Universitat Autònoma de Barcelona

Change Detection Techniques for GNSS Signal-Level Integrity

Daniel Egea-Roca

Universitat Autònoma
de Barcelona

Ph.D. Dissertation

Change Detection Techniques for GNSS Signal-Level Integrity

Author: Daniel Egea-Roca

Thesis Advisors: José A. López-Salcedo
Gonzalo Seco-Granados

Department of Telecommunications and Systems Engineering
Universitat Autònoma de Barcelona (UAB)

Bellaterra, June 2017

Abstract

The provision of accurate positioning is becoming essential to our modern society. One of the main reasons is the great success and ease of use of Global Navigation Satellite Systems (GNSSs), which has led to an unprecedented amount of GNSS-based applications. In particular, the current trend shows that a new era of GNSS-based applications and services is emerging. These applications are the so-called critical applications, in which the physical safety of users may be in danger due to a miss-performance of the positioning system. These applications have very stringent requirements in terms of integrity. Integrity is a measure of reliability and trust that can be placed on the information provided by the system. Integrity algorithms were originally designed for civil aviation in the 1980s. Unfortunately, GNSS-based critical applications are usually associated with terrestrial environments and original integrity algorithms usually fail. The main impairments are due to local effects such as interference, multipath or spoofing, which are assumed to be controlled in civil aviation but they are not in terrestrial environments. Thus, a new methodology for integrity is necessary in order to detect local effects and provide the additional level of integrity needed for GNSS-based critical applications; the so-called signal-level integrity.

This thesis investigates novel detection algorithms with the aim of providing a new generation of integrity techniques in GNSS. For this purpose, the framework of Statistical Change Detection (SCD) is considered. This framework is of particular interest because its optimal criterion target the temporal dimension. This is an indispensable requirement for critical applications, in which a prompt detection is necessary. This approach is in contrast to classical detection schemes in which the probability of missing a detection is minimized, thus disregarding the temporal dimension. Thereby, the first part of this dissertation deals with the study of the field of SCD, including both Quickest Change Detection (QCD) and Transient Change Detection (TCD). Novel contributions are provided in the field of TCD, including the finite moving average solution and its statistical characterization. Numerical results show the superiority of our contributions. Finally, to conclude our study of SCD we compare it with classical detection schemes under the same mathematical framework. This comparison shows the appropriateness of SCD when dealing with timely detections.

The main contribution of this thesis is provided in the second part by applying the framework of SCD to threat detection and integrity in GNSS. To this end, we first investigate several properties of the received GNSS signal that may be useful for local threat detection. This leads us to move a step forward in the field of threat detection by proposing a novel QCD-based framework to interference and multipath detection in GNSS. QCD aims at detecting a threat as soon as possible, which is of interest to monitor whether a threat is present or not so that a prompt alert may be raised. Nonetheless, for integrity purposes a bounded delay is desirable, and

it is here where TCD is of interest. For this reason, a novel TCD-based framework is considered for both multipath detection and integrity algorithms in GNSS, thus leading to the provision of signal-level integrity. Finally, the second part of this dissertation also investigates the practical implementation of signal-level and current integrity algorithms. A notable improvement is shown by the proposed TCD-based solutions considered in this thesis with respect to current solutions.

In the last part of the thesis, the goal is to validate the proposed threat detectors and signal-level integrity algorithm using real GNSS signals. For this purpose, a MATLAB-based software is developed to process gathered GNSS signals with the aim of detecting the presence of interference and multipath. Real signal gathered in the context of an EC-funded research project is processed to show the results of the implemented detectors. Two data collection campaigns are processed to validate the behavior of the proposed detectors in real working conditions. The work is completed with the validation of the navigation accuracy of the proposed signal-level integrity algorithm. The results obtained in a realistic scenario show the improvement of the accuracy and integrity by using the proposed solution for signal-level integrity, with respect to current integrity algorithms. Furthermore, the proposed solution is shown to have real-time processing capabilities, thus being very attractive to improve current integrity algorithms and easily implementable in mass-market receivers.

Resumen

El gran éxito y la facilidad de uso de los sistemas de navegación global por satélite (GNSSs) ha conducido a la definición de una gran cantidad de aplicaciones basadas en GNSS sin precedentes. De hecho, la tendencia muestra que actualmente está emergiendo una nueva era de aplicaciones basadas en GNSS, las denominadas aplicaciones críticas. En estas aplicaciones, la integridad física de los usuarios puede estar en riesgo en caso de un fallo del sistema de posicionamiento. Un requisito importante para estas aplicaciones es la integridad, definida como una medida de la fiabilidad y confianza que se tiene en la información proporcionada por el sistema. Los primeros algoritmos de integridad aparecieron en la década de los 80s y fueron diseados para trabajar en entornos aéreos, en concreto para aviación civil. Desafortunadamente, las aplicaciones críticas basadas en GNSS suelen estar asociadas con entornos terrestres, totalmente distintos a los entornos aéreos y por lo tanto los algoritmos de integridad tradicionales suelen fallar. El principal motivo por el cual suelen fallar los algoritmos de integridad actuales en los entornos urbanos son los efectos locales como interferencias, multicamino o el denominado spoofing. Estos efectos se asumen que están controlados en aviación civil, pero ese no es el caso en entornos terrestres. De este modo, se necesitan nuevas técnicas de integridad para aplicaciones críticas basadas en GNSS, la denominada integridad a nivel de seal (signal-level integrity).

Esta tesis investiga nuevos algoritmos de detección con el objetivo de proporcionar una nueva generación de técnicas de integridad en GNSS. Para ello, se considera el campo de detección de cambios estadísticos (SCD). Este campo es de interés porque considera la dimensión temporal, requisito indispensable para aplicaciones críticas ya que una detección rápida es necesaria. Por lo tanto, la primera parte de esta tesis se ocupa del estudio del campo de SCD, incluyendo tanto la deteccin rápida de cambios (QCD) como la detección de cambios transitorios (TCD). Se aportan nuevas contribuciones en el campo de TCD, incluyendo la denominada solución FMA y su caracterización estadística. Además, resultados numéricos muestran la superioridad de nuestras contribuciones con respecto otras contribuciones en la literatura de TCD. Finalmente, para concluir nuestro estudio de SCD, lo comparamos con esquemas clásicos de detección bajo el mismo marco matemático. Esta comparación muestra la conveniencia de SCD cuando se trata de detecciones rápida.

La principal contribución de esta tesis es la aplicación del campo de SCD a la detección de amenazas e integridad en GNSS. Para ello, primero investigamos varias propiedades de la seal GNSS que pueden ser de utilidad para la detección de amenazas locales. En segundo lugar, damos un paso adelante en el campo de detección de amenazas en GNSS proponiendo un nuevo marco basado en QCD. Sin embargo, para fines de integridad es deseable un retardo limitado y es aquí donde la teoría de TCD es interesante. Por esta razón, se considera un nuevo marco basado en TCD para la detección de multi-camino y algoritmos de integridad en GNSS, lo que conduce

a la provisión de la integridad de nivel de seal. Se muestra una mejora notable por la soluciones propuestas de TCD con respecto a las soluciones actuales. En la última parte de la tesis, se validan los detectores de amenazas y el algoritmo de integridad a nivel de seal propuestos. Esto se hace utilizando seles GNSS reales capturadas en el contexto de un proyecto de investigación financiado por la Comisión Europea. Los resultados obtenidos en un escenario realista muestran la mejora de la precisión y la integridad mediante el uso de la solución propuesta con respecto a los algoritmos de integridad actuales. Además, se muestra que la solución propuesta trabaja en tiempo real, siendo por lo tanto muy atractiva para mejorar los algoritmos de integridad actuales y fácilmente implementables.

Acknowledgements

This Ph.D. dissertation ends an ever-dreamed journey full of good experiences, unforgettable moments and deep learning. The next lines put an end to one of the most important periods in my life. Honestly, in the last two years I have been doing what I have always dreamed. So, before starting with the dissertation I would like to look back and express my personal acknowledgement and admiration to all those who have done possible this wonderful trip.

First of all, I would like to thank my supervisors Prof. José A. López-Salcedo and Prof. Gonzalo Seco-Granados, they have been guilty of everything I have achieved in this trip called PhD. This thesis would have not been possible without them. From José, I really appreciate the every-day examples he always had in our meetings, which made much easier to understand things. His didactic view of abstract concepts is fantastic. From Gonzalo, I appreciate his view of the world, his in another dimension. It is like when you try to catch a fly but it always runs away; they see far beyond most human being see. In general, I would like to thank their countless suggestion and detailed corrections, they have always been there for me. You always know when you come in into their office, but you never know when you go out, regardless time, day or month. Secondly, I have also to express my deepest gratitude to my supervisors for trusting on me since the very beginning, they gave me the opportunity of starting a PhD, until the very end by giving me the opportunity of going abroad, not only once or twice, but thrice; incredible. For this, I would also like to thank H. Vincent Poor and Sunwoo Kim for hosting me in Princeton University and Hanyang University, respectively. Beyond being formidable researchers they are better persons, what made me to have a wonderful and fruitful experience in the USA and South Korea.

This Ph.D. thesis has also allowed me to meet lovely people, such as my current and former colleagues of the SPCOMNAV group and of the Dept. of Telecommunications at UAB. I would like to especially thank to those colleagues that started with me this trip, Alfred, Pau, Sergi and Martí. The fruitful discussions and conversations at lunch time have made me life easier. I would also like to especially thank the Princeton and Hanyang crew for making me easier my stays and make me feel like in home. You are great guys. Finally, and particular care, I would like to express my gratitude to the current SPCOMNAV team, Vicente, Toni, Sergi, David, Angélica, Yi and Joseé del Peral, for the fruitful discussions about GNSS related (and non-related) things and for bearing me all this time.

Last but not least, I really thank the support provided by my family and friends during these years at the university, helping me to overcome any adversities and to enjoy life outside work and research.

Daniel Egea-Roca

June 29, 2017

This work was supported by the Spanish Ministry of Economy and Competitiveness project TEC2014-53656-R.

Contents

List of Figures

List of Tables

Acronyms

2D, 3D	Two-Dimensional, Three-Dimensional
(C)(F)DMA	(Code) (Frequency) Division Multiple Access
(G)(W)LR(T)	(Generalized) (Weighted) Likelihood Ratio (Test)
(H)(V)PL	(Horizontal) (Vertical) Protection Level
(I)(M)AT	(Interference) (Multipath) Analysis Tool
(L)LR(T)	(Log-) Likelihood Ratio (Test)
(M)(G)EO	(Medium) (Geostationary) Earth Orbit
(M)(V)(MV)C	(Mean) (Variance) (Mean and Variance) Change
(N)(U)DP	(Non-) (Undetectable) Dominant Path
(N)CI	(Non-) Coherent Integration
(N)LOS	(Non-) Line Of Sight
(P)(T)(G)DOP	(Position) (Time) (Geometry) Dilution of Precision
(Q)(T)CD	(Quickest) (Transient) Change Detection
(S)(G)(A)BAS	(Satellite) (Ground) (Aircraft)Based Augmentation System
(S)(M)C-MA	(Single) (Multi) Correlator- Multi Antenna
(S)LCA	(Safety) Liability-Critical Applications
(V)(H)AL	(Vertical) (Horizontal) Alert Limit
ACF-CUSUM	ACF-Based CUSUM
ACF	Auto-Correlation Function
ADAS	Advanced Driver Assistance System
ADC	Analog-to-Digital Converter
AGC	Automatic Gain Control
ANPR	Automatic Number Plate Recognition
ARAIM	Advanced RAIM
AWGN	Additive White Gaussian Noise
C/A	Coarse-Acquisition
C/N_0	Carrier-to-Noise-Density Ratio
cdf	cumulative distribution function
CLT	Central Limit Theorem
CW	Continuous Wave
D(GPS)(GNSS)	Differential (GPS) (GNSS)
DLL	Delay Lock Loop
DME	Distance Measuring Equipment
DoA	Direction of Arrival
DSP	Digital Signal Processing
DSRC	Dedicated Short Range Communication

DSSS	Direct Sequence Spread Spectrum
EGNOS	European Geostationary Navigation Overlay Service
EMSL	European Microwave Signature Laboratory
EVT	Extreme Value Theory
FD(E)(I)	Fault Detection (and Exclusion) (and Isolation)
FSS	Fixed Sample Size
GCS	Ground Control Segment
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GSM	Global System for Mobile communications
GUI	Graphical User Interface
Hist-CUSUM	Histogram-Based CUSUM
HPE	Horizontal Positioning Error
ICAO	International Civil Aviation Organization
iGNSSrx	Integrity GNSS Receivers project
iid	independent and identically distributed
INR	Interference-to-Noise Ratio
ION	International Of Navigation
JRC	Joint Research Center
KL	Kullback-Leibler
Kurt-CUSUM	Kurtosis-Based CUSUM
LBS	Location-Based Services
LNA	Low Noise Amplifier
LS(R)	Least Squares (Residual)
MDB	Minimum Detectable Bias
MLE	Maximum Likelihood Estimate
NCO	Numerically Controlled Oscillator
NP	Neyman-Pearson
O-CUSUM	Offset-CUSUM
OBU	On-Board Unit
OCR	Optical Character Recognition
P(MD)(FA)	Probability of (Missed Detection) (False Alarm)
pdf	probability density function
PFAP	PFA in a given Period
PLL	Phase Lock Loop
Power-CUSUM	Power-Based CUSUM
PRN	Pseudo-Random Noise
PUD	Probability of Unbounded Delay
PVT	Position Velocity Time
PVT+I^2	SLI and PVT integrity module
RAIM	Receiver Autonomous Integrity Monitoring
RFI	Radio-Frequency Interference
RNP	Required Navigation Performance
RUC	Road User Charging
SAR	Synthetic Aperture Radar

SCD	Statistical Change Detection
sdr	sequential detection rule
Sig-RAIM	Signal-Level RAIM
SLI	Signal-Level Integrity
SLM	Signal-Level Metric
SMR	Signal-to-Multipath Ratio
SNR	Signal-to-Noise Ratio
Spect-CUSUM	Spectrogram-Based CUSUM
sps	symbols per second
SS	Space Segment
SSE	Sum of Squared residual Errors
SVN	Satellite Vehicle Number
TACAN	TACtical Air Navigation
TFA	Time-Frequency Analysis
TOA	Time Of Arrival
TTA	Time to Alert
U.S.	United States
UERE	User Equivalent Range Error
ULA	Uniform Linear Array
US	User Segment
WAAS	Wide Area Augmentation System
WLC	Windowed-Limited CUSUM
WLS	Weighted Least Squares

Notation

In general, letters or symbols formatted in upper-case boldface denote matrices, in lower-case boldface denote vectors, in upper-case and lower-case italics denote random variables and scalars, respectively. The rest of the notation is described as follows:

$\mathbf{A}^*, \mathbf{A}^{\mathrm{T}}, \mathbf{A}^{\mathrm{H}}, \mathbf{A}^{-1}$	Complex conjugate, transpose, conjugate transpose (Hermitian), and inverse of matrix $\mathbf{A}$, respectively.
$\mathbf{A}^{\dagger}$	Moore-Penrose pseudo-inverse of matrix $\mathbf{A}$.
$[\mathbf{A}]_{n,k}$	The $[n, k]$ element of matrix $\mathbf{A}$.
$\mathbf{I}$	Identity matrix.
$\mathbf{1}$	Vector of ones.
$\|\|\mathbf{a}\|\|$	Norm 2 of vector $\mathbf{a}$.
$\|a\|$	Absolute value of scalar a.
$\mathrm{Re}\{\cdot\}, \mathrm{Im}\{\cdot\}$	Real and imaginary part operators.
$\lceil\cdot\rceil$, $\lfloor\cdot\rfloor$	Ceil and floor operators.
$\max\{a, b\}$	Maximum between a and b.
$\min\{a, b\}$	Minimum between a and b.
$\max_i\{\cdot\}$	Maximum among i.
$\min_i\{\cdot\}$	Minimum among i.
$\arg\max_x\{f(x)\}$	Value of x that maximizes $f(x)$.
$\arg\min_x\{f(x)\}$	Value of x that minimizes $f(x)$.
$\inf_i\{\cdot\}$	Infimum among i.
$\sup_i\{\cdot\}$	Supremum among i.
$\log$	Logarithm to the base 10.
$\ln$	Natural logarithm (i.e. base e).
$\delta(t)$	Dirac delta.
$\mathcal{A}$	Event or set A.
$\inf\{n : \mathcal{A}_n\}$	Infimum n such that the event $\mathcal{A}_n$ is true.
$\mathbb{1}_{\{\}}\mathcal{A}$	Indicator function of the event $\mathcal{A}$.
$\mathcal{A} \cup \mathcal{B}$	Union of event/set $\mathcal{A}$ and $\mathcal{B}$.
$\mathcal{A} \cap \mathcal{B}$	Intersection of event/set $\mathcal{A}$ and $\mathcal{B}$.
$\mathbb{R}$	Set of real numbers.
$\mathbb{Z}$	Set of integer numbers.
$\mathbb{N}$	Set of natural numbers.
$\doteq$	Defined as.
$x \approx a$	x is approximately equal to a.
$\{X_n\}_{n\geq 1}$	Sequence of random variables with index time $n \geq 1$.

$X \sim f(x)$	The random variable X follows the distribution given by the pdf $f(x)$.
$X \simeq a$	X is asymptotically equal to a.
$\mathbb{P}$	Probability measure.
$\mathrm{E}[\cdot]$	Expectation operator.
$\mathbb{P}_1$	Probability measure under hypothesis $\mathcal{H}_1$.
$\mathbb{P}_0$	Probability measure under hypothesis $\mathcal{H}_0$.
E_1	Expectation under hypothesis $\mathcal{H}_1$.
E_0	Expectation under hypothesis $\mathcal{H}_0$.
$\mathrm{D}_{\mathrm{KL}}(f_1, f_0)$	KL divergence of f_1 with respect to f_0.
P_{M}	PMD of a block-wise detection.
P_{F}	PFA of a block-wise detection.
$\mathrm{P}_{\mathrm{ud}}(T, m)$	PUD of the stopping time T and transient change length of m samples.
$\mathrm{P}_{\mathrm{fap}}(T, m_\alpha)$	PFAP for a given period of m_α samples of the stopping time T.
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean and variance given by μ and σ^2, respectively.
$\Phi(x)$, $\phi(x)$	Cdf and pdf of the standard Gaussian distribution, evaluated at x.
χ^2_k	Central chi-squared distribution with k degrees of freedom.
$\chi^2_k(\lambda)$	Non-Central chi-squared distribution with k degrees of freedom and non-centrality parameter λ.
$\Gamma_k(x; \lambda)$	Cdf of a non-central chi-squared distribution with k degrees of freedom and non-central parameter given by λ, evaluated at x.
$\xi_{X,n}$	n-th central moment of the random variable X.
$\mathrm{LR}(X_n); \mathrm{LR}(n)$	LR of the random variable X_n.
$\mathrm{LLR}(X_n); \mathrm{LLR}(n)$	LLR of the random variable X_n.
$\lambda_N(X)$	LR of N iid samples of the random variable X.
$\Lambda_N(X)$	LLR of N iid samples of the random variable X.

Chapter 1

Introduction

With the widespread deployment of terrestrial Global Navigation Satellite System (GNSSs) receivers (e.g. embedded in most smartphones) and the increasing interest in providing ubiquitous positioning, a new era of GNSS-based applications and services is emerging. Nowadays, GNSSs are the main systems to provide Position, Velocity and Timing (PVT) information anytime, anywhere around the Earth. The dependence of our modern society on GNSS technology is increasing over the years and currently many industrial sectors are actually heavily relying on GNSS. Indeed, by 2020, there will be a population of more than 6 billion GNSS receivers with revenues of nearly 50 billions of Euro worldwide, as reported in the 2015 GSA market report. In addition, an increasing interest has recently been placed on GNSS-based critical applications. These applications, which are often associated with terrestrial environments, have very stringent requirements in terms of *integrity*. The latter refers to the ability of the user receiver to guarantee the quality and trust of the information provided by the GNSS. Besides providing an accurate navigation solution, timely warnings must be provided to the user when the system should not be used. As a matter of fact, the disruption or miss-behavior of GNSS services can have a high economic impact and cause a major setback worldwide. Therefore, GNSS vulnerabilities constitute a major concern nowadays, and integrity is a key factor to drive GNSS further deployment and penetration into new markets.

The concept of GNSS integrity was originally developed for civil aviation in the 1980s. Particularly, various augmentation systems such as Satellite-Based Augmentation Systems (SBAS) and Receiver Autonomous Integrity Monitoring (RAIM) techniques have been developed to provide integrity to civil aviation since its inception. These techniques rely on information or redundant measurements coming from different satellites, which are feasible to obtain in civil aviation. Nonetheless, this is not the case in terrestrial environments, where a plethora of obstacles blocking the signal of different satellites may be present, so that redundancy coming from different satellites is compromised. Moreover, both SBAS and RAIM assume that local effects like multipath and interference have a controlled influence on the GNSS signal. This is actually the case for civil aviation but it is not for terrestrial environments. In the latter, local effects abound, so that position errors and, thus, integrity are strongly affected. This is why in terrestrial-critical applications, it is of paramount importance to promptly detect any possible anomaly that could be endangering the received GNSS signal. Otherwise, the safety and trust of the end-user position and time could be jeopardized. We refer to this capability as *signal integrity*, which is currently a concern within the GNSS community.

Multipath and interference are considered today as the major vulnerabilities on GNSS that can threaten the integrity of terrestrial users. In fact, significant efforts have been made to mitigate these impairments, leaving detection in the background. Notwithstanding, the detection process is actually as important, or even more important than mitigation. The reason is, on one hand, that before using mitigation techniques we need to know whether any impairment is present or not. Otherwise, the effects of mitigation techniques could seriously damage the GNSS signal of interest. On the other hand, mitigation is not always the best approach, since many times, just removing the affected measurements suffices. Consequently, there is a particular interest in interference and multipath detectors in order to guarantee the integrity of the GNSS signal. This is the reason for the recent interest of the GNSS community on the detection of these threats. So far, all the contributions available for threat detection in GNSS have adopted a classical block-wise detection framework, in which an entire batch of samples is processed at a time. The block-wise detection, though, is often not well suited to fulfill the requirements of critical applications, in which the key objective is to quickly detect the presence of a degrading effect instead of detecting that the previous block of samples is affected.

Unfortunately, this is something not explicitly targeted by traditional detection techniques, which focus instead on minimizing the probability of missing a detection, or maximizing the probability of detection within a block of samples, thus disregarding the temporal dimension. For this reason, a new methodology is needed when dealing with critical applications. In this thesis, a novel and attractive point of view of the detection problem is adopted; that is to focus on minimizing the detection delay. For the prompt detection of integrity threats, it is mandatory to formulate the problem under the so-called framework of Statistical Change Detection (SCD), which is a general framework that encompasses both Quickest Change Detection (QCD) and Transient Change Detection (TCD). In particular, QCD is aimed at minimizing the detection delay, and is thus a promising approach for the problem at hand. For integrity purposes, though, an acceptable detection delay is limited by the so-called time-to-alert. This kind of detection can be formulated in the framework of TCD. In this case, the criterion is to minimize the probability of late detections subject to a level of false alarms, which perfectly fits with GNSS integrity algorithms.

1.1 Motivation and Objectives

Local threats have a critical impact on the GNSS performance of conventional GNSS receivers. These effects prevent those receivers from providing the accurate and reliable positioning required in many of the emerging critical GNSS-based applications. Thus, advanced signal processing techniques for integrity monitoring are needed to guarantee the safely use of critical applications. In this line, the goal of this thesis is to investigate detection algorithms aimed at providing a new generation of integrity techniques. This new generation of techniques is intended to enhance the overall GNSS integrity, so that critical applications can be safely used in terrestrial environments. For this purpose, several signal-level metrics are studied to detect interferences and multipath, considering a new detection scheme based on SCD. This will lead to the so-called signal-level integrity, intended to provide the level of integrity currently needed for terrestrial environments. The goals of the research carried out in this dissertation are the following:

- Proposal of a set of observable metrics sensible to the presence of interference and multipath, and able to detect any kind of interference and multipath propagation. This will serve as input for the signal-level integrity detectors.
- Derivation of new bounds for the performance assessment of SCD techniques in the GNSS domain.
- Design of SCD algorithms adapted to the framework of signal-level GNSS integrity.
- Focus on low-complexity implementations, which could be feasibly implemented in current GNSS receivers in a real-time framework.
- Validation of the proposed algorithms using realistic data from a real measurement campaign.

1.2 Thesis Outline

The work of this dissertation has been disseminated in several publications, such as journals and international conference papers. These research contributions as well as the outline of this thesis are listed for every chapter.

Chapter 2 This chapter provides the introduction to the problem this thesis deals with; that is, we introduce the new era of location-based applications that are currently emerging. It is noted that these applications are heavily system-dependent and they have very stringent requirements in terms of accuracy and integrity, which need to be carefully addressed.

Chapter 3 This chapter introduces the principles of GNSS positioning and GNSS receivers. The accuracy, integrity and the positioning process in a GNSS is very sensitive to several threats, which are reviewed and analyzed in this chapter. This is important to understand the limitations of GNSS and to provide possible solutions to these threats, one of which is the adoption of the so-called integrity techniques. Once Integrity is carefully studied, we also provide a background on probability and detection theory.

Chapter 4 This chapter gives an overview of detection theory, which deals with the study of optimal decision making schemes. To understand the evolution of this theory and be able to select the proper approach for the problem at hand, a brief historical review is presented. Then, different detection schemes are summarized and investigated. The focus is placed on SCD, including both QCD and TCD, which is one of the goals of this dissertation. A comprehensive overview of QCD is provided with the aim to be used as a guideline for future applications. Regarding TCD, it is a field still actively investigated. This is why we introduce here its novel framework and we propose a solution to TCD problems, being the main theoretical contribution of this thesis. The optimality of this solution for the TCD problem is not proved, but it is shown to be the best candidate among all the available alternatives. Finally, a comparison between classical hypothesis testing and SCD is made under the same mathematical framework, showing that SCD is better when dealing with timely detections. Summarizing, the main result of this chapter is the provision of a comprehensive overview of QCD; the proposal of a novel solution,

and its performance bounds, for the new framework of TCD; and the comparison between classical detection and SCD under the same mathematical framework. Part of the results in this chapter have been published in the following international conferences and journal papers:

- D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, H. Vincent Poor, "When Statistical Change Detection Meets Classical Detection", *IEEE Transaction on Signal Processing (To be submitted), 2017.*
- D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, H. Vincent Poor, "Performance Bounds for Finite Moving Average Change Detection: Application to Global Navigation Satellite Systems", *IEEE Transaction on Signal Processing (Accepted with major changes on Jan 2017).*
- D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, "Comprehensive Overview of Quickest Detection Theory and its Application to GNSS Threat Detection", *Gyroscopy and Navigation*, vol. 8, no. 1, pp. 1–14, Jan 2017.
- D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, H. Vincent Poor, "A Finite Moving Average Test for Transient Change Detection in GNSS Signal Strength Monitoring", *Proceedings of IEEE Statistical Signal Processing Workshop (SSP)*, Jun 27 2016.
- D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, "On the Use of Quickest Detection Theory for Signal Integrity Monitoring in Single-Antenna GNSS Receivers", *Proceedings of International Conference on Localization and GNSS (ICL-GNSS)*, Jun 23 2015.

Chapter 5 This chapter proposes several signal metrics to be used for local threat detection, which is another objective of this thesis. This contribution provides insights on the design of interference and multipath detectors in GNSS receivers using either a single- or multiple-antennas. Based on the proposed metrics and their behavior, a novel QCD-based framework is proposed for threat detection in GNSS. The main result of this chapter is the derivation of several signal-level metrics for interference and multipath detection. The theoretical assessment of the behavior of these metrics is also given in this chapter. Part of the results of this chapter has been presented in the following conference papers:

- D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, "Transient Change Detection for LOS and NLOS Discrimination at GNSS Signal Processing Level", *Proceedings of International Conference on Localization and GNSS (ICL-GNSS)*, Jun 28 2016.
- D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, "A Quickest Detection Framework for Signal Integrity Monitoring in Low-Cost GNSS Receivers", *Proceedings of IEEE Vehicular Technology Conference (VTC) Fall*, Sep 06 2015.
- D. Egea, G. Seco-Granados, J. A. López-Salcedo, "Single- and Multi-Correlator Sequential Tests for Signal Integrity in Multi-Antenna GNSS Receivers", *Proceedings of International Conference on Localization and GNSS (ICL-GNSS)*, Jun 24 2014.
- D. Egea, J. A. López-Salcedo, G. Seco-Granados, "Interference and Multipath Sequential Tests for Signal Integrity in Multi-Antenna GNSS Receivers", *(Invited paper) Proceedings of IEEE Sensor Array and Multichannel (SAM) Workshop*, Jun 22 2014.

Chapter 6 This chapter investigates the practical implementation of integrity techniques at position level and its link with the proposed signal-level integrity algorithm. We first study current integrity techniques and compare them with novel techniques based on fundamentals in Chapter 4. Based on the results of this comparison we also provide the implementation of the proposed signal-level integrity algorithm, with a TCD framework. Finally, the connection of this framework and its integration with current integrity algorithms is shown. Part of the results of this chapter have been published in the following international conference papers and one of the above listed journal paper (i.e. IEEE TSP 2017):

- D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, "Sequential Change Detection for Next-Generation RAIM Algorithms", *Proceedings of International Technical Meeting of The Satellite Division of the Institute of Navigation (To be submitted at ION GNSS+)*, 2017.
- D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, M. Moriana, , E. Dominguez, E. Aguado, D. Lowe, D. Naberzhnykh, F. Dovis, I. Fernández-Hernández, J. P. Boyero, "Signal-level Integrity and Metrics Based on the Application of Quickest Detection Theory to Multipath Detection", *Proceedings of International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+) (Best Presentation/Paper Award)*, Sep 17 2015.

Chapter 7 This chapter describes the data collection campaign and software implemented to validate the proposed threat detectors and signal-level integrity algorithm with real GNSS signals gathered in the context of an EC-funded research project. Part of the results of this chapter have been published in the following international conference paper and the above listed conference paper (i.e. ION GNSS+ 2015):

- D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, E. Dominguez, E. Aguado, D. Lowe, D. Naberzhnykh, F. Dovis, I. Fernández-Hernández, J. P. Boyero, "Signal-level Integrity and Metrics Based on the Application of Quickest Detection Theory to Interference Detection", *Proceedings of International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+)*, Sep 18 2015.

Other contributions related with this dissertation During the PhD studies, other research contributions have been produced apart from the above listed. The following conference and journal paper have been published related to distributed detection and the application of the so-called Edgeworth approximation and extreme value theory:

- D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, "Inhomogeneous Quadratic Tests in Transient Signal Detection: Closed-Form Upper Bounds and Application in GNSS", *Digital Signal Processing, Elsevier (Accepted with major changes on May 2017).*
- D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, S. Kim, "Space-Time CUSUM for Distributed Quickest Detection Using Randomly Spaced Sensors Along a Path", *Proceedings of EUSIPCO 2017*, Sep 2017.

Chapter 2

A New Era of Location-Based Applications

2.1 Introduction

During the last decade the number of commercial applications that depend on some positioning system to provide accurate positioning has significantly increased. As these positioning systems are evolving, the level of requirements of location-based applications in terms of accuracy, availability and continuity are expected to become more stringent. In this line, integrity is an additional requirement to assure the validity of the positioning system data satisfying the performance requirements demanded by the user application. Integrity algorithms were initially designed to work in aviation applications, so providing integrity for land user operations becomes increasingly important. This is the case in applications associated with the physical safety of users, their privacy, and trust on sensitive transactions, the so-called Safety-Critical or Liability-Critical Applications (SCA or LCA). Examples of these applications are the services that focus on navigation guidance, tracking of objects and location-sensitive billing applications, which nowadays are emerging. Indeed, the current trend shows that positioning systems are expected to be used in a growing number of critical applications, but unfortunately, the integrity of current positioning systems is insufficient for these applications. This is due to the fact that traditional integrity algorithms are not fully applicable in situations where local environmental characteristics, such as buildings, trees, moving objects, Non-Line of Sight (NLOS) propagation, and Radio Frequency Interference (RFI) abound. These local threats degrade the positioning system leading to potentially high positioning errors and therefore may hinder the provision of a full integrity positioning service.

It is for the above reasons that land user applications require the development of integrity techniques able to promptly detect harmful effects like multipath, NLOS, RFI and spoofing. The aim is to guarantee that integrity requirements for critical applications are fulfilled, so that they can be safely used. In the rest of this chapter we introduce the foundations of the problem that motivated this thesis. This is illustrated in Figure 2.1; that is, we have some threats hindering the performance requirements of positioning systems, and then the safe operation of emerging commercial applications such as SCA and LCA is in danger. Details on each block (i.e. applications, challenges and requirements) are given in the rest of this chapter.

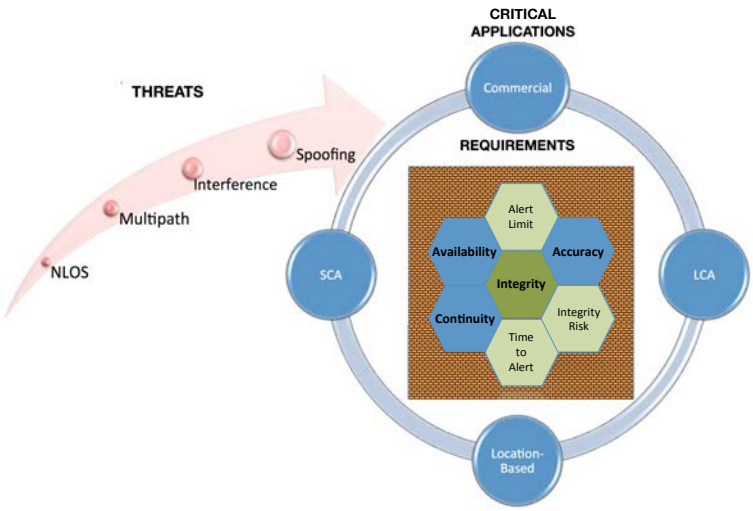

Figure 2.1: Roadmap of this dissertation.

Firstly, it is important to formally define critical applications and analyze some of them. This is done in Section 2.2, in which we analyze two new critical applications emerging nowadays, namely advanced driver assistance systems and road user charging. In addition, we also analyze the traditional SCA civil aviation navigation. Once critical applications are defined and some examples are identified, it will be apparent that positioning systems are of vital importance to the operation of these applications, and then its challenges must be recognized and analyzed. This is the objective of Section 2.4, in which we investigate the challenges of positioning systems in urban environments, basically multipath, NLOS, interference and spoofing. These challenges must be overcome so that critical applications can be safely operated with the given requirements for these applications. The definition of these performance requirements is presented in Section 2.5, giving details on the specific requirements of civil aviation navigation and road user charging. Finally, Section 2.6 introduces the concept of integrity, key requirement for critical applications, and Section 2.7 concludes this chapter.

2.2 Safety-Critical Applications (SCA)

Nowadays a lot of location-based applications have appeared with many different purposes and tolerances to system failures or errors [SG12]. Depending on the criticality that an unwarned error or failure of the location system has for a particular application, we have the following classification [Nav11]:

- **Safety-Critical Applications (SCA)**: Those applications that possess the potential of directly or indirectly causing harm to humans.
- **Liability-Critical Applications (LCA)**: When a miss-performance of the location system can generate legal or economic consequences.

Details and examples of these two types of critical applications are given next. We focus in this section on SCA, whereas LCA is considered in the next section. There are many well-known examples of SCA in areas such as medical devices, aircraft flight control, weapons and nuclear systems [Kni02]. Moreover, many modern information systems are becoming safety-critical in a

general sense because financial loss and even loss of life can result from their failure. Another clear example of SCA, where a location system plays a key role, is civil aviation, since the position of the aircraft is needed to be known in any flight phase with very stringent performance requirements [ICA06].

As a matter of fact, the scope of SCA is broad, and that breadth has to be taken into account. It is obvious that the loss of a commercial aircraft will probably kill people. Nevertheless, it may not seem that obvious that loss of a telephone system could kill people. A loss of 911 service, though, will certainly result in serious injury or death. This kind of service is an example of a critical infrastructure application, as well as transportation control, banking communications, and the management of water systems [Mul04, Med08]. All of these applications are extensively dependent of a location system, and the system failure can and does lead to extensive loss of service with consequent disruption of normal activities. In some cases, the disruption can be very serious. For instance, widespread loss of water or electricity supply has obvious implications for health and safety. Similarly, widespread loss of transportation services, such as rail and trucking, would affect food and energy distribution. Now that an overview of SCA has been given, in the following we get some insight in a new SCA which is gaining interest nowadays, that is the Advanced Driver Assistance Systems (ADAS); and in a traditional SCA as it is civil aviation navigation.

2.2.1 Advanced Driver Assistance System (ADAS)

Over the last decades, safety has become a more and more important factor within the automotive industry, leading to a large investment on the so-called *passive safety*. Nowadays, however, *active safety systems* are designed to actively assist the driver in avoiding accidents before they occur [Lin06]. These systems are also known as ADAS and they are designed to help the driver by taking the pressure of him/her in standard situations. Research shows that this kind of automation may decrease the driver's reaction time [Bro01]. Results also show that ADAS has behavioral influences on the driver as may be considering changes in speed, safety margins to the vehicle in front and frequency of lane change maneuvers [Saa04]. The support given by ADAS to the driver ranges from simple information presentation through advanced assisting and even taking over the driver's task in critical situations. The common characteristic of these ADAS is that they (compared to passive safety systems) directly intervene with the driving task with the aim of improving traffic safety. Currently, ADAS can be divided into a plethora of categories (see Figure 2.2), we summarize some of them here, referring the interested reader to [Lin06] and references therein for a more complete classification.

Among all the categories, ADAS include systems to control the lateral surroundings of the vehicle by means of cameras. These cameras are used to warn the driver about lane departures [Hea05] and blind spots [Sot08], among other warnings. Another feature is the longitudinal control and avoidance systems such as adaptive cruise control [Vah03], traffic jam assist [BMW13] or traffic sign recognition [Møg12]. These systems are sensor based technology together with cameras and positioning systems, and they are aimed at control the front and rear surroundings of the vehicle. The previous characteristics are already implemented on the market today, but more and more advanced systems are under development and expected to reach the market in the near future. Within these advanced systems we find the following [Lin06]:

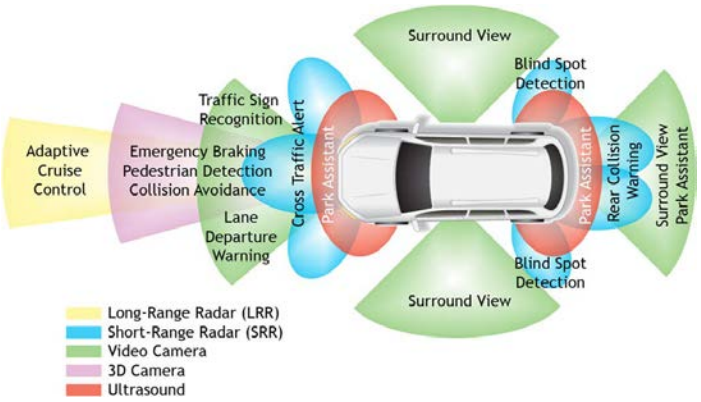

Figure 2.2: Vehicle functions and technologies likely to be used for ADAS [Bro16a].

- **Smart highways/Autonomous driving**: Future highway systems strive for safer and more efficient point-to-point travel with minimal driver participation. These smart highways are long in the experimental stage and automatic guidance lanes could be reality in select areas by the next decade [gua15, Str16, BBC13].

- **Satellite positioning**: The usage of satellite positioning can be used to receive positioning information with the aim of increasing ADASs efficiency. In fact, satellite positioning may be useful in cases where other positioning systems are not available, and it is considered one of the main components to make smart highways and autonomous driving true [TM10, Bis05].

2.2.2 Civil aviation navigation

Civil aviation navigation is a well-recognized SCA, meaning that large and undetected errors in the navigation systems may lead to unsafe situations. It is for this reason that civil aviation needs accurate and strongly reliable navigation aids. GNSS worldwide positioning has been identified by the International Aviation Organization (ICAO) as one of the possible navigation aids for aviation operations [ICA02, ICA06]. The ICAO is the authority responsible of setting the standards that radio navigation systems, including GNSS, must fulfill in order to be suitable for civil aviation. This is the framework in which the concept of GNSS integrity has been developed since its origins, defined as a measure of the trust that can be placed in the correctness of the information supplied by the navigation system, including the ability to provide alerts to the user when the requirements are not assured to be met [RTC06]. Since standalone GNSS does not usually meet the ICAO requirements, different augmentation systems have been developed to fulfill them. Consequently, most of the integrity monitoring systems, techniques and algorithms have been designed to support civil aviation operations within the ICAO specifications. The Required Navigation Performance (RNP) is the statement of the navigation performance necessary for operation within a defined airspace. GNSS is the primary navigation system to support currently defined RPN standards. The RNP is a set of requirements on the total system error, which is made of the path definition error, flight technical error and the navigation system error (see Figure 2.3).

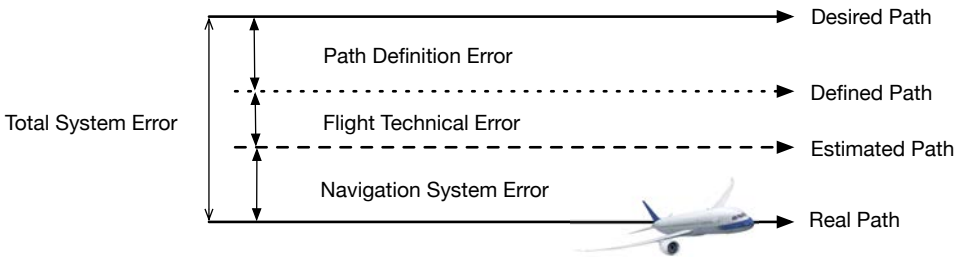

Figure 2.3: Total system error description.

2.3 Liability-Critical Applications (LCA)

The concept of LCA was first introduced in 2004 and initially presented in the Institute of Navigation (ION) 61st annual meeting [Bee05]. LCA are those location-based applications in which the computed Position, Velocity and Time (PVT) is used as the basis for legal decisions or economic transactions [CS08]. As such, an error in the PVT above certain threshold can provoke a wrong legal decision or the computation of a wrong charge. Note that economic liabilities are also associated to legal aspects due to the repercussion of potential claims. In order to keep the probability of those harmful effects below certain (very small) limit, it is essential to also bound the errors and to ensure that the probability that errors are not properly bounded is extremely small. Several examples of LCA have been identified in the last decade, as it is the case of surveillance of parolees [AS09]. However, it is in the road sector where LCA have a major added value. As an example, traffic law enforcement related to speed control have obvious legal implications [Haj09], as well as pay-as-you-drive insurance [Tro11], which is currently emerging. However, the most relevant example today in terms of potential number of users is, no doubt, Road User Charging (RUC), which is explained in detail next.

2.3.1 Road User Charging (RUC)

Charging road user for road usage has been discussed for a long time as a reasonable alternative or addition to taxes [Sma07, Tse09]. Pricing roads according to supply and demand can help to reduce congestion especially in urban areas and during peak times [Pal11]. It is for these reasons that the ability to locate and track a vehicle in space and time is becoming more and more fundamental to RUC. This ability enables a scheme to be flexible and implemented nationwide. In the beginning of electronic RUC, microwave technology was used for toll highway entrance/exit or toll cordon points. This technology relies on the so-called Dedicated Short Range Communication (DSRC) systems. These systems require tolling bridges to be installed along the roadside, as well as suitable On-Board Units (OBU) in the vehicle. The toll bridges can communicate with and identify the OBU (see left plot of Figure 2.4) [Pic06]. The charges are usually calculated by splitting the tolled roads into segments and placing a toll bridge in each segment. The prices for driving through the specific segments are then added and invoiced. Another technology used for RUC is the Automatic Number Plate Recognition (ANPR), which is based on optical character recognition (OCR) to identify the license plates of the vehicles driving through the charged road (see right plot of Figure 2.4) [Pic06]. However, to use these technologies high investments in road-side infrastructure have to be made, which does not seem economically useful to install wide-area scalable tolling systems.

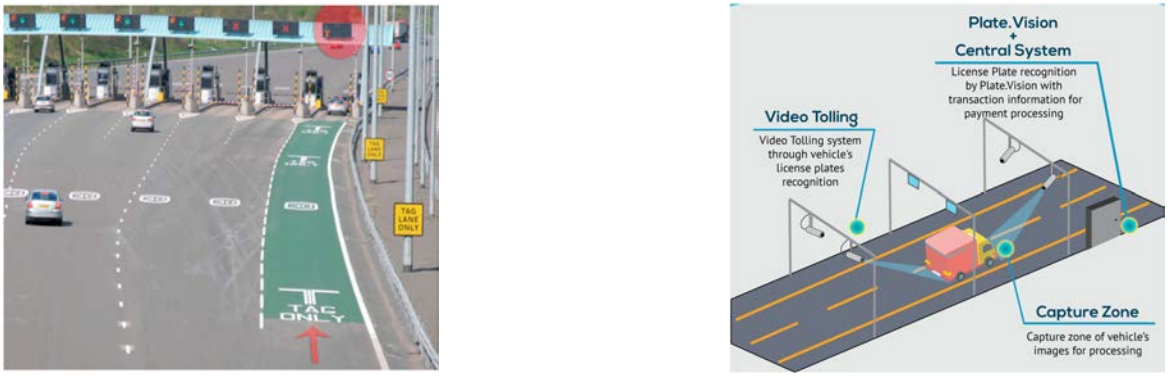

Figure 2.4: Traditional electronic tolling systems: (Left) DSRC [M6t03] and (right) ANPR [Vis17].

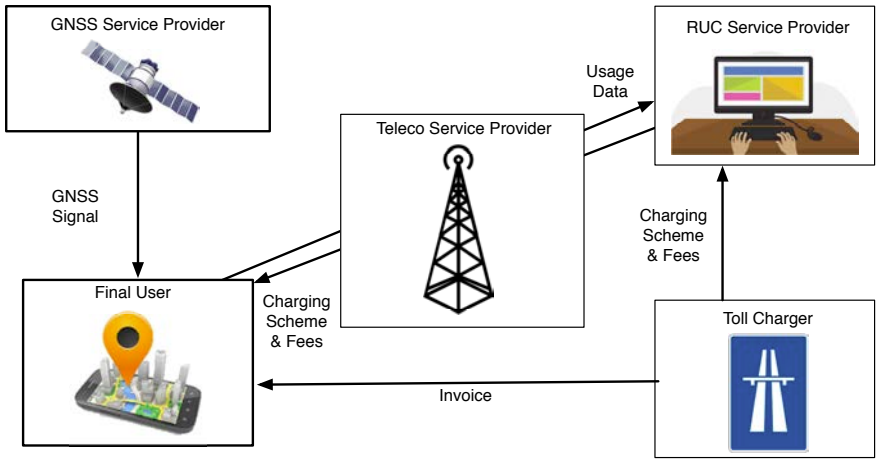

Figure 2.5: GNSS-Based road user charging scheme.

2.3.2 Modernization of RUC

For modern RUC systems, GNSS-based tolling has been studied thoroughly, developing different ways to use GNSS for RUC. All of these techniques require some kind of OBU to be installed in the vehicles, a data center, and some kind of enforcement system (see Figure 2.5) [Num12, Zab11, SA12]. One way of using GNSS for road tolling is through the use of a *thin* OBU. Those OBUs are equipped with a GNSS receiver and a Global System for Mobile Communications (GSM) module. The receiver calculates the position of the vehicle and continuously sends it to a central data center via GSM, where the correct charges are calculated and the corresponding invoices are sent to the users. Another way is to use *thick* or *smart* OBUs, which in addition to the GNSS receiver and the GSM module they contain memory and a processor, so that the correct charge can be computed on-board. This amount is then sent to the central data center via GSM, where invoices are created. As for traditional RUC systems, the tolled road is segmented into sections that constitute the basic charging units. Then, the trajectory of any user can be expressed as the list of crossed road segments, and the total toll fee can be computed as the sum of the price of the segments in the trajectory. Toll systems based on DSRC technologies usually monitor the use of road segments with a checkpoint scheme in which each segment is managed by a control point equipped with DSRC transceivers. Users are charged the whole segment's fee when they cross the control point. On the other hand, GNSS has worldwide coverage, which

allows to monitor road segment crossings by means of user positions calculated along their whole trajectory and not only at fixed checkpoints [SA12, Ch. 3].

In fact, GNSS has been identified as the most flexible and cost-efficient technology for the implementation of large RUC systems. Specifically, different studies have economically compared GNSS and DSRC technologies and have concluded that, whenever the road network in which the system is to be applied overpasses some few thousand of kilometers, GNSS technology is clearly more cost-efficient than DSRC [CS08]. In addition, if other considerations such as flexibility and expandability, or provision of additional services are taken into account the GNSS technology is clearly superior [Num12]. Currently, GPS-based RUC systems are already used to charge trucks in the German and Slovak motorway networks and certain expressways [Col09, Slo10]. Despite GNSS technology advantages, the existing position error and their potential repercussion in the charging are major concerns. RUC systems do not tolerate uncontrolled position errors as some of them may lead to incorrect charging decisions, which in turn cause loss of revenues if the amount of money charged to a user is lower than it should (undercharging), or user claims and the consequently loss of credibility on the system if the amount of money charged to the user is higher than it should (overcharging) [Sal10]. It is for this reason, that integrity monitoring play a key role in RUC systems, controlling the presence of excessively large position errors and alerting when it is not possible to estimate the position within the required specifications.

