acceptance as it operates in $R(A)^{\perp}$.

In the other limiting case, when $\|P_A^{\perp} C_y\|^2_{Q_{yy}^{-1}} = 0$, both UMPI test and SS are not able to detect any fault: in fact $C_y \subset R(A)$, and in the alternative hypothesis there is not just one solution for the maximum likelihood estimation of position and bias, but an infinite number of solutions.

I.4. CASE $q > 1$ ($\underline{T}_q$ TEST)

In case $q > 1$ things are not that simple anymore. We can refer to Figure I.3, where the detection regions in case of adoption of $\underline{T}_q$ test on one hand or SS test on the other are shown. In particular the detection regions are shown in the observation bias domain ($\hat{\nabla}$ in R^q) and in the position domain ($\hat{\nabla}\hat{x}$ in R^n) for the case $n = q = 2$.

First of all we note that, as clear also from the equivalent expressions of $\underline{T}_q$ in Equation (5.1), an upperbound to the test statistic defines an (hyper-)ellipsoid in the space $R(C_y)$. The bias size vector $\hat{\nabla}$ has dimension q, whereas in the position domain $\hat{\nabla}\hat{x}$ has dimension n. After this observation, it is clear that in the comparison between $\underline{T}_q$ and the SS $\hat{\nabla}\hat{x}$ the values q and n (n being the dimensions of $R(A)$ and also of the position solution) are important factors.

There is to consider furthermore that the q-dimensional ellipsoidal constraint lies on $R(C_y)$, but we are fundamentally interested in its projection on $R(A)$. In the same way, to move from the SS to the $\underline{T}_q$ domain we conversely first obtain the representation of the position domain in $R(A)$, then project from $R(A)$ to $R(C_y)$ and following to $R(A)^{\perp}$, with reference also to Figure I.2. The projection of $R(C_y)$ on $R(A)$ is given by the matrix $P_A C_y = A(A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1} C_y$ (in the metric defined by Q_{yy}^{-1}). Depending on whether $R(C_y)$ is perpendicular to $R(A)$ or not, the matrix $A^T Q_{yy}^{-1} C_y$ has rank smaller or equal than both q and n,

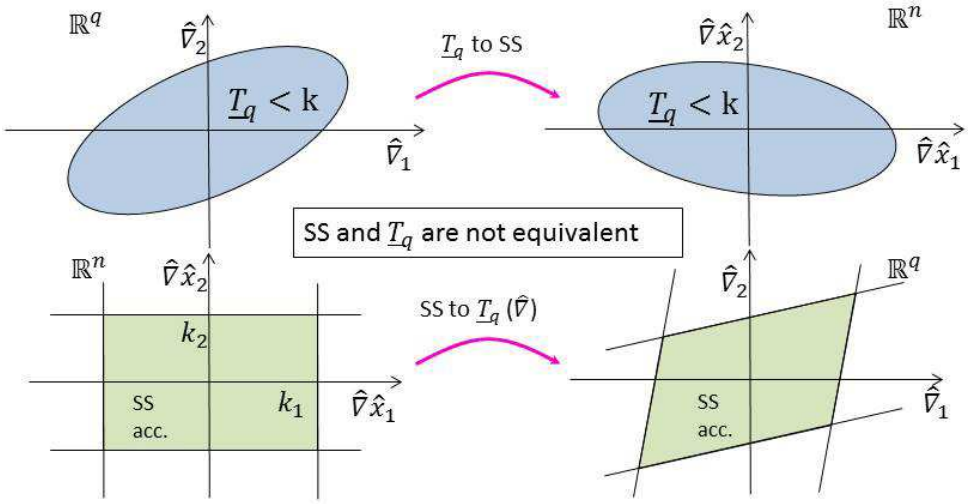


Figure I.3: Comparison between the detection regions of $\underline{T}_q$ test and the SS test statistic, case $n = q = 2$, C_y not generating a subspace perpendicular to the space $R(A)$. At the top is shown the transformation of the detection region for the $\underline{T}_q$ test from range fault domain ($\hat{\nabla}$) to position bias domain ($\hat{\nabla}\hat{x}$), whereas at the bottom is the transformation of the detection region for the SS test from position bias domain ($\hat{\nabla}\hat{x}$) to range fault domain ($\hat{\nabla}$).

$r = \text{rank}(A^T Q_{yy}^{-1} C_y) \leq \min[q, n]$. The rank of $A^T Q_{yy}^{-1} C_y$, r, defines the number of dimensions on which a direct correspondence links the constraint in $R(A)$ given by the SS test and the constraint in $R(C_y)$ given by the $\underline{T}_q$ test.

In general, an upper-bound on $\underline{T}_q$ determines a q-dimensional ellipsoid in $R(C_y)$. Its projection on $R(A)$ will be an r-dimensional ellipsoid.

Conversely, when a threshold is set in the SS approach for each component of $\hat{\nabla}\hat{x}$, this determines an hyper-rectangular constraint in the position domain, to which corresponds an hyper-parallelogram in $R(A)$. Its planes will intersect the r-dimensional subspace projection of $R(C_y)$ on $R(A)$ generating either a closed polyhedrum or an open figure. The planes generate constraints in r dimensions, therefore they will create a polyhedrum in the q-dimensional $R(C_y)$ only if $r \geq q$, otherwise the constraint does not constitute a full bound in the range bias $\hat{\nabla}$ domain (and therefore no finite threshold for $\underline{T}_q$).

In most general cases the matrix $A^T Q_{yy}^{-1} C_y$ will be of full rank $r = \min[q, n]$.

I.5. NUMERICAL EXAMPLE

Based on a specific geometry (Figure I.4), the detection regions for the two tests (Equations (5.1) and (4.56)) have been determined for the case $q = n = 2$. The detection regions are shown in Figures I.5 and I.6 for different couples of assumed faulty satellites, and with the same significance level set for the two statistics. In particular, a rectangular detection region has been chosen for the horizontal dimensions $\hat{\nabla}\hat{x}_1$ and $\hat{\nabla}\hat{x}_2$, in such a way that for each of the 2 dimensions the significance is set to $\alpha_i = 10^{-3}$. By numerical MC integration, this has been found to be equal to a total significance α, slightly different in each case but always $\alpha \approx 0.002$, to which the UMPI test threshold is based. The geometry matrix and all

the geometric relationships are developed as in the case of parameters estimation for the 4 GNSS unknown (position and clock error), only the constraint is applied to the two horizontal coordinates (in this example). As previously noted, the UMPI detection region results in an ellipse in this two-dimensional case.

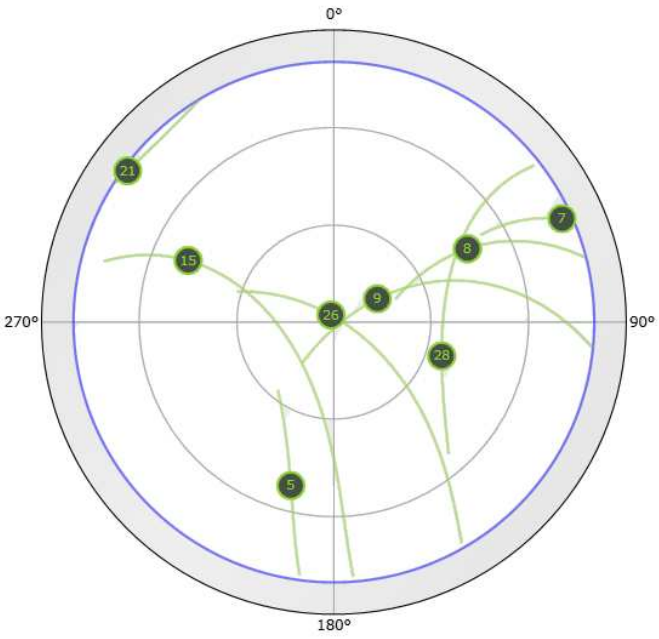


Figure I.4: Skyplot of a real GPS geometry employed for the simulation.

I.6. CONSIDERATIONS ON THE TWO METHODS AND ON IDEAL TESTING

Let us try to understand here what is the impact of the use of different test statistics with their corresponding detection regions on the performances of the tests.

The detection regions as shown in Figures I.5 and I.6 in the same domain of position bias are different and this obviously leads to different performances of the tests. The boundary of the UMPI detection region is characterized by equally likely realizations of $\underline{\hat{\nabla} \hat{x}}$ under the null hypothesis and this results in an easier monitoring of the P_{FA} and related performance parameters of the UMPI test. On the other hand, it is evident how the SS test accepts realizations of $\underline{\hat{\nabla} \hat{x}}$ fairly unlikely under the null hypothesis of no fault — allegedly, since the eventual corresponding fault would not be dangerous by an integrity view point.

Since the SS test explicitly neglects the anomalies that have no effect in the position domain, we can expect that the power saved from neglecting those anomalies will be gained to detect actually dangerous anomalies. Even though it is likely to obtain indeed some gain (though difficult to quantify), the SS test is not optimal from an integrity point of view. This can be easily seen in Figure I.7, that shows an example from the same satellite geometry in Figure I.4. In this figure the case of two different failures is plotted, and the distributions of the test statistic $\underline{\hat{\nabla} \hat{x}}$ and of the positioning error $\underline{\hat{x}}_0 - x$ are shown. In particular, the same

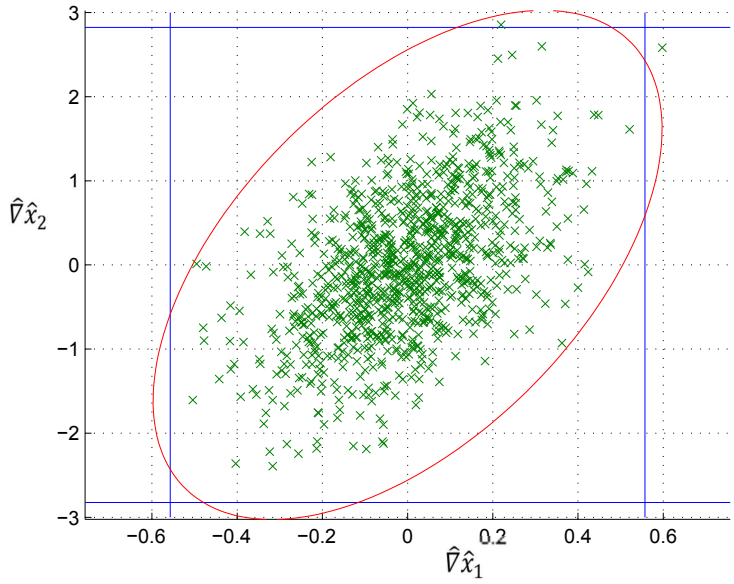


Figure I.5: Detection regions of SS test and UMPI test in the position domain, case PRN5 and PRN7 possibly faulty. On the axes the SS statistics defined in Equation (4.51). The blue rects define the SS thresholds, whereas the red ellipse define the UMPI statistic threshold ($\underline{T}_q = k$ propagated in the position domain by Equation (I.2)). The scatter represents the distribution under the null hypothesis of no failures.