2.4 Challenges

The success and development of reliable positioning systems in the last decade has caused the technology to become a standard and essential tool for several markets involving critical applications, as seen previously. However, positioning systems are vulnerable to a range of threats related to the transmitter or receiver systems, the propagation channel and external interferences. These vulnerabilities have been largely identified, analyzed and have been the subject of an intense debate to assess how they can affect critical applications [SG12, Fer16]. In particular, the possibility of unintentional or intentional interferences is becoming a critical threat to radio-frequency positioning systems. The reasons for such concern are diverse. On the one hand, unintentional interferences are expected to increase in the coming years due to the high-occupancy of the spectrum, and in particular, to the deployment of wireless systems in the neighborhood of the positioning systems band. On the other hand, intentional interference are nowadays feasible because of the increasing popularity of low-cost hand-held interference devices, which may seriously threat the reliability of critical applications based on location systems. Another kind of intentional interference is spoofing, which tries to counterfeit the signals in space emitted from the positioning system, in order to fool a victim receiver with misleading signals that lead to erroneous position fixes. In addition, when analyzing road and urban applications, one has to consider one specific effect of positioning systems, which is multipath and very in particular, the NLOS multipath. It is well known that multipath is one of the main (if not the main in many cases) sources of degradation of the position estimation in urban and, in general, non-open sky environments [SG12]. The multipath effect is produced when reflected versions of the signal used for positioning reaches the receiver along multiple propagation paths. The NLOS condition refers to the situation where the line of sight (LOS) signal is absent/obstructed or significantly weaker than one of the multipath components. Interference, spoofing and multipath

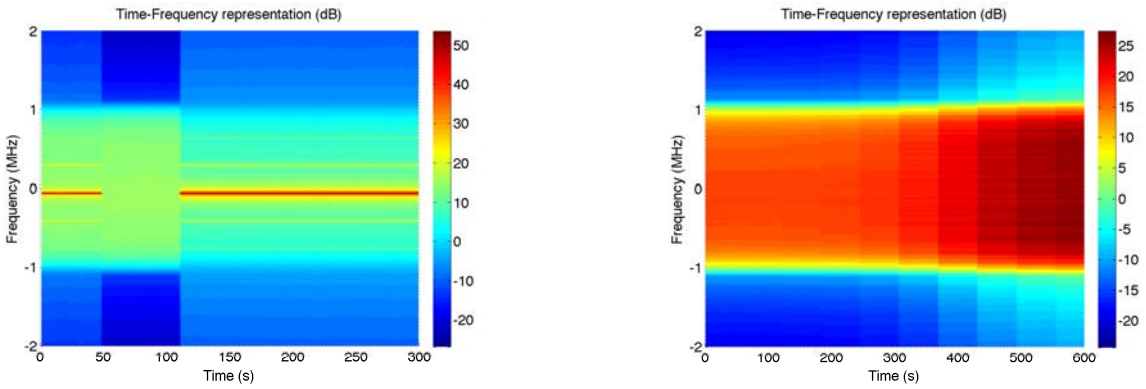

Figure 2.6: Time-Frequency representation of (left) a narrow-band and (right) a wide-band interference

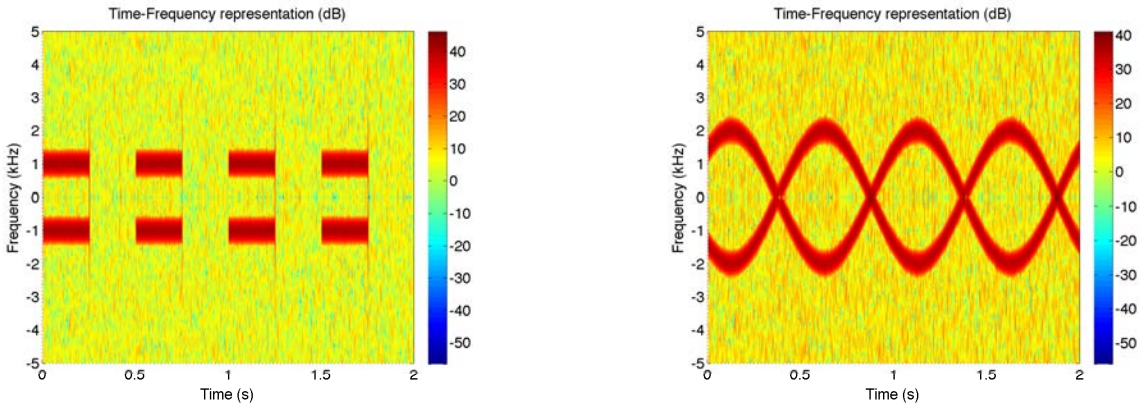

Figure 2.7: Time-frequency representation of (left) a pulsed and (right) both continuous and chirp interference.

are the three main error sources in positioning systems, and they are further analyzed in the following.

2.4.1 Radio Frequency Interference (RFI)

RFI is a major threat because positioning signals are usually received at very low power, and then undesired interferers in the same frequency band can affect the receiver performance, leading to a degraded accuracy in the position solution or even to a complete denial of service. In order to cope with this problem, there are indeed many challenges to be solved because of the wide range and different variations of possible interference sources. This is why an extensive analysis is required in order to get further insights on how interference sources behave, and how they may impact onto the ultimate performance of positioning systems. From a conventional time-frequency approach, the most common types of interferences can be classified according to [Kap05]:

- **Narrow-band and wide-band**: Depending on the interference bandwidth (see Figure 2.6).
- **Continuous and pulsed**: Continuous interference are present for all time instants, whereas pulse interferences are not present in each instant (see Figure 2.7).
- **Swept or chirp**: Signals whose instantaneous frequency varies as a function of time. The

most often chirp signals have a linear, exponential or sinusoidal variation (see right plot of Figure 2.7).

Another distinction can be made between unintentional and intentional interference. Unintentional interference can be expected practically anywhere in the world, since there are a large number of systems that our Society daily depends on. Unfortunately these systems rely on the transmission of radio-frequency energy within the positioning systems frequency bands, so interference is expected form these systems. Jamming is an example of intentional interference and it has become a big concern due to the increasing availability of low-cost jammers able to broadcast powerful signals into the system bands. Although illegal, these devices can easily be purchased for a few tens to several hundreds of dollars. The main jamming effect is either an increase of the noise component and therefore a reduction in the accuracy of the system or the denial of service due to power saturation. Jammers can be classified as follows [Des04, Ch. 4]:

- **Simple Jammers**: A simple jammer is one which generates interference signals without knowledge of the receiver design and usually generates one type of interference signal like broadband noise or CW.
- **Intelligent Jammers**: Is designed with the knowledge of the receiver and it uses pulsed or swept CW signals, making the jamming detection more difficult than the simple jammer case. However, the intelligent jammers are more complex.

2.4.2 Spoofing attacks

During a jamming attack the user may be aware of this event when the tracking is lost and position cannot be obtained. On the contrary, a spoofing attack is not so simple to detect since the victim receiver is fed with corrupted information and continues calculating faked positions. This kind of attacks poses significant challenges from the signal processing point of view, since at signal level, the spoofed signal is nothing but a synthetic copy of an authentic positioning signal. This is why in absence of any detection mechanism, the receiver will synchronize to the spoofed signal and proceed with the rest of processing steps as it would do with a true signal. The objective of the spoofer is then to degrade the navigation solution computed by the receiver, either making the solution completely inconsistent, or by causing the solution to drift away from the true position, so that the receiver believes it is located somewhere else in space and time that is not. These type of threats are increasingly becoming a world-wide concern, particularly due to the fact that some recent incidents are being speculated to have been caused by spoofing attacks [She11, CNN16]. From an implementation point of view, different variations of spoofing attacks can be distinguished, ranging from simple ones, which use a simple set-up and are easy to detect, to sophisticated ones, which are able to emulate the characteristics of the true signals [Hum08].

2.4.3 Multipath propagation

The key assumption in the design of most positioning systems is the presence of perfect LOS between the transmitters of the positioning signal and the user's receiver antenna to be located. As seen in Figure 2.8, this situation is commonly encountered in clear-sky outdoor conditions,

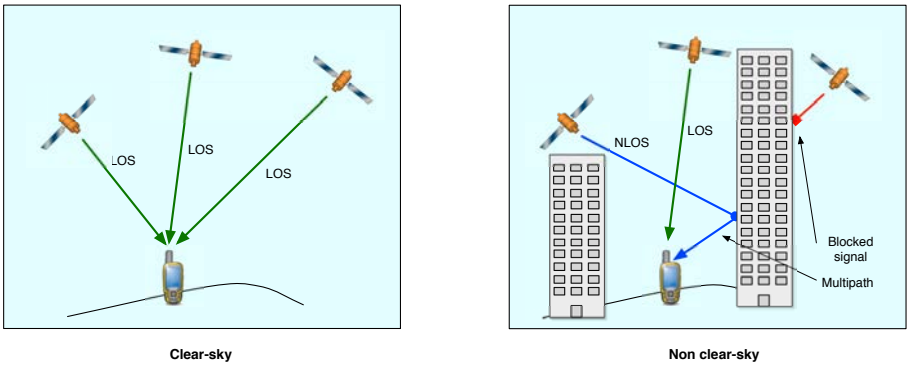

Figure 2.8: Illustration of (left) the clear-sky and (right) non clear-sky scenarios for a positioning receiver.

but progressively degrades when moving into more problematic environments such as non clear-sky environments. The main source of degradation of positioning receivers is certainly the presence of either natural or man-made obstacles, causing the transmitted signal to be reflected and refracted in its way to the receiver end. In these circumstances, the received signal is typically composed of phase-shifted, time-delayed, and attenuated signal replicas, forming the so-called multipath contribution. The impact of multipath depends on several factors such as the relative power compared to the LOS signal, and the time-spread of the overall multipath contribution. These two parameters determine the degree of distortion of the LOS signal, and thus the error in the estimate of the position measure. To provide an idea of the order of magnitude of this error, it can be said that for the case of range measurements, multipath error typically range from tens to hundreds of meters. A more dramatic situation occurs when moving to more challenging environments like urban canyon or indoor scenarios. In this case the LOS component may even not be present, and the receiver has no choice but to try to acquire the received signal parameters based on a collection of indistinguishable overlapped signal replicas, which is known as the NLOS propagation, this case is considered in the following section.

2.4.4 Non-Line of Sight (NLOS) propagation

The situation when only reflected signals are received and processed by the positioning receiver is known as NLOS. This situation challenges the basis of most positioning systems because there is no reference LOS signal that could provide truthful information on the geometric distance between the transmitter and the user's receiver. Unlike classical multipath, the resulting range measurement error in NLOS propagation is unbounded and classical multipath mitigation techniques cannot cope with it [CS08]. Everyone who has used a GPS receiver in an urban environment has suffered the effect of this NLOS propagation, which can provoke in some circumstances very large position errors. Analyses performed in different cities show that a large percentage of the signals received by a receiver in an urban environment are in fact reflected signals, so that this effect is predominant an poses a big challenge to positioning systems in urban environments. The effect of NLOS propagation in range measurements and, hence, in position errors can be simply understood by its association to the extra travel time (always positive) or phase shift of the reflected signal as compared to the theoretical direct one. This in particular provokes that the largest errors are usually produced in wide avenues rather than in narrow streets as it has been experienced along different trials. A large data campaign was performed in Madrid and London, including very different streets and building configurations,

leading to two simple but important conclusions [CS08]:

- Errors are much larger than those in an open environment, what implies that predominant errors are those associated to multipath, and in particular to NLOS propagation.
- The distribution of the errors is far from having a Gaussian behavior what makes the error bounding more difficult.

2.5 Requirements

As we have already seen, there has recently been an increasing interest in location-based SCA and LCA. These applications, which are often associated with terrestrial environments, have very stringent requirements in terms of performance of the positioning system. Although each application has its own challenges, the common challenges for the positioning system are those commented in the previous section. So, in order to be able to ensure the full operability of the above mentioned critical applications, these challenges have to be overcome in order to guarantee the stringent requirements of critical applications, given by the following parameters [SA12]:

- **Continuity**: Measures the probability that a system can fully function for a planned period without any unintended interruptions.
- **Accuracy**: Is a measure of the error between the true and estimated positions (i.e. position error).
- **Availability**: Is the percentage of the overall time that a system is available for use, understood as the overall time that the system can operate under certain performance requirements.
- **Integrity**: It measures the trust and reliability of a system on guarantee that the errors of the system are bounded with a given probability. Integrity requirements consist of the following terms:
  - *Alert Limit (AL)*: it is the maximum allowable position error without issuing an alert.
  - *Time to Alert (TTA)*: it is the maximum tolerable elapsed time to raise an alert when the AL is exceeded.
  - *Integrity Risk ($\mathcal{P}_{\text{risk}}$)*: it is the probability that the position error exceeds the AL without warning the user within the TTA.
  - *False Alarm Probability ($\mathcal{P}_{\text{fa}}$)*: it is the probability of declaring a fault when actually there is no fault within a fixed interval of time t_α.

2.5.1 Requirements for civil aviation navigation

As it has already been mentioned, the GNSS performance requirements, responsible of the navigation system error, are standardized by the ICAO for the safety use of civil aviation navigation. Civil aviation standards define GNSS performance requirements for different operations in terms of accuracy, integrity, continuity and availability [ICA06, RTC06]. In particular, the

<table>
<tr><th rowspan="2">Typical operation</th><th colspan="2">Accuracy</th><th colspan="4">Integrity</th><th rowspan="2">Continuity</th><th rowspan="2">Availability</th></tr>
<tr><th>Horizontal</th><th>Vertical</th><th>$\mathcal{P}_{\text{risk}}$</th><th>HAL</th><th>VAL</th><th>TTA</th></tr>
<tr><td>En-Route</td><td>3.7 km</td><td rowspan="3">N/A</td><td rowspan="3">10^{-7}/h</td><td rowspan="3">7.4 km (oceanic)
3.7 km (continental)
1.85 km</td><td rowspan="3">N/A</td><td>5 min</td><td rowspan="3">1–10^{-4}/h
to
1–10^{-8}/h</td><td rowspan="6">0.99
to
0.99999</td></tr>
<tr><td>En-Route, terminal</td><td>0.74 km</td><td>15 s</td></tr>
<tr><td>Non-Precision approaches</td><td>220 m</td><td>556 m</td><td rowspan="2">10 s</td></tr>
<tr><td>Approaches with vertical guidance (I)</td><td rowspan="3">16 m</td><td>20 m</td><td rowspan="3">$2 \cdot 10^{-7}$</td><td rowspan="3">40 m</td><td>50 m</td><td rowspan="3">1–$8 \cdot 10^{-6}$
per 15 s</td></tr>
<tr><td>Approaches with vertical guidance (II)</td><td>8 m</td><td>20 m</td><td rowspan="2">6 s</td></tr>
<tr><td>Precision approach (Cat. I)</td><td>4–6 m</td><td>10–35 m</td></tr>
</table>

Table 2.1: GNSS performance requirements for civil aviation navigation [ICA06].

ICAO SARPS Annex 10 [ICA06] specifies that the GNSS shall meet the performance requirements in Table 2.1 for different operations. Ranges of continuity and availability values are given for some operations because these requirement are dependent upon several factors. On the one hand, factors for the continuity are: the intended operation, traffic density, complexity of airspace and availability of alternative navigation aids. The lower value given is the minimum requirement for areas with low traffic density, while higher values are appropriate for areas with higher traffic density. On the other hand, factors for the availability are: the frequency of operations, weather, the size and duration of the outages, availability of alternate navigation aids, radar coverage and traffic density. With the lowest availability values, GNSS is considered to be practical but not adequate to replace non-GNSS navigation aids, while higher values are adequate for GNSS to be the only navigation aid for en-route navigation. It is worth mentioning, that for civil aviation two type of AL requirements are given:

- **Horizontal Alert Limit (HAL)**: The HAL is the radius of a circle in the horizontal plane, with its center being at the true position, that describes the region that is required to contain the indicated horizontal position.
- **Vertical Alert Limit (VAL)**: The VAL is half the length of a segment on the vertical axis, with its center being at the true position, that describes the region that is required to contain the indicated vertical position.

2.5.2 Requirements for RUC

As we have already mentioned, different failures in the toll system in RUC applications may cause charging errors, which are classified as overcharging when a user is levied a higher charge than the real one, and as undercharging when a user is levied a lower charge than the real one. RUC requirements are set to bound the impact of over- and undercharging errors. For instance, the Dutch ABvM project indicates that "the requirement set for the registration unit is that the amount, corresponding to the distances recorded over a one-month period, may not deviate by more than 1% from a 'perfect' measurement in 99% of cases" [ARS06], or the proposed by Expert Group 9 when indicates that "it is necessary to define a specific set of test conditions in which test road segments are guaranteed to be successfully recognized with a success rate of at least 99.99%. False recognition of a road segment should be less than 1 in 10^6" [Exp06]. With this terminology, a missed road segment recognition occurs whenever the toll system does not

		Maximum invoice error (M_e%)		
		1% ($N_{Go} = 99$)	0.1% ($N_{Go} = 999$)	0.01% ($N_{Go} = 9999$)
Percentage of invoices	99 %	10^{-4}	10^{-5}	10^{-6}
meeting the maximum	99.9 %	10^{-5}	10^{-6}	10^{-7}
error requirement (I%)	99.99 %	10^{-6}	10^{-7}	10^{-8}

Table 2.2: Maximum probability of missed geo-object recognition (p_{MGoR}) for different invoice requirements in the worst case [SA12, p 80].

charge a user the price of a road segment he/she has actually used (undercharging). On the other hand, a false road segment recognition occurs whenever the toll system charges a user the price of a road segment he has not used (overcharging). In this sense, RUC requirements define the maximum allowable rates (probabilities) of these two recognitions. Let consider specifications that requires a minimum percentage I% of the total invoices processed by the toll system to have an error equal to or lower than M_e% of the real charge. After some reasonable assumptions, given the invoice accuracy specifications $\{I, M_e\}$, the worst case probability of missed road segment recognition can be obtained [Sal10, Ch. 5]. Table 2.2 contains the value of this probability for various invoice high-level requirements, with N_{Go} the number of segments the user has driven through. Similarly, for the maximum allowable probability of false recognition (overcharging), we can also obtain the probability as a function of the number of potential false road segment recognitions N_{PFGoR} [SA12, Ch. 5]. Values of 10^{-4} and 10^{-6} are proposed in [Exp06]. These values assures that the undercharging error of at least $I = 99\%$ of all invoices processed by the toll system do not exceed $M = 1\%$ of the real charge, regardless the number of road segments crossed by the user. These values of probabilities of missed and false geo-object recognition can be translated to integrity requirements [SA12, Ch. 5].

2.6 Integrity Monitoring

Although failure in a well-designed positioning system will be relatively rare, the capability for detection and exclusion of failures is required in order to satisfy the stringent requirements for positioning systems in critical applications. The detection function generates an alarm in the event of a hazardous positioning failure and is, therefore, of paramount importance in the reduction of integrity risk. The exclusion function provides the potential means for specifically identifying and removing a detected failure, thereby improving continuity and availability. So, in addition to providing PVT, a positioning system used in critical applications must have the ability to provide timely warnings to users when the system should not be used. This capability is known as the integrity of the system. Integrity includes the ability of providing valid and timely warnings to the user when the system must not be used for the intended operation. Usually, standalone positioning system capabilities cannot meet certain performance requirements. Thus, different augmentation systems have to be implemented in order to enhance positioning performance and to allow its use in critical applications within the given requirements. All methods of integrity monitoring rely, in one way or another, on checking the consistency of redundant information.

Integrity monitoring focuses on two main aspects: Improvement of the accuracy and improvement of the integrity of the PVT solution of a positioning system. To do so, different integrity schemes exist and can be classified into two types according to the source from which

the user receives the integrity information:

- **Infrastructure-Based**: The integrity information is received from some regional or continental infrastructure dedicated to this task:
  - *Ground-Based Augmentation System (GBAS)*: The augmentation information is sent from a ground-based transmitter, with either local or regional coverage.
  - *Satellite-Based Augmentation System (SBAS)*: The augmentation information is sent from a satellite-based transmitter.
- **Receiver-Based**: This type of integrity scheme does not need any kind of infrastructure to obtain the integrity information, but it is autonomously calculated within the user equipment.

Regarding the accuracy improvement, this can be done in several ways, as we will see. For instance, we can remove from the PVT computation that information affected by large errors or we could aid the positioning system with on-board sensors measurements to increase the measurement accuracy. For integrity monitoring, we have two alternatives, namely fault detection or a protection level computation. The former is based on an observable whose value is compared to a threshold estimated from the statistical distribution of that observable in nominal conditions. On the other hand, the protection level computation is dependent upon the augmentation system, but in general it assures that positioning error is bounded with a certain probability. In the following, we explain with more detail both infrastructure and receiver-based integrity schemes, with special emphasis on the former.

2.6.1 Infrastructure-Based integrity

Infrastructure-Based integrity is provided by augmentation systems based on some additional dedicated equipment. This equipment basically consist on reference stations measuring differential corrections and communication stations aimed at sending the differential corrections, and integrity information to the user receivers. These differential corrections are intended to improve the accuracy of the positioning system so that accuracy requirements are fulfilled. Regarding the integrity monitoring, several observables are compared with different thresholds in order to decide whether the information provided by the GNSS is trustworthy or not. Then, a user that applies the computed corrections and only use the non-faulty transmitters (i.e. those whose observables have been monitored successfully by the reference station(s)) for the PVT calculation is assumed to be protected against any monitored integrity threat. A graphical interpretation of infrastructure-based integrity systems is given in Figure 2.9. Two different kinds of infrastructure are used to augment positioning systems, namely Satellite- and Ground-Based Augmentation Systems (SBAS and GBAS). SBAS is based on additional satellites and ground stations that provide to the user additional information to improve the performance of the positioning system. SBAS provides this information at the continental coverage. Wide Area Augmentation System (WAAS), the U.S.' SBAS, and European Geostationary Navigation Overlay Service (EGNOS), the European's SBAS, are examples of infrastructure-based augmentation systems for GPS. On the other hand, GBAS consists of a single ground station that provides to the user additional information to improve the performance of the positioning system. GBAS coverage is local,

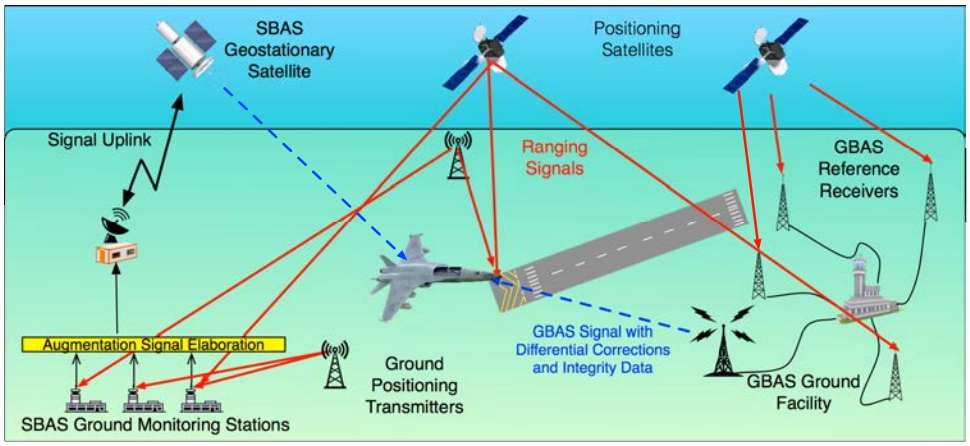

Figure 2.9: General architecture of infrastructure-based integrity monitoring systems.

	Before corrections	After corrections
$\sigma_{\text{s\&e}}$	0.85 m	0.1–0.6 mm/km × distance in km
$\sigma_{\text{T},i}$	0.2–1.5 m	1–4 cm/km × distance in km
$\sigma_{\text{I},i}$	4.6–13.7 m	0.2–4 cm/km × distance in km
σ_{MP} (m)	0.2	0.3
σ_{Noise} (m)	0.5	0.7

Table 2.3: GPS L1 C/A pseudorange measurement errors order of magnitude before and after applying differential corrections.

typically 50 km around the ground station, which is enough to cover all aircraft approaching an airport.

It is important to recall that the concept of integrity was firstly introduced in the field of civil aviation. This is why the infrastructure-based augmentation systems currently available are focused on civil aviation. For instance, both WAAS and EGNOS are designed to support non-precision approach operations [IWG14, GSA16]. On the other hand, a GBAS that already provides precision approach services is the Local Area Augmentation System (LAAS) deployed in the U.S. [FAA16]. As an example of infrastructure-based integrity we will analyze the augmentation systems used in GNSS. The main characteristic of these systems is that they provide corrections to the user to improve the accuracy of the position estimates, and it is therefore referred to as differential positioning. It is for this reason that some references use the term Differential GPS (DGPS) or Differential GNSS (DGNSS) to refer the augmentation system itself. However, as we will see, modern augmentation systems have more functionalities than the differential positioning.

Differential positioning: Accuracy improvement

The idea behind differential positioning is that pseudorange errors observed by a reference station are expected to be very similar to those experienced by a nearby user. If the reference station estimates these errors by using its known position and provides the required corrections to the user terminal, the user's position accuracy can actually be improved. Expected standard deviations of the measurement errors using differential positioning compared to a stand-alone L1 C/A GPS receiver are presented in Table 2.3 [Kap05, Ch. 8]. Values of these errors, except local effects (multipath and noise) errors, are dependent upon the distance between the user and the reference station. We have split the local errors in multipath, σ_{MP}, and receiver noise, σ_{Noise}. We also put together the effects of both satellite clock and ephemeris errors in $\sigma_{\text{s\&e}}$.

SBAS	WAAS	EGNOS
Country	United States	Europe
Number of Satellites	3	3
Central processing sites	3	4
Up-link stations	6	6
Reference stations	38	36
Control stations	2	2

Table 2.4: Architecture of two SBASs of interest: the American WAAS and European EGNOS.

These are errors encountered in any GNSS. With the values in the table we can see how for the propagation ($\sigma_{\mathrm{T},i}$ and $\sigma_{\mathrm{I},i}$), satellite clock and ephemerides errors there may be a big reduction of the residual errors. For instance, the ionospheric errors (largest errors in nominal conditions before corrections), at radius of 100 km from the reference station, are reduced to a half, in the worst-case, and to a third, in the best-case. On the other hand, it is worth remarking that the standard deviation of the errors due to multipath and noise increases after applying the differential correction, and it is independent of the distance between user and reference station. In summary, if both receivers in the reference station and in the user side are influenced by the same errors, then the differential correction will have removed all errors. However, if there is a poor spatial and time correlation of the error component the error will be different at the reference and user receivers and will not be completely compensated but even increased, as it is the case of local effects.

Integrity monitoring

The accuracy of GNSS can be improved by sending corrections to the user, but a second information, even more important in critical applications, is the quality, correctness and trust that can be placed on the corrected measurements, and then on the obtained positioning. This second information is the integrity information, and it is also provided by an augmentation system. To this end, at each epoch, the GNSS receiver computes a Protection Level (PL) by combining parameters transmitted by the augmentation ground segment (e.g. total standard deviation of pseudorange measurements), receiver parameters (e.g. number of satellites in view) and the user-satellites geometry. In other words, the PL computed by augmentation systems users assumes that corrected pseudoranges used for the PVT computation are affected by nominal errors only, and not by failures. Indeed, faulty satellites are supposed to be discarded by the fault detection process.

SBAS integrity

The general approach in SBAS is to split the total pseudorange error into its components (i.e. different error sources) and to estimate the contribution of each component over the entire region. Each SBAS's reference station continuously takes pseudorange measurements, which are sent to the central processing site in order to generate the differential corrections. These differential corrections are uploaded to a constellation of Geostationary Earth Orbit (GEO) satellites, then the satellites transmit this information to all SBAS's users in the area covered by the satellites. Some characteristics of WAAS and EGNOS are given in Table 2.4. SBAS's users also have the possibility to use SBAS signals as ranging sources to increase the number of pseudorange measurements. In SBAS, the integrity risk allocation is usually split in two integrity risk, one for the PL and another for the fault detection. So, in addition to application of SBAS differential corrections, the GNSS receiver for SBAS must also compute position error bounds (i.e. PLs), especially for critical applications. These PLs give a bound on the position errors guaranteeing

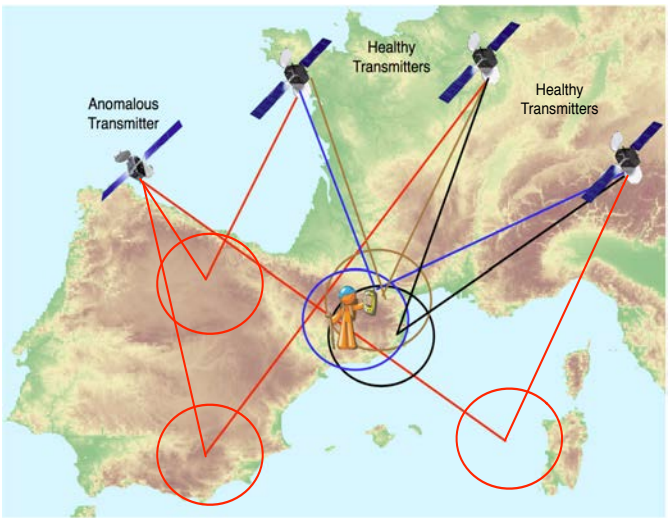

Figure 2.10: Graphical description of receiver-based integrity: Measurement redundancy and consistency check concepts.

that the user position errors will not exceed this bound without issuing an alert with the required probabilities levels and time to alerts. The PLs are continuously compared to the given alert limits (i.e. AL). A warning is issued as long as PL>AL. The computation of the PL is done by using variances broadcast from the SBAS satellites. The individual error variance are summed to form overall residual pseudorange error variances for the visible satellites. Finally, the geometry matrix and known weighting matrix with the received variances are used to bound the standard deviation of horizontal and vertical position errors.

2.6.2 Receiver-Based integrity

Receiver-based integrity consists in the integration of information provided by the positioning system, without need of any additional (augmentation) system. In particular, if redundant measurements are available (typically, more than four measurements in 3D-positioning), it is possible to perform a consistency check between the measurements in order to detect erroneous measurements. This is illustrated in Figure 2.10. Imagine we have one anomalous transmitter because its signal is affected by some of the known integrity threats. Then, suppose we obtain a position solution with each pair of available transmitters. Thereby, when we use the anomalous transmitter we see that the obtained position is far away from that obtained when this transmitter is not used, whereas when we do not use the anomalous transmitter, all the obtained positions are close to each other and to the real one. To do so, we need more than 2 transmitters (for 2D, 3 for 3D) to detect that there is an anomalous transmitter. If more than 3 (4 for 3D) are available, then we can exclude the erroneous transmitter. So, we have illustrated the idea of consistency check exploited in receiver-based integrity systems and the need of redundant measurements. This idea is used in order to perform a Fault Detection and Exclusion (FDE). The FDE performs two separate functions: first, the fault detection procedure detects the presence of unacceptable large position errors; second, the fault exclusion algorithm isolates and eliminates the source of the error, allowing navigation service to continue within the required performances without interruption. As commented previously, the integrity concept was first developed for civil aviation. It is for this reason that receiver-based integrity has traditionally been known as Aircraft Based Augmentation Systems (ABAS). Indeed, two ABAS are distin-

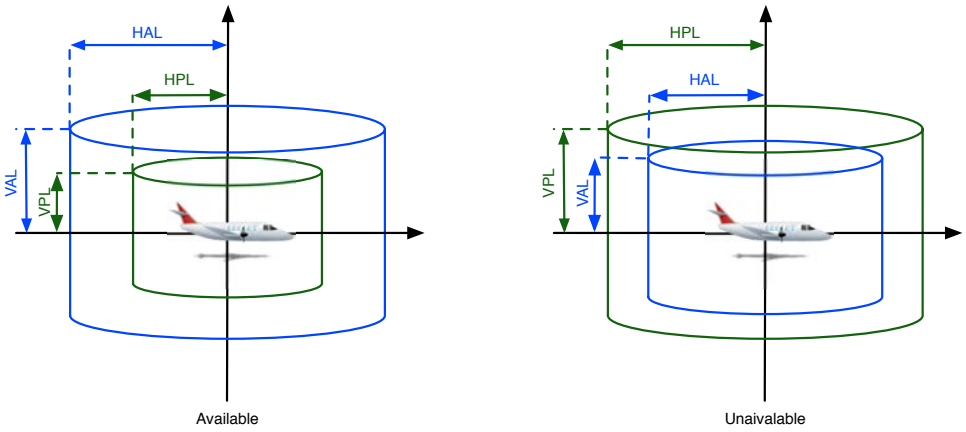

Figure 2.11: Relationship between alert and protection levels: (Left) Available integrity monitoring system and (right) unavailable integrity monitoring system.

guished: the Receiver Autonomous Integrity Monitoring (RAIM), which is based on a receiver algorithm that only uses information of the positioning system, and the Aircraft Autonomous Integrity Monitoring (AAIM), which use information of the positioning system plus on-board sensors such as baro-altimeter and inertial measurement units.

2.6.3 Integrity algorithm: Qualitative description

Integrity monitoring techniques by either means of infrastructure- or receiver-based systems check the compliance with the integrity requirements calculating statistical bounds of the position error. These bounds are denoted Protection Levels (PL) hereinafter, and are defined as follows [RTC06]:

- **Horizontal Protection Level (HPL)**: The HPL is the radius of a circle in the horizontal plane, with its center being at the true position, that describes the region assured to contain the indicated horizontal position by the positioning system within the given requirements on integrity risk and false alarms.
- **Vertical Protection Level (VPL)**: The VPL is half of the length of a segment in the vertical axis, with its center being in the true position, that describes the region assured to contain the indicated vertical position within the given probabilities requirements.

With this framework, the integrity system must assure that, in absence of an integrity alert, the estimated position is within the volume defined by the HPL and VPL in compliance with the integrity risk; that is,

$$\Pr\{[(|e_{\mathrm{H}}| > e_{\mathrm{HL}}) \text{ or } (|e_{\mathrm{V}}| > e_{\mathrm{VL}})] \text{ and } [\text{no alert within the TTA}]\} \leq \mathcal{P}_{\mathrm{risk}}, \tag{2.1}$$

where $\{e_{\mathrm{H}}, e_{\mathrm{V}}\}$ and $\{e_{\mathrm{HL}}, e_{\mathrm{VL}}\}$ stand for the actual positioning errors and AL (horizontal, vertical), respectively. When any of the PLs exceeds the AL, the integrity monitoring system is declared unavailable because it is not able to assure that the estimated position is within the volume defined by the HAL and VAL specified by the application requirements (see Figure 2.11).

A block diagram showing the operation of FDE is given in Figure 2.12. Even most of the time a position failure will not occur, the existence of a detection function, however, opens the

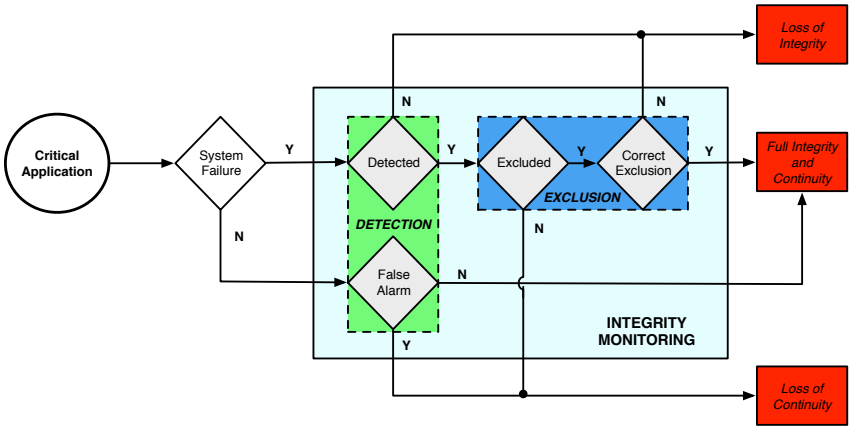

Figure 2.12: Block Diagram of an integrity algorithm with FDE.

possibility of false alarm. In this case, a failure has not occurred but the detector raises an alarm nonetheless, resulting in a loss of continuity. In the case a real failure exist, it should generally be detected, and exclusion is performed. If the correct failure is removed, both integrity and continuity are guaranteed. Nevertheless, if the failure is not detected, or if detection does occur but the wrong failure is removed, integrity is lost. The last condition occurs when a failure has been detected, but exclusion is not deemed possible (e.g. not enough measurements); in this event, continuity is lost. In summary, the following events can occur throughout the FDE process [RTC06]:

- **Positioning failure**: Occurs whenever the difference between the true position and estimated position exceeds the applicable AL, that is, when ($|e_{\mathrm{H}}| > e_{\mathrm{HL}}$) or ($|e_{\mathrm{V}}| > e_{\mathrm{VL}}$).
- **False alarm**: Is defined as the detection of a positioning failure when indeed it has not occurred. This leads to a loss of continuity.
- **Missed detection**: Occurs whenever a positioning failure is not detected.
- **Failed exclusion**: Occurs whenever a true position failure is detected but it is not eliminated within the TTA (from the onset of the positioning failure). A failed exclusion would cause an alert.
- **Wrong exclusion**: When an undetected positioning failure exists after performing an exclusion.
- **Missed alert**: Is a positioning failure that is not warmed as an alert within the TTA. Both missed detection and wrong exclusion can cause missed alerts after the TTA expires.
- **False alert**: Is defined as the indication of a positioning failure when a failure has not occurred (a result of a false detection). A false alert would cause an integrity alert.

The probability of the previous events, specially the probability of missed and false alert, define the FDE performance. For instance, in order to meet the ICAO specifications for aviation navigation, the FDE must have a maximum probability of missed alert of 10^{-3} per test and a maximum probability of false alert of $3.33 \cdot 10^{-7}$ per test for operations from en-route to non-precision approach, while approach operations with vertical guidance operations require $1.6 \cdot 10^{-5}$ per test [RTC06]. In the rest of this dissertation we will analyze detection schemes to approach the FDE for integrity monitoring.

2.7 Conclusions

We have presented the foundations and roadmap of this thesis. Basically, we have brought into attention an increase on commercial applications that heavily rely on positioning systems. Indeed, the current trend shows that more and more of these applications are expected to be used in critical sectors in which the safety or liability of the user may be in danger. It is for this reason that very stringent requirements are needed to operate these applications. Specifically, integrity is one of the most critical and demanding requirement, assuring that trust can be placed in the positioning system. Nevertheless, several challenges and threats have to be overcome in order to fulfill these requirements with the aim of safely operating critical applications. With this in mind, we have firstly defined critical applications as those applications in which the physical safety of users, their privacy, and trust on sensitive transactions are in danger in case of malfunction of the positioning system. With critical applications defined and exemplified, we have highlighted the vulnerabilities and challenges of positioning systems, specially those operating in urban environments, being multipath and NLOS the main threats. Furthermore, we have RFI produced by unintentional sources that share the positioning system band, or intentional sources like jamming or spoofing that tries either to deny the positioning service or mislead the user by faking the position solution. Next, we have defined the performance requirements of positioning systems, including continuity, accuracy, availability and most importantly integrity.

Finally, details on integrity monitoring have been provided, introducing different schemes to provide integrity monitoring. We have concluded the chapter by explaining the motions of an integrity monitoring algorithm, defining important concepts as protection levels, fault detection and exclusion and availability of the integrity algorithm. The rest of this dissertation is aimed at analyzing GNSS as positioning system for critical applications and the design of a new era of integrity techniques to be applied in urban environments. As above explained, one of the main components of integrity algorithms is the FDE. Thereby, the mathematical tool to be used for the design of integrity algorithms lies in the field of detection theory. For the above reasons, the next step in this thesis is to study the principles of GNSS and detection theory, main task of Chapter 3.

Chapter 3

Fundamentals of GNSS and Detection Theory

3.1 Introduction

One of the most recurrent problem in human history has been that of accurately determining one's position [Kar07]. An extensive way for positioning used in ancient civilizations was based on the observation of celestial bodies, dating back to the first deep sea navigation attempts. Since then, the field of positioning and navigation has experienced a dramatic evolution [Sch09], leading to the use of radio-frequency signals [For08] and driving GNSSs to be the quintessential navigation system nowadays. This has been possible thanks to their global coverage, excellent accuracy, and lack of infrastructure for the user. As most of today telecommunication systems, GNSS has military origins in the early 1960s, with the U.S. Navy's navigation satellite system, also known as Transit. Originally intended to help the navigation of U.S. submarines, Transit soon became widely adopted by commercial marine navigators, and it was the precursor of current GPS [Kap05]. GPS was developed by the U.S. Department of Defense during the 1970s, although it took over 20 years to make it operational. At the same time GPS was under development, the Soviet Union started the development of its own GNSS system, GLONASS. Similarly, Europe approved in 2002 the development of Galileo, a European GNSS system. In this line, other countries such as Japan, China, and India are developing their own global or regional navigation systems. The success of GNSS still continues, now reaching the arena of small portable devices, with a expected population of more than 6 billion GNSS receivers by 2020, as reported in the 2015 GSA market report. GNSS applications in today society range from typical GNSS applications like transportation (e.g. fleet management) to time reference [SG12]. The influence of GNSS into our modern society has lead to a new era of location-based services (LBSs) and critical applications such as SCA and LCA. The impact of these application into our future society is expected to grow in the coming years thanks to the new concept of Cloud GNSS receivers, which will play a prominent role in the definition of GNSS-based Internet of Things, Machine-to-Machine, and Smart City applications [LS16b].

For the above reasons and the widespread deployment of GNSS in our society, in this chapter we will introduce GNSSs, the positioning system considered in the rest of this dissertation. We start Section 3.2 explaining the architecture and principle of operation of any GNSS. Then,

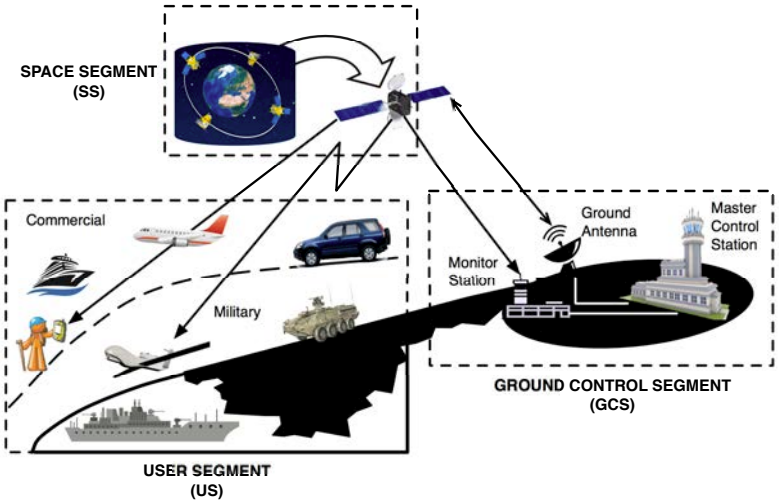

Figure 3.1: GNSSs architecture: Space Segment (SS), Ground Control Segment (GCS), and User Segment (US).

details on GNSS receivers are given in Section 3.3. Next, GNSS vulnerabilities are introduced in Section 3.4, so that their effects may be known and possible countermeasures can be though. Due to these vulnerabilities, standalone GNSS receivers performance cannot meet some requirements of some critical applications. As a matter of fact, nowadays, one of the major challenges to be solved for the GNSS community is the provision of integrity, specially to users beyond the civil aviation. GNSS integrity is presented in Section 3.5. One of the main functionalities of integrity algorithms is the fault detection and exclusion, being the detection of any integrity fault of vital interest. In order to understand the detection process from a mathematical point of view, the second part of this chapter is devoted to the introduction of the field of detection theory. This field has its foundations on the disciplines of probability theory and statistics. This is why in Section 3.6 we provide a brief technical review of the main concepts needed to understand the rest of the theoretical developments in this dissertation. Once we have introduced the background on probability theory, we will start our study of the detection theory. This theory is fundamental to the design of decision making systems and it is the subject of Section 3.7. Finally, Section 3.8 concludes this chapter.

3.2 Global Navigation Satellite Systems (GNSS)

The main purpose of GNSS is to allow users to estimate their PVT anytime anywhere in the Earth. In this section we expose GNSS concepts related to its architecture, used signals and the main idea of the PVT computation. The presented discussion in this section is mainly based on the references [SG12, Par96a, Mis11] and [Kap05].

3.2.1 GNSS architecture

GNSSs are composed of three different complementary segments, as shown in Figure 3.1:

- **The Space Segment (SS)**: Is the set of satellites in orbit that provide the ranging signals and data messages (or navigation messages) to the user equipment.

Constellations	GPS	Galileo	GLONASS	COMPASS
Country	United States	Europe	Russia	China
Altitude	20200 km	23222 km	19100 km	21528 km
Period	≈12 h	≈14 h	≈11 h	12 h 38 min
Orbital planes	6	3		
Number of satellites	24	30	24	27

Table 3.1: Space segment for different GNSSs.

- **The Ground Control Segment (GCS)**: It tracks and maintains the satellites in space. The GCS monitors the satellite health and it maintains the satellite orbits as desired. Furthermore, the GCS updates the satellite clock corrections and ephemerides, among other parameters, encapsulated in the navigation message.
- **The User Segment (US)**: Composed by the GNSS receivers, in charge of performing the navigation, timing, or other related functions (e.g. surveying).

In particular, the SS consists of a constellation of artificial satellites orbiting around the Earth. The main goal of these satellites is to transmit radio-navigation signals with a specific signal structure, and to store and re-transmit the navigation message sent by the GCS. These transmissions are controlled by highly stable atomic clocks on board the satellites. A constellation of generally around 30 Medium Earth Orbit satellites (MEO) are used, orbiting at an altitude close to 20000 km. This allows the system to provide a worldwide service, meaning that any user on Earth would see at least 4 satellites of a given constellation in an open sky environment. Typical GNSS constellations such as GPS, Galileo, GLONASS and BeiDou are described in Table 3.1 [ESA17, Nav17, GPS17].

The GCS is composed by a network of monitoring stations that store and process the signals received by the satellites in the SS. One of the main goals of the GCS is to estimate the orbit parameters, referred to as ephemeris, the satellite clock error and other parameters such as ionospheric corrections. Finally, the navigation messages are generated and uploaded to the satellites through ground antennas. The navigation message contains the information needed by any user to compute its position, such as the satellite position and the satellite clock corrections. The GCS is also responsible of maintaining the satellites and their proper functioning. Each system has its own control segment, whose stations are strategically placed around the Earth. Within the components of a GCS we find the following:

- Sensor stations, with known location, distributed over the world with the aim of monitoring the satellite positions and the transmitted signals of all satellites.
- A master control station that collects all the information from the sensor stations and extracts all the necessary information to build the navigation message.
- Communication stations to upload the messages to the satellites in view.

Finally, the US consist of all GNSS receivers including space, air, ground and marine. Early receivers were designed for military applications, they were heavy and large compared to recent devices. Receiver's cost and dimensions change rapidly according to the intended application. For instance, a receiver may be embedded in a cell phone like an integrated chip or it can be

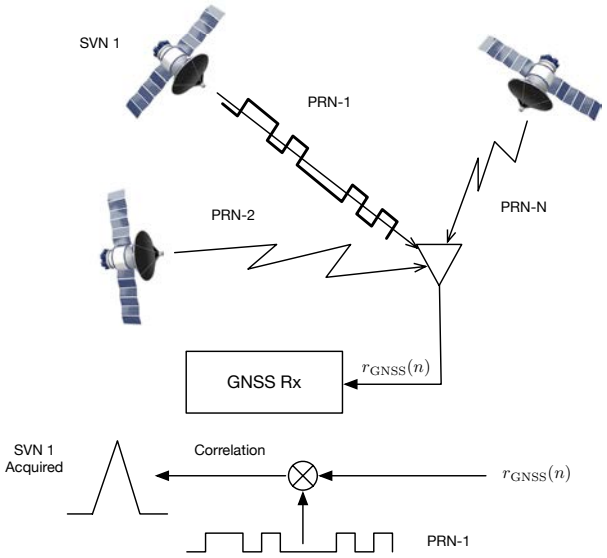

Figure 3.2: Basic principles of DSSS signals.

placed in a vessel as a big box. The main function of the GNSS receiver is to gather and process GNSS signals, in order to determine the PVT solution of the user. More details on GNSS receivers will be given in Section 3.3.

3.2.2 GNSS signals

GNSS signals fall into a broad category of signals known as spread spectrum signals. Specifically, in general, GNSS use Direct Sequence Spread Spectrum (DSSS) signals. These signals use orthogonal spreading codes, which serve as the basic tool for measuring the propagation time from satellites to receiver. Different satellites are generally associated to different orthogonal spreading codes. The orthogonality of these codes implies that the cross-correlation between two different sequences is ideally zero, or almost zero in practice. Thus, different satellites are identified by correlating the received GNSS signal by the associated spreading code, which are known and stored in the receiver. When this correlation is not equal to zero we say the corresponding satellite is acquired. Examples of systems using DSSS signals are the GPS, the future European Galileo and the Chinese BeiDou. The Russian GLONASS instead adopts a Frequency Division Multiple Access (FDMA) to differentiate different satellites and uses the same spreading sequence in all satellite signals. The DSSS signaling is illustrated in Figure 3.2, and it is composed of the following terms:

- **A carrier** to support the signal information at a suitable frequency. The frequency bands used for GNSS are chosen to limit the impact of the signal propagation channel (i.e. atmosphere) and to limit the size of antennas.
- **The spreading code** given by a code with a large number of bits that statistically behaves as white noise. A bit of this code is known as chip and its rate as chipping rate. The chipping rate or chip rate is higher than the data rate, therefore, the name of spreading code, also called Pseudo Random Noise (PRN).
- **The useful data**, which contains the navigation message. Typically, GNSS data uses BPSK modulation with a bit rate always much smaller than the chipping rate.

The signal at the input of a GNSS receiver, in an additive Gaussian noise environment, can be written as the sum of N useful signals, broadcast by N different satellites, and a noise term. The input signal is captured by the receiver antenna, down converted and filtered by the receiver front-end, then the signal after the Analog to Digital Converter (ADC) can be written as

$$r_{\text{GNSS}}(n) = \sum_{i=1}^{N} A_i s_i(n-\tau_i) e^{j2\pi(f_{\text{RF}}+f_i)n+\varphi_i} + w(n), \tag{3.1}$$

with $\{A_i, \tau_i, f_i, \varphi_i\}$ denoting the amplitude, delay, Doppler frequency, and random phase of the received i-th useful signal; f_{RF} the receiver intermediate frequency, and s_i the useful signal transmitted by the i-th satellite, which can be expressed as

$$s_i(n) = \sum_{l=-\infty}^{\infty} d_l \sum_{k=0}^{N_{\text{r}}-1} c_i\left(n - kT_{\text{code}} - lT_d\right), \tag{3.2}$$

where $d_l = \{-1, 1\}$ are the data symbols transmitted at a rate of $R_d \doteq 1/T_d$ bits per second, constituting the navigation message; $c_i(n)$ is the spreading code or PRN sequence used by the i-th satellite. This code is composed of a sequence of pulses with a total time duration (i.e. code period) of T_{code} and it is repeated N_{r} times within each bit interval. This is the case for instance of GPS. In Galileo, instead, each N_{r} code is modulated by using amplitudes $u_k = \{-1, +1\}$, generated by the so-called secondary code with length N_{r}. Actually, s_i is the DSSS signal with $c_i(n)$ the composite code formed by the PRN chips with duration T_{chip}, defining the signature of the i-th satellite with a chipping rate of $R_c \doteq 1/T_{\text{chip}} = N_c/T_{\text{code}}$. The term $w(n)$ is assumed to be Additive White Gaussian Noise (AWGN). At this point, it is worth noting that each GNSS satellite transmits several signals with the same structure as in (3.2). For instance, traditionally, GPS has offered two signals for civil use and several encrypted military signals. The so-called Coarse-Acquisition (C/A) signal, is the one used by all GPS commercial receivers, whereas the precision (P) signal is only available for military purposes. Moreover, different frequencies are used by the satellite to transmit at the same time signals at different frequency bands, for instance L1 and L2 bands. For more details on the properties of GNSS signals see [Mis11] and [Kap05].