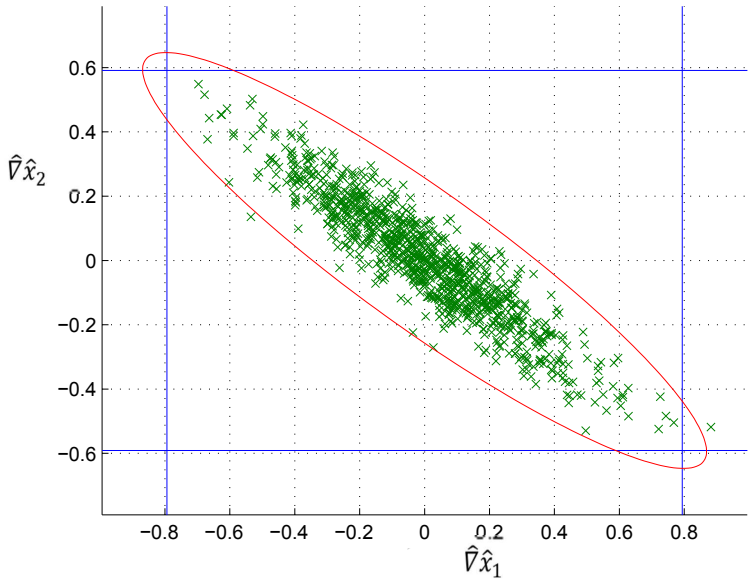


Figure I.6: Detection regions of SS test and UMPI test in the position domain, case PRN26 and PRN28 possibly faulty. On the axes the SS statistics defined in Equation (4.51). The blue rects define the SS thresholds, whereas the red ellipse defines the UMPI statistic threshold ($\underline{T}_q = k$ propagated in the position domain by Equation (I.2)). The scatter represents the distribution under the null hypothesis of no failures.

couple of satellites failing is considered, but two different vector biases are analyzed, a first one such that $\nabla\hat{x} = [k_1 \quad k_2]^T$ and a second one $\nabla\hat{x} = [-k_1 \quad k_2]^T$, where k_1 and k_2 are the threshold values for the SS tests. Therefore shown are the distribution in specific alternative

hypothesis of fault. The black lines in figure are indeed the thresholds k_1 and k_2 of the SS tests, whereas the red lines may represent the Alert Limits. The pink ellipse and scatter represent the distribution of the corresponding statistic $\hat{\nabla}\hat{x}$, whereas the blue ellipse and scatter represent the distribution of the positioning error $\hat{x}_0 - x$.

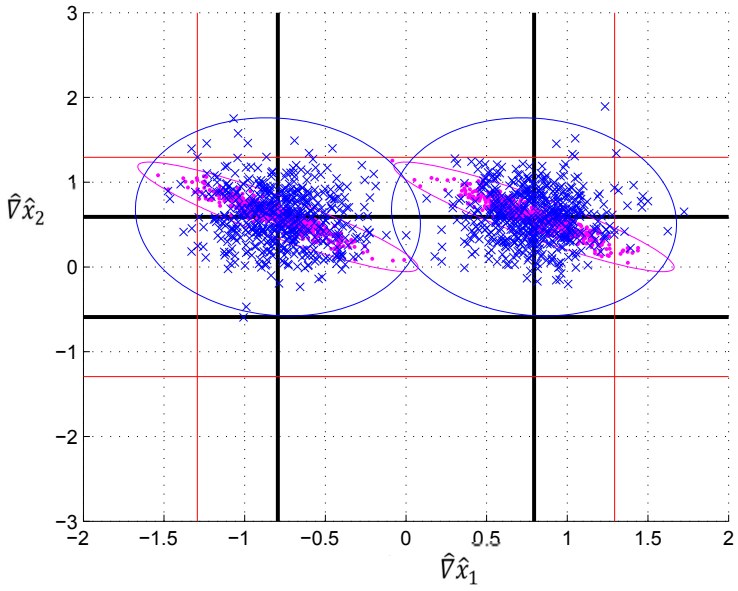


Figure I.7: Distributions of SS test statistics and of position error $\hat{x}_0 - x$, case PRN26 and PRN28 faulty by two different bias vectors, $\nabla\hat{x} = [k_1 \;\; k_2]^T$ and $\nabla\hat{x} = [-k_1 \;\; k_2]^T$. On the axes the SS statistics defined in Equation (4.51). The black straight lines define the SS thresholds, whereas the red straight lines define the ALs. The pink scatters represent the distributions of the SS statistic $\hat{\nabla}\hat{x}$ whereas the blue scatter represents the distribution of the error $\hat{x}_0 - x$.

The probability P_{PF} that the position error is bigger than the AL (occurrence named 'Positioning Failure') is the area of the blue ellipses that exceeds the red lines AL. The probability P_{MD} that a test leads to acceptance (so that no Alert is raised) is in case of SS test the area of the pink ellipses that is inside the black lines (thresholds of the SS test).

The P_{HMI}, under assumption of occurrence of an alternative hypothesis, is the product of the two, $P_{\text{HMI}} = P_{PF} \times P_{MD}$ (see also Equation (3.30)), as long as the two events, positioning failure and acceptance of the test, are uncorrelated. This is always the case when using the UMPI or the SS test statistics, since both statistics are linear combinations of $\hat{e}_0$, which is uncorrelated with $\hat{x}_0$.

It is evident how the simple rectangular shape of the detection region of the SS test cannot take into account the different shapes of blue and pink ellipses, and the P_{HMI} associated to the two faults, characterized by very similar P_{PF}, are very different (in particular, the P_{MD} are very different). In fact, in the case on the left almost half of the pink ellipse and scatter is inside the detection region, whereas on the right the fraction inside is much smaller. With an optimal test instead, we would expect to detect an equally dangerous failure occurrence with the same power. From this point of view, also the UMPI test is not optimal. In fact for integrity purposes we may be interested not in detecting a whatever fault but only a fault which is actually dangerous in terms of its effect on the position domain.

I.7. NOTE ON TU DELFT DIA PROCEDURE

The UMPI test for GNSS applications was introduced by the TU Delft school and employed in the DIA procedure. As mentioned before in Section I.2, the Delft theory includes the reliability theory (Equations (I.5) and (I.8)), that allows to monitor the effects of observations biases on the position domain. This allows to tailor the thresholds choice and de-weight anomalies that have only little effect on the position domain, in similar way as the SS algorithms. The DIA procedure is described in detail in Chapter 5, as well as in [5], [97] and [101].

I.8. CONCLUSIONS

The comparison between SS and UMPI tests led to the following results:

- In case of $q = 1$, one-dimensional threat, SS and UMPI tests are equivalent in the strong majority of the cases. Testing on each of the SS component is in reality equivalent to testing only one of them, the one creating the tightest bound, since the bias has only one dimension. The only cases in which SS and UMPI tests can lead to very different results occur when the vector C_y is perpendicular to the space $R(A)$, in which cases the SS would always accept the null hypothesis whereas the UMPI w-test can still reject it. On the other hand these cases of exact perpendicularity are supposed to be very rare, and they can be furthermore considered a limiting case for the use of the w-test, a case for which the threshold for the w-test should be set to infinity (to achieve equivalence). Therefore, when reserving the possibility of setting the threshold for the w-tests to infinity, UMPI and SS can be considered equivalent in the one-dimensional threat case.

- In case of $q > 1$, SS and UMPI tests lead to different results. The main difference between the two approaches lies in the fact that the SS statistics looks only at the effect of possible outliers in the position domain, in such a way that the outliers or faults that have no influence on the position solution will be completely neglected. This difference is especially sensible when C_y spans any of the base vectors of the space perpendicular to $R(A)$, but also when $q > n$. In these cases biases that have no impact on the position solution can grow indefinitely in some 'directions' without being detected by the SS, while being spotted by the UMPI. In the other cases the difference between the two tests lies in the shape of the detection region, which is ellipsoidal (for instance in the position bias domain) in the UMPI case while bounded by pair of parallel planes in case of the SS.

- The numerical examples presented show that both the two simple detection regions determined by the two tests cannot take into account the distributions of both test statistic and positioning error, failing to monitor the product $\beta \cdot P_{PF}$, the contribution of an alternative hypothesis to the P_{HMI}. This means they are sub-optimal in identifying situations of dangerous biases in the position domain. The determination of the optimal test statistic and detection region for the integrity problem is not a trivial task and can be the subject of further research. UMPI and SS can be adopted as simplyfied tests as long as the integrity risk is monitored consistently within.

- The UMPI test presents the advantage of monitoring more effectively the probability of False Alarm and related performance parameters of the test. On the other hand it

is less selective regarding the anomalies monitored. Being even more selective than the UMPI test, the SS test may suffer more in case of testing for fault signatures C_y which are not perfectly modelling the actual anomaly (for instance when testing for a $C_y \perp R(A)$ when in fact $C'_y \sim C_y$ is occurring). Therefore it is recommended (both in case of adoption of SS or UMPI), to always couple the specific faults tailored tests with an Overall Model or Omnibus test (F-test) of general consistency. The TU Delft theory couples the UMPI test with the Overall Model Test (OMT) and with a consistent reliability monitoring: by making a design with acceptable external reliability (for the application at hand), it is assured that the adoption of the UMPI test automatically satisfies the testing for the model errors that are most hazardous.

To summarize, whereas the two tests are equivalent for the one-dimensional threat case, for the multi-dimensional threat case the SS focuses only on the detection of faults that affect the position solution, whereas the UMPI test is instead most powerful in detecting any kind of fault. When adopting the SS test therefore, one needs to beware that some faults can pass completely undetected because of small or no influence in the position domain, and measure the consequences of this in the specific application considered.

J

ARAIM P_{HMI} COMPUTATION

In this Appendix we explain the derivation of the ARAIM formula for the $P_{\text{HMI}_{ver}}$ (and VPL) given in Equation (4.40). The basic concepts for the ARAIM P_{HMI} computation were presented in Section 4.2.1.

We start from the upperbound given in the first line of Equation (4.39):

$$P_{\text{HMI}|H_i} = P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} \cap \underline{T}_{SS_i} \in \Omega_{T_i} | H_i) \leq P(\underline{\hat{x}}_0 - x \notin \Omega_{\text{AL}} | \underline{T}_{SS_i} \in \Omega_{T_i}, H_i)$$

If we consider only one component, say the vertical one (3^{rd} component of the position vector), we can write:

$$\begin{aligned}
&P_{\text{HMI}_{ver}|H_i} = P(|\underline{\hat{x}}_{0,3} - x_3| > AL \cap \underline{T}_{SS_i} \in \Omega_{T_i} | H_i) \leq \\
&P(|\underline{\hat{x}}_{0,3} - x_3| > AL \mid \underline{T}_{SS_i} \in \Omega_{T_i}, H_i) \leq \\
&P(|\underline{\hat{x}}_{0,3} - x_3| > AL \mid |\underline{\hat{x}}_{0,3} - \underline{\hat{x}}_{i,3}| < k_{i,3}, H_i) = \\
&P(|\underline{\hat{x}}_{0,3} - x_3 + \underline{\hat{x}}_{i,3} - \underline{\hat{x}}_{i,3}| > AL \mid |\underline{\hat{x}}_{0,3} - \underline{\hat{x}}_{i,3}| < k_{i,3}, H_i) \leq \\
&P(|\underline{\hat{x}}_{i,3} - x_3| > AL - k_{i,3} \mid |\underline{\hat{x}}_{0,3} - \underline{\hat{x}}_{i,3}| < k_{i,3}, H_i) = \\
&P(|\underline{\hat{x}}_{i,3} - x_3| > AL - k_{i,3} \mid H_i)
\end{aligned}$$

the next to the last passage being valid because the condition sets a bound on the possible values of $|\underline{\hat{x}}_{0,3} - \underline{\hat{x}}_{i,3}|$; last passage instead employs the fact that the statistics $\underline{\hat{x}}_{i,3}$ and $\underline{\hat{x}}_{0,3} - \underline{\hat{x}}_{i,3}$ are uncorrelated. This way it is now possible to compute this contribution to the P_{HMI} just using the distribution of $\underline{\hat{x}}_{i,3}$, which is unbiased when the alternative hypothesis H_i holds.

K

EXAMPLE ARAIM VS DIA IN P_{HMI} COMPUTATION

A simple example is developed to compare the ARAIM upperbound to the P_{HMI} in Equation (4.40) (described in Section 4.2.1) with the worst-case bias upperbound adopted in the (adapted) DIA procedure, discussed in Section 6.2 and given in Equation (6.8) (both approaches are developed to tackle the issue of the unknown nature of the bias size in the alternative hypothesis). Consider the null hypothesis:

$$E(\underline{y}) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} x \tag{K.1}$$

and four alternative hypotheses foreseeing an anomaly in one of the four measurements (one-dimensional faults only). Furthermore $\underline{y} \sim N(E(\underline{y}), I_4)$. Each alternative hypothesis has a prior probability of occurrence p_a whereas the null hypothesis has prior probability $1-4p_a$.

For this simple configuration we compute the $P_{\text{HMI}_{araim}}$ using the ARAIM upperbound discussed in Appendix J and the $P_{\text{HMI}_{wc}}$ using the worst-case bias approach as developed in Chapter 6. We set the thresholds for $\underline{T}_{SS}$ and w-tests in such a way to guarantee the same P_{FA} (x is one-dimensional so SS tests and w-tests are equivalent).

The $P_{\text{HMI}_{araim}}$ is computed as in Equation (4.40), simplified for this special case (and without approximating the probability of occurrence of the null hypothesis to 1):

$$P_{\text{HMI}_{araim}} \leq 2(1-4p_a)\Phi\left(\frac{-AL}{\sigma_{\hat{x}_0}}\right) + \sum_{i=1}^{4} p_a \Phi\left(\frac{-AL+k_i}{\sigma_{\hat{x}_i}}\right) \tag{K.2}$$

The $P_{\text{HMI}_{wc}}$ is computed as (see Section 6.4, Equation (6.8)):

$$\begin{gathered} P_{\text{HMI}_{wc}} \leq 2(1-4p_a)(1-2\Phi(-k_w))\Phi(\tfrac{-AL}{\sigma_{\hat{x}}}) + \textstyle\sum_{i=1}^{4} p_a \cdot \\ \max_{\nabla_i} \left\{ [\Phi(k_w - \nabla w_i) - \Phi(-k_w - \nabla w_i)] \cdot \left[\Phi\left(\frac{SC_{y_i}\nabla_i - AL}{\sigma_{\hat{x}}}\right) + \Phi\left(\frac{-SC_{y_i}\nabla_i - AL}{\sigma_{\hat{x}}}\right) \right] \right\} \end{gathered} \tag{K.3}$$

with $S = (A^T Q_{yy}^{-1} A)^{-1} A^T Q_{yy}^{-1}$ and C_{y_i} the canonical unit vectors. Note that $1-2\Phi(-k_w)$ is an upperbound for the probability of not rejecting the null hypothesis in case the null hypothesis holds true (a positioning failure can occur also under the null hypothesis). Table K.1

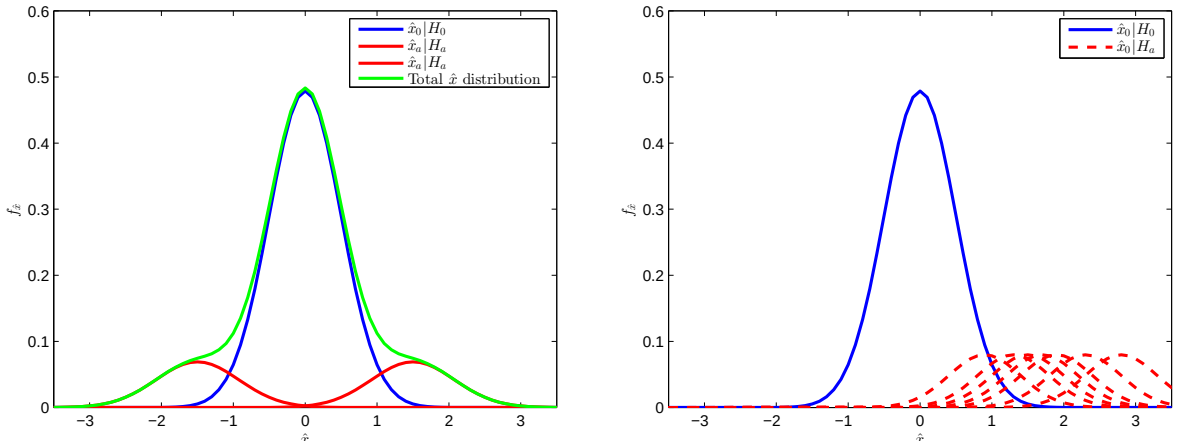


Figure K.1: ARAIM (left) and worst-case bias (right) concepts for the computation of the P_{HMI}. In the ARAIM case we can see the distribution of $\underline{\hat{x}}_0$ under the null hypothesis in blue and the distributions of $\underline{\hat{x}}_a$, in red, shifted from the espectation of $\underline{\hat{x}}_0$ by the most extreme value of the SS test statistic that still lead to acceptance of the null hypothesis. When we assume a-priori that such threshold value for the test statistic will certainly occur, then we can employ the combined distribution in green to compute the P_{HMI}, as the integral outside the AL (centered at 0). The graph for the worst-case bias approach simply shows instead the distribution of $\underline{\hat{x}}_0$ under the null hypothesis in blue and some of the infinite number of distributions of $\underline{\hat{x}}_0$ under the alternative hypotheses in dashed red, to indicate that such distribution is in fact unknown. In this approach we compute P_{MD}, P_{PF} and consequently P_{HMI} for all the distributions in the alternative hypotheses, and choose the one that maximizes the P_{HMI} (which is not just the integral of the distribution outside the AL, but has to take into account also the probability of detecting the fault).

shows the results for different input values of p_a and P_{FA} (P_{FA} refers to each test, not to the combination of all 4 tests), setting the AL always to ±3. The results show that the upperbound obtained through worst-case bias approach is sensibly tighter than the ARAIM upperbound (one order of magnitude smaller). Figure K.1 shows the conceptual difference between the two approaches. On the left is the ARAIM concept whereas on the right is the worst-case bias concept. In the ARAIM case the sum of the distributions of $\underline{\hat{x}}_0$ and $\underline{\hat{x}}_a$ is considered, with $f_{\underline{\hat{x}}_a}$ biased by an amount equal to the threshold for the SS test statistic, and the P_{HMI} is the integral of such sum of distributions outside the AL. In the worst-case bias approach instead only the distribution of $\underline{\hat{x}}_0$ is analyzed: in the null hypothesis as it is fully known, and in the alternative hypotheses considering many possible bias sizes, for each bias size computing the integral outside the AL (P_{PF}) and the P_{MD} (not represented). The bias size that yields the larger product $P_{MD} \cdot P_{PF}$ is finally chosen as worst-case bias, to compute the P_{HMI}.

Table K.1: P_{HMI} computed through ARAIM approach and worst-case bias approach, AL always set to ±3.