3.2.3 GNSS positioning principle

The simplest of the principles of radio wave propagation is that these waves travel at a known speed, so that if the transit time of a signal is known, the distance between transmitter and receiver can be measured. Position estimation based on distance measurements is referred to as *trilateration*, when three measurements are used, or *multilateration*, when more than three measurements are used. The basis of GNSS positioning is the measurement of distance by means of the Time of Arrival (TOA). To do so, each satellite transmits signals with a time stamp associated with it. At the user's side, the GNSS receiver measures the time of reception of the transmitted stamp, and then it estimates the propagation delay for each satellite as the difference between the measured time of reception and the transmitted stamp time. With this difference τ_i, the distance between the i-th satellite and the user can be calculated as $d_i = c\tau_i$, with c being the speed of light. In practice, the time of reception is measured by aligning the incoming signal with a local replica of the transmitted signal, which is known at the receiver. The precise

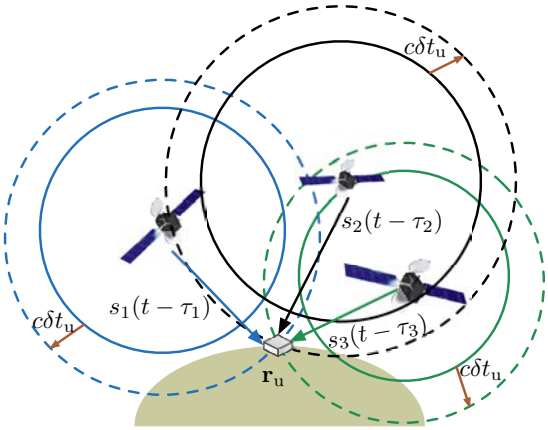

Figure 3.3: Illustration of the positioning principle behind GNSS and the effect of unknown clock bias.

estimation of this time is made possible by the use of DSSS signals. Recall that multilateration requires measurement of ranges to three (or more) position references. In GNSS, the position references are the satellites, whose position is known thanks to the navigation message created by the GCS and transmitted by the SS.

The estimated distance provides the radius of the spherical surface centered at the satellite and containing the user's location. Actually, this would be true if the clocks in the satellite and receiver are synchronized, but this is not usually the case. Hence, all distance measurements d_i are actually shifted by an unknown bias $c\delta t_u$, where δt_u denotes the bias between the receiver and satellites clock. This bias affects the measured distances equally. The corresponding measurements are thus all too short, or too long, by a common amount, thus the name *pseudoranges*. The receiver clock bias thus becomes an additional unknown to be estimated, in addition to the position coordinates. The case of two-dimensional positioning is shown in Figure 3.3, for the sake of illustration, including the effect of the unknown clock bias. It can be seen that the intersecting points of any pair of circumferences change for different values of δt_u. The key concept underlying GNSS positioning is that the three circumferences share a unique intersection point for an appropriate value of δt_u. This means that, in a two-dimensional problem, measurements to at least three satellites are needed to determine the user's position. Applying the same arguments, a user needs a minimum of four satellites in view to estimate a three-dimensional (3D) position.

3.3 GNSS Receivers

A GNSS receiver is manly aimed at determining the user's position based on a set of received signals coming from the satellites in view. Figure 3.4 shows the block diagram of a standard GNSS receiver, which is composed of the following parts:

- **Front-end**: In (a), the front-end module is a generic block that carries out the analog signal conditioning. That includes band-pass filtering, low-noise amplification, base-band conversion, and analog-to-digital conversion. The rest part of a GNSS receiver is based on digital processing.
- **Signal Processing Module**: In (b), it is a set of specific signal processing modules required to reach the ultimate goal of obtaining the user's position. It includes:

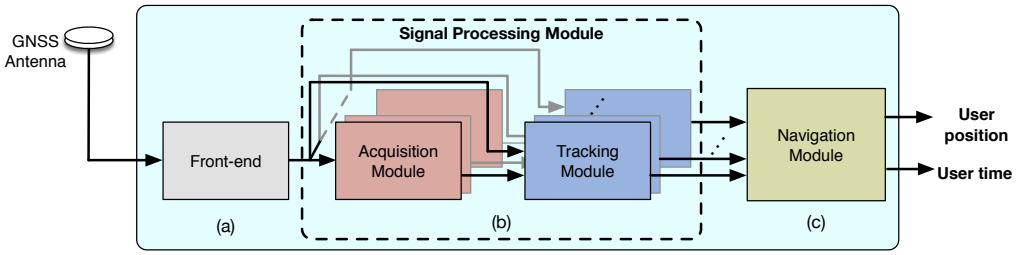

Figure 3.4: General architecture of a standard GNSS receiver.

  - *Acquisition Module*: Aimed at detecting and identifying the satellite signals in view.
  - *Tracking Module*: Aimed at tracking any variations on the acquired signals properties.

- **Navigation Module**: In (c), the navigation module is aimed at solving the user's position, based on the conceptual procedure presented in the previous section.

The signal processing module is the core of a GNSS receiver, performing several functions such as splitting the received signal into multiple channels to process multiple satellites simultaneously and demodulating the system data; generating the reference PRN code of the signals; acquiring the satellite signals and tracking the code and carrier of these signals. The outputs of the signal processing module are basically pseudoranges, carrier phase measurements, and data from the navigation message, and they are used by the navigation module in order to obtain the position and time solution of the user (sometimes referred as to PVT). In this section, we will focus on the digital processing part of a GNSS receiver (i.e. acquisition, tracking and navigation module). This is so because it is in this part where signal processing techniques can be applied to get the ultimate goal of this thesis; namely to provide signal-level integrity. It is important to say that the front-end has an impact on the received signal and it will depend on its design. Nevertheless, we do not study the front-end design here and we consider it given.

3.3.1 Acquisition module

Once the received signal has been properly conditioned by the front-end, the next task of a GNSS receiver is to detect the satellites that are currently in view [SG12]. These satellites will be used as anchor nodes to obtain the user's position, which will be calculated by measuring the propagation time from the satellites to the receiver. The main goal of the acquisition module is to calculate a rough estimate of this time-delay from satellite to receiver. Moreover, the presence of a Doppler shift in the received signal (due to receiver or satellite movement) forces the acquisition module to perform a two-dimensional (time-frequency) search. To do so, the correlation between the received signal and a locally generated replica of the PRN sequence is performed with several code-delay and Doppler shift values. This process is repeated for different values of code-delay and Doppler shift. The signal is then declared to be acquired when the magnitude of the obtained correlation exceeds a given threshold. Thereby, the acquisition code-delay and Doppler shift estimates become those values where the magnitude of the correlation gives the largest peak, depicted in its time-frequency representation in Figure 3.5. The minimum Coherent Integration (CI) time of these correlations is thus N_{scode} samples (i.e. one code period) but it can be extended in a non-coherent way by summing non-linearly together up to N_{NC} code periods as in [GC16].

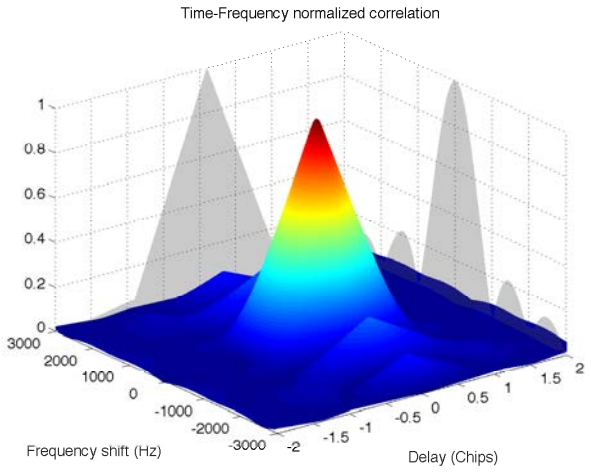

Figure 3.5: Time-Frequency representation of the correlation peak to be acquired.

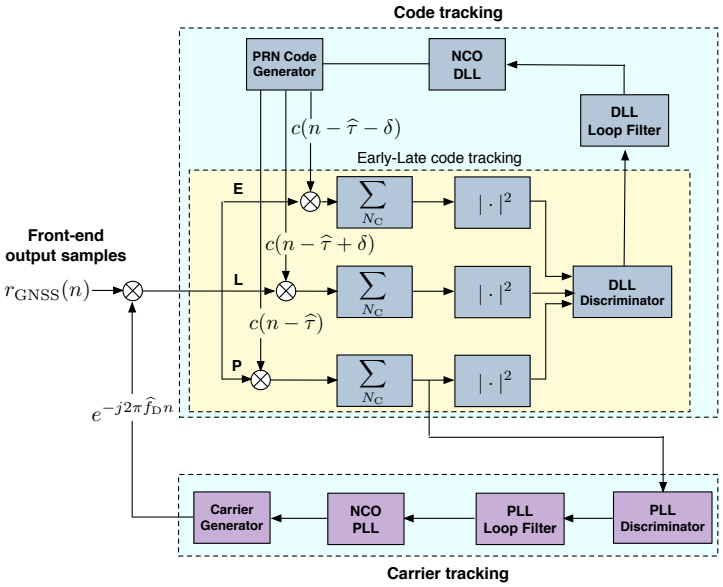

Figure 3.6: General architecture of the tracking module of a traditional GNSS receiver [SG12].

3.3.2 Tracking module

The acquisition stage provides rough estimates of the code-delay, τ, and the Doppler frequency, f_{D}, of the satellites in view (at least those declared acquired). The main goal of the tracking stage is to refine these initial estimates of $\{\tau, f_{\mathrm{D}}\}$ and to continuously track changes into these magnitudes. To do so, the tracking loop tries to keep the local PRN replica aligned with the received code, so that the coarse estimates in the acquisition stage are refined, and possible variations due to movement (either in the receiver or satellite) are tracked. For this purpose, two parallel closed-loop architectures are implemented in the tracking module, which are referred to as the Delay-Lock Loop (DLL) for the code-delay tracking, and the Phase-Lock Loop (PLL) for the carrier tracking. Three important concepts have to be introduced in order to explain the closed-loop tracking stage presented in Figure 3.6:

1. **Discriminator functions**, which use the correlator outputs in order to provide measurable values of errors in the code-delay and Doppler frequency estimates.

2. **Numerical Controlled Oscillator (NCO)**, which converts the filtered discriminator output into a frequency that controls the generation of the local replica. Two different NCOs are used, one for the PRN code and another for the carrier generation.

3. **Loop filters**: The output of the discrimination function is filtered with the aim of reducing the noise at the input of the NCO and avoid instabilities.

As closed-loop schemes, the goal of the DLL and the PLL is to obtain accurate estimates of the code-delay and Doppler frequency, respectively, by correlating the input signal with the local replica and driving the resulting error signal to zero. In both cases, when the replica is actually correlated with the input signal, the resulting correlation is fed to a discriminator that will output a signal roughly proportional to the error in the parameters to be tracked. As shown in Figure 3.6, three correlations are indeed performed for code tracking: one computed at the prompt correlation (i.e. located at the code-delay estimate), and the other two located symmetrically before and after the prompt one, which are referred to as the *early* and *late* correlators, respectively. This kind of configuration leads to the family of early-late methods, where the discriminator output is obtained by comparing in some way or another the early and late correlators [Kap05, Ch. 5]. The discriminator output is then smoothed with a loop filter before driving the corresponding NCO, thus closing the loop. This kind of closed-loop architecture has been widely analyzed in past decades, with several contributions regarding coherent and non-coherent discriminators, and closed-loop performance analyses [Par96a, Ch. 7], [Mis11, Ch. 11]. More recent tracking schemes are based on the use of adaptive Kalman filter-based techniques as in [LS16a], with the aim of providing an enhanced robustness compared to the conventional DLL/PLL-based techniques under harsh environments.

3.3.3 Navigation module

The pseudoranges computed by the GNSS receiver are a biased and noisy version of the distance between satellites and the user. From these biased and noisy measurements, the GNSS receiver has to estimate the PVT in an optimal way, or at least in a smart way. In this section we briefly review this process, which is performed by the navigation module depicted in Figure 3.4. There exist two families of navigation solutions. The ones in which the latter is obtained in closed-form, based on the Bancroft algorithm [Ban85]. These techniques provide worse accuracy than those in the second family, which are composed of iterative methods based on the linearization with respect the position solution, $\mathbf{r}_{\mathrm{u}} \doteq [x_{\mathrm{u}}, y_{\mathrm{u}}, z_{\mathrm{u}}]^T$. In this section and for the sake of illustration, we present the iterative solution based on the Least-Squares (LS) estimation.

Let us start with the basic pseudorange measurement equation, which will be the base model to be fitted by LS. The pseudorange measurement from the i-th satellite can be modeled as

$$d_i = \rho_i + c\left(\delta t_{\mathrm{u}} + \delta t_{\varepsilon,i}\right), \tag{3.3}$$

where ρ_i denotes the true distance between the i-th satellite and the receiver, δt_{u} is the user clock bias, and $\delta t_{\varepsilon,i}$ is the bias contribution of all the disturbing effects that may be present, and c is the speed of light. A user might correct each measured pseudorange for the known errors taking advantage of the corrections embedded into the navigation message broadcast by

the GNSS satellites [Mis11, Ch. 6]. Thus, we can write

$$\tilde{d}_i = \rho_i + c\delta t_{\mathrm{u}} + \epsilon_{\rho,i}, \tag{3.4}$$

with $\tilde{d}_i$ the corrected pseudorange after subtracting all available error corrections, and $\epsilon_{\rho,i}$ the combined effect of the residual errors. Now, the geometric distance ρ_i is linearized around a tentative value of the user's position, $\mathbf{r}_{\mathrm{u},0} \doteq [x_{\mathrm{u},0}, y_{\mathrm{u},0}, z_{\mathrm{u},0}]^T$, by keeping the first-order term of the Taylor series

$$\rho_i = ||\mathbf{r}_{\mathrm{s},i} - \mathbf{r}_{\mathrm{u}}|| \approx ||\mathbf{r}_{\mathrm{s},i} - \mathbf{r}_{\mathrm{u},0}|| - \frac{(\mathbf{r}_{\mathrm{s},i} - \mathbf{r}_{\mathrm{u},0})^T}{||\mathbf{r}_{\mathrm{s},i} - \mathbf{r}_{\mathrm{u},0}||}\Delta\mathbf{r}_{\mathrm{u}}, \tag{3.5}$$

where $\Delta\mathbf{r}_{\mathrm{u}} = \mathbf{r}_{\mathrm{u}} - \mathbf{r}_{\mathrm{u},0}$ and $\mathbf{r}_{\mathrm{s},i} \doteq [x_{\mathrm{s},i}, y_{\mathrm{s},i}, z_{\mathrm{s},i}]^T$ the position of the i-th satellite, extracted from the navigation message.

Using this approximation and collecting all of the corrected pseudoranges together, for N_{sat} satellites, we can write in matrix notation

$$\mathbf{y} = \mathbf{H}\Delta\mathbf{u} + \boldsymbol{\epsilon}_{\rho}, \tag{3.6}$$

with $\boldsymbol{\epsilon}_{\rho}$ the pseudorange measurement error vector, composed by the measurement errors of the N_{sat} satellites in view, $\mathbf{y}$ the $(N_{\mathrm{sat}} \times 1)$ corrected pseudorange measurement vector after linearization and $\Delta\mathbf{u}$ the navigation solution vector after linearization (including the receiver clock bias), respectively given by

$$\mathbf{y} = \begin{bmatrix} \tilde{d}_1 \\ \vdots \\ \tilde{d}_{N_{\mathrm{sat}}} \end{bmatrix} - \begin{bmatrix} ||\mathbf{r}_{\mathrm{s},1} - \mathbf{r}_{\mathrm{u},0}|| \\ \vdots \\ ||\mathbf{r}_{\mathrm{s},N_{\mathrm{sat}}} - \mathbf{r}_{\mathrm{u},0}|| \end{bmatrix}, \tag{3.7}$$

$$\Delta\mathbf{u} = \begin{bmatrix} \Delta\mathbf{r}_{\mathrm{u}} \\ c\delta t_{\mathrm{u}} \end{bmatrix} = \begin{bmatrix} x_{\mathrm{u}} - x_{\mathrm{u},0} \\ y_{\mathrm{u}} - y_{\mathrm{u},0} \\ z_{\mathrm{u}} - z_{\mathrm{u},0} \\ c\delta t_{\mathrm{u}} \end{bmatrix}, \tag{3.8}$$

and $\mathbf{H}$ the following $N_{\mathrm{sat}} \times 4$ observation matrix

$$\mathbf{H} = \begin{bmatrix} \frac{(x_{\mathrm{u},0} - x_{\mathrm{s},1})}{||\mathbf{r}_{\mathrm{s},1} - \mathbf{r}_{\mathrm{u},0}||} & \frac{(y_{\mathrm{u},0} - y_{\mathrm{s},1})}{||\mathbf{r}_{\mathrm{s},1} - \mathbf{r}_{\mathrm{u},0}||} & \frac{(z_{\mathrm{u},0} - z_{\mathrm{s},1})}{||\mathbf{r}_{\mathrm{s},1} - \mathbf{r}_{\mathrm{u},0}||} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{(x_{\mathrm{u},0} - x_{\mathrm{s},N_{\mathrm{sat}}})}{||\mathbf{r}_{\mathrm{s},N_{\mathrm{sat}}} - \mathbf{r}_{\mathrm{u},0}||} & \frac{(y_{\mathrm{u},0} - y_{\mathrm{s},N_{\mathrm{sat}}})}{||\mathbf{r}_{\mathrm{s},N_{\mathrm{sat}}} - \mathbf{r}_{\mathrm{u},0}||} & \frac{(z_{\mathrm{u},0} - z_{\mathrm{s},N_{\mathrm{sat}}})}{||\mathbf{r}_{\mathrm{s},N_{\mathrm{sat}}} - \mathbf{r}_{\mathrm{u},0}||} & 1 \end{bmatrix}. \tag{3.9}$$

Clearly, four equations are required at least to solve (3.6). In order to minimize the impact of the errors in $\boldsymbol{\epsilon}_{\rho}$, the standard LS solution is computed as

$$\widehat{\Delta\mathbf{u}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{y}. \tag{3.10}$$

The user's position is updated using (3.10), and the process can be iterated taking the resulting position as the new tentative value around which the equations are linearized again. In practice, two iterations are usually enough to converge to the real solution, even if the first tentative position has an error on the order of kilometers [SG12].

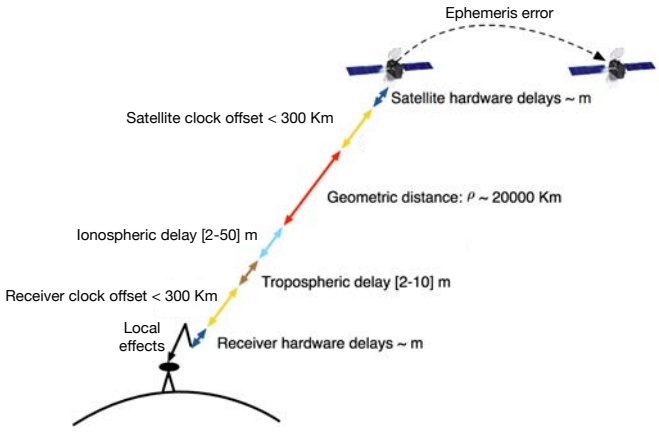

Figure 3.7: Pseudorange measurement errors due to GNSS threats.

3.4 GNSS Errors

Even though GNSSs are known to be the most accurate global navigation systems, they still can experience significant errors. By understanding these errors, designers can both try to reduce them and to realize the limitations of the GNSSs systems. To do so, a thorough review of the possible threats and their effects to GNSS is needed. In fact, a plethora of references are available dealing with the characterization and analysis of GNSS threats, grouping them as follows:

- **Control segment errors**: Those errors in the parameter values broadcast by a satellite in its navigation message for which the GCS is responsible.
- **Propagation errors**: Those uncertainties associated with the propagation medium of the GNSS signal, which affect the propagation time of the signal from the satellite to the receiver.
- **Local errors**: Those errors that are dependent on the location of the receiver. They also include those errors in the receiver's or transmitter's instrumentation.
- **Measurement errors**: All the previous error sources affect the quality of the pseudorange measurement and the ultimate accuracy of the navigation solution. It is for this reason that different models are available to model these error sources and their effects.

Figure 3.7 summarizes these threats and gives orders of magnitude of the errors introduced by each threat. This section provides and overview of these error sources, with particular emphasis on local errors. The main goal is to understand the nature of these errors, and to analyze their effects into the pseudorange measurements. For details on GNSS threats see [Mis11, Ch. 5], [Kap05, Ch. 7] and [Par96a].

3.4.1 Control segment errors

Satellite clocks are highly stable, but unfortunately they cannot remain synchronized forever. Similarly, ephemeris transmitted by the satellites might be affected by some inaccuracies due to the difficulty to predict the dynamic of the satellite. It is for this reason that one of the main purposes of the GCS is to generate satellite ephemerides and clock corrections, which

are regularly updated to the satellites. These parameters are then included as part of the navigation message transmitted by the satellites, so that user's receivers can use them to apply real-time pseudorange corrections, needed for the accurate computation of the PVT solution. Unfortunately, there are some residual errors that typically vary from 0.3–4 m and around 1–6 m for clock and ephemeris imperfection errors.

3.4.2 Propagation errors

The GNSS signals are affected by the medium through which they propagate from the satellites to a receiver. The 95% of the time the signal propagates through vacuum or free space, in which the electromagnetic signals travels with a constant speed. Before arriving the surface of the Earth, though, the signals enter into the atmosphere, including the ionosphere and troposphere. In general, the atmosphere changes the velocity (speed and direction) of propagation of electromagnetic signals, effect known as refraction. The change in speed of propagation modifies the signal propagation time, which is the basic measurement in GNSS, thus affecting the accuracy of the system. In particular, the ionosphere extends from a height of about 50–1000 km above the Earth, and it is composed of different layers at different heights, each with different densities of free electrons (due to the sun's radiation). Actually, as the sun rises, its radiation starts breaking free electrons and creating ions from the gas molecules in the ionosphere. At night, there is no further ionization and the ions and electrons start recombine, reducing the free-electron density. The speed of propagation of radio signals in the ionosphere depends upon the number of free electrons encountered in the path of a signal. Fortunately, the ionosphere is a dispersive medium (i.e. dependent on the frequency of the signal), so the ionospheric delay is frequency-dependent, a property that can be exploited to mitigate the ionospheric error using measurements with two different frequencies. Once the GNSS signal has gone through the ionosphere it is also refracted by the lower part of the Earth's atmosphere, which is composed of dry gases and water vapor. Unlike the ionosphere, the troposphere is non-dispersive for GNSS frequencies. Still, the speed of propagation of the GNSS signal in the troposphere is lower than that in free-space and, thus, the apparent range to a satellite seems longer, typically by 2.5–25 m depending upon the satellite elevation angle. Unfortunately, this delay cannot be mitigated by means of dual-frequency measurements, as it is the case of the ionospheric delay, and a user has to resort to models or differential corrections in order to correct it.

3.4.3 Local errors

Pseudorange errors discussed so far are global effects in the sense that they do not depend on the characteristics of the environment where the user's receiver is located. Now we analyze a different kind of errors that actually depend upon the local characteristics of the receiver environment such as buildings, temperature, or possible radio frequency signals sharing the frequency band used by GNSSs. The main difference of local errors with respect to global errors is that the former cannot be mitigated using differential corrections. Since those threats has already been defined in Chapter 2, we present here the effects of these threats on GNSS receivers and its performance.

Multipath

Multipath is the phenomenon whereby a signal arrives at a receiver via multiple paths, at-

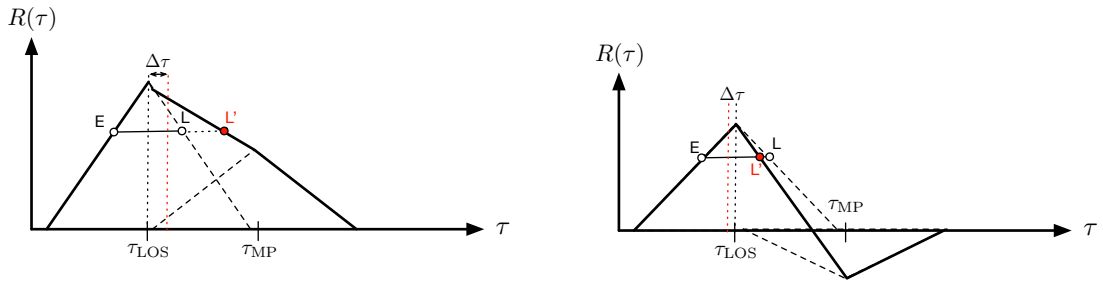

Figure 3.8: Correlation curve $R(\tau)$ (solid line) with (left) constructive and (right) destructive multipath from a reflected ray (dashed line).

tributable to reflection and diffraction. Indeed, multipath represents the dominant error source in today's GNSSs. This is because multipath is a highly localized phenomenon, and then it cannot be corrected by means of differential or augmentation systems. Moreover, multipath distorts the correlation curve used to obtain the pseudorange measurements needed to compute the navigation solution. Typically, an antenna receives the direct (i.e. LOS) signal and one or more of its reflections from structures in the vicinity of the receiver and the ground. When the LOS component is not present, the multipath is classified into the NLOS multipath condition. The main contribution of multipath into the GNSS receiver is a distortion of the correlation peak due to the additional components in the correlation curve due to multipath. In the absence of multipath we assume the correlation curve is a pristine triangle. If additional signals arrive into the receiver due to multipath, they will introduce additional correlation peaks and the early and late correlator samples may not be centered on the true arrival time of the direct ray, as desired under nominal conditions (see Section 3.3). The reflected ray will always arrive after the direct signal and creates a delayed correlation curve centered at τ_{MP}, as shown in Figure 3.8. If the multipath is constructive (i.e. in-phase multipath), then the multipath component of the correlation curve is added to the direct one, and thus the measured pseudorange will be longer than that due to the direct path, as shown in the left plot with the red doted line.

On the other hand, if the multipath is destructive (i.e. counter-phase multipath), then the multipath component is subtracted from the direct component, and the measured pseudorange will be shorter than that due to the direct path, as shown in the right plot. The relative delay of the multipath plays a major role in determining its effect. If the delay is long compared to a chip width, T_{chip}, then the auto-correlation properties of the code will eliminate the effect of multipath. This is due to the intrinsic multipath immunity of spread spectrum signals, used in GNSS. No errors exist when the rising edge of the delayed curve does not touch the direct correlation curve. This is why the vulnerability against multipath can be reduced by using a narrow correlator spacing. The primary defense against multipath is to locate the antenna away from reflectors, but that is not always possible in practice. The effect of multipath can be also reduced in antenna design process by lowering the contribution of some types of reflections (e.g. from the ground). Another alternative to reduce multipath is in the signal processing stage, where several mitigation techniques can be applied. Typical multipath error in pseudorange measurements vary from 1 m in a benign environment to more than 5 m in highly reflective environments.

Interference

A GNSS signal travels around 20000 km from MEO satellites, and then the received signal power at the Earth's surface is very low (approximately 10^{-16} W). This power is comparable

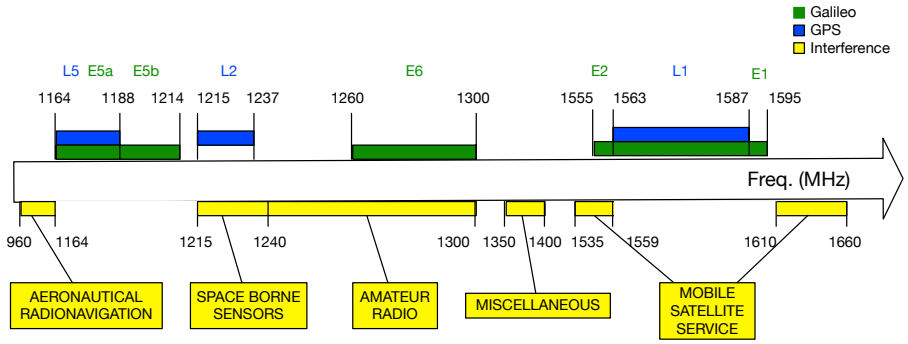

Figure 3.9: Terrestrial radio frequency services sharing frequency bands with GNSSs.

to the natural noise contained in a 1 MHz bandwidth at the GPS L1 or L2 frequency. It is for this reason that any other man-made radio frequency signal falling in the same band is very likely to interfere the GNSS signal, becoming a great challenge due to the very saturated radio spectrum. Although the radio regulatory agencies work hard to protect the GNSS spectrum, and most terrestrial radio transmissions limit the amount of energy they send in the vicinity of GNSS bands, properly operating equipment can still result in interference if it is close to a GNSS receiver. Unfortunately, some RFI is intentional like jamming and spoofing, trying to completely deny the GNSS service or fake the position provided by the GNSS receiver, respectively. The characteristics and types of jamming and spoofing attacks are similar for any positioning system, so the discussion given in Chapter 2 is valid for GNSS as well. In this section we focus on unintentional interference sources and their effects into GNSS receivers. A comprehensive overview of interference sources and effects in GNSS is provided in [Kap05, Ch. 6]. Now, in order to have an idea of possible interference sources in GNSS nowadays, it is interesting to review the frequency plan near the band where GNSSs are placed (see Figure 3.9).

First of all, GPS L5 and Galileo E5 are located within a portion of the 960-1164 MHz band, which is used worldwide for electronic aids to air navigation, including Distance Measuring Equipment (DME) and TACAN (TACtical Air Navigation) ground beacons emitting on frequencies fallen in the L5 band, with pulsed signals. Others services extensively used today falling in some GNSS band are space borne sensors, which uses Synthetic Aperture Radar (SAR) to obtain remote images of celestial bodies; Amateur radio; Mobile satellite services; and miscellaneous services like geostationary communication satellites services (Radio, TV, Internet,...) and mobile communications. All of the services falling in some GNSS band are listed in Table 3.2 with some interference sources, frequency band occupation and a brief description of the source. With this information, it becomes evident the high density of interference sources on GNSS receivers. Finally, it is worth commenting that some out-of-band energy from the signals in adjacent bands of GNSSs will sometimes fall within the range of frequencies processed by GNSS receivers, as it may be the case of harmonics or intermodulation products of these neighboring bands. Among the possible effects of RFI to GNSS are: the loss of receiver tracking due to a high power interference that saturates the ADC levels of the GNSS receiver, denying the use of GNSS; or a reduction of the measured signal quality due to an effective increment of the noise power due to the spreading of the interference in the GNSS receiver, this can cause instabilities in the code delay estimation, and thus, errors in the pseudorange measurements.

Service	Interference Source	Frequencies (MHz)	Description
Aeronautical Navigation	DME	960–1164	Located near airfields transmitting pulsed signals to obtain distance measurements.
	TACAN		System that uses DME and gives bearing information using pulsed signals.
	ADS		System used in airplanes to transmit position and altitude information among others airplanes.
	SSR		Radar used in air traffic control for detecting and measuring aircraft position with pulsed signals.
Space Borne Sensors	SAR	1215–1240	Radar used in remote sensing an mapping of the surface of the Earth and other planets.
Amateur Radio	Amateur radio users	1240–1300	Persons interested in radio technique with a personal aim for purposes of private recreation.
	Miscellaneous	1350–1400	Include geostationary communication satellites, mobile communications and radio-location.
Mobile Satellite	Mobile communications	1535–1559 1610–1660.5	Mobile satellite communications down-link/up-link

Table 3.2: Unintentional Interference Sources in GNSS receivers.

3.4.4 Measurement errors

Let us come back to the pseudorange measurements equation in (3.6), given by $\mathbf{y} = \mathbf{H}\boldsymbol{\Delta}\mathbf{u} + \boldsymbol{\epsilon}_\rho$, where $\boldsymbol{\epsilon}_\rho$ denotes the pseudorange measurement error vector, composed by the measurement errors of the N_{sat} satellites in view. Errors may occur by different reasons: a satellite clock malfunction, high power and large (delay) NLOS multipath, large and punctual ionospheric delays (e.g. due to magnetic storms), ... These errors are the addition of nominal errors and other faults not included in the nominal model, that is, $\boldsymbol{\epsilon}_\rho = \boldsymbol{\epsilon}_{\text{n}} + \mathbf{b}$, where $\{\boldsymbol{\epsilon}_{\text{n}}, \mathbf{b}\}$ are the nominal and fault error vectors, respectively, classified as *noise* and *bias*, respectively. Noise generally refers to a quickly varying error with zero mean, whereas a bias tend to persist over a period of time. This distinction is helpful in understanding the effects of the measurement errors. Indeed, the fault error vector is modeled as a bias added to the nominal error, including the bias in each satellite, ($\mathbf{b} = [b_1, b_2, \ldots, b_{N_{\text{sat}}}]^T$).

Nominal errors

Nominal errors characterize the pseudorange measurement errors that are present under nominal operation of the whole GNSS system. They are typically modeled as zero-mean independent Gaussian errors,

$$\boldsymbol{\epsilon}_{\text{n}} \sim \mathcal{N}(0, \boldsymbol{\Sigma}), \tag{3.11}$$

where $\boldsymbol{\Sigma}$ denotes the covariance matrix of the nominal error vector, given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2_{\text{UERE},1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\text{UERE},2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & 0 \quad \sigma^2_{\text{UERE},N_{\text{sat}}} \end{bmatrix}. \tag{3.12}$$

Nominal errors are modeled as the combination of several independent error sources as follows:

$$\epsilon_{\text{n},i} = \epsilon_{\text{s},i} + \epsilon_{\text{e},i} + I_i + T_i + \epsilon_{\text{l},i}, \tag{3.13}$$

where $\epsilon_{\text{s},i}$ is the error resulting from the satellite's clock bias, $\epsilon_{\text{e},i}$ is the error incurred due to

	GPS L1 C/A	GPS L5	Galileo E1C	Galileo E5a
σ_{l} (m)	0.55	N/A	N/A	N/A
σ_{T} (m)	0.2–1.5	0.2–1.5	0.2–1.5	0.2–1.5
σ_{I} (m)	4.6–13.7	8.2–24.5	4.6–13.7	8.2–24.5
σ_{s} (m)	0.6	N/A	N/A	N/A
σ_{e} (m)	0.6	0.6	0.6	0.6
σ_{UERE} (m)	5–14	8–25	5–14	8–25

Table 3.3: Order of magnitude of pseudorange measurement errors for different GNSS bands.

imperfections in the ephemerides, I_i and T_i are the errors due to propagation effects, and $\epsilon_{\mathrm{l},i}$ represents the contribution of local effects. Hence, the nominal error of the i-th satellite is modeled as

$$\epsilon_{\mathrm{n},i} \sim \mathcal{N}\left(0, \sigma^2_{\mathrm{UERE},i}\right), \tag{3.14}$$

and

$$\sigma^2_{\mathrm{UERE},i} \doteq \sigma^2_{\mathrm{s},i} + \sigma^2_{\mathrm{e},i} + \sigma^2_{\mathrm{I},i} + \sigma^2_{\mathrm{T},i} + \sigma^2_{\mathrm{l},i}, \tag{3.15}$$

the total variance of the pseudorange nominal error, the so-called User Equivalent Range Error (UERE). Several analyses have been performed in order to evaluate the UERE of different receivers, analyzing the contributions of the different error sources. An example of the order of magnitude of pseudorange measurement errors is given in Table 3.3, for different GNSS frequency bands [Pag16]. We see how GPS L1 and Galileo E1C frequency bands provide the best results with σ_{UERE} values between 5 and 15 meters, while GPS L5 and Galileo E5a offer a slightly poorer performance. We clearly see that the largest nominal error source is the ionospheric residual error, being larger for the L5/E5a frequency bands. In general, this error is higher in Galileo than in GPS because the residual error model obtained for Galileo corrections is greater than that obtained with GPS corrections [SA12]. A user would refine this error budget to take into account the knowledge of an actual measurement scenario. For instance, the atmospheric propagation error in the equatorial region or at high latitudes may double for a single-frequency receiver during high solar activity. On the other hand, this error, for a dual-frequency receiver, can be completely canceled. Similarly, a user with a receiver located near reflectors may triple the error due to multipath.

Geometric Dilution of Precision (GDOP)

So far we have analyzed the errors of the pseudorange measurements, but it is also important to determine the impact the satellite geometry has on the position estimates. Figure 3.10 illustrates how the measurement error translates into different position uncertainty for different geometries. While the quality of the measurement is the same, clearly the quality of the position estimates is not. The region of uncertainty, shown as red area, varies depending on the geometry of the satellites. For GNSS positioning, the quality of the position estimates is characterized in terms of the user-satellite matrix $\mathbf{H}$. From (3.10) and based on the presented model for nominal conditions, and assuming that $\mathbf{\Sigma} = \sigma^2_{\mathrm{UERE}}\mathbf{I}$, we have

$$\mathrm{Cov}\left[\Delta\mathbf{u}\right] = \sigma^2_{\mathrm{UERE}}\left(\mathbf{H}^T\mathbf{H}\right)^{-1} = \sigma^2_{\mathrm{UERE}}\mathbf{G}, \tag{3.16}$$

where $\mathbf{G} \doteq (\mathbf{H}^T\mathbf{H})^{-1}$. Denoting the i-th entry on the diagonal of matrix $\mathbf{G}$ as $g_{ii} = [\mathbf{G}]_{i,i}$, it follows from the above equation and the definition of $\Delta\mathbf{u}$ that

$$\sigma^2_x = \sigma^2_{\mathrm{UERE}}g_{11}; \quad \sigma^2_y = \sigma^2_{\mathrm{UERE}}g_{22}; \quad \sigma^2_z = \sigma^2_{\mathrm{UERE}}g_{33}; \quad \sigma^2_b = \sigma^2_{\mathrm{UERE}}g_{44}, \tag{3.17}$$

Figure 3.10: Example of position estimation with range measurements and the quality of the position estimate for different satellite geometries.

with $\{\sigma_x^2, \sigma_y^2, \sigma_z^2, \sigma_b^2\}$ the variances of the x-, y-, z-, and clock-component of the navigation solution estimate. Thus, the 3D root mean squared position error is given by

$$\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2} = \sigma_{\text{UERE}} \sqrt{g_{11} + g_{22} + g_{33}}. \tag{3.18}$$

Equations (3.17) and (3.18) show that the different components of the position estimates depend upon two factors: (i) variance of user range error (σ_{UERE}^2), and (ii) a term which depends entirely on the user-satellite geometry, given by the elements of $\mathbf{G}$ which basically depend on $\mathbf{H}$. In base of the geometry matrix, the so-called Dilution of Precision (DOP) parameters are defined in order to characterize the contribution of the user-satellite geometry. These parameters include Position-, Time-, and Geometric-DOP (PDOP, TDOP, GDOP). For instance, PDOP is defined as $\sqrt{g_{11} + g_{22} + g_{33}}$, and the 3D position estimation error is given by $\sigma_{\text{UERE}} \cdot \text{PDOP}$. The DOPs provide a simple characterization of the user-satellite geometry. The more favorable geometry, the lower DOP. The lower the DOP and σ_{UERE}^2, the better the quality of the position estimate.

3.5 GNSS Integrity: Receiver Autonomous Integrity Monitoring (RAIM)

Often a stand-alone GNSS cannot fulfill the requirements needed for some applications. To overcome this issue differential and augmentation systems were developed. Augmentation systems use additional complex and thus costly infrastructure, and they need a communication link between the system and receiver, which may not be available in some situations (e.g. urban canyon). It is for this reason that RAIM algorithms are implemented within the receiver, hence the term autonomous, so that they do not need external information provided by the augmentation system. The latter have already been studied in Chapter 2 and further details can be found scattered in [Par96b], [Par96b] and [Kap05]. In this section we will focus on RAIM, providing a comprehensive review of RAIM based on [Par96b] and [Kap05]. Work on RAIM emerged in the latter half of the 1980s and a myriad of papers have appeared in the navigation literature since then. The acronym RAIM was first suggested in 1987 in [Kal87], and it has been used almost universally ever since. Two early papers presented at an ION meeting in 1986 are of special interest because they illustrate two different approaches to RAIM [Lee86, Bro86]. Lee's paper

[Lee86] is a good example of what is nowadays referred to as a *snapshot* scheme, in which only current redundant measurements are used in the self-consistency check. On the other hand, [Bro86] presents a scheme where both past and present measurements are used in the RAIM decision. Such schemes are usually referred to as averaging or filtering schemes.

Three RAIM methods gained special attention during the inception of GNSS integrity: (i) The range comparison method first introduced in [Lee86], (ii) the LS Residuals (LSR) method suggested by Parkinson and Axelrad [Par88], and (iii) the parity method as described in [Stu88]. In particular, LSR RAIM is the most used RAIM algorithm nowadays, specially in civil aviation. The LSR RAIM has been proven in [Wal95] to be equivalent to the parity matrix method and the range comparison method, and it has become the benchmark for RAIM-based integrity and modernized evolutions. For instance, in order to account for different variances in the pseudorange measurements (from different satellites), the so-called WLSR RAIM was proposed in 1995 in [Wal95]. Current improvements are based on the availability of external information such as UEREs to implement RAIM-based integrity. Similarly to augmentation systems, but without needing complex and costly infrastructure. This evolution is what today is known as Advanced RAIM (ARAIM) [Bla12]. We will describe the LSR RAIM algorithm in detail (although not entering into all computational details) since it illustrates a number of concepts sufficient to understand other RAIM algorithms.

3.5.1 LSR integrity metric

Since positioning errors cannot be directly measured, RAIM algorithms calculate a measurable scalar parameter that provides information about pseudorange measurement errors. Let t denote such parameter, called integrity metric or just metric henceforth, with a known distribution. Specifically, LSR RAIM computes this metric from the pseudorange residual vector, $\mathbf{e}$, defined as the difference between the measured pseudoranges, $\mathbf{y}$, and its LS estimation obtained from (3.10); that is

$$\begin{aligned}\mathbf{e} \doteq \mathbf{y} - \widehat{\mathbf{y}} = \mathbf{y} - \mathbf{H}\widehat{\Delta\mathbf{u}} &\overset{(a)}{=} \left(\mathbf{I} - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right)\mathbf{y} \\ &\overset{(b)}{=} \left(\mathbf{I} - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right)\left(\mathbf{H}\Delta\mathbf{u} + \boldsymbol{\epsilon}_\rho\right),\end{aligned} \tag{3.19}$$

where the equality (a) follows by applying (3.10), and equality (b) is due to the measurement error model in (3.6). Equality (b) is important because it connects the residual vector with the measurement error, so it will be useful for the statistical characterization of the integrity metric and further design of the fault detection and exclusion schemes. Indeed, the integrity metric used in LSR RAIM is defined as the squared norm of the LS error

$$t \doteq ||\mathbf{e}||^2 = \mathbf{e}^T\mathbf{e}, \tag{3.20}$$

known as the Sum of Squared residual Errors (SSE). If pseudorange measurement errors have equal variance for all satellites, the SSE follows a chi-squared distribution in the nominal case (i.e. under $\mathcal{H}_0$), and a non-central chi-squared distribution in the faulty case (i.e. $\mathcal{H}_1$). This is so because $\mathbf{e}$ has zero-mean under $\mathcal{H}_0$, but its mean departs from zero under $\mathcal{H}_1$ due to the bias included by the fault.

For the sake of simplicity, in the following we assume the pseudorange errors have equal

variance. Summarizing,

$$t = \mathbf{e}^T\mathbf{e} = \mathbf{y}^T\left(\mathbf{I} - \mathbf{P_H}\right)\mathbf{y} \sim \begin{cases} \mathcal{H}_0: & \chi^2_k \\ \mathcal{H}_1: & \chi^2_k\left(\lambda\right) \end{cases}, \tag{3.21}$$

with $\mathbf{P_H} \doteq \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ the projection matrix onto $\mathbf{H}$, $\mathbf{P}_{\mathbf{H},\perp} \doteq (\mathbf{I} - \mathbf{P_H})$ the orthogonal projection matrix, and $k = N_{\text{sat}} - 4$ the number of degrees of freedom of the chi-squared distribution, given by the number of redundant measurements. Furthermore, we denote by λ the non-centrality parameter of the non-central chi-squared distribution, which actually contains the information regarding the bias being caused by some faulty measurements. That is,

$$\lambda_i = \frac{\mathbf{b}^T\mathbf{P}_{\mathbf{H},\perp}\mathbf{b}}{\sigma^2_{\text{UERE}}} = \frac{\left[\mathbf{P}_{\mathbf{H},\perp}\right]_{i,i}\left[\mathbf{b}\right]_i^2}{\sigma^2_{\text{UERE}}}, \tag{3.22}$$

where we have assumed that only the i-th satellite contains a bias; that is, $[\mathbf{b}]_j = 0$ except for $j = i$, and $[\cdot]_{i,j}$ denotes the (i,j) element of a matrix. Here we have used the assumption that all pseudoranges have the same variance, and then $\sigma^2_{\text{UERE},i} = \sigma^2_{\text{UERE}} \quad \forall i$. If this is not the case, and each pseudorange has different variance, which is the case in practice, we have to weight the SSE by the covariance matrix of the pseudoranges given in (3.12). This normalization leads to the so-called weighted SSE (WSSE), and then the same procedure as above follows but substituting σ^2_{UERE} in the previous expression by the corresponding variance of satellite i.

3.5.2 Fault Detection (FD)

Fault detection is aimed at detecting the presence of biases, that is, measurement errors which do not satisfy the assumed error statistics for the nominal errors (i.e. $\mathcal{N}(0,\mathbf{I})$). For LSR RAIM, fault detection is based on a chi-squared test on the LSR. Indeed, FD is carried out by comparing the test statistic t with a detection threshold, h, so that a fault is declared if $t \geq h$. The value of the detection threshold h is set in order to fix a constant false alarm rate. Although other alternative criteria is possible, a constant false alarm rate is often chosen. In this way, the RAIM is set to provide a fixed false alarm rate that is always equal to the value of $\mathcal{P}_{\text{fa}}$ given by the performance requirements, and a variable probability of missed detection (i.e. probability that undetected failures exceed the HAL) equal or lower than the value of $\mathcal{P}_{\text{risk}}$ given by the requirements. In particular, the probability of false alarm is defined as

$$\mathcal{P}_{\text{fa}} \doteq \Pr\left\{\text{fault detected}|\mathcal{H}_0\right\} = \Pr\left\{t \geq h | t \sim \chi^2_{N_{\text{sat}}-4}\right\}, \tag{3.23}$$

and h is thus a function of the required $\mathcal{P}_{\text{fa}}$ and the number of redundant measurements (i.e. $N_{\text{sat}}-4$), which means that we need more visible satellites than the minimum needed to compute the navigation solution (i.e. 4). If this is not the case, integrity is declared unavailable even though the navigation module may still be able to provide a position fix. When $t \geq h$ an alarm is raised, then two possibilities arise:

- If $N_{\text{sat}} > 4 + 1$ then the so-called fault exclusion function is applied in order to identify and isolate the wrong measurement.
- If $N_{\text{sat}} = 5$ the fault exclusion cannot be applied, and hence the position solution is declared

untrustworthy.

3.5.3 Fault Exclusion (FE)

Whenever $t \geq h$ and the number of measurements is greater than 5 (in the usual 3D positioning problem), the fault exclusion function is called. This function aims at determining which measurement or group of measurements is faulty. For instance, for the LSR RAIM fault exclusion, the idea is to rate satellites on how likely they are of causing the fault. In order to rate the i-th satellite, a unit bias is assumed for its measurement while nominal errors are neglected, leading to an error vector given by $\boldsymbol{\epsilon}_{\rho,i} = [0, \cdots, 1, \cdots, 0]^T$. The residual vector $\mathbf{e}$ generated by such an error is obtained as

$$\boldsymbol{\zeta}_i = \mathbf{P}_{\mathbf{H},\perp}\boldsymbol{\epsilon}_{\rho,i}. \tag{3.24}$$

Hence, if $\boldsymbol{\zeta}_i$ resembles the actual residual vector $\mathbf{e}$, the i-th satellite is likely to be the faulty one and hence it should be highly rated. The resemblance between these vectors can be measured by the normalized scalar product; the more parallel they are the larger (closer to 1) the scalar product. For convenience, in the LSR RAIM, the ratings are chosen as

$$\text{rating}_i = \frac{\left(\mathbf{e}^T\boldsymbol{\zeta}_i\right)^2}{\boldsymbol{\zeta}_i^T\boldsymbol{\zeta}_i}. \tag{3.25}$$

It is worth noting that when $N_{\text{sat}} = 5$ there is only one degree of freedom for residuals, meaning that all possible residual vectors are parallel. In terms of projection matrices; if $N_{\text{sat}} = 5$, $\mathbf{P}_{\mathbf{H},\perp} \doteq (\mathbf{I} - \mathbf{P}_{\mathbf{H}})$ has rank 1 because $\mathbf{H}$ has rank 4. Thereby, in (3.24) we are projecting $\boldsymbol{\epsilon}_{\rho,i}$ onto a line. Hence, $\boldsymbol{\zeta}_i$ is projected onto the same vector for all i. In this case, all the satellites would be rated equally and there would be no means of deciding which ones to exclude. This is why the condition $N_{\text{sat}} > 5$ must hold before performing fault exclusion. Once all satellites have been rated, there are different implementation choices. The simplest one is to flag the satellite with the highest rating as faulty and repeat the process again after excluding the satellite. Other choices include the possibility to flag several satellites as faulty based on their ratings (e.g. the two highest rated satellites or those with rating above some threshold). The different choices result in different balances between integrity and availability.