P_{FA}	p_a					
	$p_a = 0.1$		$p_a = 0.01$		$p_a = 0.001$	
	$P_{\text{HMI}_{araim}}$	$P_{\text{HMI}_{wc}}$	$P_{\text{HMI}_{araim}}$	$P_{\text{HMI}_{wc}}$	$P_{\text{HMI}_{araim}}$	$P_{\text{HMI}_{wc}}$
0.05	$2.2 \cdot 10^{-5}$	$9.3 \cdot 10^{-7}$	$2.2 \cdot 10^{-6}$	$9.5 \cdot 10^{-8}$	$2.2 \cdot 10^{-7}$	$1.1 \cdot 10^{-8}$
0.01	$7.7 \cdot 10^{-5}$	$3.7 \cdot 10^{-6}$	$7.7 \cdot 10^{-6}$	$3.7 \cdot 10^{-7}$	$7.7 \cdot 10^{-7}$	$3.9 \cdot 10^{-8}$
0.001	$2.9 \cdot 10^{-4}$	$1.6 \cdot 10^{-5}$	$2.9 \cdot 10^{-5}$	$1.6 \cdot 10^{-6}$	$2.9 \cdot 10^{-6}$	$1.7 \cdot 10^{-7}$

L

CONDITIONAL DISTRIBUTION OF THE W-TESTS

In the DIA algorithm, described in Chapter 5, the w-tests are run only after the OMT led to rejection. The distribution of the w-tests that are run after detection, therefore, will be a conditional distribution, different from the a-priori distribution in Equation (5.3). Here we want to investigate this conditional distribution, to determine probabilities of a missed *identification* or wrong identification (Wrong Detection) after the Detection step.

First of all we note that each $\underline{\hat{\nabla}}_{w_i}$ of any w-test can be expressed as a linear combination of the general bias vector estimator for the OMT, $\underline{\hat{\nabla}}$ (obtained from a full rank $m \times (m-n)$ matrix C_y), or as a linear combination of the misclosures vector $\underline{t}$:

$$\underline{\hat{\nabla}}_{w_i} = L_i \underline{\hat{\nabla}} = K_i \underline{t} \tag{L.1}$$

where L_i and K_i are $1 \times (m-n)$ matrices. $\underline{w_i}$ can be written then:

$$\underline{w_i} = \frac{1}{\sigma_{\hat{\nabla}_{w_i}}} L_i \underline{\hat{\nabla}} = \frac{1}{\sigma_{\hat{\nabla}_{w_i}}} K_i \underline{t} \tag{L.2}$$

Let us consider first the distribution of $\underline{\hat{\nabla}}$ in case the null hypothesis holds true (the distribution obtained can be used to determine the probability of False Alarm). We have:

$$H_0: \quad \underline{\hat{\nabla}} \sim N(0, Q_{\hat{\nabla}\hat{\nabla}}) \tag{L.3}$$

that is, a multivariate normal distribution. The conditional distribution $f_{\underline{\hat{\nabla}}|T_{m-n}>k_{\text{OMT}}}$ under H_0 is obtained simply by:

$$f_{\underline{\hat{\nabla}}|T_{m-n}>k_{\text{OMT}}} = \begin{cases} \frac{f_{\underline{\hat{\nabla}}}}{P(T_{m-n}>k_{\text{OMT}})} & \forall \hat{\nabla}: \; T_{m-n} > k_{\text{OMT}} \\ 0 & \forall \hat{\nabla}: \; T_{m-n} \leq k_{\text{OMT}} \end{cases} \tag{L.4}$$

which fundamentally is a truncated multivariate normal distribution, with zero probability density in the region of all $\hat{\nabla}$ for which $T_{m-n} < k_{\text{OMT}}$ (an ellipsoid centered at the origin). The probability $P(T_{m-n} > k_{\text{OMT}})$ at the denominator is computed exploiting the known distribution of $\underline{T}_{m-n}$ given in Equation (3.23). We refer to this probability as α_{OMT}, since it is the significance of the OMT (we are considering the null hypothesis holding true).

Supposed we want to find the conditional distribution of a w-test statistic $\underline{w}_i$ as defined in Equation (L.2), we can first build a vector of $m-n$ independent w-test statistics, $\underline{w}$, extending the $1 \times (m-n)$ row matrix L_i to a full square matrix L with the addition of $m-n-1$ linearly independent rows such that:

$$\underline{w} = DL\underline{\hat{\nabla}} = G\underline{\hat{\nabla}} \tag{L.5}$$

where D is a diagonal matrix with diagonal elements the $\sigma^{-1}_{\hat{\nabla}_{w_i}}$, $D = \text{Diag}(\sigma^{-1}_{\hat{\nabla}_{w_i}})$, and L is a full rank square matrix. G is therefore a full rank invertible square matrix as well.

For this w-test statistics vector we have, similarly as for $\underline{\hat{\nabla}}$:

$$H_0: \quad \underline{w} \sim N(0, Q_{ww}) \tag{L.6}$$

with $Q_{ww} = GQ_{\hat{\nabla}\hat{\nabla}}G^T$. This multivariate normal distribution is also $m-n$ dimensional. Following, the w-tests vector conditional distribution in case of rejection of the OMT is:

$$f_{\underline{w}|T_{m-n}>k_{\text{OMT}}} = \begin{cases} \frac{f_{\underline{w}}}{\alpha_{OMT}} & \forall w: \ T_{m-n} > k_{\text{OMT}} \\ 0 & \forall w: \ T_{m-n} \leq k_{\text{OMT}} \end{cases} \tag{L.7}$$

again a truncated multivariate normal distribution, with zero probability density in the region of all w for which $T_{m-n} < k_{\text{OMT}}$ (an ellipsoid centered at the origin).

Inverting Equation (L.5) and combining with the definition of the OMT statistic in Equation (5.6), the OMT statistic can be written also as:

$$\underline{T}_{m-n} = \underline{w}^T G^{-T} Q^{-1}_{\hat{\nabla}\hat{\nabla}} G^{-1} \underline{w}$$

which means that the relation $T_{m-n} < k_{\text{OMT}}$ defines an ellipsoidal shape also for the $\underline{w}$ (a degenerate ellipsoid).

Given the (conditional) joint distribution function of the w-test statistics in Equation (L.7), the (conditional) marginal distribution of a single w-test statistic $\underline{w}_i$ can be obtained by partial integration of the joint distribution over the other w-tests variables $w_{j\neq i}$. That is:

$$f_{\underline{w}_i|T_{m-n}>k_{\text{OMT}}} = \int_{R^{m-n-1}} f_{\underline{w}|T_{m-n}>k_{\text{OMT}}} \, \text{d}w_{j\neq i} \tag{L.8}$$

With $w_{j\neq i}$ being the other w-tests different from w_i.

As a result of the above integration, the posterior distribution of the w-tests when H_0 holds true is:

$$f_{\underline{w}_i|T_{m-n}>k_{OMT}} = \begin{cases} \frac{N_{(0,1)}(w_i)}{\alpha_{OMT}} & |w_i| > \sqrt{k_{OMT}} \\ \frac{N_{(0,1)}(w_i)(1-\chi^{2cdf}_{(0,m-n-1)}(k_{OMT}-w_i^2))}{\alpha_{OMT}} & |w_i| \leq \sqrt{k_{OMT}} \end{cases} \tag{L.9}$$

The above formula can be explained as follows. Recall that the w-tests have a degenerate multivariate normal distribution with $m-n$ degrees of freedom as in Equation (L.6), and the distribution conditional on the rejection of the OMT is the same distribution but with null value in the centered ellipsoid (and opportunely normalized by α_{OMT}) as in Equation (L.7). To compute the marginal distribution of $\underline{w}_i$ a partial integration is made over the other $w_{j\neq i}$, as in Equation (L.8). This means that at each value w_i we want to integrate a multivariate

normal distribution with $m-n-1$ degrees of freedom, scaled by $\frac{N_{(0,1)}(w_i)}{\alpha_{OMT}}$ (with $N_{(0,1)}(w_i)$ the value of a standard unit normal distribution in w_i), outside the centered ellipsoid acceptance region of the OMT, that we indicate with Ω:

$$\frac{N_{(0,1)}(w_i)}{\alpha_{OMT}} \int_{R^{m-1}-\Omega} N_{m-n-1}(0, Q_{ww_{j\neq i}}) \tag{L.10}$$

The acceptance area of the OMT is characterized in R^m by the equation:

$$w^T G^{-T} Q^{-1}_{\hat{\nabla}\hat{\nabla}} G^{-1} w \leq k_{OMT}$$

When a value for w_i is given, the above relation traduces to, for the remaining $m-1$ w-tests $w_{j\neq i}$ in R^{m-1}:

$$w^{-T}_{j\neq i}(G^T Q^{-1}_{\hat{\nabla}\hat{\nabla}} G^{-1})_{j\neq i} w_{j\neq i} \leq k_{OMT} - \frac{w_i^2}{(G^{-T} Q^{-1}_{\hat{\nabla}\hat{\nabla}} G^{-1})_{ii}} = k_{OMT} - w_i^2$$

where we employed the fact that $(G^{-T} Q^{-1}_{\hat{\nabla}\hat{\nabla}} G^{-1})_{ii} = 1$. This is readily obtained from $L Q_{\hat{\nabla}\hat{\nabla}} L^T = Q_{\hat{\nabla}_w \hat{\nabla}_w}$, from $G = DL$ and $D = \text{Diag}(\sigma^{-1}_{\hat{\nabla}_{w_i}})$. Now we can solve the integral in Equation (L.10). In fact the integral of a multivariate normal distribution over an elliptic shape with the same proportions as the variance matrix of the distribution is simply obtained as a χ^2 cumulative distribution function (with same number of degrees of freedom as the normal distribution) computed at the threshold value of the inner product defining the ellipse, i.e. $k_{OMT} - w_i^2$. Finally the result as in Equation (L.9) is obtained.

An example of w-test-like conditional distribution obtained both with the formula in Equation (L.9) and numerically through Monte Carlo integration is shown in Figure L.1, compared with the standard normal distribution of an unconditioned w-test statistic. This conditional distribution is obtained starting from an hypothetical $\underline{\hat{\nabla}}_w$ with a multivariate standard normal distribution $N_3(0, I_3)$ truncated (null) in the sphere centered at the origin containing probability mass 0.6 ($\alpha_{OMT} = 0.4$). This distribution (which is a conditional distribution, following the rejection of the OMT) could be eventually obtained for instance from a linear system with zero unknown parameters and 3 observations.

As a result, the conditional distribution has the shape of a normal distribution that in the central region, corresponding to the projection of the sphere excluded from the multivariate distribution onto one of the 3 axes, keeps a constant value. This is anyway only a special case, since the function assumes a constant value in the central area only when $m-n=3$. In the general case the distribution will *not* be flat in the central area but will anyway reach lower values than the normal distribution.

An example of w-test conditional distribution obtained from an actual geometry is shown in Figure L.2. Also in this case the truncated area of the original joint distribution of the w-tests has a total probability mass of 0.6. The result is based on MC integration with 10^5 samples and on the formula in Equation (L.9).

In Figure L.3 it is shown instead an example of multivariate normal distribution (bivariate in this case) truncated in the central ellipsoid characterized by probability mass of 0.6. This is the type of joint distribution from which the w-test conditional distribution is obtained as a marginal distribution.

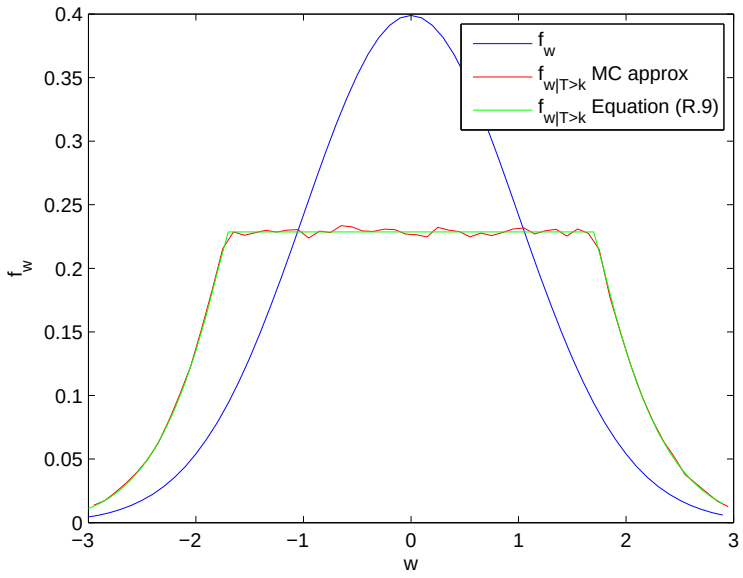


Figure L.1: Marginal distribution of w-test obtained from a truncated multivariate normal distribution, setting the significance of the OMT to 0.4 ($\alpha_{OMT} = 0.4$), through MC integration (10^6 samples) in red and through the formula in Equation (L.9) in green.

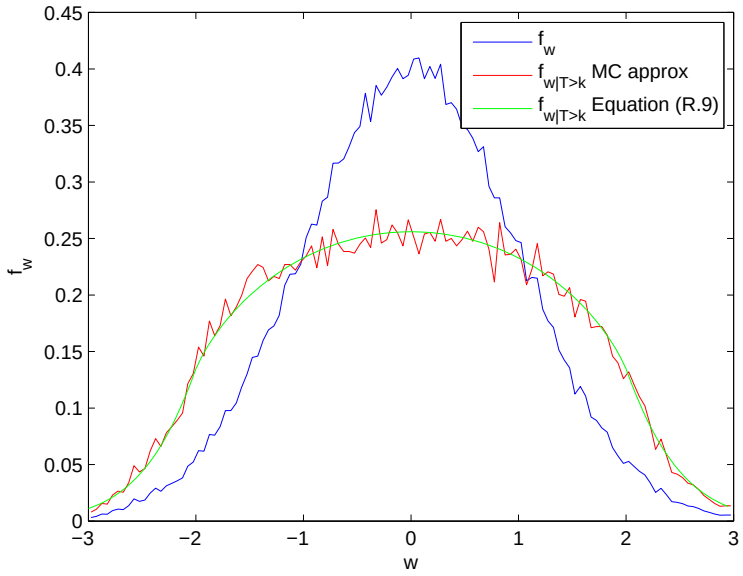


Figure L.2: Marginal distribution of w-test obtained from an actual satellite geometry, setting the significance of the OMT to 0.4 ($\alpha_{OMT} = 0.4$), through MC integration (10^5 samples) in red and through the formula in Equation (L.9) in green.

To determine the Probability of Missed Detection, we have now to consider the case of alternative hypothesis holding true. In this case we have:

$$H_a: \quad \underline{w} \sim N(\nabla w, Q_{ww}) \tag{L.11}$$

where ∇w is the unknown bias vector affecting the w-test statistics. The conditional distri-

bution after detection form the OMT is again a truncated multivariate normal distribution with zero probability density in the region of all w for which $T_{m-n} < k_{\mathrm{OMT}}$, the ellipsoid centered at the origin. An example of this type of distribution is shown in Figure L.4. In this case the problem will be more complex to solve because we have to compute the integral of a multivariate normal distribution over an ellipsoid area that is not centered at the origin (at the mean of the distribution).

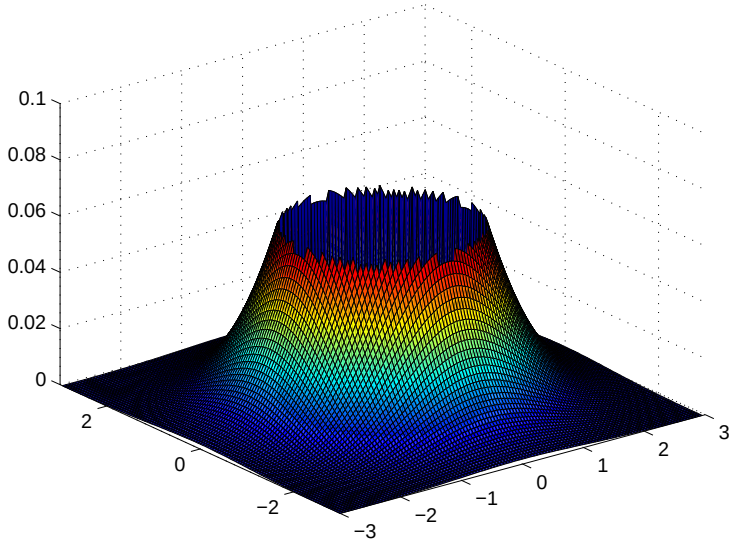


Figure L.3: Example of multivariate normal distribution centered at the origin truncated in the central ellipsoid (detection region of the OMT).

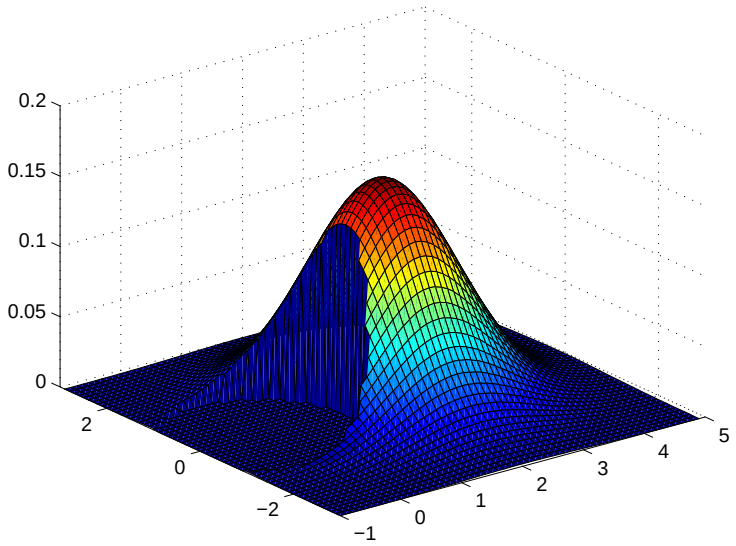


Figure L.4: Example of multivariate normal distribution displaced from the origin (representing the case of bias present in the measurements), truncated in the central ellipsoid (detection region of the OMT).

M

DIA INTEGRITY RISK CURVES

The geometry in Figure 5.8 is here considered to show the behaviour of the integrity performance parameters for the adapted DIA algorithm analyzed in Chapter 6. The same Q_{yy} as in Section 5.3.3 is assumed.

In Figures M.1 and M.2 we plot the probability of a PF, P_{PF}, however only considering the vertical dimension, $P(|\underline{\hat{x}}_3 - x_3| > AL)$, against the probability of missed detection β (for the w-tests). One satellite only, PRN7 and PRN21 respectively in Figure 6.1 and 6.2, is considered in each graph. Each blue curve stands for a different choice of test threshold, that is for a different choice of α for the w-test. The significance α, for each w-test individually, ranges from 0.0357 to 2×10^{-9}. The red curves are iso-curves, that is curves for which the product $\beta \times P(|\underline{\hat{x}} - x| > AL)$ is constant. If $\underline{\hat{x}}$ and $\underline{w}$ are independent, which is the case when $\underline{y}$ is normally distributed, it holds:

$$P_{\text{HMI}|H_i} = P(|\underline{\hat{x}} - x| > AL|H_i) \times \beta \tag{M.1}$$

where $P_{\text{HMI}|H_i}$ is the probability of Hazardous Misleading Information (HMI) given the fault in satellite i has occurred. Therefore the iso-curves are characterized by constant conditional P_{HMI}.

In Figures M.3 the conditional $P_{\text{HMI}|H_i}$ as computed in Equation (6.1) is plotted against the size of the bias in the faulty satellite, for different choices of the threshold (each blue line represents a different threshold), for satellite PRN21 (equivalent graph as in Figure 6.2, where PRN7 was shown instead).

In Figure M.4 each w-test's α is fixed at the value $\alpha = 10^{-3}$ and all the 8 satellites in view are considered. For each satellite the integrity risk $P_{\text{HMI}|H_i}$ is shown as function of the bias ∇_i. Each curve refers hypothesis H_i, failure in the corresponding i satellite.