3.5.4 Protection Level (PL)

Whenever $t < h$ and the navigation solution is based on at least 5 measurements, the PL is calculated. Recall that the LSR RAIM relies on the assumption that there is only one faulty measurement at a time. It can be generalized to support multiple faults, but we will focus herein on the single fault case, referring the interested reader to [Bro97] for the multiple faults case. As discussed in Section 3.4.4, due to the satellite-user geometry, the same amount of error in different measurements contributes differently to positioning error. Similarly, the same amount of measurement error contributes differently to the residual vector. So, in order to compute a PL we have to ask the following question: How large the position error can be for a given vector $\mathbf{e}$ and user-satellite geometry? To answer the question, we have to firstly find the worst-case measurement bias that can be detected with the given probabilities requirements. This bias, called Minimum Detectable Bias (MDB), represents the theoretical lower bound for the

pseudorange error that can be detected and isolated. Once the MDB is obtained, it must be translated into positioning error in order to finally obtain a PL. To do so, let us define the integrity risk as

$$\mathcal{P}_{\text{risk}} \doteq \Pr\{\text{no detection within the TTA}|\mathcal{H}_1\} = \left(\Pr\left\{t < h|t \sim \chi^2_k(\lambda_{\text{det}})\right\}\right)^m, \tag{3.26}$$

with m the number of samples of t within the TTA, and λ_{det} the non-centrality parameter that results in an integrity risk equal to the given requirements.

Now let us fix $\lambda = \lambda_{\text{det}}$ in (3.22) for a given integrity risk equal to $\mathcal{P}_{\text{risk}}$. Then we can find the MDB as

$$\text{MDB}_i = [\mathbf{b}]_i|_{\lambda=\lambda_{\text{det}}}, \tag{3.27}$$

which substituting in (3.22) leads to

$$\text{MDB}_i = \sigma_{\text{UERE}}\sqrt{\frac{\lambda_{\text{det}}}{[\mathbf{P}_{\mathbf{H},\perp}]_{i,i}}}. \tag{3.28}$$

The LSR RAIM must protect against horizontal or vertical positioning errors in such a way that an alert is raised whenever the actual alert limit exceeds the corresponding PL, within the required $\mathcal{P}_{\text{risk}}$ and $\mathcal{P}_{\text{fa}}$. Hence, in order to compute the PL, the relationship between the integrity metric t and what RAIM protects is the key information to be used. To find this relation it is important to analyze the horizontal (or vertical) contribution of the pseudorange bias present in the i-th satellite into the horizontal (or vertical) position error domain. Particularly, the HPL is calculated as the contribution in the position domain of the pseudorange bias in the most-difficult-to-detect satellite that generates a test statistic with a non-centrality parameter equal to λ_{det}. This contribution is given by the norm of the horizontal components of $\widehat{\Delta u}$ multiplied by the bias as

$$|b_{\text{H}}|_i = \sqrt{[\mathbf{A}]^2_{1,i} + [\mathbf{A}]^2_{2,i}} \cdot [\mathbf{b}]_i\,, \tag{3.29}$$

with $\mathbf{A} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$, $|b_{\text{H}}|$ the bias component into the horizontal error, and $[\mathbf{b}]_i$ the bias in the i-th satellite.

Since the bias $[\mathbf{b}]_i$ is unknown, the HPL is obtained by using the MDB of satellite i instead, so that

$$|b_{\text{H}}|_i \geq \sqrt{[\mathbf{A}]^2_{1,i} + [\mathbf{A}]^2_{2,i}} \cdot \text{MDB}_i, \tag{3.30}$$

and then the HPL is determined as the worst possible bias,

$$\text{HPL} = \max_i \left\{\sqrt{[\mathbf{A}]^2_{1,i} + [\mathbf{A}]^2_{2,i}} \cdot \text{MDB}_i\right\}; \tag{3.31}$$

and we have

$$\sqrt{[\mathbf{A}]^2_{1,i} + [\mathbf{A}]^2_{2,i}} \cdot \text{MDB}_i \leq |b_{\text{H}}|_i \leq \text{HPL}. \tag{3.32}$$

Equivalently, the HPL can be defined as

$$\text{HPL} = \text{slope}_{\text{MAX}} \cdot \sqrt{\lambda_{\text{det}}}, \tag{3.33}$$

with

$$\text{slope}_{\text{MAX}} \doteq \max_i \{\text{slope}_i\} = \max_i \left\{ \sigma_{\text{UERE}} \sqrt{\frac{[\mathbf{A}]_{1,i}^2 + [\mathbf{A}]_{2,i}^2}{[\mathbf{P}_{\mathbf{H},\perp}]_{i,i}}} \right\}. \tag{3.34}$$

The slope slope_i is a measure of the coupling between the effect of a pseudorange bias into the integrity metric, and what the RAIM wishes to protect, the real horizontal position error e_{H}. Each satellite has its own slope, and the one with the highest slope is the most difficult to detect. Thereby, the RAIM detects the presence of this bias with an integrity risk equal to $\mathcal{P}_{\text{risk}}$, regardless the positioning error. Hereafter, pseudoranges biases are assumed to be in the maximum slope satellite. This is a worst-case scenario because pseudorange biases that generate a non-centrality parameter equal to λ_{det} in other satellites with lower slopes provide lower errors.

3.6 Background on Probability Theory

We have seen that the detection (and possible exclusion) of integrity faults is of vital importance for safely operating critical GNSS-based applications. For this reason, in the rest of this chapter, we switch our focus on the field of detection theory, with the aim of providing a framework to design optimal detectors to be applied latter on into GNSS integrity. We start with the needed background on probability theory, and we will continue in Section 3.7 with the classical hypothesis testing problem. In particular, probability theory provides a useful mathematical formulation for detection problems. This section provides a brief overview of the concepts from this framework, and they will be used in the following chapters. Moreover, they will serve to introduce and define the notation used throughout the rest of this thesis. Specifically, we firstly introduce the basic setting of the mathematical theory of probability. Secondly, a collection of results about martingales and stopping times are given. These results are very useful to state the *Optional sampling theorem*, leading to *Wald's identity*. Such identity will be key in the rest of the disertation to analyze the performance of sequential algorithms. Finally, we briefly introduce the formulation of Markov optimal stopping problems, helpful to analyze the optimal properties of statistical change detectors. The discussion provided in this section uses [Poo09] as reference.

3.6.1 The mathematical theory of probability

Here, we define some essential notions from probability theory that will be useful in the following sections. The main idea in a probabilistic scheme is that of *random experiment*, in which outcomes are generated according to some chance mechanism. From a mathematical point of view, this notion is contained in an abstraction, namely a *probability space*, which is composed of a triple $(\Omega, \mathcal{F}, \mathbb{P})$ consisting of the following elements:

- a *sample space* Ω with all possible outcomes ω of the random experiment;
- an *event class* $\mathcal{F}$ defined as a non-empty collection of subsets of Ω to which we wish to assign probabilities; and
- a *probability measure* (or *probability distribution*) $\mathbb{P}$, which is a real-valued set function that assigns probabilities to the events in $\mathcal{F}$.

In order to be able to manipulate probabilities, we assume that the event class $\mathcal{F}$ is a *σ-field*; that is, we assume that $\mathcal{F}$ is restricted to complementation and finite unions or intersections. The elements of $\mathcal{F}$ are called *events* and a pair $(\Omega, \mathcal{F})$ is called a *measurable space*. Regarding the probability measure $\mathbb{P}$, it is constrained to have the following properties, known as the *axioms for probability theory*:

$$\mathbb{P}(\Omega) = 1, \tag{3.35}$$

$$\mathbb{P}(F) \geq 1, \ \forall F \in \mathcal{F}, \tag{3.36}$$

meaning that it is constrained to be normalized and non-negative; and

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(F_n\right), \tag{3.37}$$

for all sequences $\{F_k; k = 1, 2, \dots\}$ of elements of $\mathcal{F}$ satisfying $F_m \cap F_n = \emptyset$, $\forall m \neq n$. That is, $\mathbb{P}$ is also constrained to be countably additive for independent events.

The abstraction of probability space allows us to think of some measurable phenomena that we wish to model as being random, referred to as *random variables*. Mathematically, a random variable is defined to be a mapping from the sample space Ω to the real line $\mathbb{R}$; that is, $X : \Omega \to \mathbb{R}$. The underlying probabilistic behavior of a random variable X is usually described in terms of the *cumulative distribution function* (cdf) of X, defined as

$$F_X(x) = \mathbb{P}\left(X \leq x\right) = \mathbb{P}_X\left((-\infty, x]\right), x \in \mathbb{R}, \tag{3.38}$$

with $\mathbb{P}_X(B) = \mathbb{P}(F(B)), \forall B \in \mathcal{B}$ and $\forall F(B) \in \mathcal{F}$ a probability measure on $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ stands for the Borel σ-field in $\mathbb{R}$, which is defined as the smallest σ-field that contains all intervals in $\mathbb{R}$. Due to the axioms in (3.35)–(3.37) the cdf satisfies the following properties: is non-decreasing, right-continuous with left limit zero and right limit one. Random variables are classified according to the nature of their cdfs, namely *continuous* and *discrete* random variables. In this dissertation we will focus on continuous random variables. A *stochastic process* (or *random process*) on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is an indexed collection of random variables on $(\Omega, \mathcal{F})$ denoted by $\{X_k\}_{k \in \mathcal{K}}$, where $\mathcal{K} \subseteq \mathbb{R}$ and where, for each $k \in \mathcal{K}$, X_k is a random variable on $(\Omega, \mathcal{F})$. In this thesis, we are interested in stochastic processes with the index set equal to the non-negative integers (discrete-time processes, or *random sequences*).

Expectation and related quantities

The cdf of a random variable X completely describes its probabilistic behavior. A coarser description of this behavior, though, can be given in terms of the *expected value* of the random variable, $\mathrm{E}[X]$, which is defined as the following integral

$$\mathrm{E}\left[X\right] \doteq \int_{\Omega} X(\omega)\, \mathbb{P}(\mathrm{d}\omega) = \int_{\Omega} X \, \mathrm{d}\mathbb{P}\,. \tag{3.39}$$

This integral can be rewritten as

$$\mathrm{E}\left[X\right] = \int_{\mathbb{R}} x\, \mathbb{P}_X(\mathrm{d}x) = \int_{-\infty}^{\infty} x \, \mathrm{d}F_X(x), \tag{3.40}$$

and for a continuous random variable we thus have

$$\mathrm{E}\left[X\right] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x, \tag{3.41}$$

where $f_X(x) \doteq \mathrm{d}F_X(x)/\,\mathrm{d}x$ is the *probability density function* (pdf) of X. A simple random variable, but very important for further theoretical developments, is the indicator function of an event, say F, which is defined as

$$\mathbb{1}_{\{F\}}(\omega) = \begin{cases} 1, & \text{if } \omega \in F \\ 0, & \text{if } \omega \notin F \end{cases}. \tag{3.42}$$

It is straightforward to see that $\mathrm{E}[\mathbb{1}_{\{F\}}] = \mathbb{P}(F)$, that is, the expectation of the indicator function of an event is equal to the probability of the same event. For an event F and a random variable X, we write

$$\mathrm{E}\left[X\mathbb{1}_{\{F\}}\right] = \int_F X(\omega)\,\mathbb{P}(\mathrm{d}\omega) = \int_F X\,\mathrm{d}\mathbb{P}\,. \tag{3.43}$$

This leads to a fundamental concept known as conditional expectation, which given the event $F : \mathbb{P}(F) > 0$ is the constant

$$\mathrm{E}\left[X \mid F\right] \doteq \frac{\mathrm{E}\left[X\mathbb{1}_{\{F\}}\right]}{\mathbb{P}(F)} = \frac{\int_F X\,\mathrm{d}\mathbb{P}}{\mathbb{P}(F)}. \tag{3.44}$$

Note that this constant satisfies the condition

$$\int_F \mathrm{E}\left[X \mid \mathcal{F}\right]\mathrm{d}\mathbb{P} = \int_F X\,\mathrm{d}\mathbb{P}; \tag{3.45}$$

that is, $\mathrm{E}\left[X \mid F\right]$ has the same $\mathbb{P}$-weighted integral that X does over the event F. This notion can be generalized for conditioning in particular σ-fields (groups of events). We say a random variable X is *measurable with respect to the sub-σ-field* $\mathcal{G}$ if $X^{-1}(B) \in \mathcal{G}, \forall B \in \mathcal{B}$. The σ-field generated by a random variable X (denoted by $\sigma(X)$) is the smallest σ-field with respect to which X is measurable. Then, we write $\mathrm{E}\left[X \mid \mathcal{G}\right]$ to denote the conditional expectation given $\mathcal{G}$, which can be thought of as a projection of X onto $\mathcal{G}$. In this sense, if X itself is $\mathcal{G}$-measurable, then $\mathrm{E}\left[X \mid \mathcal{G}\right] = X$.

Conditional expectations are closely related to the notion of (statistical) *independence*, defined in various circumstances as follows: (i) Two events F and G are said to be independent if $\mathbb{P}(F \cap G) = \mathbb{P}(F)\,\mathbb{P}(G)$. (ii) Two σ-fields are said to be independent if all elements of these fields are independent between each other; and (iii) two random variables X and Y are said to be independent if the smallest σ-field with respect to which these variables are measurable, $\sigma(X)$ and $\sigma(Y)$, are independent. If X is independent of $\mathcal{F}$, then $\mathrm{E}\left[X \mid \mathcal{F}\right] = \mathrm{E}\left[X\right]$; and so if X and Y are independent, then $\mathrm{E}\left[X \mid Y\right] = \mathrm{E}\left[X\right]$. All the above definitions and properties are key points to understand most of the theoretical developments in Chapter 4. Having defined conditional expectations allows for the definition of conditional probabilities, and the following interesting quantities use in Chapter 5:

Moments: The moments of a random variable are given by $\mathrm{E}[X^n]$ for $n = 1, 2, \ldots$ The first moment (which is the expected value) is called the *mean* of X.

Central Moments: The central moments of a random variable are defined as

$$\xi_{X,n} \doteq \mathrm{E}\left[(X - \mathrm{E}\left[X\right])^n\right], n = 1, 2, \ldots \tag{3.46}$$

The second central moment is the *variance* of X.

A useful result involving expectations and functions of random variables is Jensen's inequality:

Jensen's Inequality: Let X be a random variable with expectation $\mathrm{E}[X]$, and let $g(X)$ be a convex function of X. Therefore:

$$\mathrm{E}\left[g(X)\right] \geq g\left(\mathrm{E}\left[X\right]\right). \tag{3.47}$$

Radon-Nikodym derivatives and the Likelihood ratio

Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a measurable space $(\Omega, \mathcal{F})$. We say that $\mathbb{P}$ *dominates* $\mathbb{Q}$ if $\mathbb{P}(F) = 0$ implies that $\mathbb{Q}(F) = 0$, denoted by $\mathbb{Q} \ll \mathbb{P}$. Now, suppose $\mathbb{Q} \ll \mathbb{P}$, then there exists a random variable λ such that

$$\mathbb{Q}(F) = \int_F \lambda \, \mathrm{d}\mathbb{P}, \forall F \in \mathcal{F}. \tag{3.48}$$

This random variable λ is called the *Radon-Nikodym derivative* of $\mathbb{Q}$ with respect to $\mathbb{P}$, and we write

$$\lambda(\omega) = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(\omega). \tag{3.49}$$

The Radon-Nikodym derivative will be very helpful to change between expectations under different probabilities measures. This can be understood by looking at the following simple example: suppose $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$ and $\{\mathbb{P}, \mathbb{Q}\}$ have associated pdfs $\{p, q\}$, respectively. Then, the condition $\mathbb{Q} \ll \mathbb{P}$ is equivalent to the condition that the domain of q is contained in the domain of p, and in this case we have

$$\lambda = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \frac{q}{p}. \tag{3.50}$$

Let E_p and E_q stand for the expectations under $\mathbb{P}$ and $\mathbb{Q}$, respectively, then

$$\mathrm{E}_p\left[X\right] = \int_\Omega X \, \mathrm{d}\mathbb{P} = \int_\Omega X\lambda \, \mathrm{d}\mathbb{Q} = \mathrm{E}_q\left[X\lambda\right]. \tag{3.51}$$

This example can be regarded as the case when dealing with hypothesis testing, in which p and q denote the pdfs of two different hypothesis. In this case the Radon-Nikodym derivative is equivalent to the well-known concept of Likelihood Ratio (LR) in statistics, henceforth denoted as LR and LLR for the Log-LR (LLR). The LR is the cornerstone in the design of optimal detectors. The following notation will be used when dealing with random sequences $\{X_k\}$:

$$\begin{aligned} &\mathrm{LR}(X_k) \doteq \frac{q(X_k)}{p(X_k)}; \quad \mathrm{LLR}(X_k) \doteq \ln\left(\mathrm{LR}\left(X_k\right)\right), \\ &\lambda_k \doteq \prod_{i=1}^{k} \mathrm{LR}\left(X_i\right); \quad \Lambda_k \doteq \ln\left(\lambda_k\right) = \sum_{i=1}^{k} \mathrm{LLR}(X_i). \end{aligned} \tag{3.52}$$

This notation will be used in the rest of this thesis with p and q denoting the pdfs under the null (i.e. $\mathcal{H}_0$) and alternative hypothesis (i.e. $\mathcal{H}_1$), respectively. Often, for the sake of notation simplicity, the dependence upon the random variable X_k will be denoted simply by its time

index k (e.g. $\mathrm{LLR}(X_k) = \mathrm{LLR}(k)$).

3.6.2 Martingales and stopping times

Next, we define the notions of martingales and stopping times, and we give some general properties relating to these mathematical objects. These entities will lead us to the optional sampling theorem and Wald's identity, which are essential results to understand the optimal properties of sequential detectors studied in Chapter 4. We focus on the discrete-time case, which is the case of interest in this thesis. Thus, the time set for all sequences in this section (and in the rest of this dissertation) is the set of non-negative integers. A good source for this material is [Nev75]. We consider throughout a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and we begin with some definitions.

- A *filtration* $\{\mathcal{F}_k\}_{k=0,1,\dots}$ is an increasing sequence of sub-σ-fields of $\mathcal{F}$.
- A random sequence $\{X_k\}$ is *adapted* to $\{\mathcal{F}_k\}$ if, for each k, X_k is $\mathcal{F}_k$-measurable.
- Suppose $\{X_k\}$ is adapted to $\{\mathcal{F}_k\}$. Then, $\{X_k, \mathcal{F}_k\}$ is a *martingale* if

$$\mathrm{E}\left[X_k \,|\, \mathcal{F}_l\right] = X_l, \forall l \leq k. \tag{3.53}$$

  In the following we will simply say that such a sequence $\{X_k\}$ is a martingale. Similarly, if the equality is an inequality, instead, $\{X_k\}$ is said to be either a *submartingale* or *supermartingale* if the inequality is $\geq$ or $\leq$, respectively.

One intuitive way of thinking of a martingale is as the fortune accumulated by a gambler in playing a sequence of fair games. Let X_l be the gambler's fortune after the l-th play. This fortune can be assumed to be the sum of winnings (or losses) on each played game up to time l. Therefore, the gambler's expected fortune at any time k in the future, given all that has happened playing until time l, is just equal to the fortune at time l; that is $\mathrm{E}\left[X_k \,|\, \mathcal{F}_l\right] = X_l$. This is because as a fair game we will loose the same times we gain, so that the expected gain at any time k in the future is equal to 0. This is in the case of fair games in which the probability of winning and loosing a game is the same. In this context, a submartingale can be thought of as the gambler's fortune when the game is biased in favor of the gambler, and a supermartingale otherwise. Other examples of martingale and submartingale are the LR under the null and alternative hypothesis, respectively.

On the other hand, a *stopping time* is a random variable T taking values in the set $\{0, 1, 2, \dots\} \cup \{\infty\}$, with the property that it can assume the value k only on events that are measurable with respect to $\sigma(\{X_k\})$; that is whether $T = k$ or not can be determined by observing $X_0, X_1, \dots, X_k$. Stopping times play an important role in the theory of martingales, this is mainly due to the following result.

Theorem 1 (Optional Sampling Theorem). Suppose $\{X_k, \mathcal{F}_k\}$ is a submartingale, and that there is a random variable X such that $X_k \leq \mathrm{E}\left[X \,|\, \mathcal{F}_k\right], \forall k = 0, 1, \dots$ Therefore, for any stopping time T, the random variable X_T has finite expectation; and for any two stopping times S and T satisfying $\mathbb{P}(S \leq T) = 1$, we have

$$X_S \leq \mathrm{E}\left[X_T \,|\, \mathcal{F}_S\right], \tag{3.54}$$

with equality when $\{X_k\}$ is a martingale. In particular, for any stopping time T we have $\mathrm{E}[X_T] = \mathrm{E}[X_0]$.

An application of the Optional Sampling Theorem leads to a set of results collectively known as *Wald's identities* and they are very important to evaluate the performance of sequential detectors. The following identity is one of the most important, and it will be used throughout the rest of this dissertation.

Wald's identity. Suppose $\{X_k\}_{k=1,2,\ldots}$ is an independent and identically distributed (iid) sequence adapted to a filtration $\{\mathcal{F}_k\}$, and let $\{Y_k\}$ denote the sequence of cumulative sums, $Y_k = \sum_{i=1}^{k} X_i$. Suppose $\mu = \mathrm{E}[Y_1]$ is finite. Then the sequence $\{Y_k - k\mu\}$ is a martingale. So, for any stopping time T satisfying $\mathrm{E}[T] < \infty$, from (3.54), we have

$$\mathrm{E}[Y_T] = \mu\,\mathrm{E}[T]\,. \tag{3.55}$$

3.6.3 Markov optimal stopping problems

Next, we will briefly introduce the theory of Markov optimal stopping, which is the main mathematical tool to analyze the optimal properties of SCD algorithms. Most of the results of this theory can be found in [Cho71, Shi78]. Let us consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with a filtration $\{\mathcal{F}_k\}_{k\geq 0}$ and an adapted sequence $\{Y_k\}_{k\geq 0}$ of random variables on $(\Omega, \mathcal{F})$. Such a pair, $\{Y_k, \mathcal{F}_k\}$, is called a *stochastic sequence*. Now, let the random variable Y_k denote a reward that can be claimed at time k. For a stopping time T, $\mathrm{E}[Y_T]$ is the expected reward that we will get if we stop according to T. An *optimal stopping problem* is that of finding a stopping time T that maximizes $\mathrm{E}[Y_T]$. In particular, we define the *payoff*

$$V(\mathcal{S}) = \sup_{T\in\mathcal{S}} \{\mathrm{E}[Y_T]\}\,, \tag{3.56}$$

with $\mathcal{S}$ a class of stopping times for which $\mathrm{E}[Y_T]$ exists for all $T \in \mathcal{S}$. Thus, the aim is to compute the payoff $V(\mathcal{S})$ and, when possible, finding a stopping time in $\mathcal{S}$ that achieves it. It is of particular interest the so-called *Markov optimal stopping problems*, which can be described as follows: Consider a measurable space $(E, \mathcal{A})$ and a sequence $\{X_k\}$ of measurable functions from $(\Omega, \mathcal{F})$ to $(E, \mathcal{A})$ Then, X_k is $\mathcal{F}_k$-measurable for each k, and the sequence $\{X_k, \mathcal{F}_k\}$ is said to be a *Markov process* if, for each $A \in \mathcal{A}$, we have

$$\mathbb{P}(X_{k+l} \in A \,|\, \mathcal{F}_k) = \mathbb{P}(X_{k+l} \in A \,|\, X_k)\,\forall k, \quad l \geq 0; \tag{3.57}$$

that is, the probabilistic behavior of a Markov process at a given time in the future does not depend on past information, but only on current information. A stochastic sequence $\{Y_k, \mathcal{F}_k\}$ is said to be a *Markovian representation* if there is a Markov process $\{X_k, \mathcal{F}_k\}$ and a sequence of functions y_k such that $Y_k = y_k(X_k)$ for all k. A *Markov optimal stopping problem* is an optimal stopping problem whose reward sequence has a Markovian representation.

Let $\mathcal{T}$ denote the set of all stopping times with respect to the filtration $\{\mathcal{F}_k\}$. To examine the optimal stopping problem, let us consider for each $k = 0, 1, \ldots, n$, the random variables $\gamma_0^n, \gamma_1^n, \ldots, \gamma_n^n$ defined by

$$\gamma_k^n \doteq \sup_{T\in\mathcal{T}:k\leq T\leq n} \{\operatorname{esssup} \mathrm{E}[Y_T \,|\, \mathcal{F}_k]\}\,, k = 0, 1 \ldots, n, \tag{3.58}$$

with esssup the essential supremum operator (i.e. supremum with respect the filtration $\mathcal{F}_k$). That is, γ_k^n is the worst-case payoff over all stopping times in $\mathcal{T}$ that do not stop before time k, conditioned on the information up to time k. A natural candidate of stopping time is given as follows: stop the first time k such that the current reward Y_k satisfies $Y_k = \gamma_k^n$. The optimality of this stopping time is stated in the so-called backward induction theorem [Poo09, Theorem 3.1], which characterizes the solution to the case of a finite observation interval, in terms of the random variables $\{\gamma_k^n\}_{k=0,1,\ldots,n}$. One consequence of the backward induction is that $\{\gamma_k^n\}$ is a supermartingale, thus

$$\mathrm{E}\left[Y_T\right] \leq \mathrm{E}\left[\gamma_T^n\right] \leq \mathrm{E}\left[\gamma_0^n\right], \quad \forall T \in \mathcal{T} : k \leq T \leq n, \tag{3.59}$$

where the left-hand inequality becomes an equality for the stopping time above mentioned.

Thereby, the general problem $\sup_{T\in\mathcal{T}}\{\mathrm{E}\left[Y_T\right]\}$ can be solved if we can find a sequence $\{\gamma_k\}$ such that for the stopping time $T_0 = \inf\{k \geq 0 : \gamma_k = Y_k\}$ we have that $\mathrm{E}\left[\gamma_{T_0}\right] = \mathrm{E}\left[\gamma_0\right]$. This is stated by the following theorem [Poo09].

Theorem 2. Suppose $\mathrm{E}\left[\sup_k\{Y_k\}\right] < \infty$ and let $\mathcal{T}_k$ be the subset of $\mathcal{T}$ satisfying $\mathbb{P}\left(T \geq k\right) = 1$ for all $T \in \mathcal{T}_k$. Then, the sequence $\{\gamma_k\}$ with

$$\gamma_k = \mathrm{esssup}_{T\in\mathcal{T}_k} \mathrm{E}\left[Y_T \,|\, \mathcal{F}_k\right], \quad k = 0, 1, \ldots, \tag{3.60}$$

satisfies $\mathrm{E}\left[\gamma_{T_0}\right] = \mathrm{E}\left[\gamma_0\right]$, with

$$T_0 = \inf\left\{k \geq 0 : \gamma_k = Y_k\right\} \tag{3.61}$$

the optimal stopping time over all stopping times $T \in \mathcal{T}$.

Proof. The proof can be found in [Poo09, Ch. 3]. □

Remark 1. The derivation of the sequence $\{\gamma_k\}$ is greatly simplified in the Markovian case, otherwise the problem is usually mathematically intractable.

3.7 Hypothesis Testing: Classical detection

Once we have introduced the background on probability theory, this section introduces the main concepts of detection theory by showing its classical framework and performance measures. All the theory and techniques developed by the 20-th century statisticians for detection theory are available in the comprehensive references [Ken48] and [Cra46]. For a more practical point of view of the detection problem we have [Kay98b, Poo94, Lev08]. The discussion in this section is based on the previous mentioned references. The study of detection theory deals with the decision among a finite number of possible situations or states based on the observation of random variables, whose statistical behavior is influenced by the situation or state to be detected. This situation can be modeled by means of a family of pdfs, with members corresponding to the different statistical conditions that may be observed. Mathematically, in classical detection an N-samples data set $\{X_1, \ldots, X_N\}$ is assumed available for the decision process, thus the term block-wise framework. To get the ultimate decision, a function of the data, $D(X_1, \ldots, X_N)$, is applied and then a decision is made based on the value of this function. Determining the function D and mapping it into a decision is the central problem addressed in detection theory.

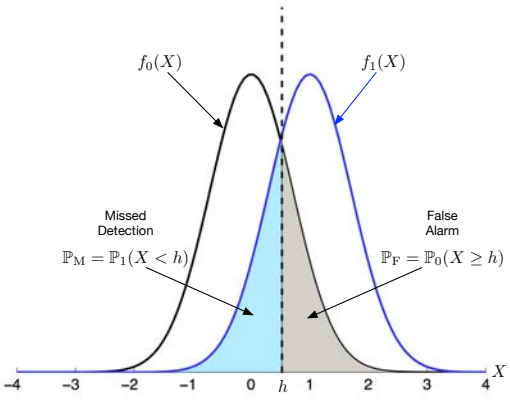

Figure 3.11: Pdfs and possible error types for the binary hypothesis testing problem in the single sample example.

3.7.1 Considerations of block-wise framework

In general, most detection problems can be cast in the framework of M-ary hypothesis testing, in which we have an observation on the basis of which we wish to decide among M possible statistical situations. In this work, we consider the particular problem of binary ($M = 2$) hypothesis testing, modeled as

$$\mathbf{x} \sim \begin{cases} \mathcal{H}_0 : f_0(\mathbf{x}) \\ \mathcal{H}_1 : f_1(\mathbf{x}) \end{cases}, \tag{3.62}$$

where $\mathbf{x} = [X_1, \ldots, X_N]^T$ denotes the $N \times 1$ vector with the N-samples data set, $\{\mathcal{H}_0, \mathcal{H}_1\}$ are the hypotheses (statistical states) corresponding to two possible pdfs $\{f_0, f_1\}$ associated to two different probability measures $\{\mathbb{P}_0, \mathbb{P}_1\}$, respectively. These hypotheses are often referred to as the null and alternative hypothesis, respectively.

In testing $\mathcal{H}_0$ versus $\mathcal{H}_1$ there are two types of errors that can be made: either $\mathcal{H}_0$ or $\mathcal{H}_1$ can be falsely rejected, leading to error types called a *false alarm* or a *missed detection*, respectively. Correct acceptance of $\mathcal{H}_1$ is usually called a *detection*. A *decision rule* δ is formed by the comparison of the function $D(\mathbf{x})$ with a threshold h. For a decision rule $\delta \doteq \{D(\mathbf{x}) \geq h\} \in \{0, 1\}$, the probabilities of false alarms (PFA) and missed detection (PMD) of δ will be denoted as $\mathrm{P_F}$ and $\mathrm{P_M}$, respectively. It is also often used the term *detection probability* of δ, $\mathrm{P_D} = 1 - \mathrm{P_M}$. With this setting a false alarm will occur as long as the threshold h is exceeded when actually the observed samples are under $\mathcal{H}_0$ (i.e. distributed as f_0), so that the PFA is defined as

$$\mathrm{P_F}\left(D(\mathbf{x})\right) \doteq \Pr\left\{D(\mathbf{x}) \geq h | \mathcal{H}_0\right\} = \mathbb{P}_0\left(D(\mathbf{x}) \geq h\right) = F_{D,0}(h), \tag{3.63}$$

where $F_{D,i}(h)$, with $i = \{0, 1\}$, stands for the cdf of $D(\mathbf{x})$ when either $\mathcal{H}_0$ or $\mathcal{H}_1$ is true, respectively. Similarly, a missed detection will occur when the threshold h is not exceeded, but the observed samples are actually distributed as f_1. Thus the PMD is given by

$$\mathrm{P_M}\left(D(\mathbf{x})\right) \doteq \Pr\left\{D(\mathbf{x}) < h | \mathcal{H}_1\right\} = \mathbb{P}_1\left(D(\mathbf{x}) < h\right) = F_{D,1}(h). \tag{3.64}$$

When the decision function D is known, we will directly write $\mathrm{P_F}$ and $\mathrm{P_M}$.

This framework is illustrated in Figure 3.11 for the case of a single observation of a Gaussian

distribution with variance $\sigma^2 = 1$ and mean $\mu = \{0, 1\}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. In this example we have defined the detection function as $D(X_1) = X_1$, which is compared with h. The selection of the threshold h will fix the amount of false alarms and missed detections. As can be inferred from (3.63)–(3.64), we can modify these probabilities by varying the detection threshold h, so that either the PFA or PMD is reduced. Notwithstanding, it is not possible to reduce both of them simultaneously, and a trade-off appears. Another way to write the PFA and PMD is through the region R_1 defined by set of values in $\mathbb{R}^N$ that map into the decision $\mathcal{H}_1$; that is $R_1 \doteq \{\mathbf{x} : \text{ decide } \mathcal{H}_1\} = \{\mathbf{x} : D(\mathbf{x}) \geq h\}$. With this notation, the PFA and PMD respectively become

$$\begin{aligned} \mathrm{P_F} &= \int_{R_1} f_0(\mathbf{x})\, \mathrm{d}\mathbf{x}, \\ \mathrm{P_M} &= 1 - \int_{R_1} f_1(\mathbf{x})\, \mathrm{d}\mathbf{x}. \end{aligned} \tag{3.65}$$

Thereby, a decision will be made in favor of $\mathcal{H}_1$ as long as the observed samples pertain to the region R_1.

3.7.2 Neyman-Pearson (NP) criterion

There are several useful definitions of optimality for the detection problem, we will consider here the Neyman-Pearson (NP) criterion, referring the interested reader to [Kay98b, Poo94, Lev08] for alternative criteria. Indeed, the NP criterion is one of the most extended optimality criterion used for hypothesis testing. A typical approach in designing an optimal detector is to hold one error probability fixed while minimizing the other. Usually, $\mathrm{P_F}$ is constrained to a fixed value $\widetilde{\alpha}$, and then the optimal detector is designed to minimize $\mathrm{P_M}$ or equivalently to maximize $\mathrm{P_D}$. This set-up is termed the NP criterion to hypothesis testing, which tells us how to choose the region R_1, resulting to a solution generally known as the Likelihood Ratio Test (LRT). This is stated in the so-called Neyman-Pearson theorem.

Theorem 3 (Neyman-Pearson theorem). Let λ_N be the LR between f_1 and f_0 of N observed iid samples gathered in $\mathbf{x}$, and let $\mathrm{LR}(i)$ be the same LR for the i-th sample. Therefore, to maximize $\mathrm{P_D}$ for a given $\mathrm{P_F} = \widetilde{\alpha}$ decide $\mathcal{H}_1$ if

$$\lambda_N \doteq \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} = \prod_{i=1}^{N} \mathrm{LR}(i) \geq h, \tag{3.66}$$

where the threshold h is found solving the following equation

$$\mathrm{P_F} = \int_{\{\mathbf{x}:\lambda_N \geq h\}} f_0(\mathbf{x})\, \mathrm{d}\mathbf{x} = \mathbb{P}_0\left(\lambda_N \geq h\right) = F_{\lambda,0}(h) = \widetilde{\alpha}, \tag{3.67}$$

with $F_{\lambda,0}(h)$ the cdf of λ_N under $\mathcal{H}_0$, evaluated at h.

Proof. The proof is given in [Kay98b]. □

3.7.3 Unknown parameters

The hypothesis testing problem considered so far is usually called *simple* hypothesis testing problem because it considers that the pdfs under both hypotheses are completely known. In

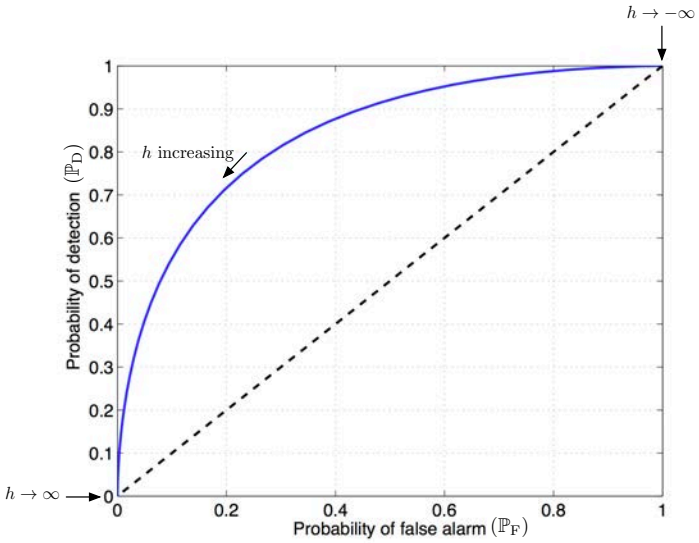

Figure 3.12: ROC for the case of a Gaussian distribution with variance $\sigma^2 = 1$ and mean $\mu = \{0, 1\}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively.

many hypothesis testing problems, though, these pdfs are unknown due to the presence of unknown parameters. These problems are known as *composite* hypothesis testing and they deal with a set of possible pdfs. There are mainly two major approaches when dealing with unknown parameters. The first is to consider the unknown parameters as realizations of random variables and to assign a prior pdf to these parameters, referred to as the Bayesian approach. Assigning these prior to the unknown parameters, the pdfs under $\mathcal{H}_0$ and $\mathcal{H}_1$ can be completely specified, no longer dependent on the unknown parameters, thus a LRT can be performed. The second is to estimate the unknown parameters and compress them into the required LRT, referred to as the Generalized Likelihood Ratio Test (GLRT). Replacing the unknown parameters by their Maximum Likelihood Estimates (MLEs), the LR can be completely specified. The Bayesian approach needs prior knowledge of the unknown parameters, whereas the GLRT does not. It is for this reason that the GLRT is more extensively used in practice.

3.7.4 Receiver Operating Characteristic (ROC)

Once the design of optimum procedures for hypothesis testing have been discussed, it is time to analyze the performance evaluation of these detectors. The performance evaluation of most detectors involves the computation of the two error probabilities $\{\mathrm{P_F}, \mathrm{P_D}\}$. A very extended way to summarize the detection performance of a detector is to plot $\mathrm{P_D}$ versus $\mathrm{P_F}$, called as the Receiver Operating Characteristic (ROC). An example of ROC is shown in Figure 3.12, where we have used the same considerations as for Figure 3.11. Each point on the curve corresponds to a value of ($\mathrm{P_F}$,$\mathrm{P_D}$) for a given value of threshold h. By adjusting h any point on the curve can be obtained. As expected, as h increases $\mathrm{P_F}$ decreases but so does $\mathrm{P_D}$ and vice-versa. The ROC should always be above the 45°(dashed line in Figure 3.12). This is because the 45° ROC can be attained by a detector that bases its decision on flipping a coin, ignoring all the data.

3.8 Conclusions

In this chapter we have introduced the basic concepts to understand the operation of GNSS and GNSS receivers. In order to compute the navigation solution from the received GNSS signal, the receiver has to perform several functions with the aim of identifying all the satellites in view, computing the distances from those satellites, track these satellites and distances, and finally compute the user's position from the calculated distances and known satellites' coordinates. Next, we have presented the error sources in GNSS, being the local effects such as multipath, NLOS propagation, interference, or spoofing the most challenging error sources in nowadays GNSS. The reason is that they cannot be reduced by means of corrections or models. Due to the presence of the above mentioned threats the integrity concept emerged, specially in critical applications where the miss-performance of GNSS may jeopardize the life or liability of the user. Integrity algorithms were developed under the umbrella of civil aviation. The main elements of these systems to provide integrity in today GNSSs is the provision of FD, FDE and the computation of PLs, being the detection process of vital importance. This is why we have studied the mathematical theory of detection. The needed background to understand detection theory has been given. Then, the traditional block-wise detection framework has been introduced, leading to the so-called LRT solution; the optimal one for the NP criterion.

The LRT is the optimal solution in the sense that it minimizes the PMD of a block-wise detection for a given level of PFA. Nevertheless, in a block-wise detection the sample size N must be fixed, without any recommendation given by the NP criterion. This is not very convenient on safety-critical applications, in which a detection should be made as soon as an integrity fault appears, instead of detecting that a threat is present in the previous batch of samples. This kind of detection can only be sought if samples are sequentially processed, without a pre-established block of samples. In such a way, a sequential detector should stop sampling when enough information is available to perform a confident detection. Furthermore, a prompt detection must be targeted for integrity algorithms, so that requirements in terms of TTA and integrity risk are fulfilled. Unfortunately, in classical detection the time dimension is not considered. It is for the above discussion that a different point of view of the detection problem must be considered for integrity algorithms. In particular, the targeted detection is that of sequentially detecting abrupt changes in the statistical behavior of a random process. This type of detection lies in the field of SCD and it will be deeply investigated in Chapter 4.

After analyzing the error sources of GNSSs we can conclude that the most challenging ones in today GNSSs are the local error produced by effects like interference or multipath. Unfortunately, GNSS integrity algorithms were designed to work in civil aviation, where local errors are controlled. This is the case in civil aviation, but it is not in terrestrial environments, where local errors abound. These errors may threaten the signal integrity, thus compromising the subsequent PVT integrity. It is for this reason that integrity checks must be carried out at the received signal level, so that integrity can be guaranteed in terrestrial environments. This will be the main goal addressed in Chapter 5.

Chapter 4

Primer on Statistical Change Detection

4.1 Brief Historical Review

One of the main functionalities of integrity algorithms is the FDE, being the detection of integrity faults of vital interest. For this reason in this chapter we study the field of detection theory, with special interest on sequential detection and particularly on the so-called Statistical Change Detection (SCD). Actually, hypothesis testing (presented in Chapter 3) and SCD have often been considered to be disparate fields, treated by statisticians and control engineers, respectively. Nevertheless, as we will investigate in this chapter, they are indeed closely linked. As a matter of fact, the field of detection theory is composed by scattered results across disparate disciplines. To get an overview and understand the spread of these results we will start with a brief historical review. The detection theory has its foundations on the field of probability theory and statistics. Actually, the mathematical theory of probability relies on the axiom system for probability theory presented by Kolmogorov in 1933. After that, the 19-th century mathematicians were led to define the theory of statistics, based on Kolmogorov's axioms. A period of intensive development of statistical methods began toward the last of the century with the work of Pearson, Fisher and Neyman. In particular, two areas were explored in their researches: parameter estimation and the testing of hypotheses. An influential synthesis of the work of this period by Cramér appeared in 1946 [Cra46], and more recent surveys are contained, for example, in [Kay98a, Kay98b, Poo94, Lev08]. A formal unification of the theories of estimation and hypothesis testing was achieved by Wald in his general theory of sequential analysis. An account of this theory is found in Wald's masterpiece work [Wal45].

At the same time as the theory of statistics was under development in the 19-th century, a new kind of problem became of key interest. That is the problem of detecting abrupt changes in the statistical behavior of an observed signal or time series. The origin of this problem dates at least to work in the 1930s on the problem of monitoring the quality of manufacturing processes [Wet91]. Since then a very diverse and vibrant field appeared, known as SCD, change-point detection, or disorder detection. Motivated by considerations of quality control, the subject of SCD was mainly considered by control engineers, being the Shewhart's charts very popular in the in the 1920-1930s. The analysis of optimal procedures, though, was not developed until the

1950-1960s, after the emergence of the branch of sequential analysis, ushered by Wald. The ideas set in motion by Shewhart and Wald have formed a platform for a vast literature on both theory and practice of SCD provided mainly by Page [Pag54], Shiryaev [Shi78], Lorden [Lor71], Siegmund [Sie85], and Moustakides [Mou86]. More recent surveys and developments are contained in [Bas93, Poo09, Tar14]. In contrast to the classical methods of mathematical statistics, according to which the number of observations is fixed in advance, the methods of SCD are characterized by the fact that the time at which the observations are terminated is random and it is defined by the observer based on the sequentially observed data.

In this chapter, we first start with some considerations of sequential detection, the discipline originated by Wald. This is provided in Section 4.2, then we will switch our focus on SCD by investigating the QCD problem in Section 4.3 by providing a comprehensive and simple discussion of the most important theoretical aspects of this theory. This is done by bringing together scattered results across disparate disciplines, leading to an unified treatment intended for a general reader. Next, Section 4.4 addresses the TCD problem, by analyzing the general problem and providing a novel solution and performance bounds. These contributions will be shown to be the best results obtained so far for the TCD problem. At this point, we will have provided a complete overview of the state of art on detection theory. Then, Section 4.5 compares the classical detection and SCD philosophies. This will be the first time that these two approaches are compared within the same framework for the general case (i.e. without assuming any particular distribution). This will shed light to the link between these two frameworks, useful to get insight into the asymptotic properties of the proposed solution for the TCD problem. Finally, Section 4.7 concludes this chapter.

4.2 Considerations of Sequential Detection

In Section 3.7 we described how to design optimal NP tests by implementing LRTs with appropriately selected thresholds. The kind of detectors presented so far are called Fixed Sample Size (FSS) detectors; that is, in each case a given fixed number of observations were given and the performance of the tests was described by the pair ($\mathrm{P_D}$,$\mathrm{P_F}$). However, we did not mention anything about how to fix the sample size N in order to achieve some target values for the error probabilities. This question was investigated in 1943 by the statistician Wald and lead to what nowadays is known as sequential hypothesis testing. In this formulation, instead of fixing the sample size, samples are analyzed sequentially until there is enough information to reach a decision with a high degree of confidence. In many ways this hypothesis testing approach represents an accurate model of human behavior, since, when confronted with uncertain situations, most individuals may be able to make a decision after only a few observations, whereas for some other observations they may wish to continue sampling to make a better decision. This kind of decision process is analyzed below.

The error probabilities of a NP test decay exponentially with the sample size N, when it tend to infinity, as shown in [Lev08, Ch. 3]. However, is this exponential decay the maximum rate we can achieve? or in other words, is there any smaller N that gives the same error probabilities? We will answer these questions in this section. In particular, in the case of a FSS test the sample size is fixed, and may be redundantly large for making a reasonably good inference on which of the two hypotheses is true. With sequential testing, though, no observations should be

wasted. Indeed, as soon as we can declare that one of the two hypotheses is true with reasonable certainty, the sequential test should stop. In order to find the optimal sequential test, let us consider a sequence $\{X_n\}_{n\geq 1}$ of iid observations that obey one of two statistical hypotheses

$$X_n \sim \begin{cases} \mathcal{H}_0: & f_0\left(X_n\right), n=1,2,\ldots \\ \mathcal{H}_1: & f_1\left(X_n\right), n=1,2,\ldots \end{cases}, \tag{4.1}$$

where f_0 and f_1 are two different pdfs associated to hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Let us assume now that $\mathcal{H}_1$ occurs with prior probability π and $\mathcal{H}_0$ with prior $1-\pi$. With this setting we would like to decide between these hypotheses in a way that minimizes an appropriate measure of error probability and sampling cost.

4.2.1 Bayesian optimal stopping formulation

With the aim of finding an appropriate measure of error probability and cost, let us consider the probability space $(\Omega, \mathcal{F}, \mathbb{P}_\pi)$, with

$$\mathbb{P}_\pi = (1-\pi)\,\mathbb{P}_0 + \pi\,\mathbb{P}_1, \tag{4.2}$$

and where $\mathbb{P}_0$ and $\mathbb{P}_1$ are two probability measures on $(\Omega, \mathcal{F})$ describing the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. In addition, suppose that there is a cost for sampling, namely $c>0$ per sample observed. Furthermore, assume we observe $\{X_n\}_{n\geq 1}$ sequentially, generating the filtration $\{\mathcal{F}_n\}$, with $\mathcal{F}_n = \sigma(\{X_n\})$, $n=1,2,\ldots$, and $\mathcal{F}_0 = (\Omega, \emptyset)$. The error probabilities $\mathrm{P_F}$ and $\mathrm{P_M}$ will improve with increasing the number of samples. Nevertheless, the cost of sampling introduce a trade-off between the error probabilities and the cost of sampling. To examine the optimization of such trade-off, let us define a *sequential decision rule* (sdr) as a duple consisting of a stopping time T and a sequence $\delta = \{\delta_n\}$ of *terminal decision rules*, with $\delta_n \in \{0,1\}$. The stopping time T declares the time to stop sampling, and once the value of T is given, δ_T takes the value 0 or 1 declaring which of the two hypotheses to accept. Two performance metrics are of interest for an sdr: the cost of sampling, $c\,\mathrm{E}_\pi[T]$, and the *average cost of errors*,

$$c_{\mathrm{e}}(T,\delta) = (1-\pi)c_0\,\mathbb{P}_0\left(\delta_T=1\right) + \pi c_1\,\mathbb{P}_1\left(\delta_T=0\right), \tag{4.3}$$

where $\mathbb{P}_0\left(\delta_T=1\right) = \mathrm{P_F}$ and $\mathbb{P}_1\left(\delta_T=0\right) = \mathrm{P_M}$, $c_j>0$ is the cost of falsely rejecting hypothesis $\mathcal{H}_j$, and $\mathrm{E}_\pi[\cdot]$ stands for the expectation under measure $\mathbb{P}_\pi$. So, we wish to find the sdr solving

$$\inf_{T,\delta}\left\{c_{\mathrm{e}}(T,\delta) + c\,\mathrm{E}_\pi[T]\right\}. \tag{4.4}$$

This problem is treated in many sources [Poo94, Lev08] as a traditional convex optimization problem. However we treat it here as an optimal stopping problem that can be solved using the tools of Markov optimal stopping theory, introduced in Section 3.6. This vision is intended to serve as the basis to evaluate the optimal properties of the related problem of SCD. To do so, first, the above optimization problem should be converted into a Markov optimal stopping problem. This can be done via the following proposition given in [Poo09]. We will follow a similar proof, but we include it here because it will be helpful to prove further theoretical results in following sections.