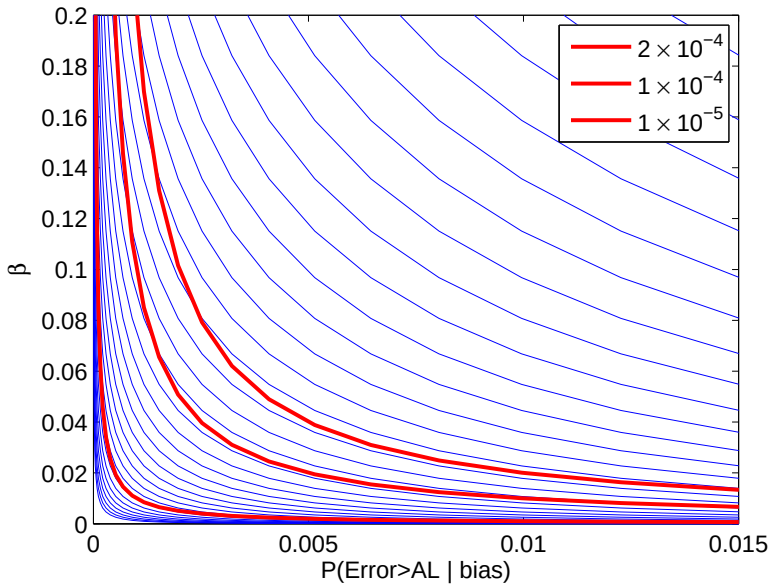


Figure M.1: Probability of a Positioning Error (PE) in the vertical dimension, $P(|\underline{\hat{x}} - x| > AL|H_i)$, with respect to the probability of MD β, for satellite PRN7 (in case of its fault). Each blue curve is for a different threshold, that is for a different choice of α, which determines the continuity of the algorithm. The significance α, for the w-test individually, ranges from 0.0357 to 2×10^{-9}. The red curves are iso-curves, for which the product $\beta \times P(|\underline{\hat{x}} - x| > AL|H_i)$ is constant.

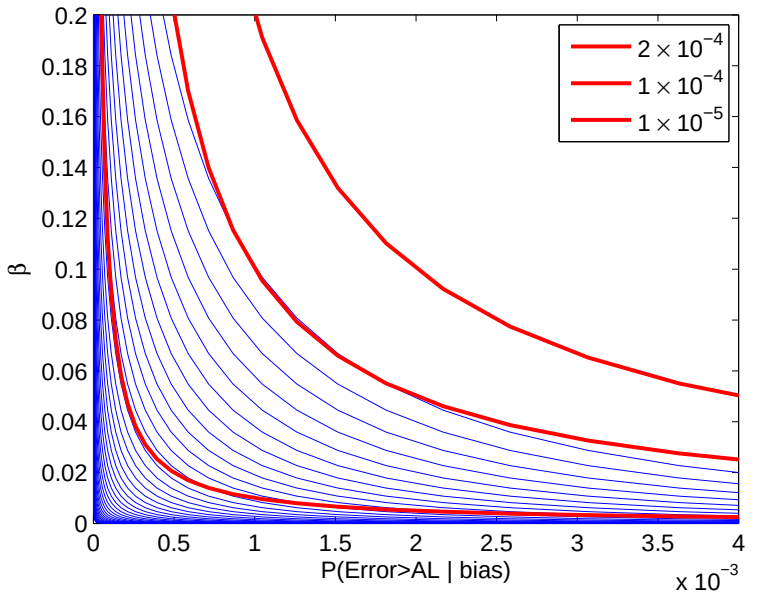


Figure M.2: Probability of a PE in the vertical dimension, $P(|\underline{\hat{x}} - x| > AL|H_i)$, with respect to the probability of MD β, for satellite PRN21 (in case of its fault). Each blue curve is for a different threshold, that is for a different choice of α, which determines the continuity of the algorithm. The significance α, for the w-test individually, ranges from 0.0357 to 2×10^{-9}. The red curves are iso-curves, for which the product $\beta \times P(|\underline{\hat{x}} - x| > AL|H_i)$ is constant.

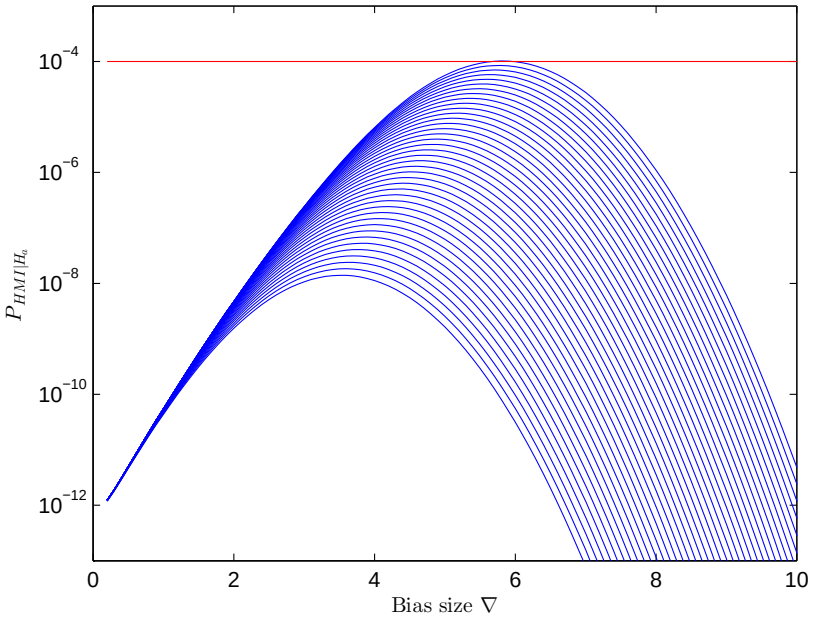


Figure M.3: $P_{\text{HMI}|H_i}$ as a function of the size of the bias in faulty satellite, for different choices of the threshold, for satellite PRN21. AL=4m. Each blue line represents a different threshold, i.e. a different choice for α, which determines the continuity of the algorithm. The significance α, for the w-test individually, ranges from 0.0357 to 2×10^{-9}. Each curve clearly exhibits a maximum.

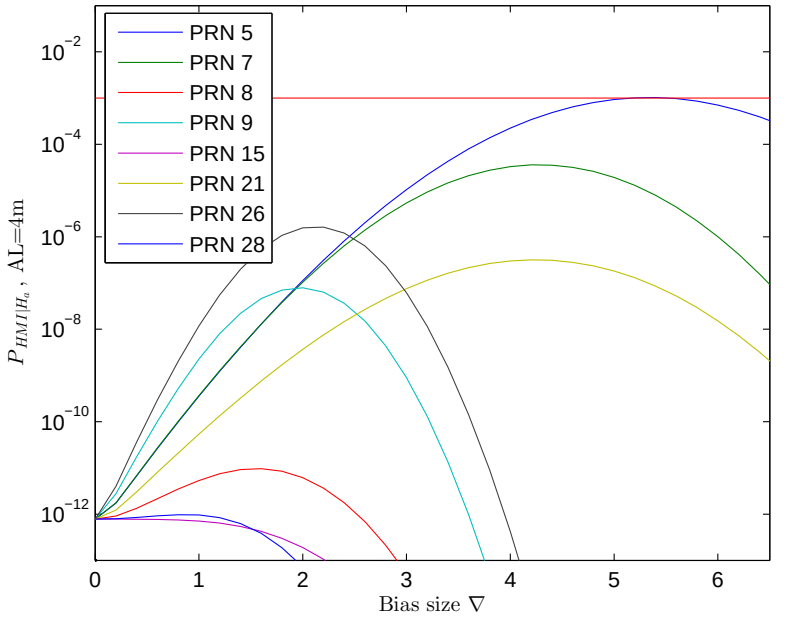


Figure M.4: $P_{\text{HMI}|H_i}$ as a function of the size of the bias in the faulty satellite, with a fixed $\alpha = 10^{-3}$ and therefore fixed threshold for the w-tests, for the 8 satellites in view, for the Up direction. Approximation considering one alternative hypothesis at a time, in line with Equation (5.37). AL=4m. Each curve clearly exhibits a maximum.

BIBLIOGRAPHY

REFERENCES

[1] Akaike H., *A New Look at the Statistical Model Identification*, IEEE Transactions on Automatic Control, Vol. AC-19, No. 6, pp. 716-723, 1974.

[2] Anderson T. W., *The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities*, Proc. Almer. Math. Soc. No. 6, pp. 170-176, 1955.

[3] Arnold S. F., *The theory of linear models and multivariate analysis*, Series in probability and mathematical statistics, John Wiley & Sons, New York, 1981.

[4] Baarda W., *Statistical Concepts in Geodesy*, Netherlands Geodetic Commission, Publications on Geodesy, New Series, vol. 2, no. 4, 1967.

[5] Baarda W., *A testing procedure for use in geodetic networks*, Netherlands Geod-etic Commission, Publications on Geodesy, New Series, vol. 2, no. 5, 1968.

[6] Barber J. J. and K. Ogle, *To P or not to P?*, Ecology, no. 95 (3), pp. 621-626, the Ecological Society of America, 2014.

[7] Berger J. O., *Statistical Decision Theory and Bayesian Analysis*, Springer Series in Statistics Second Edition, Springer-Verlag New York, 1985.

[8] Blanch J., A. Ene, T. Walter and P. Enge, *An Optimized Multiple Hypothesis RAIM Algorithm for Vertical Guidance*, in proceedings of the 20^{th} International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2007), Fort Worth Convention Center, Fort Worth, TX, 2007.

[9] Blanch J., T. Walter and P. Enge, *RAIM with Optimal Integrity and Continuity Allocations Under Multiple Failures*, IEEE Transactions on Aerospace and Electronic Systems, Vol. 46, No. 3, 2010.

[10] Blanch J., T. Walter, P. Enge et al., *A Proposal for Multi-constellation Advanced RAIM for Vertical Guidance*, in proceedings of the 24^{th} International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2011), Oregon Convention Center, Portland, OR, 2011.

[11] Blanch J., T. Walter, P. Enge et al., *Advanced RAIM User Algorithm Description: Integrity Support Message Processing, Fault Detection, Exclusion, and Protection Level Calculation*, in proceedings of the ION Institute of Navigation Global Navigation Satellite Systems Conference, Nashville, TN, 2012.

[12] Blanch J., T. Walter and P. Enge, *Optimal Positioning for Advanced Raim*, in proceedings of the 2013 International Technical Meeting of The Institute of Navigation, in Navigation, Vol. 60, Issue 4, pp. 279-289, 2013.

[13] Blanch J., T. Walter and P. Enge, *Results on the Optimal Detection Statistic for Integrity Monitoring*, in proceedings of the 2013 International Technical Meeting of The Institute of Navigation, San Diego, California, pp. 262-273, 2013.