Proposition 1. Let π_k^π be the posterior probability that $\mathcal{H}_1$ is true, given $\{X_1, \ldots, X_k\}$, defined by the recursion

$$\pi_k^\pi = \frac{\pi_{k-1}^\pi f_1(X_k)}{\pi_{k-1}^\pi f_1(X_k) + (1-\pi_{k-1}^\pi) f_0(X_k)}, \quad k = 1, 2, \ldots; \text{ and } \pi_0^\pi = \pi. \tag{4.5}$$

Hence, for any T, we have

$$\inf_\delta \{c_{\mathrm{e}}(T,\delta)\} = \mathrm{E}_\pi \left[\min\left\{c_1 \pi_T^\pi, c_0 (1-\pi_T^\pi)\right\}\right], \tag{4.6}$$

and the infimum is achieved by the terminal decision rule

$$\delta_k^0 = \begin{cases} 1, & \text{if } \pi_k^\pi \geq c_0/(c_0+c_1) \\ 0, & \text{if } \pi_k^\pi < c_0/(c_0+c_1) \end{cases}. \tag{4.7}$$

Proof. Since the proposition is straightforward if $\pi = \{0, 1\}$, we restrict ourselves to the case $\pi \in (0, 1)$. Now fix T, then for any δ we can write

$$c_{\mathrm{e}}(T,\delta) = \sum_{k=1}^{\infty} \left[(1-\pi) c_0 \int_{\{\delta_k=1, T=k\}} \mathrm{d}\mathbb{P}_0 + \pi c_1 \int_{\{\delta_k=0, T=k\}} \mathrm{d}\mathbb{P}_1 \right]. \tag{4.8}$$

This is because $\mathbb{P}_0(\delta_T = 1) = \int_{\{\delta_T=1\}} \mathrm{d}\mathbb{P}_0 = \sum_{k=1}^\infty \int_{\{\delta_k=1,T=k\}} \mathrm{d}\mathbb{P}_0$, and similarly for $\mathbb{P}_1(\delta_T = 0)$. Consider now the second summand in this expression. Using the Radon-Nikodym derivative and the fact that $\mathbb{P}_\pi$ dominates $\mathbb{P}_1$ for $\pi > 0$, we have

$$\int_{\{\delta_k=0,T=k\}} \mathrm{d}\mathbb{P}_1 = \int_{\{\delta_k=0,T=k\}} \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_\pi} \, \mathrm{d}\mathbb{P}_\pi \,. \tag{4.9}$$

Now, since the events $\{\delta_k = 0\}$ and $\{T = k\}$ are in $\mathcal{F}_k$, the second summand becomes

$$\pi c_1 \int_{\{\delta_k=0,T=k\}} \frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_\pi} \, \mathrm{d}\mathbb{P}_\pi = \pi c_1 \int_{\{\delta_k=0,T=k\}} \mathrm{E}_\pi \left[\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_\pi} \,\middle|\, \mathcal{F}_k \right] \mathrm{d}\mathbb{P}_\pi, \tag{4.10}$$

with

$$\pi_k^\pi = \pi \, \mathrm{E}_\pi \left[\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_\pi} \,\middle|\, \mathcal{F}_k \right] = \frac{\pi \prod_{i=1}^k f_1(X_i)}{\pi \prod_{i=1}^k f_1(X_i) + (1-\pi) \prod_{i=1}^k f_0(X_i)}. \tag{4.11}$$

Similar arguments can be given for the first summand in (4.8), so that

$$c_{\mathrm{e}}(T,\delta) = \sum_{k=0}^{\infty} \left[c_0 \int_{\{\delta_k=1,T=k\}} (1-\pi_k^\pi) \, \mathrm{d}\mathbb{P}_\pi + c_1 \int_{\{\delta_k=0,T=k\}} \pi_k^\pi \, \mathrm{d}\mathbb{P}_\pi \right]. \tag{4.12}$$

This expression achieves its minimum over δ at the terminal decision rule

$$\delta_k^0 = \begin{cases} 1, & \text{if } \pi_k^\pi \geq c_0/(c_0+c_1) \\ 0, & \text{if } \pi_k^\pi < c_0/(c_0+c_1) \end{cases}, \tag{4.13}$$

which is the same as (4.7). That is, δ_T^0 will choose the hypothesis that will produce the smallest

average cost of error. Thus,

$$\begin{aligned}\min_{\delta}\{c_{\mathrm{e}}(T,\delta)\} &= \sum_{k=0}^{\infty}\int_{\{T=k\}}\min\left\{c_1\pi_k^{\pi}, c_0\left(1-\pi_k^{\pi}\right)\right\}\mathrm{d}\mathbb{P}_{\pi} \\ &= \mathrm{E}_{\pi}\left[\min\left\{c_1\pi_T^{\pi}, c_0\left(1-\pi_T^{\pi}\right)\right\}\right],\end{aligned} \tag{4.14}$$

and (4.6) thus follows. Finally, it is straightforward to check from (4.11) that $\{\pi_k^{\pi}\}$ satisfies the recursion (4.5), completing the proof. □

4.2.2 Optimal solution

Proposition 1 reduces the problem (4.4) to the alternative problem

$$\inf_{T}\left\{\mathrm{E}_{\pi}\left[\min\left\{c_1\pi_T^{\pi}, c_0\left(1-\pi_T^{\pi}\right)\right\}+cT\right]\right\}. \tag{4.15}$$

Because of the recursivity of $\{\pi_k^{\pi}\}$ this new problem can be solved using the tools of Markov optimal stopping problems. In particular, the following result stating the solution of (4.15) can be proved.

Theorem 4. There are constants π_{L} and π_{U} satisfying $0 \le \pi_{\mathrm{L}} \le c_0/(c_0+c_1) \le \pi_{\mathrm{U}} \le 1$, such that problem (4.15) is solved by the stopping time

$$T_{\mathrm{SPRT}} \doteq \inf\left\{k \ge 1 : \lambda_k \notin (A,B)\right\}, \tag{4.16}$$

where $\{\lambda_k\}_{k\ge1}$ is the sequence of LRs of k samples, and the thresholds A and B given by

$$A = \frac{1-\pi}{\pi}\,\frac{\pi_{\mathrm{L}}}{1-\pi_{\mathrm{L}}} \text{ and } B = \frac{1-\pi}{\pi}\,\frac{\pi_{\mathrm{U}}}{1-\pi_{\mathrm{U}}}. \tag{4.17}$$

Proof. A formal proof follows from applying Theorem 2 to the problem in (4.15), which can be found in [Poo09, Proposition 4.3]. We give here an intuitive proof based on the result obtained in Proposition 1. From (4.7) we see that the optimal terminal decision rule decides $\mathcal{H}_1$ at time k if $\pi_k^{\pi} \ge c_0/(c_0+c_1)$ or $\mathcal{H}_0$ otherwise. So, since the term cT in (4.15) is linear with T, the optimal sdr will stop as soon as a decision with enough confidence can be made. This confidence is given by the error probabilities, which will lead to given upper and lower thresholds π_{L} and π_{U}, in such a way that when π_k^{π} exceeds one of these thresholds there is enough confidence for doing a decision. This decision will be either $\mathcal{H}_0$ or $\mathcal{H}_1$ depending if the threshold π_{L} or π_{U} is exceeded, respectively. Next, it is worth noting that from (4.11) π_k^{π} can be written as

$$\pi_k^{\pi} = \frac{1}{1+(1-\pi)/(\pi\lambda_k)}, \tag{4.18}$$

so the comparison $\pi_{\mathrm{L}} < \pi_k^{\pi} < \pi_{\mathrm{U}}$ is equivalent to $A < \lambda_k < B$, completing the proof. □

Proposition 1 and Theorem 4 specify the optimal sdr for the optimal stopping problem in (4.15), summarized as follows: The optimal sdr given by $(T_{\mathrm{SPRT}},\delta^0)$, in (4.16) and (4.7), respectively, continues sampling until $\lambda_k \notin (A,B)$, at which time it chooses hypothesis $\mathcal{H}_1$ if $\lambda_{T_{\mathrm{SPRT}}} \ge A$, and it chooses $\mathcal{H}_0$ otherwise. Figure 4.1 illustrates the operation of this sdr. This test is referred to as the Sequential Probability Ratio Test (SPRT) with boundaries A and B. A

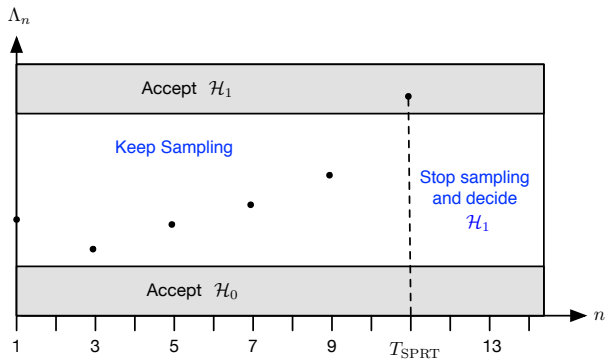

Figure 4.1: An illustration of the SPRT stopping time T_{SPRT}.

partial converse to the optimality of the SPRT leads to the conclusion that it exhibits minimal expected stopping time (i.e. minimal *runlength*) among all sdr for some given error probabilities (i.e. PFA and PMD). A consequence of this result is that SPRTs require smaller sample sizes than those of FSS tests with the same error probabilities.

4.2.3 Sequential Probability Ratio Test (SPRT): Performance analysis

As a solution of a Markov optimal stopping problem, the performance of an SPRT can be computed by means of the tools for this kind of problem. However, a more convenient analysis can be obtained by examining the relationships among the thresholds $\{A, B\}$, the two conditional error probabilities $\{\mathrm{P_F}, \mathrm{P_M}\}$, and the two expected runlengths $\{\mathrm{E}_0[T_{\text{SPRT}}], \mathrm{E}_1[T_{\text{SPRT}}]\}$. Such relationships are considered in the following two propositions. These results and their proofs will be very useful in Section 4.3 to analyze the optimal properties and performance of the CUSUM algorithm.

Proposition 2. Let $(T_{\text{SPRT}}, \delta)$ denote the sdr for the SPRT with thresholds $0 < A \leq 1 < B < \infty$. Let $\mathrm{P_F} \doteq \mathbb{P}_0(\delta_T = 1)$ and $\mathrm{P_M} \doteq \mathbb{P}_1(\delta_T = 0)$ denote the PFA and PMD of this SPRT, respectively. Therefore, the thresholds and error probabilities of the SPRT are related as

$$B \leq \frac{1 - \mathrm{P_M}}{\mathrm{P_F}} \quad \text{and} \quad A \geq \frac{\mathrm{P_M}}{1 - \mathrm{P_F}}. \tag{4.19}$$

Proof. From the definition of error probabilities, we can write

$$\mathrm{P_F} = \mathbb{P}_0(\lambda_T \geq B) = \sum_{k=0}^{\infty} \int_{\{\lambda_k \geq B, T=k\}} \mathrm{d}\mathbb{P}_0 \leq \frac{1}{B} \sum_{k=0}^{\infty} \int_{\{\lambda_k \geq B, T=k\}} \lambda_k \, \mathrm{d}\mathbb{P}_0 \,. \tag{4.20}$$

Now, since the events of the integral are both in $\mathcal{F}_k$, we can write $\lambda_k = \mathrm{d}\mathbb{P}_1 / \mathrm{d}\mathbb{P}_0$, so that

$$\mathrm{P_F} \leq \frac{1}{B} \sum_{k=0}^{\infty} \int_{\{\lambda_k \geq B, T=k\}} \mathrm{d}\mathbb{P}_1 = \frac{\mathbb{P}_1(\lambda_T \geq B)}{B} = \frac{1 - \mathrm{P_M}}{B}. \tag{4.21}$$

A similar argument gives $\mathrm{P_M} \leq A(1 - \mathrm{P_F})$, and thus the proposition follows. □

Proposition 3 (Wald's inequalities). Let $d_0 = \mathrm{E}_0[\mathrm{LLR}(n)]$ and $d_1 = \mathrm{E}_1[\mathrm{LLR}(n)]$ be finite numbers denoting the mean of the $\mathrm{LLR}(n) \doteq \ln(f_1(X_n)/f_0(X_n))$ under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Let T_{SPRT} be the SPRT stopping time with error probabilities fixed as $\mathrm{P_F}$ and $\mathrm{P_M}$.

Therefore

$$\begin{aligned} \mathrm{E}_0\left[T_{\mathrm{SPRT}}\right] &\geq d_0^{-1}\left[\mathrm{P_F}\cdot\ln\left(\frac{1-\mathrm{P_M}}{\mathrm{P_F}}\right)+(1-\mathrm{P_F})\cdot\ln\left(\frac{\mathrm{P_M}}{1-\mathrm{P_F}}\right)\right], \\ \mathrm{E}_1\left[T_{\mathrm{SPRT}}\right] &\geq d_1^{-1}\left[(1-\mathrm{P_M})\cdot\ln\left(\frac{1-\mathrm{P_M}}{\mathrm{P_F}}\right)+\mathrm{P_M}\cdot\ln\left(\frac{\mathrm{P_M}}{1-\mathrm{P_F}}\right)\right]. \end{aligned} \tag{4.22}$$

Proof. With similar arguments as given in the proof of Proposition 2, and letting Γ be the event $\{\lambda_{T_{\mathrm{SPRT}}} \geq B\}$, we can write

$$1-\mathrm{P_M} = \sum_{k=1}^{\infty}\int_{\{\lambda_k, T_{\mathrm{SPRT}}=k\}} \lambda_k \,\mathrm{d}\mathbb{P}_0 = \mathrm{E}_0\left[\lambda_{T_{\mathrm{SPRT}}}\mathbb{1}_{\{\Gamma\}}\right] = \mathrm{E}_0\left[\lambda_{T_{\mathrm{SPRT}}}\,|\,\Gamma\right]\mathbb{P}_0\left(\Gamma\right). \tag{4.23}$$

Now, from the definition of $\mathrm{P_F}$ and using Jensen's inequality, we have

$$1-\mathrm{P_M} \leq \mathrm{P_F}\cdot\exp\left(\mathrm{E}_0\left[\Lambda_{T_{\mathrm{SPRT}}}\,|\,\Gamma\right]\right) = \mathrm{P_F}\cdot\exp\left(\mathrm{E}_0\left[\Lambda_{T_{\mathrm{SPRT}}}\mathbb{1}_{\{\Gamma\}}\right]/\mathrm{P_F}\right), \tag{4.24}$$

or equivalently

$$\mathrm{E}_0\left[\Lambda_{T_{\mathrm{SPRT}}}\mathbb{1}_{\{\Gamma\}}\right] \geq \mathrm{P_F}\ln\left(\frac{1-\mathrm{P_M}}{\mathrm{P_F}}\right). \tag{4.25}$$

A similar expression can be obtained for $\mathrm{E}_0\left[\Lambda_{T_{\mathrm{SPRT}}}\mathbb{1}_{\{\Gamma^{\mathrm{c}}\}}\right]$, with $\Gamma^{\mathrm{c}} = \{\lambda_{T_{\mathrm{SPRT}}} < B\}$, which when added to (4.25) yields

$$\mathrm{E}_0\left[\Lambda_T\right] \geq \mathrm{P_F}\cdot\ln\left(\frac{1-\mathrm{P_M}}{\mathrm{P_F}}\right)+(1-\mathrm{P_F})\cdot\ln\left(\frac{\mathrm{P_M}}{1-\mathrm{P_F}}\right). \tag{4.26}$$

Now, using Wald's identity in (3.55) we have that $\mathrm{E}_0\left[\Lambda_T\right] = d_0\,\mathrm{E}_0\left[T\right]$, which when applied to (4.26) leads to the first inequality in (4.22). A similar proof can be given for the second inequality, thus completing the proof. □

It is worth noting that these inequalities turn out to be equalities when $\lambda_T = B$ on the event $\Gamma = \{\lambda_T \geq B\}$ and $\lambda_T = A$ on the event $\Gamma = \{\lambda_T \leq A\}$. These two conditions are sufficient for equality in Jensen's inequality when used in (4.19). In general, the conditions for equality will not be met. However, these inequalities can be considered to be approximations if the "excess over the boundaries" (i.e. $\lambda_T - B$ or $A - \lambda_T$) is negligible, which is the case when the error probabilities goes to 0, case of practical interest. In such a case, the above results are collectively known as *Wald's approximations*.

4.3 Quickest Change Detection (QCD)

So far, we have considered the problem of optimally deciding between two statistical models for a set of observed data. Either when the data is block-wise or sequentially observed, all the analyzed schemes dealt with the decision between two homogeneous statistical models; that is, the data obey only one of the two hypotheses during the entire period of observation. In the rest of this dissertation we turn to a generalization of this problem in which it is possible for the statistical behavior to change from one model to another at some unknown time. Thus, the goal will be to detect such a change as soon as possible. This objective may be formulated in different ways and it must be balanced with a desire to minimize false alarms. Such problems are known as SCD

problems, including Quickest Change Detection (QCD) and Transient Change Detection (TCD). Two different formulations for the SCD problem exist. On the one hand, there is the Bayesian formulation, in which the unknown change point is assumed to be a random variable with a given, geometric, prior distribution [Shi78]. Nevertheless, the assumption of a prior on the change point is unrealistic in most practical applications. This is why, on the other hand, there is the non-Bayesian formulation, in which the change point is modeled as unknown but deterministic. In this thesis, we will focus on the non-Bayesian formulation. In particular, we will deal with QCD in this section, and TCD in the following. The QCD framework has been extensively studied in the past decades, being applied to many fields such as quality process control [Mer01], health-care and public-health surveillance [Woo06], detecting of infectious diseases outbreaks [Hut97], or in wireless communications for signal detection [Osk05] and spectrum sensing for cognitive radio [Axe12], just to mention a few.

One of the most popular techniques in QCD is the CUSUM algorithm, which was first proposed in 1954 by Page [Pag54] for a continuous inspection scheme in quality control processes. However, it was not until 1971 when Lorden [Lor71] actually showed the optimality of the CUSUM for the QCD problem from an asymptotic point of view (i.e. when the mean time between false alarms goes to infinity). More than a decade later, in 1986, Moustakides [Mou86] proved the optimality of the CUSUM in a non-asymptotic way. The long period between the CUSUM proposal by Page and the optimal results by Lorden and Moustakides makes evident the difficulties in analyzing the optimality theory of QCD in particular, and SCD in general. As a matter of fact, even nowadays, there is scarce literature addressing the problem in a rigorous and comprehensive manner, being exceptions the work by Nikiforov [Bas93, Tar14] and Poor [Poo09], which are not intended for a general reader. In the rest of this section, we provide an overview of the most important theoretical results of QCD. In particular, this overview is based on our own publication [ER17a], which was aimed at introducing this theory to the GNSS community. It is worth pointing out that the presented flow in this section is based on how QCD has historically been treated. This is for the non-Bayesian case. Notwithstanding, we will see that optimal stopping theory plays a major role in specifying the optimal procedure for QCD. This connection was first shown by Shiryaev in his pioneering work [Shi78] devoted to the Bayesian change detection problem.

4.3.1 Lorden's Problem

We focus on the situation in which a change in distribution appears at a time v, which can be either ∞ or any finite positive integer value. This situation is modeled by considering a measurable space $(\Omega, \mathcal{F})$ and a family $\{\mathbb{P}_v : v \in [1, 2, \ldots, \infty]\}$ of probability measures on $(\Omega, \mathcal{F})$. We further consider that independent observations $\{X_n\}_{n\geq 1}$ are sequentially taken. Thus, under $\mathbb{P}_v$, it is assumed that before v (what we will consider $\mathcal{H}_0$) the observation X_n follows a given pdf f_0, whereas after the change (or $\mathcal{H}_1$) it follows a different one $f_1 \neq f_0$:

$$X_n \sim \begin{cases} \mathcal{H}_0 : f_0(X_n), & n < v \\ \mathcal{H}_1 : f_1(X_n), & n \geq v \end{cases}. \tag{4.27}$$

With this setting, we would like to obtain procedures that can detect the change time, if it occurs (i.e. $v < \infty$), as quickly as possible. To do so, we consider the stopping time $T \in \mathcal{T}$ with

respect to the filtration $\{\mathcal{F}_k\}_{k\geq 1}$, where $\mathcal{F}_k = \sigma(\{X_k\})$. Following [Lor71], it is of interest to penalize the detection delay via its worst-case value

$$\mathfrak{D}(T) \doteq \sup_{v\geq 1}\{\mathfrak{D}_v(T)\} = \sup_{v\geq 1}\left\{\operatorname{esssup} \mathrm{E}_v\left[(T-v+1)^+ \,\middle|\, \mathcal{F}_{v-1}\right]\right\}, \tag{4.28}$$

where $(x)^+ = \max\{0, x\}$, and $\mathrm{E}_v[\cdot]$ denotes expectation under the probability measure $\mathbb{P}_v$. Note that $\mathfrak{D}_v$ is the worst-case average delay under $\mathbb{P}_v$, where the worst case is taken over all realizations of the sequence $\{X_1, \ldots, X_{v-1}\}$ (i.e. esssup). That is, the worst-case sequence $\{x_{v-1}\}$ is the one providing the largest delay.

The desire to make $\mathfrak{D}(T)$ small must be balanced with a constraint on the rate of false alarms. In the non-Bayesian case instead of adopting a constraint on the PFA, as in classical hypothesis testing, it is more practical to adopt the philosophy that false alarms will occur, but to impose some limit on the rate at which they occur. That is, we should use the mean time between false alarms, defined as

$$\mathfrak{N}(T) \doteq \mathrm{E}_\infty[T]; \tag{4.29}$$

and a useful optimal design criterion is then given by

$$\inf_{T\in\mathcal{T}}\{\mathfrak{D}(T)\} \text{ subject to } \mathfrak{N}(T) \geq N_{\mathrm{fa}}, \tag{4.30}$$

where N_{fa} is a positive, finite constant. That is, we seek a stopping time that minimizes the worst-case delay within a lower-bound constraint on the mean time between false alarms. This problem is known as Lorden's problem and will be considered the optimal detection criterion.

4.3.2 CUSUM algorithm

QCD aims at finding the strategy that minimizes the detection delay, while keeping the mean time between false alarms larger than a conveniently set value. For this purpose, the CUSUM algorithm was proposed based on a key concept in statistics, namely the LLR. A fundamental property of this ratio, consequence of Jensen's inequality, is as follows: Let E_∞ and E_1 denote the expectations under the two distributions f_0 and f_1 associated with the two probability measures $\mathbb{P}_\infty$ and $\mathbb{P}_1$, respectively. Hence,

$$\mathrm{E}_\infty[\mathrm{LLR}(n)] < 0 \text{ and } \mathrm{E}_1[\mathrm{LLR}(n)] > 0; \tag{4.31}$$

that is, a change in the distribution of the random sequence $\{X_n\}_{n\geq 1}$ is reflected as a change in the sign of the mean value of the LLR. This is depicted in Figure 4.2, which shows the cumulative LLR $\Lambda_n \doteq \sum_{i=1}^n \mathrm{LLR}(i)$ in the lower plot for the case of a Gaussian set of observations illustrated in the upper plot with pdfs

$$f_{0,1}(X_n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left(X_n-\mu_{0,1}\right)^2}{2\sigma^2}}, \tag{4.32}$$

so that

$$\mathrm{LLR}(n) = \frac{\mu_1-\mu_0}{\sigma^2}\left(X_n - \frac{\mu_1+\mu_0}{2}\right), \tag{4.33}$$

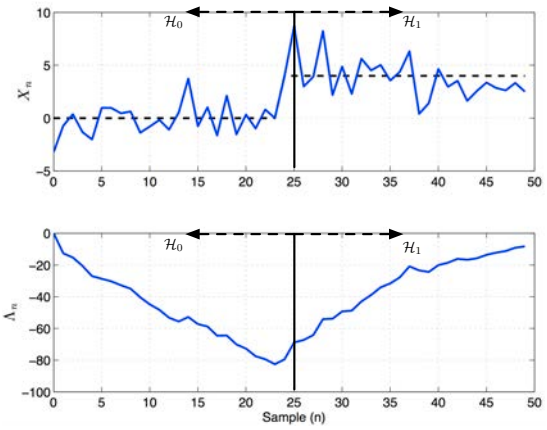

Figure 4.2: (Up) Samples corresponding to a Gaussian mean change with constant variance. (Down) Behavior of the cumulative LLR Λ_n. The example is done for a mean change from $\mu_0 = 0$ to $\mu_1 = 4$ and constant variance $\sigma^2 = 2.5$.

with $\mu_{0,1}$ the mean value of the incoming observations before and after the change, respectively. Therefore,

$$\Lambda_n = \frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=0}^{n} \left(X_i - \frac{\mu_1 + \mu_0}{2} \right). \tag{4.34}$$

Next, we describe three different derivations of the CUSUM algorithm. The first comes intuitively from the behavior of the LLR. The second derivation is based on a more formal online statistical approach, based upon a repeated use of the SPRT. This was the original idea of Page's derivation of the CUSUM. Finally, we introduce a derivation based upon the concept of open-ended tests, idea exploited by Lorden.

Intuitive approach

As depicted in Figure 4.2, the typical behavior of Λ_n shows a negative drift before the change and a positive drift afterwards. With this behavior, the relevant information to detect a change lies in the difference between the value of Λ_n and the minimum value collected up to time n. This difference is zero under $\mathcal{H}_0$, whereas it drifts upwards when the change appears. Hence, the corresponding decision rule is, at each instant, to compare this difference to a threshold h, namely

$$g_n = \Lambda_n - m(n) \geq h, \tag{4.35}$$

with $\Lambda_n \doteq \sum_{i=1}^{n} \mathrm{LLR}(i)$ and $m(n) = \min_{1 \leq j \leq n} \{\Lambda_j\}$. The typical behavior of g_n is shown in Figure 4.3, for the same values considered in the example of Figure 4.2. We can see how the values of g_n remain close to zero under $\mathcal{H}_0$ and they increase when the change appears. The stopping time thus becomes

$$T_{\mathrm{C}} = \inf \{ n \geq 1 : \Lambda_n \geq m(n) + h \}. \tag{4.36}$$

Page's approach

Next, we show the derivation of the CUSUM algorithm based upon a repeated use of the SPRT, previously analyzed in Section 4.2. We remind that the SPRT is defined by means of the

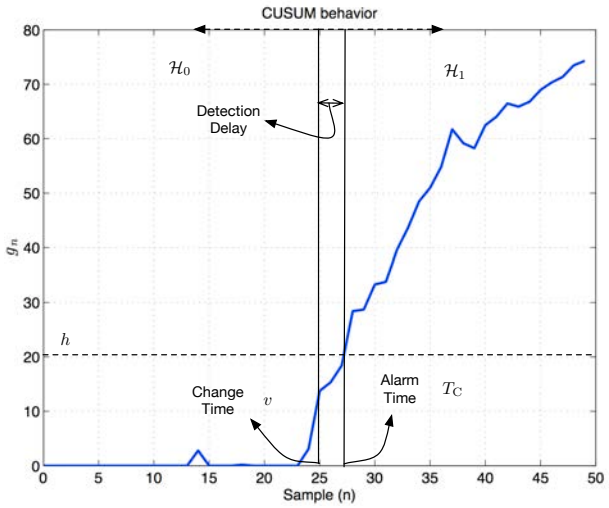

Figure 4.3: Typical behavior of the CUSUM algorithm g_n and stopping time T_{C} for $v = 25$.

following terminal decision rule

$$\delta = \begin{cases} 0 & \text{if } \Lambda_{T_{\mathrm{SPRT}}} \leq a \\ 1 & \text{if } \Lambda_{T_{\mathrm{SPRT}}} \geq h \end{cases}, \tag{4.37}$$

where T_{SPRT} is the SPRT stopping time defined as

$$T_{\mathrm{SPRT}} \doteq \inf \{n \geq 1 : (\Lambda_n \geq h) \text{ or } (\Lambda_n \leq a)\}. \tag{4.38}$$

The key idea of Page was to restart the SPRT whenever the taken decision was $\delta = 0$, and to stop the first time at which the decision is $\delta = 1$, thus becoming the *stopping time* at which the change is detected. To do so, Page suggested to adopt a lower threshold a equal to zero (i.e. $a = 0$) and to repeat the SPRT until $\delta = 1$. The resulting decision function can be written as

$$g_n = \begin{cases} g_{n-1} + \mathrm{LLR}(n) & \text{if } g_{n-1} + \mathrm{LLR}(n) > 0 \\ 0 & \text{if } g_{n-1} + \mathrm{LLR}(n) \leq 0 \end{cases}, \tag{4.39}$$

with $g_0 = 0$. This function can be compacted into

$$g_n = (g_{n-1} + \mathrm{LLR}(n))^+, \tag{4.40}$$

where $(x)^+ = \max(0, x)$. Thereby, the stopping time is defined by the well-known expression for the CUSUM algorithm

$$T_{\mathrm{C}} \doteq \inf \{n \geq 1 : g_n \geq h\}. \tag{4.41}$$

Lorden's approach

Finally, let us introduce an idea due to Lorden [Lor71] that turns out to be helpful to analyze change detectors. Let us define the LLR from sample i to sample n as

$$S_i^n \doteq \sum_{k=i}^{n} \mathrm{LLR}(k), \tag{4.42}$$

so that (4.35) can be rewritten as

$$g_n = \max_{1\leq j\leq n} S_j^n. \tag{4.43}$$

With this result in mind, the CUSUM stopping time can be interpreted as a set of *parallel* so-called open-ended SPRT, which are activated at each possible change time $j = 1, \ldots, n$. An open-ended SPRT have upper threshold h and lower threshold $a = -\infty$. Each of the parallel SPRT stops at time n if, for some $j \leq n$, the observations $X_j, \ldots, X_n$ are significant to accept the change. This can be formalized as follows: Let T_j be the stopping time for the SPRT activated at time j

$$T_j = \inf\left\{n \geq j : S_j^n \geq h\right\}, \tag{4.44}$$

with the convention that $T_j = \infty$ when the infimum is never reached. Now, let

$$T^* \doteq \min_{j=1,2,\ldots}\left\{T_j\right\}. \tag{4.45}$$

Comparing (4.41) and (4.45), with the definition of g_n in (4.43), we see that $T_{\mathrm{C}} = T^*$.

4.3.3 Performance evaluation

In this section, we describe the statistical properties of the CUSUM. The main purpose is to describe several criteria for the performance evaluation of the CUSUM, so that the detection threshold h can be fixed to obtain a desired performance. The performance of QCD is evaluated by means of the worst-case delay for detection (i.e. $\mathfrak{D}(T) \propto \mathrm{E}_1[T]$) and the mean time between false alarms (i.e. $\mathfrak{N}(T) = \mathrm{E}_\infty[T]$). In fact, it would be interesting to have a specific function that contains all the information related to both values. Let θ be a parameter describing the distribution of X_n, then such function is the ARL, defined as

$$L_\theta(T) \doteq \mathrm{E}_\theta\left[T\right], \tag{4.46}$$

where $\theta = \{\theta_0, \theta_1\}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively, and T is the stopping time under analysis. Thereby, the ARL function defines $\mathfrak{N}(T)$ at θ_0, and $\mathfrak{D}(T)$ at θ_1. Let us now define two important concepts useful for the analysis of the statistical properties of the SPRT and further computation of the ARL.

Average Sample Number (ASN): The ASN of a SPRT, $\mathrm{E}_\theta[T_{\mathrm{SPRT}}]$, is the mean number of samples necessary to test the hypotheses with acceptable error probabilities (see Wald's inequalities, (4.22)).

Operating Characteristic (OC): The probability $\mathbb{P}_\theta\left(T_{\mathrm{SPRT}}\right)$ of accepting hypothesis $\mathcal{H}_0$ (i.e. the SPRT reaches the lower threshold a) is called the OC.

Average Run Length (ARL)

Now, remembering that the CUSUM can be derived as a repeated SPRT, the ARL of the CUSUM can be linked to the statistical properties of the SPRT with lower threshold $a = 0$ and upper threshold h. The result of this link is the expression for the ARL of the CUSUM stopping time, given by the following result. This expression is important to set a given performance, then fixing the detection threshold h in practice.

Proposition 4. Let $\mathrm{E}_\theta[T_{0,h}]$ and $\mathbb{P}_\theta(T_{0,h})$ be the ASN and OC, respectively, of a SPRT with lower threshold $a = 0$ and upper threshold h. Therefore, the ARL for the CUSUM algorithm, $L_\theta(T_{\mathrm{C}})$,

is

$$L_\theta(T_{\mathrm{C}}) = \frac{\mathrm{E}_\theta\left[T_{0,h}\right]}{1-\mathbb{P}_\theta(T_{0,h})}. \tag{4.47}$$

Proof. With the formulation of the CUSUM as a repeated SPRT with lower threshold $a = 0$ and upper threshold h, it is possible to write the ARL function, by definition, as

$$L_\theta(T_{\mathrm{C}}) \doteq \mathrm{E}_\theta\left[T_{\mathrm{C}}\right] = \mathrm{E}_\theta\left[T_{0,h} \,|\, \Lambda_{T_{0,h}} \leq 0\right] \cdot \mathrm{E}_\theta\left[c-1\right] + \mathrm{E}_\theta\left[T_{0,h} \,|\, \Lambda_{T_{0,h}} \geq h\right] \cdot 1, \tag{4.48}$$

where $\mathrm{E}_\theta\left[T_{0,h} \,|\, \Lambda_{T_{0,h}} \leq 0\right]$ is the conditional ASN of one cycle of SPRT when the cumulative sum reaches the lower threshold $a = 0$, and then the SPRT is restarted; and $\mathrm{E}_\theta\left[T_{0,h} \,|\, \Lambda_{T_{0,h}} \geq h\right]$ is the conditional ASN of one cycle of SPRT when the cumulative sum reaches the upper threshold h, and then a final decision is taken. In (4.48), $\mathrm{E}_\theta[c-1]$ is the mean number of cycles that the SPRT is repeated before the cumulative sum reaches the upper threshold h. The random variable $y = c-1$ is a geometrical random variable with distribution $f(y) = (1-p)p^y$ for $y = 0, 1, 2, \ldots$, where $p = \mathbb{P}_\theta(T_{0,h})$ is the OC. Thus,

$$\mathrm{E}_\theta[c-1] = \frac{p}{1-p}, \tag{4.49}$$

and it results from (4.48) that

$$L_\theta(T_{\mathrm{C}}) = \frac{\mathrm{E}_\theta\left[T_{0,h} \,|\, \Lambda_{T_{0,h}} \leq 0\right] p + \mathrm{E}_\theta\left[T_{0,h} \,|\, \Lambda_{T_{0,h}} \geq h\right] (1-p)}{1-p}. \tag{4.50}$$

Now, considering that p is the OC of the stopping time $T_{0,h}$, we have that the summation in the numerator in (4.50) is equal to $\mathrm{E}_\theta[T_{0,h}]$, and (4.47) thus follows. □

Remark 2. The computation of $L_\theta(T_{\mathrm{C}})$ from (4.47) needs the OC and ASN of the stopping time $T_{0,h}$ (i.e. $\mathbb{P}_\theta(T_{0,h})$ and $\mathrm{E}_\theta[T_{0,h}]$, respectively). These quantities are solution to the following Fredholm integral equations:

$$\begin{aligned} \mathbb{P}_\theta(T_{0,h}) &= \int_0^h \mathbb{P}_\theta\left(T_{0,h}\right) f_\theta\left(x\right) \mathrm{d}x, \\ \mathrm{E}_\theta[T_{0,h}] &= 1 + \int_0^h \mathrm{E}_\theta\left[T_{0,h}\right] f_\theta(x)\, \mathrm{d}x, \end{aligned} \tag{4.51}$$

where f_θ is the kernel of the integral equation. These equations have to be solved numerically, and two approaches have been proposed to do so (see [Bas93, Ch. 5]).

Performance Bounds

The ASN and OC are solutions of the Fredholm integral equation, which has to be solved numerically. In order to avoid this numerical computation, and with the aim of making easier the design of the CUSUM algorithm, some approximations for the ARL based on Wald's identities are available [Bas93, Ch. 5]. Usually, approximations are complex to compute and bounds on the ARL function are more desirable. This is because, in practice, it is often important to fix a conveniently chosen performance of the change detection algorithm (given by the fixed threshold h) and then be sure that this performance will always be achieved within some limits. Relevant bounds for the problem at hand become an upper bound for the worst-case detection delay $\mathfrak{D}$, and a lower bound for the mean time between false alarms $\mathfrak{N}$. In some sense, these bounds play

the same role in the QCD theory as the Cramer-Rao lower bound in estimation theory, or as a ROC lower bound in hypothesis testing. Such bounds for the CUSUM are provided in the following theorem.

Theorem 5. Let $d_1 = \mathrm{E}_1\left[\mathrm{LLR}(n)\right] \doteq \mathrm{D}_{\mathrm{KL}}(f_1, f_0)$ be the Kullback-Leibler (KL) divergence between f_1 and f_0. Therefore, possible bounds for $\mathfrak{N}$ and $\mathfrak{D}$ of the CUSUM algorithm, with threshold h, are

$$\begin{aligned} \mathfrak{N}(T_{\mathrm{C}}) &\geq e^h, \\ \mathfrak{D}(T_{\mathrm{C}}) &\leq \frac{h}{d_1}. \end{aligned} \tag{4.52}$$

Proof. A lower bound for $\mathfrak{N}$ can be obtained from Wald's inequalities (4.19)

$$h \leq \ln\left(\frac{1-\mathrm{P_M}}{\mathrm{P_F}}\right) \quad \text{and} \quad a \geq \ln\left(\frac{\mathrm{P_M}}{1-\mathrm{P_F}}\right), \tag{4.53}$$

with $\mathrm{P_M}$ and $\mathrm{P_F}$ the PMD and PFA, respectively, of the SPRT. It is worth noting that we use here the logarithm of (4.19) because the CUSUM algorithm uses the LLR, whereas (4.19) was derived by using the LR. Furthermore, for the CUSUM algorithm the lower threshold is equal to zero (i.e. $a = 0$). Regarding the PMD, the CUSUM algorithm will always detect the presence of a change. It would take more or less time, but it will detect the change. Therefore the PMD of the CUSUM algorithm is equal to zero (i.e. $\mathrm{P_M} = 0$). So, using these two facts into (4.53), we have that $h \leq -\ln(\mathrm{P_F})$. With these considerations, taking into account that $\mathfrak{N}$, for an infinite observation period, can be regarded as the inverse of the PFA (i.e. $\mathfrak{N} = 1/\mathrm{P_F}$), the lower bound for $\mathfrak{N}(T_{\mathrm{C}})$ in (4.52) follows. On the other hand, we can obtain an upper bound for the mean detection delay from Wald's approximations (see (4.22)):

$$\mathrm{E}_1\left[T_{0,h}\right] \approx d_1^{-1} \cdot \left((1-\mathrm{P_M}) \cdot h - \mathrm{P_M} \cdot a\right), \tag{4.54}$$

with $T_{0,h}$ the SPRT stopping time with lower threshold $a = 0$ and upper threshold h. Now, it can be shown that $\mathfrak{D}(T_{\mathrm{C}}) \leq \mathrm{E}_1(T_{0,h})$, so that taking into account the previous considerations (i.e. $\mathrm{P_M} = 0$ and $a = 0$), the upper bound for the detection delay in (4.52) follows, completing the proof. □

4.3.4 Optimal properties

The results in Theorem 5 have been proved to be the optimal ones for the QCD problem. In particular, Lorden [Lor71] showed that the CUSUM minimizes the detection delay, attaining the bounds in (4.52) for an asymptotically large mean time between false alarms (i.e. $N_{\mathrm{fa}} \to \infty$). Lorden's method was based on linking the CUSUM with the SPRT, similarly as we did to prove Theorem 5. On the other hand, instead of studying the optimal detection problem via SPRT, Moustakides [Mou86] was able to formulate the problem as an optimal stopping problem and to prove that the CUSUM is indeed the optimal solution. Next, we briefly state the asymptotic results due to Lorden, and then we introduce the formulation of the QCD problem as an optimal stopping problem due to Moustakides. The presented discussion is based in the pioneering works of Lorden [Lor71] and Moustakides [Mou86], but without entering into details on proving the results. We include them here because they are helpful to understand further results in this thesis. The original works are not intended to the general reader and they are out of the scope

of this thesis. We refer the interested reader to [Poo09, Ch. 6] for a comprehensive analysis of both proofs.

Asymptotic optimality

The procedure followed by Lorden to prove the asymptotic optimality of the CUSUM was to first prove that the CUSUM attains the bounds in (4.52). This was done by proving the following lemma. We present it here because it will be useful in the discussion of Section 4.3.5, but its proof will be avoided.

Lemma 1 (Lorden's lemma). Suppose T is a stopping time with respect to the filtration $\mathcal{F}_k = \sigma(\{Z_k\})$ for $k \geq 1$, such that

$$\mathbb{P}_\infty (T < \infty) \leq a, \quad 0 < a < 1. \tag{4.55}$$

For each $k = 1, 2, \ldots$, let T_k denote the stopping time obtained by applying T to $Z_k, Z_{k+1}, \ldots$, and define

$$T^* = \inf \{T_k + k - 1\}. \tag{4.56}$$

Then T^* is a stopping time that satisfies

$$\mathfrak{N}(T^*) \geq \frac{1}{a} \quad \text{and} \quad \mathfrak{D}(T^*) \leq \mathrm{E}_1 [T]. \tag{4.57}$$

Proof. The proof can be found in [Lor71] or [Poo09, Ch. 6]. □

Once the CUSUM bounds are known, Lorden established that these bounds are asymptotically the best one can do. That is, Lorden showed that the minimum $\mathfrak{D}$ over all stopping times T, satisfying $\mathfrak{N}(T) \geq N_{\mathrm{fa}}$, is given by

$$\min_T \{\mathfrak{D}(T)\} \sim \frac{\ln(N_{\mathrm{fa}})}{d_1}, \text{ as } N_{\mathrm{fa}} \to \infty. \tag{4.58}$$

Exact optimality

Here, we examine Moustakides' method of proof for the exact optimality of the CUSUM algorithm in the sense of (4.30). Moustakides took a different perspective from Lorden's proof and reworked $\mathfrak{D}(T)$ into a form that can be optimized by means of the Markov optimal stopping theory. This form is given by the following proposition. The proof is given in Appendix 4.A, and it will be useful in Section 4.4 to formulate the problem of TCD in an optimal stopping time framework.

Proposition 5. Suppose $0 < \mathrm{E}_\infty [T] < \infty$, and for $k \geq 1$ let

$$\widetilde{g}_k \doteq \max_{1 \leq i \leq k} \left\{ \prod_{l=i}^{k} \mathrm{LR}(l) \right\} = \max \{\widetilde{g}_{k-1}, 1\} \, \mathrm{LR}(k). \tag{4.59}$$

Therefore, we have the following inequality, with equality if $T = T_{\mathrm{C}}$,

$$\mathfrak{D}(T) \geq \frac{\mathrm{E}_\infty \left[\sum_{m=0}^{T-1} \max \{\widetilde{g}_m, 1\} \right]}{\mathrm{E}_\infty \left[\sum_{m=0}^{T-1} (1 - \widetilde{g}_m)^+ \right]} \doteq \mathfrak{D}^*(T). \tag{4.60}$$

Proof. The proof is given in Appendix 4.A. □

Proposition 5 provides the means to convert the QCD problem in (4.30) into a traditional optimal stopping problem. This allows the proof of optimality of the CUSUM by showing that T_{C} solves the problem

$$\min_{T\in\mathcal{T}}\{\mathfrak{D}^*(T)\} \text{ subject to } \mathfrak{N}(T_{\mathrm{C}}) = N_{\mathrm{fa}}. \tag{4.61}$$

This proof can be found in [Mou86] and [Poo09, Ch. 6], asserting the optimality of the stopping time based on the first exit of $\widetilde{g}_k$ from the interval $[0, h)$. This time is particularly interesting when $h \geq 1$, in which case the stopping time can be written equivalently as the CUSUM stopping time T_{C} in (4.41).

4.3.5 Offset-CUSUM (O-CUSUM)

The optimal results shown so far are met when the true distributions before and after the change are completely known. In this section, we discuss the case when the LLR is not completely known, that is, the case in which the parameters under $\mathcal{H}_1$, $\boldsymbol{\theta}_1$, are unknown. Without loss of generality, the parameters under $\mathcal{H}_0$ are assumed to be known. Thus, similar approaches as for composite hypothesis testing (see Section 3.7) can be considered. For known $\boldsymbol{\theta_1}$, QCD algorithms are based on the cumulative LLR Λ_n. For unknown $\boldsymbol{\theta_1}$, Λ_n must be replaced by other statistic like the Weighted LR (WLR), $\widetilde{\Lambda}_n$, or the GLR $\widehat{\Lambda}_n$. On the one hand, WLR weights the LLR with respect to the probability of occurrence of all possible values of $\boldsymbol{\theta_1}$. On the other hand, the GLR replaces $\boldsymbol{\theta}_1$ by its MLE. Nonetheless, for the WLR we need information about the unknown parameters, which is not a common situation. Moreover, in general, the GLR-CUSUM cannot be written in a recursive form since it depends upon the maximization over the unknown time change (i.e. we need all the collected samples). This gives rise to a big computational burden. For these reasons there is an alternative CUSUM-based approach in order to avoid the previous practical issues. This approach will be referred to as the Offset-CUSUM (O-CUSUM).

In general, when the LLR is not completely known, it can be replaced by any other function of the observations X_n, namely $\rho(n) \doteq q(X_n)$, with negative mean before the change and positive mean after the change, in line with the key idea of the CUSUM (see (4.31)). That is [Osk05],

$$g_n^{(\mathrm{OC})} \doteq \left(g_{n-1}^{(\mathrm{OC})} + \rho(n)\right)^+, \tag{4.62}$$

for $n \geq 1$ and $g_0^{(\mathrm{OC})} = 0$; and then, the stopping time is defined by

$$T_{\mathrm{OC}} \doteq \inf\left\{n \geq 1 : g_n^{(\mathrm{OC})} \geq h\right\}. \tag{4.63}$$

In this case, the stopping time is no longer guaranteed to be optimal. Nevertheless, it is still a very good candidate, provided that an appropriate function $\rho(n)$ is chosen. As for the CUSUM, bounds for the detection delay and mean time between false alarms are desirable. These bounds are given in the following theorem.

Theorem 6. Let $\omega_0 > 0$ be the non-zero root of the equation $\mathrm{E}_\infty\left[e^{\omega\rho(n)}\right] = 1$, then the bounds for $\mathfrak{N}(T_{\mathrm{OC}})$ and $\mathfrak{D}(T_{\mathrm{OC}})$ of the O-CUSUM are given by

$$\begin{aligned} \mathfrak{N}(T_{\mathrm{OC}}) &\geq e^{\omega_0 h}, \\ \mathfrak{D}(T_{\mathrm{OC}}) &\leq \frac{h}{\mathrm{E}_1\left[\rho(n)\right]}, \end{aligned} \tag{4.64}$$

Proof. The proof is found in Appendix 4.B. □

Remark 3. Usually, the mean under $\mathcal{H}_0$ of $\rho(n)$ is not negative and then the idea proposed above is not applicable. In that case, we propose the modified metric

$$\rho(n) \doteq \rho_b(n) - b, \tag{4.65}$$

with ρ_b the function of the observations X_n, and b a proper offset selected so that the mean of $\rho(n)$ under $\mathcal{H}_0$ will be negative, but it will become positive under $\mathcal{H}_1$. Specifically, the choice of the offset b should be large enough to ensure a negative mean before change and to provide a certain false alarm rate. But at the same time, b should be small enough to maintain a positive mean after change. From (4.64) and using the expression of $\rho(n)$ in (4.65), we are able to adjust the false alarm rate through the non-zero root, ω_0, of the following equation

$$\mathrm{E}_\infty\left[e^{\omega\cdot(\rho_b(n)-b)}\right] = \phi(\omega)e^{-\omega\cdot b} = 1, \tag{4.66}$$

where $\phi(\omega) \doteq \mathrm{E}_\infty\left[e^{\omega\rho_b(n)}\right]$ is the characteristic function of $\rho_b(n)$ under $\mathcal{H}_0$.

4.4 Transient Change Detection (TCD)

In this section we focus on TCD, which deals with a change that appears and remains present only for a finite amount of time. This is in contrast to QCD problems in which an infinite change duration is assumed. As analyzed in the previous section, the optimal criterion in QCD is to minimize the detection delay subject to a level of false alarms. In contrast, in TCD problems a bounded detection delay is desired, so that the change is detected before it disappears. Unfortunately, the traditional QCD criterion does not completely fit into this problem. In this case, it makes more sense to minimize the Probability of Unbounded Delay (PUD) subject to a level of false alarms. This criterion was first adopted in the 1980s for the Bayesian approach [Boj79], but without controlling the false alarm rate. Recently, in 2013, the authors of [Pol13] considered a semi-Bayesian approach imposing a suitable constraint on the false alarm rate. But, it was not until 2014, that the first optimal results for the non-Bayesian case were provided again by Moustakides [Mou14]. Notwithstanding, all of these works have considered the very particular case of a bounded delay of a single sample, which has very limited practical application. The only optimal result is the one provided by Moustakides for the case of a bounded delay of one sample, but the problem is still open for finite delays greater than one sample. It is for this reason that the related literature for the fundamentals of TCD is scarce. Nonetheless, this dearth of theory is in contrast with the recent boom of SCAs in areas as diverse as navigation monitoring [Bak00], drinking water quality monitoring [Gué12], or cyber attacks on networked control systems [Do15], to mention a few.

In the above mentioned applications, it is desirable to detect abnormal situations with an established maximum tolerable delay m_d, so that detections declared after this delay are actually considered as missed. In these applications, the above mentioned TCD criterion is appropriate. Several approaches to deal with the TCD problem in these SCAs have been proposed relying on the CUSUM algorithm. Unfortunately, almost all available results are applicable to off-line detection on finite observation intervals [Por92, Lee95, Str99, Wan01], or they adopt the

traditional criterion of QCD [Han98, Han99, Wan05]. Exceptions, adopting the probability minimization criterion above mentioned, are [Bak00] and [Gué12]. Specifically, [Gué12], which proposes a Windowed-Limited CUSUM (WLC) solution (i.e. using a number of samples equal to m_{d}), is of interest. The reason is that it was shown that a Finite Moving Average (FMA) stopping time is the solution of the Gaussian mean change when the WLC is optimized. The optimality criterion was to find the threshold of the WLC that provides, for a Gaussian mean change, the minimum bound on the probability of unbounded delay with a constraint on the false alarms. Since the FMA is the optimal solution for the Gaussian mean change, we propose its use for the general TCD problem, although its optimality with respect to the above criterion is no longer guaranteed. In the rest of this section, we introduce the problem of TCD. In particular, most of the discussion of this section is due to our own publication [ER17d], which deals with the general problem when $m_{\mathrm{d}} \geq 1$.

4.4.1 A probability minimizing approach

In general, a SCD algorithm can be modeled as follows: Let $\{X_n\}_{n\geq 1}$ be a random sequence, and let v be the time instant at which a statistical change appears in this sequence. We consider a family $\{\mathbb{P}_v : v \in [1, 2, \ldots, \infty]\}$ of probability measures, such that under $\mathbb{P}_v$

$$X_n \sim \begin{cases} \mathcal{H}_0 : f_0(X_n) & \text{if } n < v \text{ or } n \geq v + m \\ \mathcal{H}_1 : f_1(X_n) & \text{if } v \leq n < v + m \end{cases}, \tag{4.67}$$

with f_0 and f_1 two different pdfs describing the behavior when the change is not present ($\mathcal{H}_0$) and when it is present ($\mathcal{H}_1$), respectively. Unlike QCD, in which the change duration m is assumed to be infinite, the change duration in TCD problems is assumed to be finite. There are two types of TCD problems. The first type involves the detection of signals arriving suddenly with an unknown and random duration. In such a case m denotes the unknown duration of the change. The second type involves SCA where a maximum tolerable detection delay is a priori fixed to a given value m_{d}. We will focus in this thesis on this second class of TCD problems; in this case a detection with a delay greater than m_{d} is considered as unbounded (i.e. missed), even if $m > m_{\mathrm{d}}$. On the other hand, if the duration of the change m is smaller than m_{d}, then such a change is considered less dangerous because its impact on the system is limited or negligible. Thus, in these cases we consider $m = m_{\mathrm{d}}$ to be the worst-case transient change we can allow.