[14] Blanch J., T. Walter and P. Enge, *Exclusion for Advanced RAIM: Requirements and a Baseline Algorithm*, in proceedings of the Institute of Navigation (ION) International Technical Meeting, January 2014, San Diego, California, 2014.

[15] Brenner M., *Implementation of a RAIM Monitor in a GPS Receiver and an Integrated GPS-IRS*, proceedings of the Third Technical Meeting of the Satellite Division of The Institute of Navigation, Colorado Springs, Colorado USA, pp. 397-406, 1990.

[16] Brown R. G. and P. Y. C. Hwang, *GPS Failure Detection by Autonomous Means Within the Cockpit*, proceedings of the Annual Meeting of The Institute of Navigation, Seattle, Washington USA, pp.5-12, June 24-26, 1986.

[17] Brown R. G. and P. W. McBurney, *Self-Contained GPS Integrity Checks Using Maximum Solution Separation as the Test Statistic*, in proceeding of the First Technical Meeting of the Satellite Division of the Institute of Navigation, Colorado Springs, Colorado USA, pp. 263-268, 1987.

[18] Brown R. G., *A Baseline GPS RAIM Scheme and a Note on the Equivalence of Three RAIM Methods*, Navigation: Journal of The Institute of Navigation, Vol. 39, No. 3, 1992.

[19] Chan, F.-C., M. Joerger, S. Khanafseh and B. Pervan, *Bayesian Fault-Tolerant Position Estimator and Integrity Risk Bound for GNSS Navigation*, Journal of Navigation, Vol. 67, No. 5, September 2014.

[20] Chang L., *Monitoring civil infrastructure using satellite radar interferometry*, Ph.D. thesis, Delft University of Technology, 2015. `http://repository.tudelft.nl/`.

[21] Chatfield C., *Model uncertainty, data mining and statistical inference (with discussion)*, J. Roy. Statist. Soc., A, No. 158, pp. 419Ű466, 1995.

[22] Datta-Barua S., *Ionospheric Threats to the Integrity of Airborne GPS Users*, Ph.D. dissertation, Stanford University, Stanford, CA, 2007.

[23] DeCleene B., *Defining pseudorange integrity — Overbounding*, Proceedings of the ION GPS-2000.

[24] De Heus H. M., M. H. F. Martens, H. M. E. Verhoef, *Stability-Analysis as Part of the Strategy for the Analysis of the Groningen Gas Field Levellings*, Proceedings of the Perlmutter Workshop on Dynamic Deformation Models, Haifa, Israel, pp. 259-272, 1994.

[25] Edgar C., F. Czopek and B. Barker, *A Co-operative Anomaly Resolution on PRN-19*, proceedings of the ION GPS, 1999.

[26] Ene A., J. Blanch and J. D. Powell, *Fault Detection and Elimination for Galileo-GPS Vertical Guidance*, in proceedings of the 2007 National Technical Meeting of The Institute of Navigation, San Diego, CA, pp. 1244-1254, 2007.

[27] Ene A., *Utilization of Modernized Global Navigation Satellite Systems for Aircraft-based Navigation Integrity*, Dissertation, Department of Aeronautics and Astronautics, Stanford University, California USA, March 2009.

[28] Enge P., *Local Area Augmentation of GPS for the Precision Approach of Aircraft*, Proceedings of the IEEE, vol. 87, no. 1, pp. 111-132, 1999.

[29] Federal Aviation Authority (FAA) Technical Center, *Global Positioning System (GPS) Standard Positioning Service (SPS) performance analysis report*, Rep. 42. [Online] Available: `http://www.nstb.tc.faa.gov/REPORTS/PAN42_0703.pdf`, 2003.

[30] Federal Aviation Authority (FAA), *Next Generation Air Transportation System (NextGen)*, `http://www.faa.gov/about/initiatives/nextgen`

[31] Fontana R. D. et al., *The new L2 civil signal*, Proceedings of the ION GPS, 2001.

[32] Forsythe G. E., G. H. Golub, *On the stationary values of a second-degree polynomial on the unit sphere*, SIAM J. No. 13, pp. 1050-1068, 1965.

[33] Forum, *P values, hypothesis testing, and model selection: it's déjà vu all over again*, Ecology, 95(3), pp. 609-653, bt The Ecological Society of America, 2014.

[34] Förstner W., *Reliability and Discernability of Extended Gauss-Markov Models*, DGK Reihe A Heft 98, pp. 79-103, Munich, 1983.

[35] Gordon S., C. Sherrell and B. J. Potter, *WAAS Offline Monitoring*, Proceedings of the 23^{rd} International Technical Meeting of The Satellite Division of The Institute of Navigation (ION GNSS 2010), Portland, OR, pp. 2021-2030, September 2010.

[36] Gratton L., M. Joerger, and B. Pervan, *Carrier Phase Relative RAIM Algorithms and Protection Level Derivation*, Journal of Navigation, Vol. 63, No.2, April 2010.

[37] Gratton L., R. Pramanik, H. Tang and B. Pervan, *Ephemeris Failure Rate Analysis and its Impact on Category I LAAS Integrity*, Proceedings of the 20th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2007), Fort Worth, TX, pp. 386-394, September 2007.

[38] Hansen A., T. Walter, D. Lawrence and P. Enge, *GPS Satellite Clock Event of SV#27 and Its Impact on Augmented Navigation Systems*, Proceedings of ION GPS-98, Nahville, TN, September 1998.

[39] Hatch R., *The Synergism of GPS code and carrier measurements*, Proceedings of the 3rd international Geodetic Symposium on Satellite Doppler Positioning, Vol. 2. pp. 1213-1231, Las Cruces - New Mexico, 1982.

[40] Hegarty C. J. and E. Chatre, *Evolution of the Global Navigation Satellite System (GNSS)*, Proceedings of the IEEE, Vol. 96, No. 12, pp. 1902-1917, December 2008.

[41] Hekimoglu S., Koch K. R., *How can reliability of the test for outliers be measured?*, Allgemeine Vermessungsnachrichten. VDE Verlag Beril Offenbach, S. 247-253, 2000.

[42] Heng L., G. X. Gao, T. Walter and P. Enge, *GPS Signal-in-Space Anomalies in the Last Decade: Data Mining of 400,000,000 GPS Navigation Messages*, Proceedings of the 23^{rd} International Technical Meeting of The Satellite Division of The Institute of Navigation (ION GNSS 2010), pp. 3115-3122, Portland, OR, September 2010.

[43] `http://www.esa.int/Our_Activities/Navigation/The_future_-_Galileo/Galileo_satellites`

[44] `http://www.sesarju.eu`

[45] `http://waas.stanford.edu/metrics.html`

[46] *ICAO*, Annex 10, Aeronautical Telecommunications, Volume 1 (Radio Navigation Aids), Amendment 84, published 20 July 2009, effective 19 November 2009. GNSS standards and recommended practices (SARPs) are contained in Section 3.7, Appendix B and Attachment D, 2009.

[47] Imparato D., *Detecting Multi-Dimensional Threats: a Comparison of Solution Separation Test and Uniformly Most Powerful Invariant Test*, in proceedings of the ENC-GNSS 2014 conference, Rotterdam, The Netherlands, 2014.

[48] *Interface Specification IS-GPS-200H*, 24/09/13 (2013). `http://www.gps.gov/technical/icwg/`

[49] *Interface Specification IS-GPS-705D*, 24/09/13 (2013). `http://www.gps.gov/technical/icwg/`

[50] Joerger M. and B. Pervan, *A Practical Approach to Optimal Estimator Design in RAIM*, Proceedings of the ION Institute of Navigation Global Navigation Satellite Systems Conference, September 8-9, Tampa, Florida, 2014.

[51] Joerger M. and B. Pervan, *Solution Separation and Chi-Squared ARAIM for fault detection and exclusion*, in Position, Location and Navigation Symposium — PLANS 2014, 2014 IEEE/ION, pp.294-307, 5-8 May 2014.

[52] Joerger M., F. Chan and B. Pervan, *Solution Separation Versus Residual-Based RAIM*, Journal of The Institute of Navigation, Vol. 61, No. 4, pp. 273-291, Winter 2014.

[53] Joerger, M., S. Stevanovic, S.Langel and B. Pervan, *Integrity Risk Minimization in RAIM, Part 1: Optimal Detector Design*, Journal of Navigation, FirstView article published online 6 January 2016, http://dx.doi.org/10.1017/S0373463315000983.

[54] Kenselaar F., *A Testing Procedure for Subsidence Analysis*, The 10th FIG International Symposium on Deformation Measurements, Orange, California, USA, pp.40-49, 2001.

[55] Koch K. R., *Introduction to Bayesian Statistics*, Springer Second Edition, Springer-Verlag Berlin Heidelberg, 2007.

[56] Lawrence D. et al., *Wide Area Augmentation System (WAAS) — Program Status*, Proceedings ION GNSS, 2007.

[57] Le, A. Q., and P. J. G. Teunissen, *Optimal Recursive Least-Squares Filtering of GPS Pseudorange Measurements*, VI Hotine-Marussi Symposium on Theoretical and Computational Geodesy, Springer Berlin Heidelberg, 2008.

[58] Lee W. C., *Analysis of Range and Position Comparison Methods as a Means to Provide GPS Integrity in the User Receiver*, proceedings of the Annual Meeting of The Institute of Navigation, Seattle, Washington USA, June 24-26, pp. 1-4, 1986.

[59] Lee Y. C., M. P. McLaughlin, *A Position Domain Relative RAIM Method*, Proceeding of IEEE/ION PLANS 2008, Monterey, CA, pp. 271-284, May 2008.

[60] Lee Y. C., *New Techniques Relating Fault Detection and Exclusion Performance to GPS Primary Means Integrity Requirements*, Proceeding of the 8th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1995), Palm Springs, CA, September 12-15, 1995.

[61] Lee Y., K. Van Dyke, B. DeCleene, J. Studenny and M. Beckmann, *Summary of RTCA SC-159 GPS Integrity Working Group Activities*, NAVIGATION, Journal of The Institute of Navigation, Vol. 43, No. 3, Fall 1996, pp. 307-362, 1996.

[62] Lehmann E. L., G. Casella, *Theory of Point Estimation*, Springer Texts in Statistics Second Edition, Springer-Verlag New York, 1998.

[63] Lehmann R., *On the formulation of the alternative hypothesis for geodetic outlier detection*, Journal of Geodesy (2013) no. 87 pp. 373-386, Springer-Verlag Berlin Heidelberg, 2012.

[64] Mallows C. L., *Some Comments on* $\underline{C}_p$, Technometrics 15 (4), pp. 661-675, 1973.

[65] Massat P. and M. Zeitzew, *The GPS constellation design — current and projected*, in proceedings of ION National Technical Meeting 'Navigation 2000', Long Beach, California, January 21-23, pp. 435-445, 1998.

[66] Maxwell S. E., H. D. Delaney, *Designing Experiments and Analyzing Data: A Model Comparison Perspective*, Second Edition, Lawrence Erlbaum Associates, ISBN 0-8058-3718-3, 2004.

[67] Miller A., *Subset Selection in Regression*, Chapman and Hall/CRC, Second Edition, 2002.

[68] Miller R. G., *Simultaneous Statistical Inference*, Springer Series in Statistics, Springer-Verlag New York Heidelberg Berlin, Second Edition, 1981.

[69] Misra P. and P. Enge, *Global Positioning System Signals, Measurements, and Performance*, Ganga-Jamuna Press, Second Edition, P.O. Box 692 Lincoln, Massachusetts 01773, 2006.

[70] Montenbruck O., A. Hauschild, P. Steigenberger and R. B. Langley, *Three's the Challenge*, GPS World, July 2010.

[71] Montgomery D. C., G. C. Runger, *Applied Statistics and Probability for Engineers*, Wiley 6th Edition, 2014.

[72] *MOPS for Global Positioning System/Wide Area Augmentatation System Airborne Equipment*, RTCA DO-229D, version D-13/12/06, prepared by SC-159, 2006.

[73] Murphy T. et al., *Implementation and Operational Use of Ground Based Augmentation Systems (GBASs) — A Component of the Future Air Traffic Management System*, Proceedings of IEEE, vol. 96, no. 12, pp. 1936-1957, Dec. 2008.

[74] Murphy T., *GNSS Positioning and Future trends in Air Traffic Management*, Proceedings of ION-GNSS 2009, September 22-25, Savannah, Georgia, pp. 3560-3578, 2009.

[75] Murtaugh P. A., *In defense of P values*, Ecology, no. 95 (3), pp. 611-617, the Ecological Society of America, 2014.

[76] Nadarajah S., *A generalized normal distribution*, Journal of Applied Statistics, no. 32(7), pp. 685-694. doi:10.1080/0266470500079464, 2005.

[77] Neyman J. and E. S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 231, pp. 289-337, 1933.

[78] Ober P. B., *Integrity Prediction and Monitoring of Navigation Systems*, Ph.D. thesis (TU Delft), Integricom Publishers, Leiden, The Netherlands, 2006.

[79] Oehler V., F. Luongo, H. L. Trautenberg, J-P Boyero, J. Krueger and T. Rang, *The Galileo Integrity Concept and Performance*, in proceedings of ENC-GNSS 2005, July 19-22, Munchen, Germany, pp. 11, 2005.

[80] Tiberius C. C. J. M. and D. Odijk, *Does the HPL bound the HPE?*, In J-L Issler & M Meurer (Eds.), NAVITEC 2008, 4^{th} ESA workshop on satellite navigation user equipment technologies (pp. 1-8), Noordwijk (ESA), 2008.

[81] Pervan S. B., S. P. Pullen and J. R. Christie, *A Multiple Hypothesis Approach to Satellite Navigation Integrity*, Navigation, Vol. 45, Issue 1, pp. 61-71, 1998.

[82] Pesonen H., *Bayesian Estimation and Quality Monitoring for Personal Positioning Systems*, Julkaisu-Tampere University of Technology, Publications 112, 2013.

[83] Porretta M., W. Schuster, A. Majumdar and W. Ochieng, *Strategic Conflict Detection and Resolution Using Aircraft Intent Information*, The Journal of Navigation, Vol. 63, No. 1, pp. 61-88, 2010.

[84] Rife J., S. Pullen, B. Pervan, and P. Enge, *Paired overbounding and application to GPS augmentation*, Position Location and Navigation Symposium, 2004. PLANS 2004, pp. 439-446, 26-29 April 2004.

[85] Rife J., P. Enge, B. Pervan, *Paired overbounding for nonideal LAAS and WAAS error distributions*, IEEE Transactions on Aerospace and Electronic Systems, Vol. 42, Issue 4, pp. 1386 - 1395, October 2006.

[86] Rivers M. H., *2 SOPS Anomaly Resolution on an Aging Constellation*, Proceedings of the 13^{th} International Technical Meeting of the Satellite Division of The Institute od Navigation (ION GNSS 2000), pp. 2547-2550, Salt Lake City, UT, September 2000.

[87] Salabert F., A. Hendriks, R. Rawlings and R. Farnworth, *EUROCONTROL Policy on GNSS*, Proceedings of ION-GNSS 2008, September 16-19, Savannah, Georgia, 2008.

[88] Salabert F., *Operational Benefits of Multi Constellation Dual Frequency GNSS for Aviation*, Coordinates, March 2015. `http://mycoordinates.org/operational-benefits-of-multi-constellation-dual-frequency-gnss-for-aviation/`

[89] Salgado G., S. Abbondanza, R. Blondel, S. Lannelongue, *Constellation Availability Concepts for Galileo*, in proceedings of ION National Technical Meeting 'Navigation 2001', Long Beach, California, January 22-24, pp. 778-786, 2001.

[90] Scheffé H., *The Analysis of Variance*, John Wiley & Sons, Inc., New York, 1959.

[91] Shank C. M. and J. Lavrakas, *GPS Integrity: An MCS Perspective*, Proceedings of ION GPS-1993, Salt Lake City, UT, September 1993.

[92] Speidel J., M. Tossaint, S. Wallner, J.Á. Ávila-Rodrìguez, *Integrity for Aviation*, InsideGNSS, July/August 2013, pp. 54-64, 2013.

[93] Spilker J. J., *GPS Signal Structure and Theoretical Performance*, in Global Positioning System: Theory and Applications, vol. I, Reston, VA: AIAA, ch. 3, 1996.

[94] Spjøtvoll E., *Multiple Comparison of Regression Functions*, Institute of Mathematical Statistics, The Annals of Mathematical Statistics, Vol. 43, No. 4, pp. 1076-1088, 1972.

[95] Sturza M. A., *Navigation System Integrity Monitoring Using Redundant Measurements* Navigation: Journal of The Institute of Navigation, Vol. 35, No. 4, 1988-89.

[96] Tarantola A., *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, 2005.

[97] Teunissen P.J.G., *An Integrity and Quality Control Procedure for use in Multi Sensor Integration*, Proceedings of the 3rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1990), September 19-21, The Broadmoor Hotel Colorado Spring, CO 1990.

[98] Teunissen P.J.G., *Adjustment Theory* Series on Mathematical Geodesy and Positioning, VSSD, 2000.

[99] Teunissen P.J.G., *Testing Theory* Series on Mathematical Geodesy and Positioning, VSSD, 2000.

[100] Teunissen P.J.G. , *Integer estimation in the presence of biases*, Journal of Geodesy (2001) 75: pp. 399-407, Springer-Verlag, 2001.

[101] Teunissen P.J.G., D.G. Simons, C.C.J.M. Tiberius, *Probability and Observation Theory*, Lecture Notes, TU Delft, 2009.

[102] Teunissen P.J.G., *Mixed Integer Estimation and Validation for Next Generation GNSS*, Handbook of Geomathematics by W. Freeden, M.Z. Nashed, T. Sonar (Eds.), Springer-Verlag Berlin Heidelberg, pp. 1101-1128, 2010.

[103] US Department of Defense, *Global Positioning System Standard Positioning Service Performance Standard*, September 2008.

[104] Van Dierendonck A. J. et al., *Signal Specification for the Future GPS Civil Signal at L5*, Proceedings of the ION Annual Meeting, 2000.

[105] Verhagen S. and P. Joosten, *Algorithms for design computations for integrated GPS-Galileo.*, in proceedings of ENC-GNSS 2003, April 22-25, Graz, Austria, pp. 14, 2003.

[106] Volpe National Transportation System Center, *Vulnerability assessment of the transportation infrastructure relying on the Global Positioning System*, final report, August 29, 2001.

[107] Walter T. and P. Enge, *Weighted RAIM for Precision Approach*, Proceedings of the ION GPS-95, Palm Springs, FL, 1995.

[108] Walter T. and M. B. El-Arini, Eds., *Selected Papers on Satellite Based Augmentation Systems*, ION Redbook, vol. VI, 1999.

[109] Walter T., J. Blanch, P. Enge, B. Pervan, L. Gratton, *Future Architecture to Provide Aviation Integrity*, Proceedings of the ION National Technical Meeting, 2008.

[110] Walter T., P. Enge, J. Blanch and B. Pervan, *Worldwide Vertical Guidance of Aircraft Based on Modernized GPS and New Integrity Augmentations*, Proceedings of the IEEE, Vol. 96, No. 12, December 2008.

[111] Walter T., J. Blanch and P. Enge, *Evaluation of Signal in Space Error Bounds to Support Aviation Integrity*, Journal of The Institute of Navigation Vol. 57, No. 2, Summer 2010.

[112] Warren D. L. M. and J. F. Raquet, *Broadcast Vs. Precise GPS Ephemerides: A Historical Perspective*, GPS Solutions, vol. 7, no. 3, December 2003.

About the Author

Davide Imparato was born in Ortona (CH), Italy, 1985. He obtained both Bachelor's and Master degree in Aerospace Engineering (with specialization in Electrical and Chemical Propulsion) at the University of Pisa, respectively in 2008 and 2010. He completed his Master degree thesis out of a six-month intership in Rolls-Royce plc, at its headquarters in Derby, UK, during the year 2010. After the completion of his Master degree, he worked for 8 months for Assystem UK, in Derby, as a consultant for Rolls-Royce plc. From July 2011 till April 2016 he worked as a Ph.D. student at TU Delft, on the topic of Aircraft Navigation Integrity, in the Geoscience and Remote Sensing Department of the Faculty of Civil Engineering and Geosciences.