We observe from the optimality criterion in QCD, given in (4.30), that no hard limit is imposed on the detection delay; consequently, this quantity can become arbitrarily large. In this sense, the optimality criterion for the TCD problem should be modified, with respect to the one used in QCD. This modified criterion should aim at seeking a small PUD given an acceptable false alarm rate. In other words, we wish to have $v \leq T < v+m$, with T the stopping time. Stopping within the prescribed interval constitutes a desirable event while stopping at $T \geq v + m$ is considered an unbounded detection. That is, the probability minimizing criterion should involve the minimization of the following worst-case PUD:

$$\mathrm{P}_{\mathrm{ud}}(T, m) \doteq \sup_{v\geq 1} \{\mathbb{P}_v(T \geq v + m \,|\, T \geq v)\}. \tag{4.68}$$

This minimization should be balanced with a given constraint on the following worst-case prob-

ability within a period of m_α samples:

$$\mathrm{P}_{\mathrm{fap}}\left(T, m_\alpha\right) \doteq \sup_{l \geq 1} \left\{ \mathbb{P}_\infty \left(l \leq T < l + m_\alpha \right) \right\}. \tag{4.69}$$

We will refer this probability to as the worst-case PFA in a given period (PFAP) with the aim of avoiding ambiguities with the PFA used in classical detection. The above presented probabilities are the performance metrics used in TCD problems. Consequently, we need to replace the minimization of the worst-case detection delay used in the classical QCD problem with the minimization of the PUD. In particular, the criterion of optimality for the TCD problem involves the following minimization problem:

$$\inf_{T \in C_\alpha} \left\{ \mathrm{P}_{\mathrm{ud}}\left(T, m\right) \right\}, \text{ with } C_\alpha = \left\{ T : \mathrm{P}_{\mathrm{fap}}(T, m_\alpha) \leq \widetilde{\alpha} \right\}, \tag{4.70}$$

where $\widetilde{\alpha} \in (0, 1)$ is a given constant value. The optimal solution of this class of TCD problem, for $m = 1$, was shown to be the Shewhart test [Mou14]

$$T_{\mathrm{S}}(h) \doteq \inf \left\{ n \geq 1 : \mathrm{LLR}(n) \geq h \right\}. \tag{4.71}$$

4.4.2 Optimal properties

Next, we analyze the optimal properties of the Shewhart stopping time for the TCD problem. Specifically, Moustakides formulated and solved the problem of TCD with $m = 1$ as an optimal stopping problem. We will present here the generalization of this formulation for the case $m > 1$. Unfortunately, the resulting problem cannot be solved by means of the optimal Markov stopping time theory, so that the problem remains open for $m > 1$.

General Problem

The procedure taken here is similar to that used in Proposition 5, that is, we try to rework the PUD $\mathrm{P}_{\mathrm{ud}}(T, m)$ into a form that can be treated by the optimal stopping theory. This is given by the following result.

Proposition 6. Let $\mathrm{P}_{\mathrm{bd}}(T, m) \doteq 1 - \mathrm{P}_{\mathrm{ud}}(T, m)$, with $T \in \mathcal{T}$ and $1 \leq m < \infty$. Also, let $\mathrm{LR}(n)$ be the LR of the observation X_n. Therefore,

$$\mathrm{P}_{\mathrm{bd}}\left(T, m\right) \leq \frac{\mathrm{E}_\infty \left[\sum_{v=T-m+1}^{T} \prod_{i=v}^{v+m-1} \mathrm{LR}(i) \right]}{\mathrm{E}_\infty \left[T \right]}. \tag{4.72}$$

Proof. From the definition of $\mathrm{P}_{\mathrm{ud}}(T, m)$ in (4.68) we have that for all $v \geq 1$

$$\mathrm{P}_{\mathrm{bd}}\left(T, m\right) \leq \mathbb{P}_v \left(v \leq T < v + m \,|\, T \geq v \right). \tag{4.73}$$

Multiplying both sides by $\mathbb{1}_{\{T \geq v\}}$ and taking expectation with respect to $\mathbb{P}_\infty$,

$$\mathrm{P}_{\mathrm{bd}}\left(T, m\right) \mathrm{E}_\infty \left[\mathbb{1}_{\{T \geq v\}} \right] \leq \mathbb{P}_v \left(v \leq T < v + m \,|\, T \geq v \right) \mathrm{E}_\infty \left[\mathbb{1}_{\{T \geq v\}} \right], \tag{4.74}$$

yields

$$\begin{aligned} \mathrm{P}_{\mathrm{bd}}\left(T, m\right) \mathbb{P}_\infty \left(T \geq v \right) &\leq \mathbb{P}_v \left(v \leq T < v + m \,|\, T \geq v \right) \mathbb{P}_\infty \left(T \geq v \right) \\ &\leq \mathbb{P}_v \left(v \leq T < v + m \,|\, T \geq v \right) \mathbb{P}_v \left(T \geq v \right). \end{aligned} \tag{4.75}$$

Applying the Bayes rule to the right-hand side we have

$$\begin{aligned} \mathrm{P_{bd}}\left(T,m\right)\mathbb{P}_{\infty}\left(T\geq v\right) &\leq \mathbb{P}_{v}\left(T\geq v\,|\,v\leq T<v+m\right)\mathbb{P}_{v}\left(v\leq T<v+m\right) \\ &\overset{(a)}{=} \mathbb{P}_{v}\left(v\leq T<v+m\right), \end{aligned} \tag{4.76}$$

where the equality $\overset{(a)}{=}$ follows because $\mathbb{P}_{v}\left(T\geq v\,|\,v\leq T<v+m\right)=1$. Then applying the Radon-Nikodym derivative, similarly as for proving Proposition 5, we get

$$\mathrm{P_{bd}}\left(T,m\right)\mathbb{P}_{\infty}\left(T\geq v\right)\leq \mathrm{E}_{v}\left[\mathbb{1}_{\{v\leq T<v+m\}}\right]=\mathrm{E}_{\infty}\left[\prod_{i=v}^{v+m-1}\mathrm{LR}(i)\mathbb{1}_{\{v\leq T<v+m\}}\right]. \tag{4.77}$$

Now, summing over all $v\geq 1$ we obtain the following relation

$$\mathrm{P_{bd}}\left(T,m\right)\sum_{v=1}^{\infty}\mathbb{P}_{\infty}\left(T\geq v\right)\leq \mathrm{E}_{\infty}\left[\sum_{v=1}^{\infty}\prod_{i=v}^{v+m-1}\mathrm{LR}(i)\mathbb{1}_{\{v\leq T<v+m\}}\right], \tag{4.78}$$

so that

$$\mathrm{P_{bd}}\left(T,m\right)\mathrm{E}_{\infty}\left[T\right]\leq \mathrm{E}_{\infty}\left[\sum_{v=T-m+1}^{T}\prod_{i=v}^{v+m-1}\mathrm{LR}(i)\right], \tag{4.79}$$

and (4.72) thus follows, completing the proof. □

Remark 4. With the above inequality we might try to optimize the right-hand side by the theory of optimal stopping time, and wish that with the optimal solution the inequality turns to be an equality. Unfortunately, the right-hand side term cannot be described as a Markovian optimal stopping problem. The problem thus become analytically intractable.

Case $m=1$

Moustakides' method of proof for the optimal solution of the TCD problem for $m=1$, was to show that the Shewhart stopping time satisfies with equality the equation (4.72) for the particular case of $m=1$. Actually, Moustakides showed that the Shewhart stopping time maximizes the right-hand side term in (4.72), satisfying the equation with equality. This is stated in the following theorem.

Theorem 7. The optimal stopping time that solves the TCD problem in (4.70) for $m=1$ is the Shewhart stopping time T_{S}, given in (4.71), with constant threshold h obtained by solving the equation $\mathbb{P}_{\infty}\left(\lambda_1\geq h\right)=1/\gamma$, where γ is a constant satisfying $\mathrm{E}_{\infty}\left[T_{\mathrm{S}}\right]\geq\gamma$.

Proof. A sketch of Moustakides' proof is shown, for a detailed discussion see [Mou14]. Let T be a stopping time satisfying $\mathrm{E}_{\infty}\left[T\right]\geq\gamma$, and let $\mathrm{P_{bd}}(T)=\mathrm{P_{bd}}(T,m=1)$. Therefore, if $m=1$ we have, from (4.72), that

$$\mathrm{P_{bd}}\left(T\right)\leq\gamma^{-1}\,\mathrm{E}_{\infty}\left[\mathrm{LR}(T)\right], \tag{4.80}$$

and then we can write

$$\sup_{T:\mathrm{E}_{\infty}[T]\geq\gamma}\left\{\mathrm{P_{bd}}(T)\right\}\leq\gamma^{-1}\sup_{T:\mathrm{E}_{\infty}[T]=\gamma}\left\{\mathrm{E}_{\infty}\left[\mathrm{LR}(T)\right]\right\}. \tag{4.81}$$

The constrained maximization in the right-hand side can be reduced into the unconstrained

maximization of

$$\mathrm{P}^*_{\mathrm{bd}}(T) = \mathrm{E}_\infty \left[\mathrm{LR}(T) - \lambda T\right], \tag{4.82}$$

with λ the Lagrange multiplier. Now, let $Y_k = \mathrm{LR}(k) - \lambda k$, then it is easy to show that the stochastic sequence $\{Y_k, \mathcal{F}_k\}$ follows a Markovian representation, so that, applying Theorem 2, we can easily conclude that $\mathrm{P}^*_{\mathrm{bd}}(T)$ is optimized by the Shewhart stopping time T_{S}, with threshold h selected so that $\mathbb{P}_\infty(\lambda_1 \geq h) = 1/\gamma$. Moreover, from the same theorem, we also have

$$h = \mathrm{E}_\infty \left[\mathrm{LR}(T_{\mathrm{S}})\right] - \lambda\, \mathrm{E}_\infty \left[T_{\mathrm{S}}\right]. \tag{4.83}$$

Furthermore, since

$$\mathrm{E}_\infty \left[T_{\mathrm{S}}\right] = \frac{1}{\mathbb{P}_\infty(\mathrm{LR}(T_{\mathrm{S}}) \geq h)} = \gamma, \tag{4.84}$$

and

$$\mathrm{E}_\infty[\mathrm{LR}(T_{\mathrm{S}})] = \frac{\mathbb{P}_1(\lambda_1 \geq h)}{\mathbb{P}_\infty(\lambda_1 \geq h)} = \gamma\, \mathbb{P}_1(\lambda_1 \geq h), \tag{4.85}$$

we can write

$$\mathrm{P}^*_{\mathrm{bd}}(T_{\mathrm{S}}) = \gamma\, \mathbb{P}_1 \left(\lambda_1 \geq h\right). \tag{4.86}$$

Finally, it is worth observing that for every $v \geq 1$ we have $\mathbb{P}_v(T_{\mathrm{S}} = v) = \mathbb{P}_1(\lambda_1 \geq h)$, and consequently $\mathrm{P}_{\mathrm{bd}}(T_{\mathrm{S}}) = \mathbb{P}_1(\lambda_1 \geq h)$. Using these two facts in (4.81) leads to

$$\mathrm{P}_{\mathrm{bd}}(T_{\mathrm{S}}) \leq \sup_{T:\mathrm{E}_\infty[T] \geq \gamma} \left\{\mathrm{P}_{\mathrm{bd}}(T)\right\} \leq \gamma^{-1} \sup_T \left\{\mathrm{P}^*_{\mathrm{bd}}(T)\right\} = \mathbb{P}_1 \left(\lambda_1 \geq h\right) = \mathrm{P}_{\mathrm{bd}}(T_{\mathrm{S}}), \tag{4.87}$$

proving that (4.80) and thus (4.81) are equalities for $T = T_{\mathrm{S}}$, concluding the proof. □

4.4.3 Windowed solutions

Since there is no optimal solution available in the literature of TCD for a finite $m > 1$, windowed solutions are proposed based on the following idea: We know that the optimal solution for QCD, that is for $m = \infty$, is the CUSUM test, which uses information about all the past samples. On the other hand, the Shewhart test, which uses information of one sample, is established to be optimal for the non-Bayesian TCD problem with $m = 1$. Hence, it is intuitive to think that the optimal solution for $1 < m < \infty$ would be some test statistic between these two techniques, and particularly, a test statistic using information about m samples (i.e. windowed).

Window-Limited CUSUM (WLC)

The first proposed windowed solution for the TCD problem was the WLC, suggested in [Gué12] as a CUSUM solution, but using at each moment the m last observations only; that is

$$T_{\mathrm{WLC}}\left(h\right) \doteq \inf \left\{ n \geq m : \max_{n-m+1 \leq k \leq n} \left\{S_k^n\right\} \geq h \right\}, \tag{4.88}$$

with $S_k^n \doteq \sum_{i=k}^n \mathrm{LLR}(i)$ the cumulative LLR from sample i to sample n, defined as in (4.42). It is assumed that the WLC is not operational during the first $m-1$ observations. It is worth noting, that this expression is similar to the one for the CUSUM given in (4.43), but carrying out the maximization over the past m samples instead of all the past samples, then the name WLC.

Finite Moving Average (FMA)

It was also shown in [Gué12] that after certain optimization process the WLC, for a Gaussian mean change, leads to

$$T^*_{\mathrm{WLC}}\left(\widetilde{h}\right) = \inf\left\{n \geq m : \sum_{i=n-m+1}^{n} X_i \geq \widetilde{h}\right\}, \tag{4.89}$$

where $\widetilde{h}$ denotes the chosen threshold after the optimization. The optimal criterion considered was to find the threshold h of the WLC that provides the minimum bound, for the Gaussian mean change, on the PUD with a constraint on the PFAP. It is worth saying that this does not mean that the FMA is the optimal solution for TCD. It just means that the FMA is the optimal WLC. It is important to see that the stopping time in (4.89) is equivalent to an FMA test (i.e. comparison of a moving average of m LLRs to a threshold) with the LLR of a Gaussian mean change. Inspired by this result and the idea of windowed solution, in [ER17d] we proposed the use of an FMA stopping time for any general TCD problems; that is,

$$T_{\mathrm{FMA}}(h) \doteq \inf\{n \geq m : \Upsilon_n \geq h\}, \tag{4.90}$$

with

$$\Upsilon_n \doteq S^n_{n-m+1} \doteq \sum_{i=n-m+1}^{n} \mathrm{LLR}(i). \tag{4.91}$$

Furthermore, the use of the FMA stopping time is motivated by the fact, as we will show next, that we can obtain very tight performance bounds.

4.4.4 Performance bounds

The goal of this section is to theoretically investigate the statistical performance of the FMA stopping time $T_{\mathrm{FMA}}(h)$; that is, to determine the worst-case PUD, $\mathrm{P_{ud}}\left(T_{\mathrm{FMA}}(h), m\right)$, and the worst-case PFAP for a given duration m_α, $\mathrm{P_{fap}}\left(T_{\mathrm{FMA}}(h), m_\alpha\right)$. The exact calculation of these probabilities is mathematically intractable, and it is here where the existence of tight-enough bounds is of interest. Similar arguments apply to the CUSUM-based solutions. For this reason we analyze here performance bounds for these solutions. We show first the novel bounds proposed for the FMA, and then we introduce the bounds for the CUSUM and WLC. The theoretical analysis will be only shown for the FMA, which is based on [ER17d]. For the CUSUM-based solutions, a similar proof as for the FMA for the bound of the PUD can be given; for the bound of the PFAP we will refer to [Lai98].

FMA bounds

We introduce here the bounds of the FMA stopping time, stated in the next theorem.

Theorem 8. Let $\Lambda_m \doteq \sum_{i=1}^{m} \mathrm{LLR}(i)$, and let $T_{\mathrm{FMA}}(h)$ be the FMA stopping time in (4.90). Hence, we have that the worst-case PFAP and PUD are respectively bounded as

$$\mathrm{P_{fap}}\left(T_{\mathrm{FMA}}(h), m_\alpha\right) \leq \alpha_{\mathrm{f}}\left(h, m_\alpha\right), \tag{4.92}$$

$$\mathrm{P_{ud}}\left(T_{\mathrm{FMA}}(h), m\right) \leq \beta(h, m), \tag{4.93}$$

with

$$\alpha_{\mathrm{f}}\left(h, m_{\alpha}\right)=1-\left[\mathbb{P}_{\infty}\left(\Lambda_{m}<h\right)\right]^{m_{\alpha}}, \tag{4.94}$$

$$\beta(h, m)=\mathbb{P}_{1}\left(\Lambda_{m}<h\right). \tag{4.95}$$

Proof. The proof is given in Appendix 4.C. □

In practice, values $\widetilde{\alpha}$ for $\mathrm{P}_{\mathrm{fap}}$ are imposed, so that we have to guarantee that $\mathrm{P}_{\mathrm{fap}} \leq \widetilde{\alpha}$. Thus, the threshold h has to be selected in order to satisfy this constraint, and then P_{ud} turns out to be a function of the fixed $\widetilde{\alpha}$ (i.e. $\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{FMA}}(h(\widetilde{\alpha})), m\right)$). In some sense, $\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{FMA}}(h(\widetilde{\alpha})), m\right)$ plays the same role in the TCD theory as the Cramér-Rao lower bound in estimation theory, as the ROC in classical detection, or as the ARL in QCD. Moreover, this kind of ROC allows us to compare the performance of different algorithms in terms of the optimality criterion in (4.70). This relation, for the FMA, is given in the following corollary.

Corollary 1. Let F_i, with $i=\{0,1\}$, be the cdf of Λ_m under $\mathcal{H}_i$ and fix h so that

$$\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{FMA}}(h), m_{\alpha}) \leq \widetilde{\alpha}. \tag{4.96}$$

A possible threshold h, satisfying (4.96), is given by

$$h\left(\widetilde{\alpha}\right)=F_{0}^{-1}\left[(1-\widetilde{\alpha})^{1/m_{\alpha}}\right], \tag{4.97}$$

where F_0^{-1} is the inverse of F_0, and thus

$$\beta\left(h\left(\widetilde{\alpha}\right), m\right)=F_{1}\left[F_{0}^{-1}\left[(1-\widetilde{\alpha})^{1/m_{\alpha}}\right]\right]. \tag{4.98}$$

Moreover,

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{FMA}}\left(\tilde{h}\right), m\right) \leq \beta\left(h\left(\widetilde{\alpha}\right), m\right), \tag{4.99}$$

with $\tilde{h}$ the threshold for which the exact PFAP fulfills $\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{FMA}}\left(\tilde{h}\right), m_{\alpha}\right)=\widetilde{\alpha}$.

Proof. It is worth noting that $\mathbb{P}_j(\Lambda_m<h)=F_i(h)$, with $j=\{\infty,1\}$ and $i=\{0,1\}$, is the cdf of Λ_m under $\mathcal{H}_i$, respectively, evaluated at h. Hence, solving the equation $\alpha_{\mathrm{f}}(h, m_{\alpha})=\widetilde{\alpha}$ for h from (4.94), a possible threshold can be selected as

$$h(\widetilde{\alpha})=F_{0}^{-1}\left[(1-\widetilde{\alpha})^{1/m_{\alpha}}\right], \tag{4.100}$$

which from (4.92) guarantees that $\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{FMA}}(h), m_{\alpha}\right) \leq \widetilde{\alpha}$, and (4.97) thus follows. The proof of (4.98) follows immediately by the definition of the cdf F_1 and by substituting (4.100) into (4.95). In order to prove (4.99) it is important seeing that the threshold $\tilde{h}$, such that $\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{FMA}}\left(\tilde{h}\right), m_{\alpha}\right)=\widetilde{\alpha}$, is lower than $h(\widetilde{\alpha})$ (i.e. $\tilde{h}<h(\widetilde{\alpha})$) and that

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{FMA}}\left(\tilde{h}\right), m\right) \leq \beta\left(\tilde{h}, m\right) \leq \beta\left(h, m\right), \tag{4.101}$$

where the last inequality follows because $\beta(h, m)$ in (4.95) is a monotonically increasing function of h, so that (4.99) follows, completing the proof. □

Remark 5. Results in Theorem 8 and Corollary 1 are valid for the general FMA stopping time; that is, they are not restricted to the Gaussian mean change, as in [Gué12], but they are valid for any kind of change. Moreover, as we will see later, these bounds are tighter than bounds for other available methods in the literature. This is beneficial for the selection of the detection threshold in practice. Unfortunately, we cannot establish the optimality of the proposed FMA stopping time. To do so we should analyze the speed of convergence of the term $F_0^{-1}[(1-\widetilde{\alpha})^{1/m_\alpha}]$, which is beyond the scope of this thesis. Notwithstanding, we will show how the proposed FMA stopping time outperforms other available methods in the literature, thus making evident that the FMA stopping time is a good candidate for TCD problems.

CUSUM-based

We focus here on bounds for the CUSUM-based solutions. We show the results for the CUSUM algorithm, being the same for the WLC, given by the following result.

Corollary 2. Let $\Lambda_m \doteq \sum_{i=1}^{m} \mathrm{LLR}(i)$, and let $T_{\mathrm{C}}(h)$ be the CUSUM stopping time defined as in (4.41); then the worst-case PFAP for a given duration m_α and the worst-case PUD are respectively bounded as

$$\begin{aligned} \mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{C}}(h), m_\alpha\right) &\leq \alpha_{\mathrm{c}}\left(h, m_\alpha\right) \\ \mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{C}}(h), m\right) &\leq \beta\left(h, m\right) \end{aligned}, \tag{4.102}$$

with

$$\begin{aligned} \alpha_{\mathrm{c}}\left(h, m_\alpha\right) &= m_\alpha e^{-h}, \\ \beta\left(h, m\right) &= \mathbb{P}_1\left(\Lambda_m < h\right). \end{aligned} \tag{4.103}$$

Proof. For the bound of PUD, a similar proof as for Theorem 8 can be given. For PFAP, it was shown in [Lai98] that the CUSUM stopping time T_{C} asymptotically minimizes (as $\widetilde{\alpha} \to 0$) the detection delay in QCD, over all stopping times satisfying $\mathrm{P}_{\mathrm{fap}}(T, m_\alpha) \leq \tilde{\alpha}$, if h fulfills the equation $\alpha_{\mathrm{c}}(h, m_\alpha) = \tilde{\alpha}$, with $\alpha_{\mathrm{c}}(h, m_\alpha)$ given by (4.103), thus completing the proof. □

4.4.5 Goodness assessment

This section is intended to assess the goodness of the theoretical results in Theorem 8 by making use of numerical simulations. To do so, we statistically characterize the random variable Λ_m in order to obtain (4.94) and (4.95). This is done by assuming different distributions for the LLR. Specifically, we present two cases quite prevalent in practice, further analyzed later, namely a mean- or variance-Gaussian change on the distribution of X_n, leading to a Gaussian and chi-squared distribution for the LLR, respectively. Finally, we present the case of a LLR with a Rician distribution. This is maybe a not common case in practice, but it is a good example where the distribution of Λ_m is not a common one, but the bounds are still fulfilled. We will focus here on the FMA stopping time, in Chapter 6 we will show results comparing different stopping times, showing the better performance of the FMA stopping time than other solutions. Moreover, let us consider the following general parameters: $m = 6$, $m_\alpha = 60$ (motivated by an application of GNSS integrity monitoring) and 10^6 Monte-Carlo runs for the results presented herein.

CASE 1: Gaussian distribution

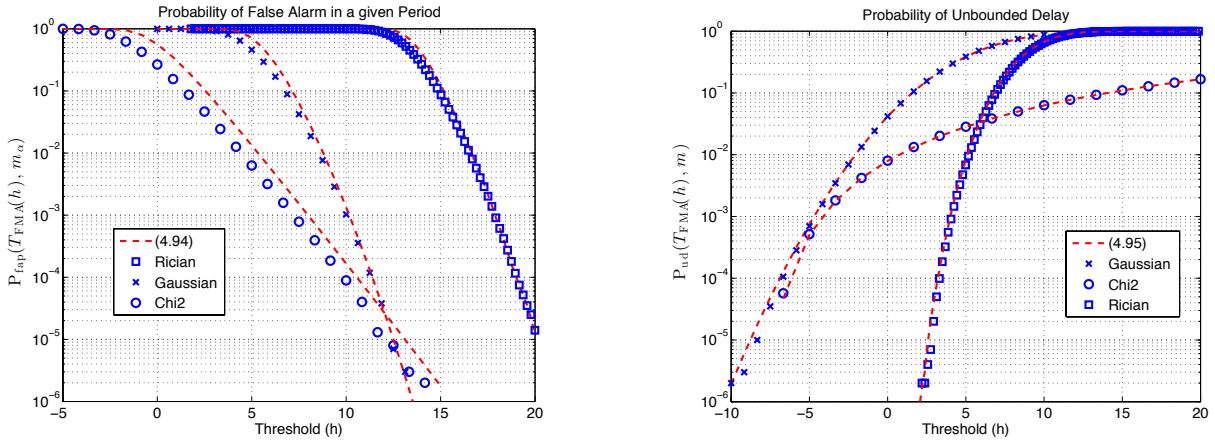

Figure 4.4: (Left) PFAP for a given duration of $m_\alpha = 60$ samples, $\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{FMA}}, m_\alpha)$, and (right) worst-case PUD, $\mathrm{P}_{\mathrm{ud}}(T_{\mathrm{FMA}}, m)$, of the FMA stopping time for the considered distributions of the LLR, with $m = 6$. Comparison between theoretical expressions in (4.94) and (4.95) (dashed lines), respectively, and simulated probabilities (markers). The following parameters are considered: for the Gaussian distribution $\{\mu_0, \mu_1\} = \{0, 1\}$, $\{\sigma_0^2, \sigma_1^2\} = \{1, 2\}$; for the chi-squared distribution $\{a_0, a_1, b\} = \{0.5, 10, -1.3\}$; and for the Rician distribution a scale and non-central factor equal to 1.

Let us consider a Gaussian distribution for the LLR, that is

$$\mathrm{LLR}(n) \sim \begin{cases} \mathcal{N}\left(\mu_0, \sigma_0^2\right) & \text{if } n < v \text{ or } n \geq v + m \\ \mathcal{N}\left(\mu_1, \sigma_1^2\right) & \text{if } v \leq n < v + m \end{cases}, \tag{4.104}$$

so that Λ_m is Gaussian as well with mean $\mu_{\Lambda,i} = m\mu_i$ and variance $\sigma_{\Lambda,i}^2 = m\sigma_i^2$, where $i = \{0, 1\}$. Thus, we have that

$$\mathbb{P}_j\left(\Lambda_m < h\right) = \Phi\left(\frac{h - \mu_{\Lambda,i}}{\sigma_{\Lambda,i}}\right), \tag{4.105}$$

with $j = \{\infty, 1\}$ and $i = \{0, 1\}$, respectively, and $\Phi(x)$ the cdf of the standard normal distribution. It is worth mentioning that if $\sigma_0 = \sigma_1$ this is equivalent to a mean-Gaussian change on X_n. The worst-case PUD and PFAP of T_{S}, for a Gaussian distribution for the LLR, is shown in Figure 4.4 (cross markers) as a function of the threshold h. These probabilities are compared with the theoretical results in Theorem 8, using (4.105). It follows from the figure that both results satisfy those stated in Theorem 8, with a very tight match between theoretical and simulated results.

CASE 2: Chi-squared distribution

Here, we present the case of a LLR distributed as a chi-squared distribution, specifically

$$\mathrm{LLR}(n) \sim \begin{cases} a_0\chi_1^2 + b & \text{if } n < v \text{ or } n \geq v + m \\ a_1\chi_1^2 + b & \text{if } v \leq n < v + m \end{cases}, \tag{4.106}$$

where χ_1^2 is the chi-squared pdf with one degree of freedom. A short calculation yields to

$$\mathbb{P}_j\left(\Lambda_m < h\right) = \Gamma_m\left(\frac{h - m \cdot b}{a_i}\right), \tag{4.107}$$

with $j = \{\infty, 1\}$ and $i = \{0, 1\}$, respectively, and $\Gamma_m(x)$ the cdf of the chi-squared distribution with m degrees of freedom. It is worth pointing out that this case is equivalent to a variance-

Gaussian change on X_n. Figure 4.4 (circle markers) shows the simulation of $\mathrm{P_{ud}}(T_{\mathrm{FMA}}, m)$ and $\mathrm{P_{fap}}(T_{\mathrm{FMA}}, m_\alpha)$, for the case of a chi-squared distributed LLR, as well as the theoretical results presented in Theorem 8, applying (4.107), as a function of h. Again, the results are satisfactory.

CASE 3: Rician distribution
Finally, we assume that the LLR is distributed as a standard Rician distribution (i.e. scale and non-central factor equal to 1). Thereby, it is difficult to find a closed-form expression for the distribution of Λ_m, so that the cdf of Λ_m must be computed numerically, and then the theoretical results for $\mathrm{P_{ud}}(T_{\mathrm{FMA}}, m)$ and $\mathrm{P_{fap}}(T_{\mathrm{FMA}}, m_\alpha)$ are numerically computed. This is shown in Figure 4.4 together with the Monte-Carlo simulations (square markers) satisfying the results stated in Theorem 8.

4.5 Classical and SCD Comparison

So far, we have given an overview of the different criteria we can find in the field of detection theory, including both classical hypothesis testing and SCD. Traditionally, these two fields have been considered to be apart each other. Hypothesis testing has usually been adopted by statisticians and signal processing engineers, whereas SCD has usually been adopted by control engineers and probability theorists. Nevertheless, as we made apparent in previous sections, the whole detection theory has its foundations on probability theory and stopping times. It is for this reason that both classical and statistical change detection are closely related, in contrast with what has been customarily thought. In particular, the LRT has traditionally been considered the optimal detector to minimize $\mathrm{P_M}$, subject to some constraint on $\mathrm{P_F}$. A key feature of this test is that it is block-based, that is, it uses a block of samples in order to obtain the ultimate decision. The size of this block of samples is a degree of freedom that must be fixed by the designer of the detector. Unfortunately, even the optimality of the LRT to the NP criterion is an established fact, for a fixed block size, no optimal rule is available for the configuration of this size. This section is aimed at showing the link between classical hypothesis testing and SCD. This is done by investigating the configuration of the block size of a NP test for different performance criteria. Particularly, we focus on QCD and TCD in which the criteria is to minimize either the mean delay for detection or the PUD, respectively.

Specifically, we use a sequential detector based on a repeated LRT, referred to as the FSS detector henceforth. We find the optimal sample size N for the FSS, and we compare the optimal FSS detector with sequential change detectors (CUSUM, FMA) in terms of the different performance criteria. Among this line, very scarce literature is available, mainly focused on the study of the effect of the sample size on the PMD and PFA. This study has been considered in textbooks as [Poo94] and [Lev08], showing as well a comparison between the LRT and SPRT. Regarding the comparison with SCD, even less work is available for the comparison with QCD, and nothing is available for TCD. For QCD, a first comparison was made by Shiryaev for the Bayesian setting [Shi63, Shi78], but limited work is found for the non-Bayesian case, being the contributions in [Nik97b] and [Pel87] exceptions. The former shows asymptotic results for the case of a Gaussian mean change. The asymptotic results show that the delay of the optimal FSS is approximately twice the delay of the CUSUM algorithm, when the mean time between false alarm goes to infinity. The contribution in [Pel87] makes the assumption that the sample size N goes to infinity and then the Central Limit Theorem (CLT) can apply. The goal of

this contribution is to obtain the optimal sample size (and threshold) numerically. In other words, the paper provides an approximated equation that can be numerically solved to obtain the optimal sample size. This approximation uses the CLT and then is valid when the sample size is large. So, the aim of this contribution rather than comparing the LRT with QCD is to numerically obtain the optimal sample size which minimizes the detection delay.

Unfortunately, the previous contributions for the non-Bayesian case consider the particular case of a Gaussian mean change, so that they do not provide a general and complete comparison between the LRT and change detectors. Moreover, they are only focus on QCD, without considering the TCD problem. It is for these reasons that in this section we investigate the LRT and the configuration of the sample size for the general SCD problem, including both QCD and TCD. The result of this investigation is intended to merge two fields that right now seem to be disjoint, namely fixed size and sequential detection, but indeed they are quite connected. Moreover, we show that the optimality of the LRT is lost when we deal with other criterion than the minimization of the PMD. This is done for the general problem; that is, without assuming any particular distribution. We firstly discuss the QCD comparison. This comparison is interesting to link and fill the gap between the fixed and sequential change detection fields. Secondly, we discuss the TCD comparison. This comparison may be more interesting because no optimal solution for the TCD problem is available, so it is worth considering the LRT as a possible candidate. Moreover, this is also motivated by the fact that no comparison is available in the literature, so showing that the FMA is a better solution for the TCD problem contributes to the analysis of the optimal properties of the FMA test for this problem. The presented discussion in this section is based on our own work in [ER17e].

4.5.1 QCD comparison

The main contribution of this part is to find asymptotic results for the general case (i.e. not restricted to the Gaussian mean change, as in [Nik97b]) and show equations to numerically find the optimal size sample of the FSS for the QCD criterion. Let us now recall the QCD framework: Let $\{X_n\}_{n\geq 1}$ be an iid random sequence, and consider a family $\{\mathbb{P}_v : v \in [1, 2, \ldots, \infty]\}$ of probability measures, with v be the instant (in samples) when a change in distribution appears. Then we consider the following model, already stated in (4.27):

$$X_n \sim \begin{cases} \mathcal{H}_0 : f_0(X_n) & \text{if } n < v \\ \mathcal{H}_1 : f_1(X_n) & \text{if } n \geq v \end{cases}, \tag{4.108}$$

The effectiveness of QCD is quantified by means of the trade-off between worst-case mean detection delay, $\mathfrak{D}$, and mean time between false alarm, $\mathfrak{N}$, both defined in (4.28) and (4.29), respectively. Indeed, the optimal criterion of QCD, as given in (4.30), is to minimize $\mathfrak{D}(T)$ over all stopping times satisfying $\mathfrak{N}(T) \geq N_{\text{fa}}$. The optimal stopping time for the QCD problem was shown to be the CUSUM stopping time given in (4.41). Moreover, it was also shown that the optimal delay, $\mathfrak{D}^*(T)$, reached by the CUSUM is given by

$$\tau_{\text{c}} = \mathfrak{D}^*(T) \sim \frac{\ln(N_{\text{fa}})}{d_1}, \tag{4.109}$$

with $d_1 \doteq \mathrm{E}_1\left[\mathrm{LLR}(i)\right] = \mathrm{D}_{\mathrm{KL}}(f_1, f_0)$ the LLR expectation under $\mathcal{H}_1$, given by the KL divergence between f_1 and f_0.

Optimal FSS strategy

The general FSS strategy can be described as follows: sample blocks with a fixed size N are collected, and at the end of each sample block a decision between $\mathcal{H}_0$ and $\mathcal{H}_1$ is taken. We stop sampling as soon as a decision is taken in favor of $\mathcal{H}_1$. As previously analyzed, the solution to the optimal hypothesis testing problem (in the NP criterion sense) is the LRT, and then the stopping time of the FSS is given by

$$T_{\mathrm{NP}} \doteq N \cdot \inf\left\{n \geq 1 : \delta_n = 1\right\}, \tag{4.110}$$

with δ_n the decision at the n-th block of N samples of the NP test, defined as

$$\delta_n \doteq \begin{cases} 1 & \text{if } \Psi_n \geq h \\ 0 & \text{if } \Psi_n < h \end{cases}, \tag{4.111}$$

with Ψ_n the LLR corresponding to the n-th block of N samples; that is

$$\Psi_n \doteq \sum_{i=(n-1)N+1}^{nN} \mathrm{LLR}(i). \tag{4.112}$$

With this setting, $\mathfrak{D}$ and $\mathfrak{N}$ are given by the following proposition [Nik97b].

Proposition 7. Let T_{NP} be the FSS stopping time in (4.110) and $\Lambda_N = \Psi_1 \doteq \sum_{i=1}^{N} \mathrm{LLR}(i)$ be the LLR of N samples. Therefore, the worst-case mean detection delay and mean time between false alarms for T_{NP} are respectively given by

$$\tau_{\mathrm{NP}} = \mathfrak{D}(T_{\mathrm{NP}}) = N + \frac{N}{1 - \beta_{\mathrm{NP}}\left(N, h\right)} - 1, \tag{4.113}$$

$$\gamma_{\mathrm{NP}} = \mathfrak{N}(T_{\mathrm{NP}}) = \frac{N}{\alpha_{\mathrm{NP}}\left(N, h\right)}, \tag{4.114}$$

where β_{NP} and α_{NP} are the PMD and PFA of a LRT with sample size N, given by

$$\begin{aligned} \beta_{\mathrm{NP}}\left(N, h\right) &= \mathbb{P}_1\left(\Lambda_N < h\right), \\ \alpha_{\mathrm{NP}}\left(N, h\right) &= \mathbb{P}_\infty\left(\Lambda_N \geq h\right). \end{aligned} \tag{4.115}$$

Proof. We give here a sketch of the proof, for a detailed proof see [Nik97b]. The key point is to realize that $T_{\mathrm{NP}} = N \cdot \widetilde{T}$, with $\widetilde{T}$ the number of trials needed by a NP test to get a detection. This kind of random variable follows a geometric distribution with a given probability of hit defined as $\mathbb{P}(\delta_n = 1)$. The so-called probability of hit will be given by $(1 - \mathrm{P_M})$ or $\mathrm{P_F}$ under $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively. With this in mind, and using $k = v - (n_0 - 1)N$ as the change point in the n_0-th block of N samples, the proof easily follows taking into account the properties of the geometric distribution. □

It turns out from the results in (4.113)–(4.114) that τ_{NP} depends on $\{\beta_{\mathrm{NP}}, \alpha_{\mathrm{NP}}, N\}$. Following the NP criterion we can fix $\alpha_{\mathrm{NP}}(N, h)$ for a given mean time between false alarms γ_{NP} from (4.114). Thus, by virtue of the NP theorem (see Theorem 3), the LRT minimizes $\beta_{\mathrm{NP}}(N, h)$ for

a fixed value of N. With this setting, τ_{NP} boils down to depend on N only, and it can thus be easily minimized for a given value of γ_{NP}. So, we have

$$\tau_{\text{NP}}^{*} = \min_{N} \{\tau_{\text{NP}}(N)\} = N^{*} + \frac{N^{*}}{1 - \beta_{\text{NP}}\left(N^{*}, h(N)\right)} - 1, \tag{4.116}$$

with

$$\beta_{\text{NP}}(N, h(N)) = F_1(h(N); N), \tag{4.117}$$

where $F_1(h; N)$ stands for the cdf of Λ_N under $\mathcal{H}_1$, evaluated at $h(N)$, and N^{*} denotes the optimal value of the sample size minimizing $\tau_{\text{NP}}(N)$; that is

$$N^{*} = \arg\min_{N} \left\{ N + \frac{N}{\beta_{\text{NP}}(N, h(N))} \right\} - 1, \tag{4.118}$$

which should be solved numerically by selecting $h(N)$ as

$$h(N) = F_{\infty}^{-1}(1 - \widetilde{\alpha}; N) = F_{\infty}^{-1}\left(1 - \frac{N}{\gamma_{\text{NP}}}; N\right), \tag{4.119}$$

with $F_{\infty}(a; N)$ the cdf of Λ_N under $\mathcal{H}_0$ evaluated at a, so that P_{F} fulfills $\text{P}_{\text{F}} = \alpha_{\text{NP}}(N, h(N)) = \widetilde{\alpha}$.

Asymptotic analysis

Now, we would like to find an expression relating the optimal mean delay for T_{NP}, τ_{NP}^{*}, with γ_{NP} and compare it with the mean delay of the CUSUM stopping time, τ_{c}. We have shown how to find τ_{NP}^{*}, but we cannot do that in a closed-form, and then we could not explicitly compare τ_{NP}^{*} with τ_{c} for a given value of γ_{NP}. It is for this reason that it is of interest to obtain bounds or approximations for τ_{NP}^{*} so that it can be explicitly compared with τ_{c}. To do so, it is worth proving first the following result.

Proposition 8. Let $\{\alpha_{\text{NP}}, \beta_{\text{NP}}\}$ be the PMD and PFA, respectively, of a LRT with N samples and fixed threshold h guaranteeing γ_{NP}; and let $d_1 \doteq \text{D}_{\text{KL}}(f_1, f_0)$ be the KL divergence between f_1 and f_0. Therefore

$$N \geq d_1^{-1} \left[(1 - \beta_{\text{NP}}) \cdot \ln\left(\frac{1 - \beta_{\text{NP}}}{\alpha_{\text{NP}}}\right) + \beta_{\text{NP}} \cdot \ln\left(\frac{\beta_{\text{NP}}}{1 - \alpha_{\text{NP}}}\right) \right], \tag{4.120}$$

and

$$N \simeq -\frac{\ln(\alpha_{\text{NP}})}{d_1} \qquad \text{as} \quad \gamma_{\text{NP}} \to \infty. \tag{4.121}$$

Proof. Similar steps as for the proof of Proposition 3 leads to (4.120). In order to prove (4.121) it is worth evaluating when the previous inequality is an equality. As in Proposition 3, this would be the case when the excess over the boundary is negligible; that is when $h \to \infty$ (or $\alpha_{\text{NP}} \to 0$), which is equivalent to say when $\gamma_{\text{NP}} \to \infty$. In such a case, we have from (4.114) and (4.118) that $\beta_{\text{NP}} \to 0$, so that both $\{\alpha_{\text{NP}}, \beta_{\text{NP}}\} \ll 1$ when $\gamma_{\text{NP}} \to \infty$. Therefore, with the previous considerations and from (4.120), (4.121) thus follows; completing the proof. □

With this proposition we can establish the following results for the optimal sample size and minimum detection delay for the FSS stopping time.

Theorem 9. Let γ_{NP} and τ_{NP} be the mean time before a false alarm and the worst-case mean detection delay, respectively, for the FSS stopping time T_{NP} with a sample size N. Therefore,

the optimal sample size N^* for the QCD criterion is given by

$$N^* \simeq \frac{\ln(\gamma_{\mathrm{NP}})}{d_1}, \text{ as } \gamma_{\mathrm{NP}} \to \infty. \tag{4.122}$$

Now, let τ^*_{NP} be the minimum detection delay achievable by T_{NP} with size N, then

$$\tau^*_{\mathrm{NP}} \simeq 2 \cdot \frac{\ln(\gamma_{\mathrm{NP}})}{d_1}, \text{ as } \gamma_{\mathrm{NP}} \to \infty. \tag{4.123}$$

Proof. We first note that from (4.114) and applying (4.121), we have

$$\ln(\gamma_{\mathrm{NP}}) = \ln(N) - \ln(\alpha_{\mathrm{NP}}) \simeq \ln(N) + N \cdot d_1, \tag{4.124}$$

as $\gamma_{\mathrm{NP}} \to \infty$, so that

$$N \simeq \frac{\ln(\gamma_{\mathrm{NP}}) - \ln(N)}{d_1}, \tag{4.125}$$

and then

$$N \lesssim \frac{\ln(\gamma_{\mathrm{NP}})}{d_1}, \text{ as } \gamma_{\mathrm{NP}} \to \infty, \tag{4.126}$$

where the approximation condition is satisfied as $\gamma_{\mathrm{NP}} \to \infty$. Thus the maximum N in order to guarantee a mean time between false alarm of γ_{NP} is asymptotically given by $N^* \simeq \ln(\gamma_{\mathrm{NP}})/d_1$. Furthermore, by virtue of the NP theorem, a LRT minimizes β_{NP} for a given value of $\alpha_{\mathrm{NP}} = \widetilde{\alpha}$ and fixed N. Thereby, for a fixed $\widetilde{\alpha}$ and a FSS stopping time with sample size N^*, from (4.113) τ_{NP} is minimized as well, so that (4.122) follows. Now, also from (4.113), we have that as $\beta_{\mathrm{NP}} \ll 1$ (or as $\gamma_{\mathrm{NP}} \to \infty$) $\tau^*_{\mathrm{NP}} \simeq 2N^* - 1$, and (4.123) thus follows, completing the proof. □

Finally, with the previous theorem we can establish the following result explicitly comparing τ^*_{NP} and τ_{c} for a given value of mean time between false alarms.

Corollary 3. Let us compare the optimal quickest and FSS detection procedures. The worst-case mean delay for detection of these procedures is given by

$$\mathfrak{D}(T) \sim \begin{cases} \frac{\ln(\gamma_{\mathrm{NP}})}{d_1}, & \text{if } T = T_{\mathrm{C}} \\ 2\frac{\ln(\gamma_{\mathrm{NP}})}{d_1}, & \text{if } T = T_{\mathrm{NP}} \end{cases}, \text{ as } \gamma_{\mathrm{NP}} \to \infty, \tag{4.127}$$

and then the mean delay for detection with a FSS stopping time for a given value of false alarms is asymptotically greater than the delay obtained by the CUSUM stopping time:

$$\tau_{\mathrm{NP}} \simeq 2\tau_{\mathrm{c}} \geq \tau_{\mathrm{c}}, \text{ as } \gamma_{\mathrm{NP}} \to \infty. \tag{4.128}$$

Remark 6. The above investigated results coincide with those shown in [Nik97b]. Notwithstanding, Theorem 9 and Corollary 3 are valid for the general case; that is, they are not restricted to the Gaussian mean change.

4.5.2 TCD comparison

We focus now on the TCD comparison by evaluating the performance of the FSS stopping time in the TCD framework. This is with the aim of providing equations to numerically calculate the optimal sample size of the corresponding FSS. Then, we give asymptotic results that allow

us to prove that the FMA stopping time is better than the FSS in the TCD framework. Before entering into details, let us now remind that in contrast to QCD, where the change duration m is assumed to be infinite, for TCD the change duration is assumed to be finite. This is modeled as in (4.67) by

$$X_n \sim \begin{cases} \mathcal{H}_0 : f_0(X_n), & \text{if } n < v \text{ or } n \geq v + m \\ \mathcal{H}_1 : f_1(X_n), & \text{if } v \leq n < v + m \end{cases}. \tag{4.129}$$

We consider here m to be known. The effectiveness of TCD is quantified by the trade-off between $\mathrm{P}_{\mathrm{ud}}(T)$ and $\mathrm{P}_{\mathrm{fap}}(T)$, both defined in (4.68) and (4.69), respectively. Actually, the optimal criterion of TCD, as given in (4.70), is to minimize $\mathrm{P}_{\mathrm{ud}}(T)$ among all stopping times satisfying $\mathrm{P}_{\mathrm{fap}}(T) \leq \widetilde{\alpha}$. As commented, no optimal solution is available for this class of TCD problem for the general case of $1 < m < \infty$, being the FMA solution the best available one. This solution is given by the FMA stopping time T_{FMA} in (4.90). Moreover, it was also shown that T_{FMA} fulfills the following bounds, stated in Theorem 8:

$$\begin{aligned} \mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{FMA}}(h), m\right) &\leq \mathbb{P}_1\left(\Lambda_m < h\right) = \beta_{\mathrm{NP}}(m, h), \\ \mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{FMA}}(h), m_\alpha\right) &\leq 1 - \left[\mathbb{P}_\infty\left(\Lambda_m < h\right)\right]^{m_\alpha} = 1 - \left[1 - \alpha_{\mathrm{NP}}(m, h)\right]^{m_\alpha}, \end{aligned} \tag{4.130}$$

where m_α stands for the period of samples used to define the PFAP, and $\{\beta_{\mathrm{NP}}(m,h), \alpha_{\mathrm{NP}}(m,h)\}$ denote the PMD and PFA, respectively, of a LRT with sample size equal to $N = m$ and detection threshold h.

Optimal FSS strategy

As for the QCD case, we define the FSS stopping time as in (4.110) by

$$T_{\mathrm{NP}} \doteq N \cdot \inf\left\{n \geq 1 : \delta_n = 1\right\} = N \cdot \tilde{T}, \tag{4.131}$$

with δ_n the decision of a LRT at the n-th block of N samples. Thereby, the performance of the FSS stopping time in terms of a TCD problem is given by the following result.

Proposition 9. Let T_{NP} be the FSS stopping time in (4.131) with size sample $N \geq 1$ and let $m \geq 1$ be the transient change duration. Moreover, let $\alpha_{\mathrm{NP}}(N,h) \doteq \mathbb{P}_\infty(\Lambda_N \geq h)$ denote the PFA of a LRT with sample size N and detection threshold h. Therefore, the PFAP, with a given interval of samples m_α, boils down to

$$\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{NP}}\right) = 1 - \left[1 - \alpha_{\mathrm{NP}}(N,h)\right]^{\left\lceil \frac{m_\alpha}{N} \right\rceil}. \tag{4.132}$$

On the other hand, let $\beta_{\mathrm{NP}}(N,h) \doteq \mathbb{P}_1(\Lambda_N < h)$ denote the PMD of the same LRT, and let $k = \left\lfloor \frac{m}{N} \right\rfloor$. Hence, the worst-case PUD is given by

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{NP}}\right) = \begin{cases} \mathbb{P}_{(k+1)N-m}\left(\Lambda_N < h\right) \cdot \left[\beta_{\mathrm{NP}}(N,h)\right]^{k-1}, & \text{if } 1 \leq N \leq m \\ 1, & \text{otherwise} \end{cases}. \tag{4.133}$$

Proof. The proof is given in Appendix 4.D. □

Similarly to the QCD problem, it turns out that $\mathrm{P}_{\mathrm{ud}}(T_{\mathrm{NP}})$ depends on $\{\beta_{\mathrm{NP}}, \alpha_{\mathrm{NP}}, N\}$, so

that we can fix N in order to minimize $\mathrm{P_{ud}}(T_{\mathrm{NP}})$, satisfying that $\mathrm{P_{fap}}(T_{\mathrm{NP}}) \leq \widetilde{\alpha}$. Thus,

$$\mathrm{P}^*_{\mathrm{ud}}\left(T_{\mathrm{NP}}\right) \doteq \min_{1\leq N\leq m}\left\{\mathrm{P_{ud}}\left(T_{\mathrm{NP}}\right)\right\} = \mathbb{P}_{(k+1)N^*-m}\left(\Lambda_{N^*} < h(\widetilde{\alpha})\right)\cdot\left[\beta_{\mathrm{NP}}(N^*, h(\widetilde{\alpha}))\right]^{k-1}, \quad (4.134)$$

with

$$\beta_{\mathrm{NP}}(N^*, h(\widetilde{\alpha})) = \mathbb{P}_1\left(\Lambda_{N^*} < h(\widetilde{\alpha})\right) = F_1\left(h(\widetilde{\alpha}); N^*\right), \quad (4.135)$$

where N^* stands for the optimal value of the sample size N (in terms of the TCD criterion) and $F_1(h(\widetilde{\alpha}); N)$ is the cdf of Λ_N under $\mathcal{H}_1$ evaluated at $h(\widetilde{\alpha})$. The threshold $h(\widetilde{\alpha})$ is then fixed in order to guarantee a given PFAP equal to $\widetilde{\alpha}$, from (4.132), as

$$h(\widetilde{\alpha}) = F_\infty^{-1}\left((1-\widetilde{\alpha})^{\frac{1}{\left\lceil \frac{m_\alpha}{N^*}\right\rceil}}; N^*\right), \quad (4.136)$$

and then the optimal sample size of the T_{NP} to minimize $\mathrm{P_{ud}}(T_{\mathrm{NP}})$ is given by

$$N^* = \arg\min_N\left\{\mathbb{P}_{(k+1)N-m}\left(\Lambda_N < h(\widetilde{\alpha})\right)\cdot\left[\beta_{\mathrm{NP}}(N, h(\widetilde{\alpha}))\right]^{k-1}\right\}, \quad (4.137)$$

which should be numerically computed. The term $\mathbb{P}_{(k+1)N-m}(\Lambda_N < h(\widetilde{\alpha}))$ is also numerically computed as the probability that the sum of $(k+1)N-m-1$ LLRs under $\mathcal{H}_0$ and $m-kN+1$ LLRs under $\mathcal{H}_1$ exceeds the threshold $h(\widetilde{\alpha})$ (computed as in (4.136)).

Asymptotic analysis

Now we would like to compare the optimal PUD for the FSS, $\mathrm{P}^*_{\mathrm{ud}}(T_{\mathrm{NP}})$, with the PUD of the FMA stopping time, $\mathrm{P_{ud}}(T_{\mathrm{FMA}})$, which is the benchmark of comparison for the TCD problem. Previously, we showed how to find $\mathrm{P_{ud}}(T_{\mathrm{NP}})$ numerically, but we did not provide a closed-form expression. Unfortunately, we cannot directly compare $\mathrm{P}^*_{\mathrm{ud}}(T_{\mathrm{NP}})$ with $\mathrm{P_{ud}}(T_{\mathrm{FMA}})$ using the previous results. For this reason it is of interest to obtain approximations for $\mathrm{P}^*_{\mathrm{ud}}(T_{\mathrm{NP}})$, so that it can be directly compared with $\mathrm{P_{ud}}(T_{\mathrm{FMA}})$. This is provided by the following theorem.

Theorem 10. Let β_{NP} be the PMD of a LRT with N samples and let $\widetilde{\alpha}$ be a desired level of PFAP such that $\mathrm{P_{fap}}(T_{\mathrm{NP}}) \leq \widetilde{\alpha}$, then the optimal sample size N^* for the TCD problem, with transient length m, is

$$N^* \simeq \left\lceil \frac{m}{2}\right\rceil \text{ as } \widetilde{\alpha}\to 0. \quad (4.138)$$

Now, let $\mathrm{P}^*_{\mathrm{ud}}(T_{\mathrm{NP}})$ be the minimum PUD achievable by T_{NP} with sample size N and transient length m. Hence

$$\mathrm{P}^*_{\mathrm{ud}}\left(T_{\mathrm{NP}}\right) \simeq \beta_{\mathrm{NP}}\left(\left\lceil \frac{m}{2}\right\rceil, h(\widetilde{\alpha})\right) \text{ as } \widetilde{\alpha}\to 0. \quad (4.139)$$

Proof. The proof is given in Appendix 4.E. □

Finally, with the previous theorem we can establish the following result directly comparing $\mathrm{P}^*_{\mathrm{ud}}(T_{\mathrm{NP}})$ and $\mathrm{P_{ud}}(T_{\mathrm{FMA}})$ for a given value of $\widetilde{\alpha}$.

Corollary 4. Let us compare the optimal FSS and FMA stopping times. Asymptotically, as $\widetilde{\alpha}\to 0$, the worst-case PUD with a FSS stopping time for a given value of PFAP is greater than that PUD obtained by the FMA stopping time:

$$\mathrm{P}^*_{\mathrm{ud}}\left(T_{\mathrm{NP}}, m\right) \geq \mathrm{P_{ud}}\left(T_{\mathrm{FMA}}, m\right), \text{ as } \widetilde{\alpha}\to 0. \quad (4.140)$$

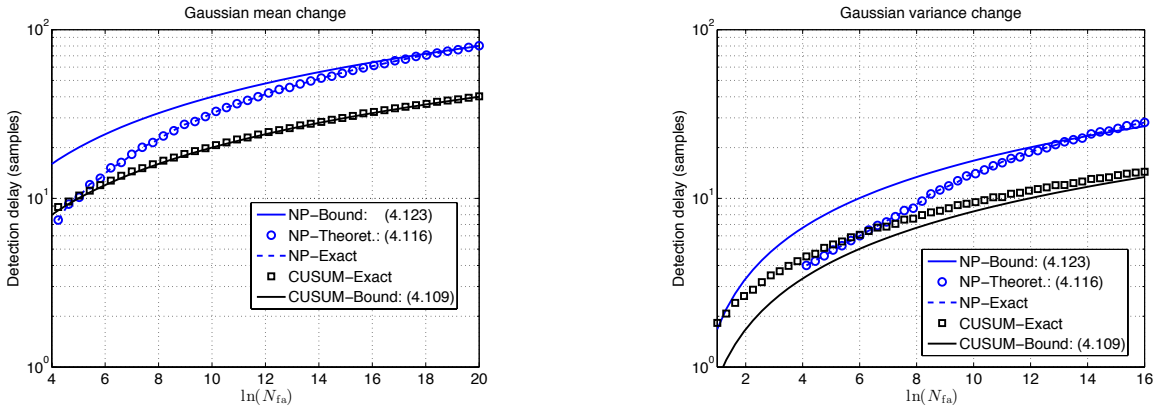

Figure 4.5: Comparison between FSS and CUSUM stopping times in terms of detection delay. (Left) Gaussian mean change and (right) Gaussian variance change with the simulation parameters listed in Table 4.1.

Proof. Let h_m be fixed so that

$$\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{FMA}}, m_\alpha\right) \leq 1 - \left[1 - \alpha_{\mathrm{NP}}(m, h_m)\right]^{m_\alpha} = \widetilde{\alpha}, \tag{4.141}$$

and assume we fix h_{N^*} so that

$$\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{NP}}, m_\alpha\right) \leq 1 - \left[1 - \alpha_{\mathrm{NP}}\left(\left\lceil \frac{m}{2} \right\rceil, h_{N^*}\right)\right]^{\left\lceil \frac{2m_\alpha}{m} \right\rceil} \leq \widetilde{\alpha}, \tag{4.142}$$

with $\widetilde{\alpha}$ the desired level of false alarms. Therefore, we have

$$\begin{aligned} \mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{FMA}}, m\right) &\leq \mathbb{P}_1\left(\Lambda_m < h_m\right) = \beta_{\mathrm{NP}}(m, h_m), \\ \mathrm{P}^*_{\mathrm{ud}}\left(T_{\mathrm{NP}}, m\right) &= \mathbb{P}_1\left(\Lambda_{N^*} < h\right) = \beta_{\mathrm{NP}}\left(N^*, h_{N^*}\right), \end{aligned} \quad \text{as } \widetilde{\alpha} \to 0, \tag{4.143}$$

with $N^* = \lceil m/2 \rceil$ and $\beta_{\mathrm{NP}}(n, h_n)$ given by

$$\beta_{\mathrm{NP}}(n, h_n) = e^{-nD_n}, \tag{4.144}$$

where D_n is increasing with n. This result can be obtained from Stein's lemma [Cov06] as shown in [Bla74]. Hence,

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{FMA}}, m\right) \leq \beta_{\mathrm{NP}}(m, h_m) \leq \beta_{\mathrm{NP}}\left(N^*, h_{N^*}\right) = \mathrm{P}^*_{\mathrm{ud}}\left(T_{\mathrm{NP}}, m\right), \text{ as } \widetilde{\alpha} \to 0, \tag{4.145}$$

and Corollary 4 thus follows. □

Actually, Corollary 4 can also be proved for any $N \leq \lceil m/2 \rceil$ in a similar fashion as we did to prove Proposition 9, without needing to prove Theorem 10. Thus a converse of Corollary 4 can alternatively be used to prove the results in Theorem 10.

4.5.3 Simulation results

Finally, we provide numerical results in order to validate the provided theoretical results for the comparison between SCD and FSS detection. This is done for several distributions for the LLR of the samples, as well as for several values of the KL divergence, emulating the variation of the change fault magnitude. Next, we analyze with details these results.

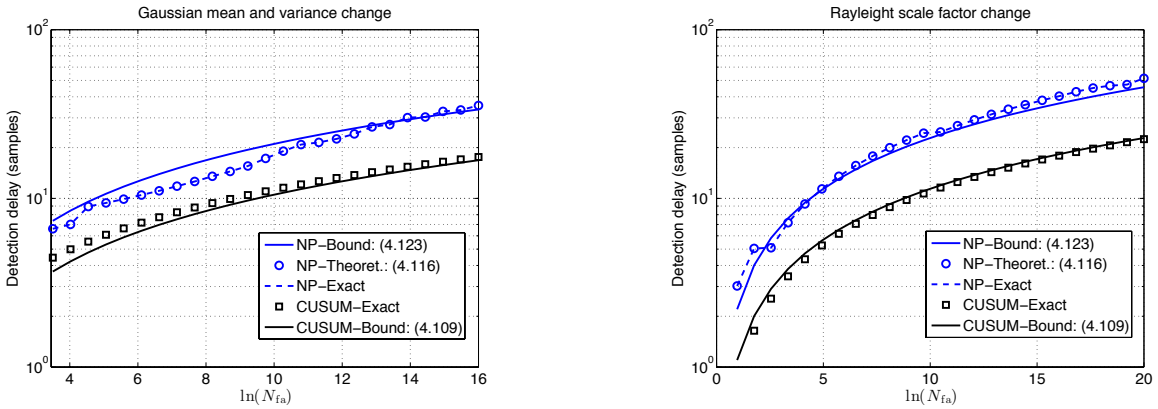

Figure 4.6: Comparison between FSS and CUSUM stopping times in terms of detection delay. (Left) Gaussian mean and variance change and (right) a LLR distributed as a Rayleigh distribution with different scale factor under $\mathcal{H}_0$ and $\mathcal{H}_1$. The simulation parameters are shown in Table 4.1

Mean change		Variance change		M&V change		Rayleight	
μ_0	0	μ	0	μ_0	0	b_0	0.4
μ_1	1			μ_1	1		
σ^2	1	σ_0^2	1	σ_0^2	1	b_1	0.7
		σ_1^2	5	σ_1^2	3		

Table 4.1: Simulation parameters for the different simulated scenarios.

CASE 1: QCD comparison

Figure 4.5 illustrates the comparison between the detection delay of the optimal FSS and CUSUM stopping times. The comparison is done for different distributions. Specifically, the left and right plots represent for the common problem of a change in the mean and variance of a Gaussian distribution, respectively. These two cases lead to a Gaussian and chi-squared distributed LLR, respectively. Moreover, Figure 4.6 accounts for different distributions for the LLR. In particular, the left plot accounts for the LLR resulting from a Gaussian mean and variance change, and the right plot considers a LLR distributed as a Rayleigh distribution with different scale factor under $\mathcal{H}_0$ and $\mathcal{H}_1$. This last figure maybe does not take into account realistic statistical change problems in practice, but it is intended to evaluate the theoretical results with a wide range of distributions to confirm the generality of the results. In each plot we include the asymptotic bound for the detection delay of both the CUSUM and FSS stopping times, given by (4.109) and (4.123), respectively. This is depicted as a function of the logarithm of the desired mean time between false alarms (i.e. $\ln(N_{\text{fa}})$). Furthermore, we include the exact detection delay for both stopping times, simulated using 10^4 Monte-Carlo runs. In addition, for the FSS we include the theoretical expression for the detection delay given in (4.116), which is numerically computed. The simulation parameters used to obtain the results are listed in Table 4.1. The simulations show a tight match between theoretical and simulated results. In particular, we see how the theoretical expression in (4.116) perfectly fits the simulated one. On the other hand, the asymptotic results in (4.123) are approached when $\gamma \to \infty$, as expected.

We also see the perfect match between simulated and theoretical results for the CUSUM, concluding that the CUSUM stopping time gives better results than the FSS one, in terms of detection delay for a fixed N_{fa}. This is true for any value of N_{fa}, but in addition, we also confirm the result that the optimal FSS delay is asymptotically twice the delay of the CUSUM. These results hold true for all the simulated distributions, thus confirming the generality of the results.

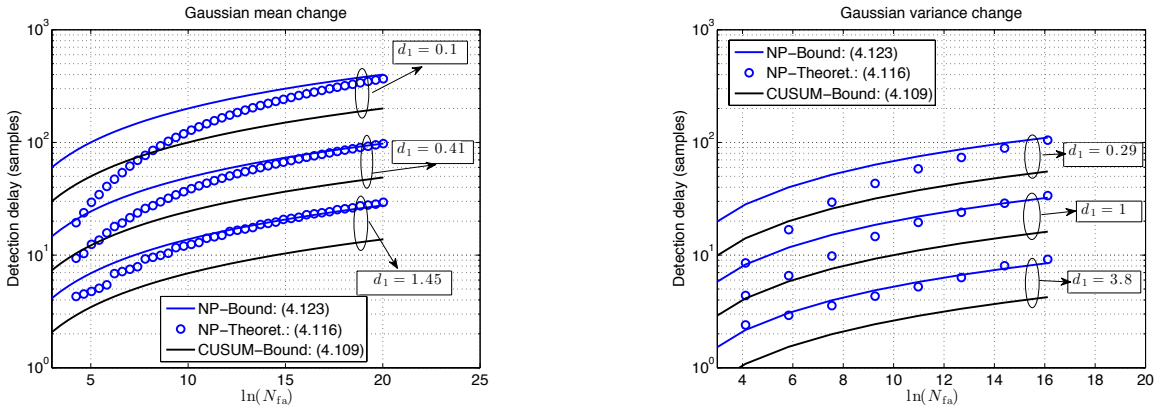

Figure 4.7: Comparison between FSS and CUSUM stopping times in terms of detection delay for (left) a change in the mean and (right) variance of a Gaussian distribution for different values of the change magnitude.

Similar results are obtained in Figure 4.7, in which we show the detection delay dependence upon the $\ln(N_{\text{fa}})$ for different values of the KL divergence of f_1 with respect to f_0, $d_1 \doteq \mathrm{E}_1[\mathrm{LLR}(n)]$. This analysis is of interest because this KL divergence depends upon the change parameters, so that it is related with the change magnitude. For instance, for a Gaussian mean change we have

$$d_1 = \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}, \tag{4.146}$$

where μ_i, with $i = \{0, 1\}$, stands for the mean before and after the change, respectively; and σ^2 the constant variance of the Gaussian distribution. So, in many applications such as communications or radar, the value of d_1 stands for the Signal-to-Noise Ratio (SNR). Results in Figure 4.7 also indicates the match between theoretical and simulated results, but in this case we see that for very low SNR (i.e. low d_1) the FSS stopping time performs better than the CUSUM for low values of N_{fa}. It is worth noting that this is not usually the case in practice, where a large mean time between false alarms is desired. So, in many practical cases (i.e. large N_{fa}) the CUSUM will still perform better than the FSS stopping time. Moreover, we see in the figure that as the SNR is slightly increased this behavior is not longer experienced. At this point, it is also important to say that the results presented in Figure 4.7 have been obtained by fixing the detection threshold with the well-known logarithmic relation (i.e. $h = \ln(N_{\text{fa}})$). However, this relation is a lower bound for the mean time between false alarms, so that the real performance (taking into account the real mean time between false alarm) of the CUSUM algorithm will be better than the one presented in the figure. We have follow this approach due to the very time consuming simulation needed to obtain the real mean time between false alarm.

CASE 2: TCD comparison

Now, we aim to asses the theoretical findings for the optimal FSS stopping time in the sense of the TCD criterion. As before, we will consider several distributions for the LLR. This is aimed at proving the generality of the obtained theoretical results. We focus now on the validation of the theoretical results in Proposition 9, given by (4.132) and (4.133). These results give an exact expression for the PFAP and PUD of the FSS stopping time with a given sample size N, respectively. The simulation of these probabilities with 10^6 Monte-Carlo runs for a change in the mean or variance of a Gaussian distribution is illustrated in Figure 4.8 and Figure 4.9, respectively. These figures also show the corresponding exact expression in (4.132) and (4.133), respectively. These results have been obtained by fixing the change parameters listed in Ta-

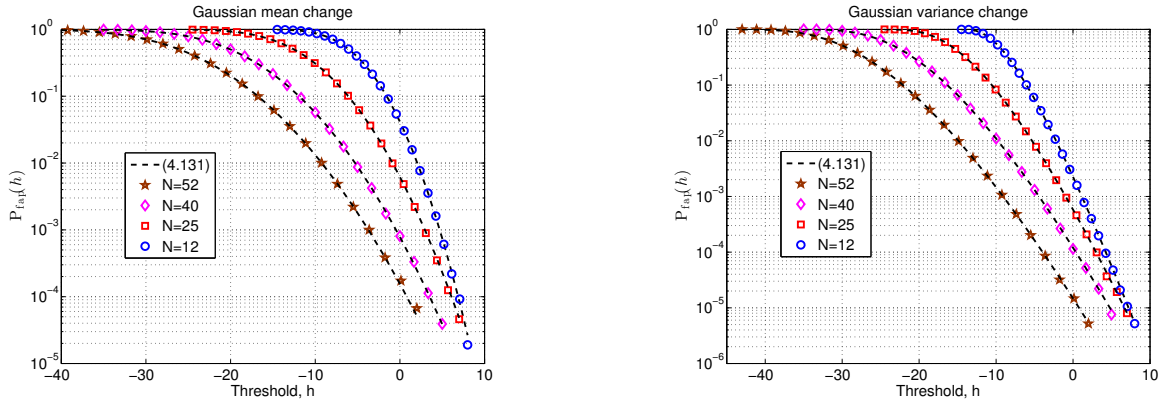

Figure 4.8: Assessment of the theoretical PFAP given in (4.132) for different values of the block sample size N. Comparison with theoretical (dashed line) and simulated (markers) PFAP as a function of the detection threshold. (Left) Gaussian mean change and (right) Gaussian variance change.

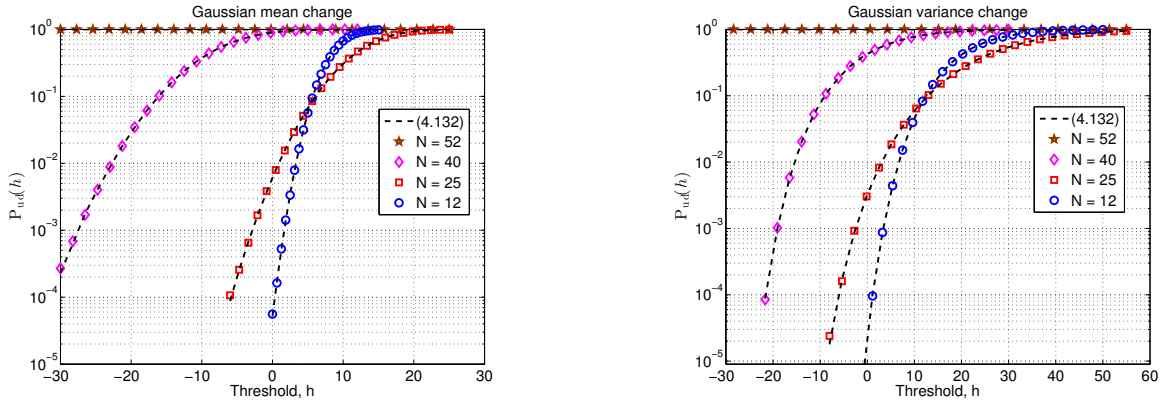

Figure 4.9: Assessment of the theoretical PUD given in (4.133) for different values of the block sample size N. Comparison between theoretical (dashed line) and simulated (markers) PUD as a function of the detection threshold. (Left) For a Gaussian mean change and (right) a Gaussian variance change.

ble 4.1, a change duration of $m = 51$ samples, $m_\alpha = 1$ sample, and using different values for the block sample size N for the FSS detection. For the Gaussian mean change, the parameters to consider are the mean before and after the change and the constant variance, $\{\mu_0, \mu_1, \sigma^2\}$; for the Gaussian variance change the variance before and after the change and the constant mean, $\{\sigma_0^2, \sigma_1^2, \mu\}$. We consider these two kind of changes because they are very prevalent in practice. For a broader variety of distributions for the LLR have been considered in Figure 4.10. This is aimed at proving the generality of the results in Proposition 9. In particular, Figure 4.8 shows a perfect fit between theoretical and simulated results for all the considered values of N. It is interesting to note that the larger the sample size N, the smaller the PFAP for a fixed threshold. This is intuitive since in this case the FSS will accumulate more samples, thus being more likely that the LLR exceeds the threshold.

Similar results are observed for the PUD in Figure 4.9; that is, the perfect match between the theoretical expression in (4.133) and the simulated results. In this case, though, as indicated in (4.133), the PUD is equal to 1 for all values of threshold when $N > m$. This is shown by the brown curve in both plots, accounting for the Gaussian mean and variance change, respectively, and a sample size of $N = 52$. Furthermore, an interesting behavior is experienced in this case when $1 \leq N \leq m$. On the one hand, a better performance is experienced as N decreases. Actually, this is so for all value of the detection threshold h as long as $N \geq \lceil m/2 \rceil$. On the

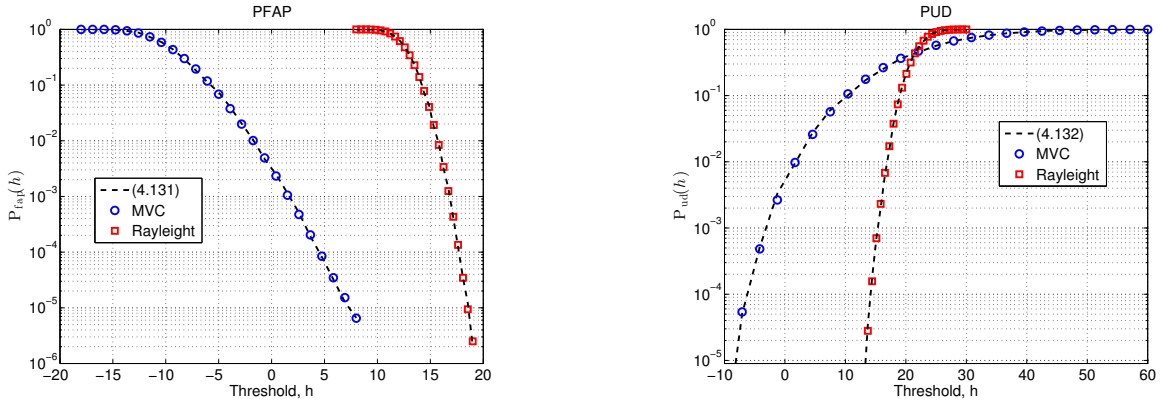

Figure 4.10: Performance evaluation for several distributions for the LLR. (Left) PFAP and (right) PUD as a function of the detection threshold.

other hand, when $N < \lceil m/2 \rceil$ the situation changes. As observed in Figure 4.9, in this case we experience two different behaviors:

1. When $h \to \infty$, the PUD gets better as the sample size N increases.
2. When $h \to -\infty$, the PUD gets better as the sample size N decreases.

This behavior confirms our asymptotic results about the optimal sample size; that is, the optimal sample size is equal to $N^* = \lceil m/2 \rceil$ as $\widetilde{\alpha} \to 0$. The asymptotic regime $\widetilde{\alpha} \to 0$ is equivalent to $h \to \infty$, so that the results in Figure 4.9 confirms the above asymptotic result. Finally, just saying that Figure 4.10 is provided to confirm that the previous discussion holds true for a variety of distributions, so that the generality of the results in Proposition 9 is proved.

CASE 3: Evaluation of the optimal FSS for the TCD criterion

Here, we assess the theoretical findings in Theorem 10, given by (4.138) and (4.139). These results give asymptotic expressions for the optimal sample size, N^* and PUD, $\mathrm{P}^*_{\mathrm{ud}}$, of the FSS stopping time, respectively. The numerical comparison of these theoretical expressions with the exact results respectively provided in (4.137) and (4.134) is carried out. Without loss of generality, we show the results for the Gaussian mean change. The same parameters as before are fixed, but different values of m are considered. In these simulations, the detection threshold has been set in order to fix a PFAP equal to α from (4.132). Then, both the optimal sample size and PUD are obtained for different values of α. The result of this process is depicted in the left and right plots of Figure 4.11, respectively. In the left plot we see a very tight match between the exact optimal sample size obtained from (4.137) and the asymptotic result in (4.138) for all the considered values of the transient length m. This is equivalent to say that the optimal sample size of the FSS stopping time for the TCD problem is given by $N^* = \lceil m/2 \rceil$. This holds true even for large values of the fixed PFAP. We see some instabilities for very large values of α, but they are due to the lack of realizations used to numerically computing N^* from the exact expression in (4.137). In these cases, in which $\alpha \to 1$, the numerical computation of N^* turns out to be equal to $\lceil m/k \rceil$, with k some integer greater than 2. This is in line with the proof of Theorem 10, in which is shown that the PUD at these points is better than in the points within. Thus, the lack of realizations in the numerical computation makes the optimization process to choose one of these points. It is worth noting that even though this lack of realizations introduces some instabilities in the match between simulated and theoretical results, this is so

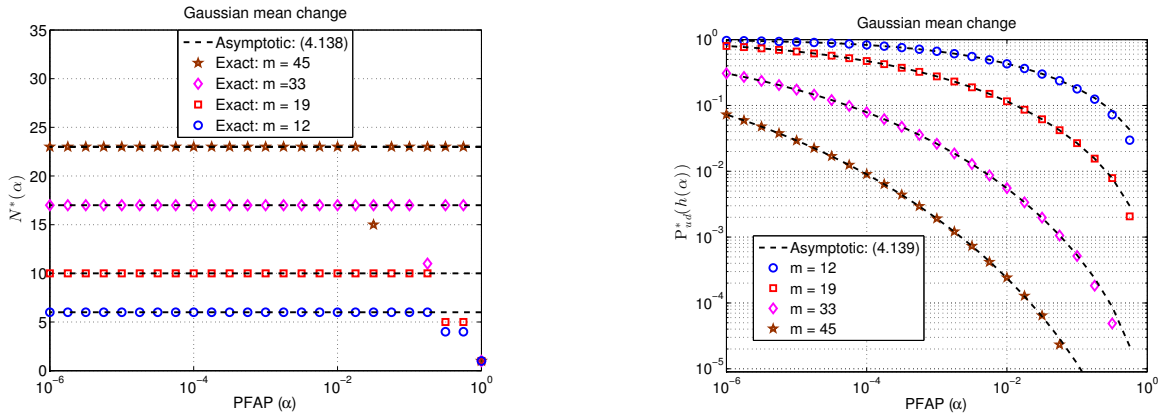

Figure 4.11: Asymptotic results of the optimal FSS stopping time for the TCD framework as a function of $\widetilde{\alpha}$ and several values of m. (Left) Optimal sample size in (4.138) and (right) optimal PUD in (4.139).

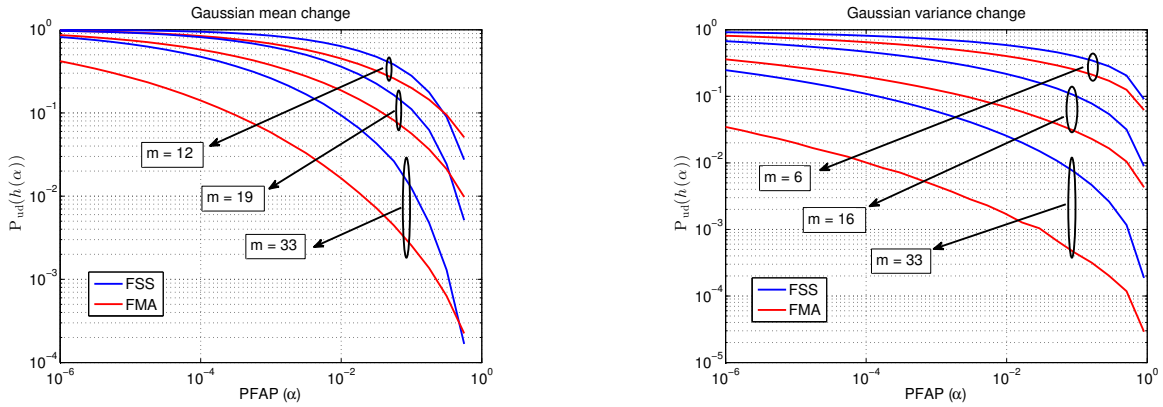

Figure 4.12: Comparison between FSS and FMA stopping times in terms of PUD for a fixed PFAP α for several values of m. (Left) Gaussian mean change and (right) Gaussian variance change.

for the optimal sample size at some large values of the fixed PFAP. Nevertheless, the match between these results is perfect for low and moderate α. This is the case of the optimal sample size. In the right plot of Figure 4.11 we see that these instabilities are not present in the optimal PUD. This is because although the optimal sample size may deviate for some vales of $\alpha \to 1$, the resulting PUD in these cases will be very similar. This is why the results in the right plot of Figure 4.11 confirms the perfect match between the simulated and asymptotic results for the optimal PUD, given in (4.133), thus proving the validity of Theorem 10, even for large α.

CASE 4: Comparison between FMA and optimal FSS for TCD

Finally, we conclude this section by showing the performance comparison between the FMA and optimal FSS stopping time in terms of the TCD criterion; that is, we compare the PUD for a given value of PFAP, henceforth referred to as ROC. This is aimed at evaluating the theoretical results of Corollary 4, given by (4.140), stating that the PUD for the FMA stopping time is asymptotically (as $\alpha \to 0$) smaller than the one given by the optimal FSS stopping time. To do so, we compare these stopping times for different distributions and values of m. This is done in terms of the simulated ROC of both stopping times. This simulations have been done by numerically computing the exact PUD with 10^6 Monte-Carlo runs. The detection threshold has been set in order to fix a given value of PFAP, given by α in the x-axis, from (4.94) and (4.132); for the FMA and optimal FSS stopping time, respectively. The same parameters as before has been used. The results for a change in the mean or variance of a Gaussian distribution are

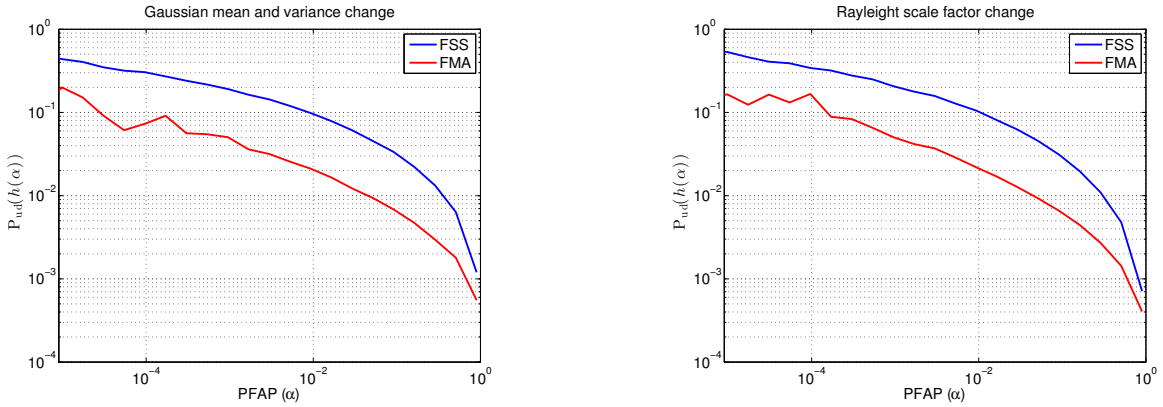

Figure 4.13: Comparison between FSS and FMA stopping times in terms of ROC for $m = 25$. (Left) Gaussian mean and variance change and (right) Rayleight scale factor change.

respectively shown in the left and right plots of Figure 4.12. We can conclude from this results that the FMA presents an improvement with respect to the optimal FSS stopping time. This is so for any of the considered transient lengths m and distributions. It is true, that for low values of m the FSS stopping time gives a slightly better performance than the FMA. Nevertheless, as for the CUSUM, we have used a bound to fix the threshold for the FMA stopping time. So, actually, when evaluating these stopping times in terms of the real PFAP, the FMA stopping time will give better performance even for low m and large α. This behavior is also corroborated with the results in Figure 4.13, in which the simulation for different distributions is shown, thus providing generality to the results. Thereby, we confirm the validity and generality of the theoretical results stated in Corollary 4.

4.6 Distributed Quickest Detection: Spatio-Temporal CUSUM

Up to this point we have investigated the field of SCD. This field deals with detecting a change in distribution of a stochastic process. We have considered the case where all information about the change is available at a single sensor, only. Nevertheless, there are situations where the information available for the decision process is decentralized (i.e. available at different sensors distributed in space). We consider this situation in this section, broadly known as decentralized or distributed detection. Conventional decentralized or distributed detection approaches often consider that the change affects either a single sensor at a given time [Tar06], or all existing sensors at the same time [Tar08, Zar09]. There are other approaches, though, that consider that the sensors are gradually affected after some propagation time. In this latter case it is often assumed that sensors are placed across a sensor array, and therefore the change will always appear first at some predetermined sensor [Rag08] or at any possible sensor [Li13]. Notwithstanding, these contributions assume that all sensors will ultimately be affected after some large enough time, and they adopt a Bayesian approach with *a-priori* knowledge on the time at which the change may appear, as well as for the propagation time. We will focus on a non-Bayesian approach (i.e. no prior knowledge) and we address a more general case with just an unknown subset of sensors may experience the change. This scenario was first considered in [Mei10] and [Mei11], and it is often the case in practice when sensors are randomly distributed over a geographical path, and just a few of these sensors become exposed to the event.

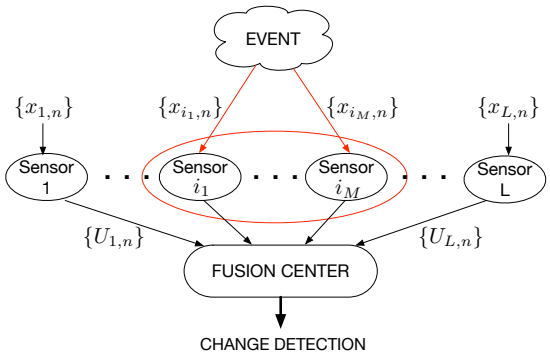

Figure 4.14: Considered scheme for decentralized change detection

Specifically, the scenario considered herein is shown in Figure 4.14, which considers a system with L sensors. At time n, an observation $X_{i,n}$ is made at sensor i and the message $U_{i,n}$, created from the available information, is sent to a central location called the fusion center. Then, based on the sensor messages, a decision about the change (event) is made at the fusion center. When the change to be detected happens, we assume that only $M < L$ consecutive sensors are affected, usually those closest to the event (i.e. $i_1, \ldots, i_M$). We will take advantage here of the fact that affected sensors are consecutive, so that the problem can be formulated as a transient change detection problem in the spatial domain. Furthermore, apart from detecting the change, we will be also able to identify the unknown set of affected sensors. This scenario happens to many real-world applications in which the sensors are randomly spaced along a path, as when deployed from an aircraft, so that the affected sensors by the change can fairly be assumed to be consecutive. The above considered scenario is of interest for applications such as environmental monitoring [Sim03], where for instance a fire will affect the closest sensors to the fire, interference monitoring [Lin07], where the affected sensors will be those close to the interference source or not obstructed by any obstacle, or intrusion detection in security monitoring systems [Sna91].

Based on this observation we firstly propose a CUSUM-based solution to the distributed quickest detection problem with unknown number of affected sensors. To do so we transform the problem to that of detecting a change within the spatial domain (i.e. from different sensors). This is in contrast to traditional solutions, which combine spatial samples for detecting a change in the temporal domain. Secondly, based on the assumption that the affected sensors are consecutive, we formulate the spatial change detection problem as a transient change detection problem within the spatial samples. Thirdly, with the previous formulation, beyond detecting the temporal change we are able to localize those sensors affected by the change without any additional cost.

4.6.1 Signal model and preliminaries

Suppose we have L randomly spaced sensors along a path communicating with a fusion center (see Figure 4.14), and each local sensor is taking measurements over time, namely $\{x_{i,n}\}_{n=1}^{\infty}$ for $i = 1, \ldots, L$. Initially, the measurements are under nominal conditions ($\mathcal{H}_0$) and $x_{i,n}$ is distributed according to the density $f_{i,0}$ at the i-th sensor. At some unknown time v, an event (change) occurs driving the measurement of an unknown subset of M sensors to different conditions ($\mathcal{H}_1$), so that if the i-th sensor is affected, the density function of its local observation $x_{i,n}$ changes from $f_{i,0}$ to $f_{i,1}$ at time v. The subset of M affected sensors will be denoted by

$\mathcal{I}_M = \{i_1, \ldots, i_M\} \in \mathcal{I}$, with $\mathcal{I}$ the set of all possible subsets of affected sensors, with a total of $|\mathcal{I}|$ elements. Both M and $\mathcal{I}_M$ are assumed to be unknown. The problem is to detect the presence of a change affecting M of the L available sensors by taking advantage of the diversity given by the information provided by the $L > 1$ sensors, in contrast to the traditional problem where we only have one sensor (i.e. $L = 1$). Let us denote by $\mathbb{E}_v^{(\mathcal{I}_M)}[X]$ the expectation of X when a change appears at time v and sensors $i \in \mathcal{I}_M$, and denote by E_∞ the same when there are no changes. Then, the previous criterion can be formulated, like in traditional QCD, as the minimization of

$$D(T) \doteq \sup_{v \geq 1} \operatorname{essup} \mathbb{E}_v^{(\mathcal{I}_M)} \left[(T - v + 1)^+ \, | \boldsymbol{X}_{v-1} \right], \tag{4.147}$$

subject to the global false alarm constraint

$$N(T) \doteq \mathrm{E}_\infty [T] \geq N_{\mathrm{fa}}, \tag{4.148}$$

where $\boldsymbol{X}_{v-1} \doteq [\boldsymbol{x}_{1,[1,v-1]} \ldots \boldsymbol{x}_{L,[1,v-1]}]$ denotes past global information (i.e. at the fusion center) at time v, with $\boldsymbol{x}_{i,[1,v-1]} = [X_{i,1}, \ldots, X_{i,v-1}]^T$ the past information at the i-th sensor; and $N_{\mathrm{fa}} > 0$ is a desired finite constant for the global false alarm rate.

The problem of monitoring a single sensor (i.e. $L = 1$) has been addressed in Section 4.3. For this problem, the optimal procedure for the non-Bayesian framework is the CUSUM algorithm, which can recursively be computed at the i-th sensor, at time n, as

$$g_{i,n} = (g_{i,n-1} + \mathrm{LLR}_{i,n})^+, \tag{4.149}$$

with $g_{i,0} = 0$ and $\mathrm{LLR}_{i,n} \doteq \ln(f_{i,1}(X_{i,n})/f_{i,0}(X_{i,n}))$ the log-likelihood ratio (LLR) of the observation $X_{i,n}$. The global problem (i.e. $L > 1$), though, in addition includes the nuisance parameter $\mathcal{I}_M \in \mathcal{I}$, and then a generalized LLR should be used instead, leading to the following stopping time [Mei10]:

$$T_{\mathrm{LR}} = \inf \left\{ n \geq 1 : \max_{\mathcal{I}_M \in \mathcal{I}} \left\{ g_n^{(\mathcal{I}_M)} \right\} \geq h \right\}, \tag{4.150}$$

with $h > 0$ the detection threshold chosen to satisfy the false alarm constraint in (4.148) and $g_n^{(\mathcal{I}_M)}$ the CUSUM algorithm applied to the subset of sensors $i \in \mathcal{I}_M$, defined as

$$g_n^{(\mathcal{I}_M)} \doteq \left(g_{n-1}^{(\mathcal{I}_M)} + \sum_{j=1}^{M} \mathrm{LLR}_{i_j,l} \right)^+. \tag{4.151}$$

Based on this recursive form, T_{LR} can be implemented as simultaneously evaluating a total of $|\mathcal{I}|$ CUSUMs with each CUSUM considering a specific subset of affected sensors $\mathcal{I}_M$.

The implementation in (4.150) is very suitable when some prior information about the number of affected sensors is available. This is the case, for instance, when we know that only one sensor is affected when the event appears (i.e. $M = 1$). In this case, from (4.150), we have the so-called Max-CUSUM

$$T_{\mathrm{mx}} \doteq \inf \left\{ n \geq 1 : \max_{1 \leq i \leq L} \{ g_{i,n} \} \geq h \right\}. \tag{4.152}$$

That is, we raise an alarm at the global level as soon as the maximum of the local CUSUMs is above the detection threshold h [Tar06]. Unfortunately, when the number M of affected sensors is completely unknown, the implementation in (4.150) is unfeasible, as it requires to evaluate

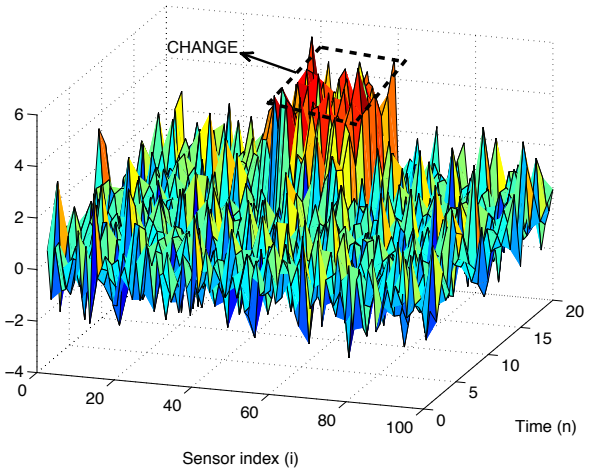

Figure 4.15: Space-Time plane with $L = 100$ sensors and a Gaussian mean change at time $v = 10$ of $M = 20$ appearing at sensors $i \in \{40, 41, \ldots, 59\}$.

too many subsets. Alternative efficient approaches were presented in [Mei11]. One of these is the henceforth referred to as Hard-CUSUM, which proposes to raise a global alarm at time

$$T_{\text{hd}} \doteq \inf \left\{ n \geq 1 : \sum_{i=1}^{L} g_{i,n} \mathbb{1}_{\{g_{i,n} \geq b\}} \geq h \right\}, \tag{4.153}$$

with $b \geq 0$. In other words, the fusion center raises an alarm as soon as the sum of those local CUSUMs above the local threshold b is too large. Results in [Mei11] show an improvement of the detection delay for a fixed false alarm rate for specific values of thresholds (i.e. b and h) of the Hard-CUSUM with respect to other approaches available in the literature.

4.6.2 Space-Time Double-CUSUM (STD-CUSUM)

Our proposed scheme for distributed quickest detection is based on the observation that physical phenomena typically affect a group of neighboring sensors, where the $M \leq L$ affected sensors can fairly be assumed to be consecutive. This can equivalently be understood as a transient change detection problem in the spatial domain. That is, for a given time instant, the spatial samples coming from each sensor can be processed sequentially as a conventional sequential detector does in the time domain. Under $\mathcal{H}_0$ all spatial measurements follow certain distribution $f_{i,0}$, whereas after the change, M of these measurements in the space domain experience a change in their distribution (i.e. from $f_{i,0}$ to $f_{i,1}$). This behavior is depicted in Figure 4.15, in which the so-called active space-time plane is highlighted. If we take a look at one of the spatial realizations under $\mathcal{H}_1$ in Figure 4.15 (i.e. for time instants $n \geq 10$) we see how the mean of some sensors measurements abruptly changes. Thus, for a given time instant, the problem can be thought as that of change detection using spatial samples. With this behavior in mind, we can think on applying the CUSUM algorithm on the spatial domain, and then use only the information of those sensors declared to be affected for making a global decision. Indeed, this is similar to the behavior of the Hard-CUSUM presented in (4.153), which only uses the information of those sensors whose local CUSUM has detected a change.

Nevertheless, the local CUSUMs used for the Hard-CUSUM are not exploiting the space

dimension, in the sense that they only consider the corresponding local sensor, thus disregarding spatial samples from neighboring sensors. As a result, it may take some time until all the local CUSUMs of the affected sensors are above the local threshold, thus incurring in a penalty in terms of detection delay. This is so because some measurements may seem not to be affected, even if they actually are. Our proposed approach is based on the idea of exploiting the space dimension in such a way that we have some statistic that increases along the space dimension as long as the sensor measurement seems to be affected and it is restarted just when the change in the space dimension disappears. Thus, the statistic will be above the local threshold even if some sensor measurements seem not to be affected.

STD-CUSUM detection rule

Assuming that the $M < L$ affected sensors are consecutive and unknown, we can model the problem of detecting the change in the space dimension as a transient change detection problem, with transient length equal to M samples. In fact, L is fixed, so this detection problem can be seen as an off-line problem. In other words, at each time we have L samples and we want to decide if there is some change within these samples. This problem is analyzed in [Kay98a], but it leads to a complex solution due to the large number of trials we need to evaluate, similarly to what happened in (4.150). Our contribution is intended to circumvent this limitation by using a simple and computationally efficient space-time CUSUM detector. Actually, in order to have the desired behavior (i.e. detecting the transient change in the space dimension) we need to use two different CUSUMs, one to detect the change from $\mathcal{H}_0$ to $\mathcal{H}_1$ and another to detect the change from $\mathcal{H}_1$ to $\mathcal{H}_0$. Let $g_{i,n}$ denote the local CUSUM to detect the former change, defined as in (4.149), and let $\tilde{g}_{i,n}$ denote the local CUSUM of the i-th sensor, but to detect the disappearance of the change, defined as

$$\tilde{g}_{i,n} \doteq \left(\widetilde{g}_{i,n-1} - \mathrm{LLR}_{i,n}\right)^+ . \tag{4.154}$$

Furthermore, let us define the CUSUM to detect the appearance and disappearance of the change in the space dimension for a given time instant n, respectively, as

$$\begin{aligned} G_{i,n} &\doteq \left(G_{i-1,n} + g_{i,n-1} + \mathrm{LLR}_{i,n}\right)^+ \mathbb{1}_{\{\widetilde{G}_{i,n-1}<b\}} \\ \widetilde{G}_{i,n} &\doteq \left(\widetilde{G}_{i-1,n} + \widetilde{g}_{i,n-1} - \mathrm{LLR}_{i,n}\right)^+ \mathbb{1}_{\{G_{i,n}\geq b\}} \end{aligned} , \tag{4.155}$$

with $G_{i,0} = 0$ and $\widetilde{G}_{i,0} = 0$ for $i = 1, \ldots, L$. It is worth pointing out that this recursive form includes both time (with $g_{i,n-1}$) and space (with $G_{i-1,n}$) integration, and it is here where our contribution exploit both time and space dimensions. With this framework the henceforth referred to as Space-Time Double-CUSUM (STD-CUSUM) is defined as

$$g_{i,n}^{(\mathrm{STD})} \doteq G_{i,n}\mathbb{1}_{\{\widetilde{G}_{i,n}<b\}}. \tag{4.156}$$

Identification of affected sensors

The introduction of the STD-CUSUM leads to the desired behavior, i.e., we have a CUSUM-based statistic that is able to detect the appearance and disappearance of the change in the space dimension. Doing so we are exploiting the space dimension in such a way that, at time n, the STD-CUSUM increases along the space dimension as long as the sensor measurements seems to be affected by the change, so that it may still be above the local threshold even if some of the measurements seem not to be affected. With this formulation, we are able to localize

those sensors affected by the change, even if they are not consecutive, at time n, as

$$\widehat{\mathcal{I}}_M(n) = \left\{\widehat{i}_1, \ldots, \widehat{i}_M\right\} = \left\{i : g_{i,n}^{(\mathrm{STD})} \geq b\right\}. \tag{4.157}$$

Global STD-CUSUM stopping time

Let us now focus on the global detection carried out in the fusion center. The idea is, at time n, to apply the STD-CUSUM for detecting those sensors that are affected by the change, so that we can use the expression in (4.150) but substituting $\mathcal{I}_M$ by their estimates at time n given by (4.157). Doing so, we get rid off the maximization over $\mathcal{I}_M \in \mathcal{I}$ in (4.150), drastically improving the computational burden. To do so, we suppose that, at time n, each sensor i sends its LLR (i.e. $U_{i,n} = \mathrm{LLR}_{i,n}$), and then the fusion center calculates the STD-CUSUM in (4.156) so that it can raise a global alarm at time

$$T_{\mathrm{sc}} = \inf\left\{n \geq 1 : \sum_{i=1}^{L} g_{i,n} \mathbb{1}_{\left\{g_{i,n}^{(\mathrm{STD})} \geq b\right\}} \geq h\right\}. \tag{4.158}$$

4.6.3 Numerical results

This section is intended to provide numerical simulations to illustrate the performance of our proposed scheme and compare it to other schemes in the literature. We consider two scenarios, one using $L = 100$ and another with $L = 20$ iid sensors. For the sake of clarity and illustration we suppose iid Gaussian observations at each sensor with mean 0 and variance 1 before the change (i.e. $\mathcal{H}_0$) and with mean 0.5 and same variance after the change (i.e. $\mathcal{H}_1$). In practice, these observations can be any detection metric related with the event of interest, like temperature, power, integrity metric, ... The results presented next under case #1 consider the affected sensors to be consecutive and case #2 considers the case when all the affected sensors are not consecutive. In particular, the presented results compare the performance of the STD-CUSUM in terms of detection delay, with a fixed value for the time between false alarm, with the Hard- and Max-CUSUM. For our proposed detector (STD-CUSUM) we use different values of the local threshold b and we find a value h, from numerical simulations, such that $N(T) \approx N_{\mathrm{fa}} = 10^4$. The selected local thresholds are chosen so that we can make a wide comparison with those selected in [Mei11] for the Hard-CUSUM. Next, using the obtained threshold h, we obtain the detection delays of the analyzed algorithms with different number of affected sensors M. This is done for both scenarios with $L = 100$ and $L = 20$. For the other schemes we use the same values for the thresholds as in [Mei11], and then the obtained results are the same.

CASE 1: Consecutive affected sensors

In this section, we consider the case when the affected sensors are consecutive, as it would be the case in a sensor array.

In Table 4.2 we can see how for $L = 100$ the STD-CUSUM is in general better than the other analyzed schemes. In particular, for $M > 5$ the STD-CUSUM is better than the Hard-CUSUM for any of the used thresholds. For lower M only with the configuration using the larger local threshold ($b = 5$) in the STD-CUSUM is better than the Hard-CUSUM (for any of its configurations), giving a gain of 10 samples in terms of detection delay. Regarding the Max-CUSUM, the STD-CUSUM outperforms it for any value of M, except for $M = 1$. It is interesting to see that, in contrast to the Hard-CUSUM, the larger the local threshold b the

γ	Detection Scheme	$L = 100$			
		# sensors affected M			
		80	20	10	1
10^4	$T_{\text{hd}}(h = 106.4, b = 0.5)$	7.3	20.2	33.8	195.5
	$T_{\text{hd}}(h = 62.3, b = 2.3)$	9.2	19.7	31.9	191.6
	$T_{\text{hd}}(h = 29.7, b = 4.6)$	14.2	21.9	29.9	152.6
	$T_{\text{sc}}(h = 76, b = 0.25)$	5.5	17.2	29.8	164.4
	$T_{\text{sc}}(h = 63.5, b = 2.6)$	4.9	14.5	24.7	170.1
	$T_{\text{sc}}(h = 47, b = 5)$	**3.8**	**11.7**	**20.1**	142.8
	$T_{\text{mx}}(h = 11.12)$	32.7	39.9	45.2	**85.5**
	Detection Scheme	$L = 20$			
		# sensors affected M			
		16	7	2	1
	$T_{\text{hd}}(h = 32.9, b = 0.5)$	12.1	24.4	69.9	122.8
	$T_{\text{hd}}(h = 25, b = 2.3)$	13.7	24.1	70.1	126.8
	$T_{\text{hd}}(h = 16.5, b = 4.6)$	19.3	27.1	59.8	112.4
	$T_{\text{sc}}(h = 24.2, b = 0.25)$	10.4	20.9	57.5	98.3
	$T_{\text{sc}}(h = 22.1, b = 2.6)$	8.9	18.3	55.5	102.9
	$T_{\text{sc}}(h = 18, b = 5)$	**7.9**	**16.2**	**50.4**	96.0
	$T_{\text{mx}}(h = 9.75)$	34.16	41.36	57.89	**74.42**

Table 4.2: Detection delays with $L = 100$ and $L = 20$ for different values of M consecutive affected sensors

γ	Detection Scheme	$L = 20$			
		# sensors affected M			
		16	7	4	2
10^4	$T_{\text{sc}}(h = 24.2, b = 0.25)$	10.3	20.5	32.3	57.6
	$T_{\text{sc}}(h = 22.1, b = 2.6)$	9.4	18.8	30.3	56.2
	$T_{\text{sc}}(h = 18, b = 5)$	**8.7**	**18.7**	**28.7**	**51.4**
	Detection Scheme	$L = 100$			
		# sensors affected M			
		80	20	10	5
	$T_{\text{sc}}(h = 76, b = 0.25)$	5.6	16.5	27.9	47.6
	$T_{\text{sc}}(h = 63.5, b = 2.6)$	5.2	15.1	25.3	44.1
	$T_{\text{sc}}(h = 47, b = 5)$	**4.1**	**14.0**	**22.8**	**37.4**

Table 4.3: Detection delays for non consecutive affected sensors

better performance of our proposed scheme in terms of detection delay for a fixed false alarm rate for all the simulated values of M. So, for the STD-CUSUM it is of interest to fix a large local threshold b independently of the number of affected sensors M. This is a very good advantage since M is unknown and then with our proposed scheme we make sure it will work well for any value of M. Similar results are obtained for $L = 20$ in Table 4.2. We see how the STD-CUSUM is better than the Hard-CUSUM for any of the used configurations for all M, confirming the superiority of our proposed solution for different configurations of the local threshold b even for low values of L. Finally, it is worth pointing out that the larger M the larger is the improvement of the STD-CUSUM with respect to the Hard-CUSUM, which is intuitive since the STD-CUSUM increases as the transient change in the space dimension is larger. So, we can conclude that the STD-CUSUM is good for large L and M, but it still is a good option otherwise.

CASE 2: Non-consecutive affected sensors

Now, we show, in Table 4.3, the obtained results when the M affected sensors are not consecutive, as it would be the case in a randomly deployed sensor network. We only show the results for the STD-CUSUM, since the results for the Hard- and Max-CUSUM are the same as in Table 4.2. Firstly, we see how the detection delay in this case, for the same configuration of thresholds, is slightly degraded compared to the case of consecutive affected sensors (see Table 4.2). This

is in line with the fact that the STD-CUSUM works better when the number of consecutive affected sensor is greater. Secondly, we see that the loss of performance of the STD-CUSUM with respect the case when all the M sensors are consecutive is greater for lower values of M. This can be seen for instance for $L = 100$ when $M = 80$ the detection delay is similar for both cases (i.e. M consecutive or not consecutive sensors), whereas when $M = 5$ we have a difference of 3 samples. This is due to the fact that for large M there will likely be some subset of affected sensors that are consecutive, and then the loss of performance is negligible, whereas for small M it may happen that no consecutive sensors are affected, thus incurring in some degradation. Nevertheless, comparing with the results in Table 4.2 for the Hard-CUSUM we see that the STD-CUSUM still gives lower detection delay, so we can conclude that the STD-CUSUM outperforms the Hard-CUSUM even when the affected sensors are not consecutive.

4.7 Conclusions

In this chapter we have investigated the mathematical theory of detection, with emphasis on SCD, including both QCD and TCD. Specifically, for QCD, we have provided a comprehensive and simple discussion of the most important theoretical aspects in this theory. Contrary to the current literature in QCD, which is formed by scattered results across disparate disciplines, we have brought together these results so that a formal and unified treatment is provided for a general reader (after the given background in Section 3.6). On the other hand, we have also investigated the TCD problem in a optimal stopping time framework. However, the problem has been left open for the general case of a finite transient duration. The problem has been solved by Moustakides only for the case of a change duration of one sample. Due to the lack of optimal solution for the general TCD problem, windowed-solutions have been proposed in the literature. In particular, our windowed solution, the FMA stopping time, have been shown to be the best available candidate for the TCD problem. Up to this point, we have given an overview of traditionally used detection criteria. Actually, classical detection (studied in Section 3.7) and SCD have often been considered to be disparate fields. It is for this reason that we have compared these two methods. The theoretical findings have shown that the LRT, the optimal solution in classical detection, looses its optimality with respect the NP criterion when dealing with QCD and TCD problems. These findings are completely general and it is the first time that classical and statistical change detection have been compared within the same framework, filling the gap between these two fields.

Once the field of SCD have been investigated and compared with classical detection, we have a complete overview of detection schemes and we can conclude that statistical change detectors should be used when dealing with timely detections, either minimizing the detection delay or the probability of unbounded delay. As concluded in Chapter 3, a timely detection is of interest for GNSS-based critical applications, and particularly for integrity algorithms. Moreover, we have also highlighted in Chapter 3 the need of promptly detecting local threats for the definition of a new era of integrity algorithms to work in terrestrial environments. For the above reasons, and the fact that a classical detection has traditionally been addressed in GNSS, Chapter 5 is devoted to the design of interference and multipath quickest change detectors in GNSS. Based on the framework and concepts presented in this chapter, the design flow will be as follows:

1. Definition of observable metrics experiencing a change in distribution with the presence of

a local threat. This will lead to the definition of the random sequence $\{X_n\}$ used thorough the present chapter.

2. Statistical characterization, leading to the definition of the pdfs $f_0(X_n)$ and $f_1(X_n)$, and ultimate definition of the LLR.

3. Application of the CUSUM algorithm.

Finally, the goal of Chapter 6 is to apply the framework of TCD presented in the current chapter to integrity algorithms.

Appendix 4.A Proof of Proposition 5

We aim here at proving the results stated in Proposition 5, following the same proof as in [Poo09, Ch. 6]. As already mentioned, we show it here because it will be useful to get some insight to formulate the problem of TCD in an optimal stopping time framework. For integers $m \geq 1$, let us start by defining

$$b_m(T) = \mathrm{E}_m \left[(T - m + 1)^+ \,\middle|\, \mathcal{F}_{m-1} \right]. \tag{4.159}$$

Since X_i, with $1 \leq i < m$, has identical distributions under both $\mathbb{P}_m$ and $\mathbb{P}_\infty$, it follows that $\mathfrak{D}(T) \geq \mathfrak{D}_m(T) \geq b_m(T)$ under $\mathbb{P}_\infty$. Thus, we can write

$$\mathfrak{D}(T) \sum_{m=1}^{\infty} \mathrm{E}_\infty \left[\mathbb{1}_{\{T \geq m\}} (1 - \widetilde{g}_{m-1})^+ \right] \geq \sum_{m=1}^{\infty} \mathrm{E}_\infty \left[b_m(T) \mathbb{1}_{\{T \geq m\}} (1 - \widetilde{g}_{m-1})^+ \right]. \tag{4.160}$$

The sum on the left-hand side of this inequality is given by

$$\begin{aligned} \sum_{m=1}^{\infty} \mathrm{E}_\infty \left[\mathbb{1}_{\{T \geq m\}} (1 - \widetilde{g}_{m-1})^+ \right] &= \mathrm{E}_\infty \left[\sum_{m=1}^{\infty} \mathbb{1}_{\{T \geq m\}} (1 - \widetilde{g}_{m-1})^+ \right] \\ &= \mathrm{E}_\infty \left[\sum_{m=1}^{T} (1 - \widetilde{g}_{m-1})^+ \right] \\ &= \mathrm{E}_\infty \left[\sum_{m=0}^{T-1} (1 - \widetilde{g}_{m})^+ \right]. \end{aligned} \tag{4.161}$$

Furthermore, we have

$$\begin{aligned} b_m(T) = \sum_{k=m}^{\infty} (k - m + 1) \, \mathbb{P}_m \left(T = k \,|\, \mathcal{F}_{m-1} \right) &= \sum_{k=m}^{\infty} \mathbb{P}_m \left(T \geq k \,|\, \mathcal{F}_{m-1} \right) \\ &= \sum_{k=m}^{\infty} \mathrm{E}_m \left[\mathbb{1}_{\{T \geq k\}} \,\middle|\, \mathcal{F}_{m-1} \right]. \end{aligned} \tag{4.162}$$

Now, similarly as in the proof of Proposition 2, we have

$$\mathrm{E}_m \left[\mathbb{1}_{\{T \geq k\}} \,\middle|\, \mathcal{F}_{m-1} \right] = \mathrm{E}_\infty \left[\prod_{l=m}^{T-1} \mathrm{LR}(l) \mathbb{1}_{\{T \geq k\}} \,\middle|\, \mathcal{F}_{m-1} \right], \tag{4.163}$$

where in this case we have $\mathrm{d}\mathbb{P}_m \,/\, \mathrm{d}\mathbb{P}_\infty = \prod_{l=m}^{T-1} \mathrm{LR}(l)$. Hence, we have

$$b_m(T) = \sum_{k=m}^{\infty} \mathrm{E}_\infty \left[\prod_{l=m}^{T-1} \mathrm{LR}(l) \mathbb{1}_{\{T \geq k\}} \,\middle|\, \mathcal{F}_{m-1} \right] = \mathrm{E}_\infty \left[\sum_{k=m}^{T} \prod_{l=m}^{k-1} \mathrm{LR}(l) \,\middle|\, \mathcal{F}_{m-1} \right]. \tag{4.164}$$

With the above expression and using the facts that both $\mathbb{1}_{\{T \geq m\}}$ and $\widetilde{g}_{m-1}$ are $\mathcal{F}_{m-1}$-

measurable, the right-hand side of (4.160) becomes

$$\sum_{m=1}^{\infty} \mathrm{E}_{\infty}\left[b_m(T)\mathbb{1}_{\{T\geq m\}}\left(1-\widetilde{g}_{m-1}\right)^{+}\right] = \mathrm{E}_{\infty}\left[\sum_{m=1}^{\infty}\mathbb{1}_{\{T\geq m\}}\left(1-\widetilde{g}_{m-1}\right)^{+}\sum_{k=m}^{T}\prod_{l=m}^{k-1}\mathrm{LR}(l)\right]$$
$$= \mathrm{E}_{\infty}\left[\sum_{k=0}^{T-1}\max\left\{\widetilde{g}_k,1\right\}\right], \tag{4.165}$$

where we have used the representation

$$\max\left\{\widetilde{g}_k,1\right\} = \sum_{m=1}^{k+1}\left(1-\widetilde{g}_{m-1}\right)^{+}\prod_{l=m}^{k}\mathrm{LR}(l), \tag{4.166}$$

which can easily be verified via induction. The inequality (4.60) thus follows from (4.160), (4.161) and (4.165). Finally, to show that $\mathfrak{D}(T_{\mathrm{C}}) = \mathfrak{D}^{*}(T_{\mathrm{C}})$ we consider the sequence $U_k = \max\{S_k, 1\}$. It is straightforward to prove that this sequence is homogeneously Markovian, so that $\mathfrak{D}_m(T_{\mathrm{C}}) = \mathfrak{D}_1(T_{\mathrm{C}}) = \mathfrak{D}(T_{\mathrm{C}}), \forall m$. Another consequence of the Markovity of $\{U_k\}$ is that $\operatorname{esssup} b_m(T_{\mathrm{C}})$ is achieved when $\{T_{\mathrm{C}} \geq m\} \cap \{\widetilde{g}_{m-1} \leq 1\}$ (i.e. when the previous sample has driven the CUSUM to 0). Hence, we can conclude that the inequality (4.160) is an equality, and the Proposition 5 thus follows.

Appendix 4.B Proof of Theorem 6

In order to prove the bounds in Theorem 6 let J be the stopping time given by

$$J = \inf\left\{n : Z_n \geq h\right\}, \tag{4.167}$$

for some test statistic Z_n, which is function of the iid observations X_n (i.e. $Z_n = q(X_n)$). For $k = 1, 2, \ldots$, let J_k be the stopping time of the same test applied to the sequence $\{Z_n\}_{n\geq k}$ and define J^{*} as

$$J^{*} = \inf_{k\geq 1}\left\{J_k + k - 1\right\}. \tag{4.168}$$

Now, we aim at using Lemma 1. To do so, let Z_n in (4.167) be equal to

$$Z_n = \sum_{i=1}^{n}\rho(i). \tag{4.169}$$

Thus, the probability that the test terminates (i.e. $\mathbb{P}_{\infty}\left(J < \infty\right)$) is the probability that the boundary h is exceeded by Z_n. For the case of $\rho(n) = \mathrm{LLR}(n)$, this probability can be computed using the theory of sequential detection [Wal45]. However, this probability cannot be computed exactly for arbitrary jumps $\rho(n)$. This is why we seek bounds based on the result in Lemma 1 to overcome this issue. In order to compute the probabilities and expectations required in that lemma for J, we consider the two-sided test with boundaries $a < 0 < h$ given by the stopping time M,

$$M = \inf\left\{m : Z_m < a \text{ or } Z_m \geq h\right\}, \tag{4.170}$$

with Z_m as in (4.169). Thus, the stopping time J in (4.167) can be written now as the limit of M in (4.170) as $-a$ tends to infinity (i.e. $\lim_{a\to-\infty} M = J$).

At this point, we need some information about the OC and ASN, so that we can obtain $\mathbb{P}_\infty(M < \infty)$ and $\mathrm{E}_1[M]$, needed to apply Lemma 1. To do so, on the one hand, we will use the following bound for the OC under $\mathcal{H}_0$ [Bas93, Ch. 4]:

$$\mathbb{P}_{\theta_0}(M) \geq \frac{e^{\omega_0 h} - 1}{e^{\omega_0 h} - \eta e^{\omega_0 a}}, \tag{4.171}$$

where $\eta \leq 1$, and $\omega_0 > 0$ is the root of the equation $\mathrm{E}_\infty[e^{\omega\rho(n)}] = 1$. Thus, the probability that J does not stop under $\mathcal{H}_0$ is given by the limit as $a \to -\infty$ of the above expression; that is,

$$\mathbb{P}_\infty(J = \infty) = \lim_{a\to-\infty} \mathbb{P}_{\theta_0}(M) \geq \frac{e^{\omega_0 h} - 1}{e^{\omega_0 h}} = 1 - e^{-\omega_0 h}, \tag{4.172}$$

so that the probability that J terminates is upper bounded by

$$\mathbb{P}_\infty(J < \infty) \leq e^{-\omega_0 h}. \tag{4.173}$$

On the other hand, we will use the following bound for the ASN under $\mathcal{H}_1$ [Bas93, Ch. 4]:

$$\mathrm{E}_1[M] \leq \frac{a\,\mathbb{P}_{\theta_1}(M) + h\,[1 - \mathbb{P}_{\theta_1}(M)]}{\mathrm{E}_1[\rho(n)]}. \tag{4.174}$$

So, we have

$$\mathrm{E}_1[J] = \lim_{a\to-\infty} \mathrm{E}_1[M] \leq \lim_{a\to-\infty} \frac{a\,\mathbb{P}_{\theta_1}(M) + h\,[1 - \mathbb{P}_{\theta_1}(M)]}{\mathrm{E}_1[\rho(n)]}. \tag{4.175}$$

The right hand side is a decreasing function of $\mathbb{P}_{\theta_1}(M)$ so that the inequality is preserved if we replace $\mathbb{P}_{\theta_1}(M)$ by zero, resulting in

$$\mathrm{E}_1[J] \leq \frac{h}{\mathrm{E}_1[\rho(n)]}. \tag{4.176}$$

Now, Lemma 1 can be applied to yield results on the O-CUSUM with a threshold h and using $\rho(n)$ instead of $\mathrm{LLR}(n)$. It is worth mentioning that J^* is equivalent to the stopping time T^* defined in (4.45), and then we can show that $T_{\mathrm{OC}} = J^*$, with $Z_n = \sum_{i=1}^n \rho(i)$. Hence, the mean time between false alarms and the mean delay for the O-CUSUM are bounded as in (4.62), and the proof of Theorem 6 thus follows.

Appendix 4.C Proof of Theorem 8

The proof of Theorem 8 is divided in two parts. Firstly, we prove the bound for the PFAP given by (4.92) and (4.94). Secondly, we show the proof of the bound for the PUD given by (4.93) and (4.95).

4.C.1 Probability of false alarm in a given period $\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{FMA}}(h), m_\alpha)$

We first introduce an important result, stated in the following lemma, that will be very helpful to prove Theorem 8.

Lemma 2. Let $\Upsilon_n \doteq \sum_{i=n-m+1}^n \mathrm{LLR}(i)$, $\Lambda_m = \Upsilon_m \doteq \sum_{i=1}^m \mathrm{LLR}(i)$, $k \geq m$ and $N > k$ be

integers; then

$$\mathbb{P}_\infty\left(\bigcap_{i=k}^{k+N-1}\{\Upsilon_i < h\}\right) \geq \left[\mathbb{P}_\infty\left(\Lambda_m < h\right)\right]^N. \tag{4.177}$$

Proof. Let $y_i = \mathrm{LLR}(i)$, then from (4.90), for $n \geq m$ we can write

$$\Upsilon_n = \sum_{i=n-m+1}^{n} y_i = \sum_{i=1}^{n} c_{n-i} y_i, \tag{4.178}$$

with

$$c_i = \begin{cases} 1 & \text{if } 0 \leq i \leq m-1 \\ 0 & \text{if } i \geq m \end{cases}, \tag{4.179}$$

so that Υ_n is written as a monotonically increasing function of $\{y_1, \ldots, y_n\}$ (since $c_i \geq 0$). Therefore, since $y_1, y_2, \ldots$ are iid under $\mathbb{P}_\infty$, from Theorem 5.1 of [Esa67], we have that

$$\mathbb{P}_\infty\left(\bigcap_{i=k}^{k+N-1}\{\Upsilon_i < h\}\right) \geq \prod_{i=k}^{k+N-1}\mathbb{P}_\infty\left(\Upsilon_i < h\right), \tag{4.180}$$

and the inequality (4.177) thus follows from the fact that the distribution of Υ_i, under $\mathbb{P}_\infty$, is the same for any $i \geq m$. □

Next, we aim to prove first another useful result to obtain (4.92) and (4.94); that is,

$$\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{FMA}}(h), m_\alpha\right) = \mathbb{P}_\infty\left(m \leq T_{\mathrm{FMA}}(h) < m + m_\alpha\right). \tag{4.181}$$

To do so, from (4.69), we can write

$$\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{FMA}}(h), m_\alpha) = \sup_{l \geq m}\left\{\sum_{k=l}^{l+m_\alpha-1}\mathbb{P}_\infty\left(T_{\mathrm{FMA}}(h) = k\right)\right\}. \tag{4.182}$$

Let $V_l = \mathbb{P}_\infty(l \leq T_{\mathrm{FMA}}(h) < l + m_\alpha)$, for $l \geq m$, and $U_k = \mathbb{P}_\infty(T_{\mathrm{FMA}}(h) = k)$; then

$$\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{FMA}}(h), m_\alpha) = \sup_{l \geq m}\{V_l\} = \sup_{l \geq m}\left\{\sum_{k=l}^{l+m_\alpha-1} U_k\right\}. \tag{4.183}$$

It is easy to verify, from the definition of $T_{\mathrm{FMA}}(h)$ in (4.90), that

$$U_m = \mathbb{P}_\infty\left(\Lambda_m \geq h\right), \tag{4.184}$$

and

$$\begin{aligned} U_{m+1} &= \mathbb{P}_\infty\left(\{\Lambda_m < h\} \bigcap \{\Upsilon_{m+1} \geq h\}\right) \\ &\leq \mathbb{P}_\infty\left(\Upsilon_{m+1} \geq h\right) \overset{(a)}{=} \mathbb{P}_\infty\left(\Lambda_m \geq h\right) = U_m, \end{aligned} \tag{4.185}$$

where $\overset{(a)}{=}$ follows because Υ_n has the same distribution under $\mathbb{P}_\infty$ for $n \geq m$ (i.e. the distribution

of the sum of m LLRs under $\mathcal{H}_0$). Similarly, for $k > m$, we have

$$\begin{aligned} U_{k+1} &= \mathbb{P}_\infty\left(\bigcap_{n=m}^{k}\{\Upsilon_n < h\}\bigcap\{\Upsilon_{k+1} \geq h\}\right) \leq \mathbb{P}_\infty\left(\bigcap_{n=m+1}^{k}\{\Upsilon_n < h\}\bigcap\{\Upsilon_{k+1} \geq h\}\right) \\ &\overset{(a)}{=} \mathbb{P}_\infty\left(\bigcap_{n=m}^{k-1}\{\Upsilon_n < h\}\bigcap\{\Upsilon_k \geq h\}\right) = U_k, \end{aligned} \tag{4.186}$$

where $\overset{(a)}{=}$ follows because $\mathbb{P}_\infty(\Upsilon_n < h) = \mathbb{P}_\infty(\Upsilon_{n-1} < h)$ for all $n > m$.

Thereby, $\{U_k\}_{k\geq m}$ is a non-increasing sequence, and then

$$V_l - V_{l+1} = \sum_{k=l}^{l+m_\alpha-1} U_k - \sum_{k=l+1}^{l+m_\alpha} U_k = U_l - U_{l+m_\alpha} \geq 0, \tag{4.187}$$

so that $\{V_l\}_{l\geq m}$ is a non-increasing sequence as well. Hence, from (4.183) and the definition of V_l,

$$\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{FMA}}(h), m_\alpha\right) = \sup_{l\geq m}\{V_l\} = \mathbb{P}_\infty\left(m \leq T_{\mathrm{FMA}}(h) < m + m_\alpha\right), \tag{4.188}$$

and (4.181) thus follows. Now, we can proceed with the calculation of $\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{FMA}}(h), m_\alpha)$. However, the exact calculation from (4.188) is complicated to obtain, and then the calculation of an upper bound is proposed instead. From (4.188) and since $T_{\mathrm{FMA}}(h) \geq m$, from its definition in (4.90), we can write

$$\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{FMA}}(h), m_\alpha\right) = 1 - \mathbb{P}_\infty\left(T_{\mathrm{FMA}}(h) \geq m + m_\alpha\right), \tag{4.189}$$

with

$$\mathbb{P}_\infty\left(T_{\mathrm{FMA}}(h) \geq m + m_\alpha\right) = \mathbb{P}_\infty\left(\bigcap_{n=m}^{m+m_\alpha-1}\{\Upsilon_n < h\}\right). \tag{4.190}$$

So, the proof of (4.92) and (4.94) follows by direct application of (4.177) to (4.190).

4.C.2 Probability of unbounded detection $\mathrm{P}_{\mathrm{ud}}(T_{\mathrm{FMA}}(h), m)$

Applying the Bayes rule in (4.68) we have

$$\mathrm{P}_{\mathrm{ud}}(T_{\mathrm{FMA}}(h), m) = \sup_{v>m}\left\{\frac{\mathbb{P}_v\left(T_{\mathrm{FMA}}(h) \geq v + m\right)}{\mathbb{P}_v\left(T_{\mathrm{FMA}}(h) \geq v\right)}\right\} \overset{(a)}{=} \sup_{v>m}\left\{\frac{\mathbb{P}_v\left(\bigcap_{n=m}^{m+v-1}\{\Upsilon_n < h\}\right)}{\mathbb{P}_v\left(\bigcap_{n=m}^{v-1}\{\Upsilon_n < h\}\right)}\right\}, \tag{4.191}$$

where $\overset{(a)}{=}$ follows from the definition of $T_{\mathrm{FMA}}(h)$ in (4.90). Due to the windowed behavior of $T_{\mathrm{FMA}}(h)$ we have assumed that $v > m$. As for $\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{FMA}}(h), m_\alpha)$, the exact calculation of $\mathrm{P}_{\mathrm{ud}}(T_{\mathrm{FMA}}(h), m)$ from (4.191) is quite difficult, and then we propose the derivation of an upper bound. Now, letting the event $\mathcal{A}_n = \{\Upsilon_n < h\}$, with $n \geq m$, it is clear that $\mathcal{A}_{v-1}$ and $\mathcal{A}_{m+v-1}$ are independent because they do not share any samples, thus

$$\mathbb{P}_v\left(\bigcap_{n=m}^{m+v-1}\mathcal{A}_n\right) \leq \mathbb{P}_v\left(\mathcal{A}_{m+v-1}\right)\mathbb{P}_v\left(\bigcap_{n=m}^{v-1}\mathcal{A}_n\right), \tag{4.192}$$

since in the left side we evaluate more events than in the right side. So, applying this result to (4.191) we have that

$$\mathrm{P_{ud}}\left(T_{\mathrm{FMA}}(h), m\right) \leq \sup_{v>m}\left\{\mathbb{P}_v\left(\Upsilon_{m+v-1} < h\right)\right\} \overset{(a)}{=} \mathbb{P}_1\left(\Lambda_m < h\right), \tag{4.193}$$

where $\overset{(a)}{=}$ follows because Υ_{m+v-1} is identically distributed under $\mathbb{P}_v$ for all $v > m$, so that (4.93) and (4.95) thus follow, completing the proof of Theorem 8.

Appendix 4.D Proof of Proposition 9

Let us start with the proof of (4.132). To do so, it is worth noting that

$$\begin{aligned}\mathbb{P}_\infty\left(T_{\mathrm{NP}} \geq N + m_\alpha\right) = \mathbb{P}_\infty\left(\widetilde{T} \geq 1 + \left\lfloor \frac{m_\alpha}{N} \right\rfloor\right) &= \mathbb{P}_\infty\left(\bigcap_{i=1}^{\lceil \frac{m_\alpha}{N} \rceil}\left\{\Psi_i < h\right\}\right) \\ &= \left[\mathbb{P}_\infty\left(\Lambda_N < h\right)\right]^{\lceil \frac{m_\alpha}{N} \rceil},\end{aligned} \tag{4.194}$$

and (4.132) thus follows from similar arguments as given in the proof of Theorem 8. Let us now focus on the proof of (4.133). Since the proof for $N > m$ is trivial, we will prove the case when $1 \leq N \leq m$. To do so, we start with the definition of the PUD

$$\mathrm{P_{ud}}\left(T_{\mathrm{NP}}, m\right) = \sup_{v\geq 1}\left\{\mathbb{P}_v\left(T_{\mathrm{NP}} \geq v + m \,|\, T_{\mathrm{NP}} \geq v\right)\right\} = \sup_{v\geq 1}\left\{\frac{\mathbb{P}_v\left(T_{\mathrm{NP}} \geq v + m\right)}{\mathbb{P}_v\left(T_{\mathrm{NP}} \geq v\right)}\right\}. \tag{4.195}$$

Now, applying the definition of T_{NP} in (4.131) we have

$$\mathrm{P_{ud}}\left(T_{\mathrm{NP}}, m\right) = \sup_{v\geq 1}\left\{\frac{\mathbb{P}_v\left(\bigcap_{i=1}^{\lceil \frac{v+m}{N} - 1 \rceil}\left\{\Psi_i < h\right\}\right)}{\mathbb{P}_v\left(\bigcap_{i=1}^{\lceil \frac{v}{N} - 1 \rceil}\left\{\Psi_i < h\right\}\right)}\right\} = \sup_{v\geq 1}\left\{\prod_{i=\lceil \frac{v}{N} \rceil}^{\lceil \frac{v+m}{N} - 1 \rceil}\mathbb{P}_v\left(\Psi_i < h\right)\right\}, \tag{4.196}$$

where the last equality follows from the fact that Ψ_i is iid. It is important to note that Ψ_i is identically distributed under $\mathbb{P}_v$ and $\mathbb{P}_{jN-v+1}$ $\forall j \in \mathbb{N}$, so the supremum in the previous equation can be evaluated on $1 \leq v \leq N$ because the rest of values ($v > N$) provide similar results. Therefore

$$\mathrm{P_{ud}}\left(T_{\mathrm{NP}}, m\right) = \sup_{1\leq v\leq N}\left\{\prod_{i=1}^{k}\mathbb{P}_v\left(\Psi_i < h\right)\right\} = \sup_{1\leq v\leq N}\left\{A_v\right\}, \tag{4.197}$$

with $k = \left\lfloor \frac{m}{N} \right\rfloor$.

Let us now analyze $A_v = \mathbb{P}_v(\Psi_i < h)$. Firstly, it is worth pointing out that A_v is composed of k products, whose terms are the probabilities that the LRT does not detect in a block (each block consecutive) of N samples. So, all of these terms except one are evaluating samples under $\mathcal{H}_1$ and only the first block evaluates some samples under $\mathcal{H}_0$ (depending on v). Thus, we can write

$$A_v = \mathbb{P}_v\left(\Lambda_N < h\right)\left[\mathbb{P}_1\left(\Lambda_N < h\right)\right]^{k-1}, \tag{4.198}$$

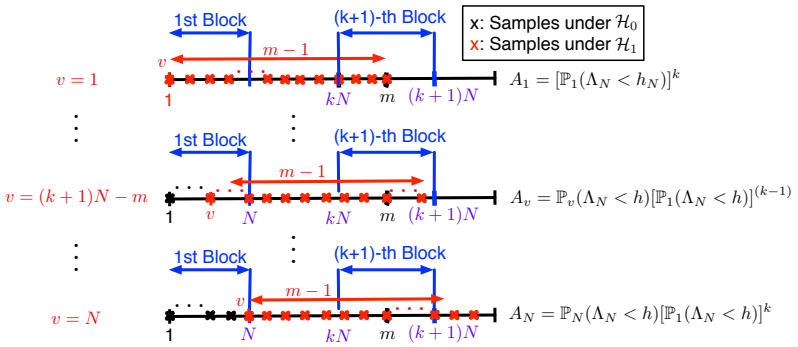

Figure 4.16: Behavior of the term $A_v = \mathbb{P}_v(\Psi_i < h)$ with respect to v.

and then the dependence of A_v with v is given by the first term. This behavior of A_v with respect to v is given in Figure 4.16. We see how for $v = 1$ the first block of samples includes all the samples under $\mathcal{H}_1$, so that $\mathbb{P}_v(\Lambda_N < h) = \mathbb{P}_1(\Lambda_N < h)$. Then, we see that A_v increases as v increases until $v = (k+1)N - m$. This is because as v increases, $\mathbb{P}_v(\Lambda_N < h)$ includes more samples under $\mathcal{H}_0$ within the block of N samples, and then Λ_N is getting smaller or equivalently $\mathbb{P}_v(\Lambda_N < h)$ increases. So, we have that A_v is increasing with v for $1 \leq v \leq (k+1)N - m$. Up to this point, the behavior of A_v changes. This is because (as seen in Figure 4.16) now we are able to evaluate an additional block of samples, with all the samples under $\mathcal{H}_1$ (i.e. k-th block), before the limit of m samples.

With the above considerations, for $(k+1)N - m < v \leq N$, we have

$$A_v = \mathbb{P}_v\left(\Lambda_N < h\right)\left[\mathbb{P}_1\left(\Lambda_N < h\right)\right]^k, \tag{4.199}$$

which is also increasing with v. This is because the larger the value of v, the more samples under $\mathcal{H}_0$ are evaluated by the probability measure $\mathbb{P}_v\left(\Lambda_N < h\right)$, thus being larger. Thereby, A_v is increasing in both intervals $1 \leq v \leq (k+1)N - m$ and $(k+1)N - m < v \leq N$, so that

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{NP}}, m\right) = \sup_{1 \leq v \leq N}\left\{A_v\right\} = \sup\left\{A_{(k+1)N-m}, A_N\right\}. \tag{4.200}$$

Finally, we also have

$$\begin{aligned} A_N &= \mathbb{P}_N\left(\Lambda_N < h\right)\left[\mathbb{P}_1\left(\Lambda_N < h\right)\right]^k \leq \left[\mathbb{P}_1\left(\Lambda_N < h\right)\right]^k \\ &\leq \mathbb{P}_{(k+1)N-m}\left(\Lambda_N < h\right)\left[\mathbb{P}_1\left(\Lambda_N < h\right)\right]^{k-1} = A_{(k+1)N-m}, \end{aligned} \tag{4.201}$$

so that

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{NP}}, m\right) = A_{(k+1)N-m} = \mathbb{P}_{(k+1)N-m}\left(\Lambda_N < h\right)\left[\mathbb{P}_1\left(\Lambda_N < h\right)\right]^{k-1}, \forall N : [1, m], \tag{4.202}$$

and (4.133) thus follows, completing the proof of Proposition 9.

Appendix 4.E Proof of Theorem 10

Let us first consider the case $N > \lceil m/2 \rceil$, for which $k = 1\ \forall N$, thus

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{NP}}\right) = \mathbb{P}_{2N-m}\left(\Lambda_N < h\right), \tag{4.203}$$

which is increasing with N (more samples under $\mathcal{H}_0$ are evaluated). Now, it is worth seeing that for $N = \lceil m/2 \rceil$

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{NP}}\right) = \mathbb{P}_N\left(\Lambda_N < h\right)\mathbb{P}_1\left(\Lambda_N < h\right) \leq \mathbb{P}_1\left(\Lambda_N < h\right) \leq \mathbb{P}_{2N-m}\left(\Lambda_N < h\right), \tag{4.204}$$

so that we have only to focus on $N \leq \lceil m/2 \rceil$ in order to prove Theorem 10. Next, we give an sketch of the proof to find the optimal sample size $N^* = \lceil m/2 \rceil$. It is important to analyze the behavior of $\mathrm{P}_{\mathrm{ud}}(T_{\mathrm{NP}})$ with respect to the sample size N. The key point is to see that for $\lceil m/(i+1) \rceil < N \leq \lceil m/i \rceil$ we have the same value of k and then

$$\mathrm{P}_{\mathrm{ud}} = \mathbb{P}_{(k+1)N-m}\left(\Lambda_N < h\right)\left[\mathbb{P}_1\left(\Lambda_N < h\right)\right]^{k-1}. \tag{4.205}$$

Now, we have to prove that the previous expression gets its minimum for $N = \lceil m/i \rceil$. The following step is to prove that among all i, the minimum P_{ud} is obtained for $i = 2$. To obtain these proofs we have to analyze the decreasing exponent of the probabilities in (4.205). This is a sketch to be followed in order to prove the theorem. Notwithstanding, work is still needed to obtain all the needed proofs. This can be done investigating the ideas in [Cov06, Pol10, Bla74].

Chapter 5

GNSS Signal-Level Integrity Metrics for SCD

5.1 Introduction

With the ubiquitous use of GNSSs in SCA and LCA, one of the major challenges to be solved is the provision of integrity to different types of users beyond the civil aviation community, where this feature is already well established. In fact, as commented in previous chapters, the concept of integrity in GNSS was first proposed under the umbrella of civil aviation. Both augmentation and RAIM-based integrity algorithms were developed to guarantee that GNSS operate within the given performance requirement in any phase of flight. However, such methods are generally not suited to provide integrity in road and urban environments due to the predominance of local degradation effects such as multipath, fading, NLOS propagation, and interference signals. This is the reason why integrity analyses on the received signal quality, henceforth referred to as *Signal-Level Integrity (SLI)*, should be considered, thus truly contributing to the subsequent provision of PVT integrity. It is worth noting that this has not been traditionally the case in civil aviation applications, where it has been assumed that local effects have a limited and controlled influence on the received signal. Actually, mitigation techniques have attracted the attention of many researchers during the past years [Bro16b, Cal01, Bhu07, Bra01], leaving the detection of these threats in a secondary place, when indeed it is even more important than mitigation, especially for NLOS. The reason is that before using mitigation techniques, the user can benefit from knowing whether these threats are present or not. For instance, we can discard the current measurements [Mub10], thus not requiring complicated mitigation techniques.

In the past years, different contributions to provide GNSS integrity in terrestrial environments have appeared. Unfortunately, current contributions are based on the use of external information like map-matching to identify local threats (e.g. known interference sources or signal blocking obstacles) [TM10], external sensors to obtain redundant information [Bha09], or fisheye cameras to obtain a sky plot and determine the geometrical distribution of satellites in view [Shy14]. Nonetheless, the use of external aid needs prior information about the user environment or external hardware, which is not always available in mass-market GNSS receivers (i.e. those mostly used in terrestrial environments). On the other hand, so far, contributions on local threat detection have been addressed either adopting a classical detection framework

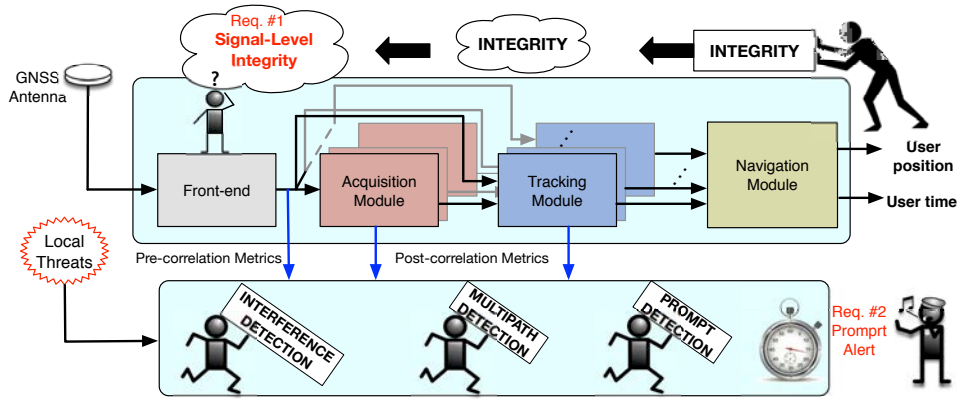

Figure 5.1: SLI philosophy scheme: Move the integrity module to the very first stages of a GNSS receiver and apply SCD so that a quickest detection is obtained.

or working at observable level (i.e. pseudoranges, PVT, ...) [Bal09, Lee04, Bro12], which is not well suited to fulfill the requirements of critical applications. The reason is that classical detection performance metrics are typically in the form of PMD and PFA, thus disregarding the temporal dimension. Notwithstanding, the temporal dimension is indeed of paramount importance for integrity applications, since a prompt detection is needed. It is for the above reasons that in this chapter we take a leap forward in the field of GNSS local threats detection, and we propose the adoption of a QCD framework, where incoming measurements are assumed to exhibit a sudden change in either their statistical parameters (e.g. mean, variance) or even in the type of its pdf (e.g. from Gaussian to Uniform). This approach fits very well into the kind of threats that a GNSS receiver may experience in real life, where sudden changes in the properties of the received signal are often common. In summary, two requirements are needed (see Figure 5.1):

1. Because of the presence of local threats, integrity must be analyzed at a *local* scale too. To do so, integrity checks have to be carried out at the receiver level, and preferably, at the received *signal* level.

2. Moreover, we want to detect these threats *as soon as possible*, in such a way that *a prompt alert* can be raised to the user. To do so, SCD techniques must be implemented, so that time is explicitly targeted, as required by SCA and LCA.

The first requirement is obtained by the use of metrics at signal level, whereas the second one is obtained by applying SCD to these observable metrics.

5.1.1 Signal-Level Metrics (SLM)

The first step to provide SLI to GNSS users is to define metrics at signal level that provide information about the presence or absence of local effects. In this dissertation we refer to these metrics as Signal-Level Metrics (SLM) and they will be introduced in this chapter. In particular, in this thesis we will focus on multipath, NLOS propagation, and interference, so we provide different SLM that react to these effects. A first classification of these metrics can be done in terms of the stage at which the metric is computed. This is equivalent to the particular effect to be measured (i.e. interference or multipath):

- **Pre-correlation metrics**: Measured just before the acquisition and tracking module of a GNSS receiver; that is, they are obtained from the signal samples at the output of the front-end. It is here where the presence of interference is measurable, because after the de-spreading process the interference effects may not be visible.
- **Post-correlation metrics**: Obtained after the de-spreading process; that is, in the acquisition or tracking module of a GNSS receiver. In this case, these metrics are intended to deal with multipath. Contrary to the interference case, multipath is only visible after the correlation process, because otherwise the multipath component is not visible due to the spread spectrum properties of GNSS signals.

Based on the effects of local threats into the GNSS signal we will introduce different SLM that can be used to detect both interferences and multipath. Specifically, in Section 5.2 and Section 5.3 we study the above mentioned SLM (i.e. pre- and post-correlation metrics). Then, Section 5.4 is intended to introduce some SLM based on multi-antenna GNSS receivers.

5.1.2 Application of QCD

As shown in Chapter 4, the first step to design a detector is to have observations X_n that experience a change in their statistical behavior under the presence of an event. In the rest of this dissertation we will assume these observations come from the considered pre- and post-correlation SLM. That is to say, we suppose that the change in distribution of the SLMs will appear with the presence of some of the local threats considered. Then, after introducing the SLMs considered in this thesis, the rest of this chapter is devoted to show the application of QCD to these SLM. This is a novel contribution because while the QCD framework has been extensively applied into a wide range of fields, it has barely been applied in the GNSS domain. This contribution is worthy to provide signal quality monitoring in a QCD framework, so that a prompt alert can be provided. Notwithstanding, for a complete integrity algorithm TCD must be considered. This will be the goal of Chapter 6. Regarding the QCD application, we addressed in previous publications the problem of detecting local degrading effects in a QCD framework. Particularly, we considered the case of multi-antenna GNSS receivers (see [Ege14a, Ege14b]), and single-antenna receivers in [ER15b] and [ER15a]. Finally, a comprehensive overview of interference and multipath QCD was provided in [ER15c] and [ER15d], respectively. We will base on these references to show the application of QCD to interference and multipath detection. Specifically, Section 5.2 and Section 5.3 deal with quickest interference and multipath detection, respectively. Finally, Section 5.5 concludes the chapter. It is worth mentioning that single-antenna receivers will be considered in this chapter, except in Section 5.4, in which we will introduce some multi-antenna metric.

5.2 Quickest Interference Detection

The very low received power of GNSS signals, typically 20 dB below the noise floor, makes them very sensitive to the presence of either intentional jamming or unintentional RFI. Jamming is usually driven by two reasons, on the one hand they try to prevent the use of GNSS to a third party, as it is the case of the GNSS outages in South Korea since 2010 generated by North

Korea [Seo13]. On the other hand, jammers are sometimes aimed at preventing people and vehicles from being tracked. This is the case at the Newark, NJ, USA, airport, which was lead to GPS disruption due to a truck driver using a so-called personal privacy device [Pul12]. Due to the increasing dependence of our modern society on GNSS technology, jamming is nowadays one of the main concerns for the GNSS community [Ami16]. This is because the disruption of GNSS services (main consequence of jamming) can have a high economic impact and cause a major setback worldwide. For the above discussion, it becomes evident the need of SLMs to detect the large amount of possible interferences in GNSS. We focus here on the detection of interference under the framework of QCD. In particular, the discussion presented in this section is mainly based on our work published in [ER15c] and references therein. To reach the goal of promptly detecting the presence of GNSS interference, we first need to statistically characterize interference metrics in order to be fed into a CUSUM-type algorithm. The statistical characterization and its use in some CUSUM algorithm variants, will be justified by means of simulation results.

Let us first define the signal model for the interference detection case. In the absence of interference (i.e. $\mathcal{H}_0$), the received signal will be dominated by noise, since the GNSS signal remains under the noise floor. This is so until the unknown time v in which an interference appears (i.e. $\mathcal{H}_1$), and then, the received signal will be dominated by the interference itself. So, in the rest of this dissertation, we consider the null hypothesis as the case when threat is not present, whereas the alternative hypothesis is considered to be the case when the threat is present. The detection problem thus becomes:

$$x(n) = \begin{cases} \mathcal{H}_0 : w(n) & n < v \\ \mathcal{H}_1 : i(n) + w(n) & n \geq v \end{cases}, \tag{5.1}$$

where $x(n)$ denotes the GNSS discrete-time base-band sample at time n, $i(n)$ models an interference signal impinging into the GNSS receiver, and $w(n)$ is the thermal noise disturbing the received samples, which can be modeled as an iid zero-mean Gaussian random process with variance σ_w^2. We assume that the interference can be modeled as a random process, but the statistical characteristics are completely unknown, which is the most realistic case that can be addressed. For this reason we will use several metrics, based on $x(n)$, for the detection process. The key point of using these metrics is that they can be statistically characterized regardless the distribution of the interference $i(n)$.

The model in (5.1) is only valid when the Automatic Gain Control (AGC) of the front-end is disabled (i.e. fixed gain) and the dynamic range of the ADC is large enough to resemble the Gaussian distribution. If the AGC is not disabled (i.e. variable gain), we could use the gain indicator evolution instead of the received samples. Henceforth we assume the AGC is disabled and the ADC has enough resolution. Different pre-correlation metrics can be used to detect the presence of interference. In this work, we restrict ourselves to the so-called statistical analysis and Time-Frequency Analysis (TFA) metrics. The former are based on the statistical properties of the samples at the GNSS front-end output, while the later are based on the time and frequency characteristics of the received samples. Both kinds of pre-correlation metrics and their detection algorithm design are introduced next. Specifically, we will consider three different statistical metrics, namely the histogram, the kurtosis, and the Auto-Correlation Function (ACF) of the front-end output samples; and two different TFA metrics, the spectrogram and received power.

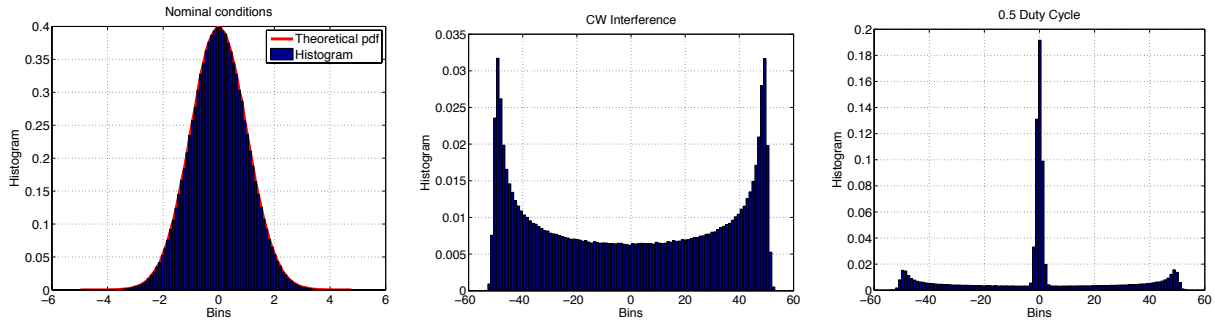

Figure 5.2: (Left) Histogram in the absence of interference, (middle) with a CW interference, and (right) a PW interference with 0.5 duty cycle.

5.2.1 Histogram detector

In nominal conditions, the signal samples at the GNSS front-end output are dominated by thermal noise, and thus, they follow a Gaussian distribution. Based on this observation, a possible metric to measure the presence of interference is the empirical pdf. Comparing the empirical pdf with the Gaussian expected one, it is possible to infer the presence of interference. Moreover, under the presence of interference, depending on the resulting shape of the empirical pdf we might infer the type and properties of the interfering signal. The empirical pdf can be obtained through the histogram, which indicates the rate of appearance of the received samples values, providing an estimate of the pdf of the received samples. Figure 5.2 shows the normalized histogram under different situations. We see the matching between the histogram shape and the theoretical Gaussian pdf in the absence of interference. On the other hand, when interference is present, the shape of the normalized histogram deviates from a Gaussian one. So, according to these results, the histogram provides graphical information on whether an interference is present or not and on its type. This highlights the capability of the histogram to detect and characterize interference sources within the received GNSS signal. Actually, the histogram-based detection problem is equivalent to a goodness-of-fit test (GoF), that is, we are interested on determining whether our received signal samples follow a Gaussian distribution or not. To do so, we will make use of the so-called chi-square test, based on the histogram. An advantage of this test is that it does not require any a-priori information about the interfering source, so that it is applicable to all types of interferences. In addition, it has a low computational burden because it is based on histogram calculations, which are simple operations.

Statistical characterization

The chi-square test evaluates the following test statistic [D'A86]:

$$X_{\text{hist},m} = \sum_{i=1}^{N_b} \frac{(O_{i,m} - E_i)^2}{E_i}, \tag{5.2}$$

with E_i the i-th bin of the reference *theoretical* histogram evaluated under $\mathcal{H}_0$ with N_b bins. $O_{i,m}$ is the i-th bin of the measured histogram at snapshot m with N_b bins, where each snapshot contains N samples. The reference histogram, E_i for $i = 1, \ldots, N_b$, may be computed once and stored in the receiver memory, or it might be eventually updated to take into account possible variations in the receiver conditions. Pearson [D'A86] claimed that for large N, the random variable $X_{\text{hist},m}$ is approximately chi-squared distributed with $N_b - 1$ degrees of freedom. Under

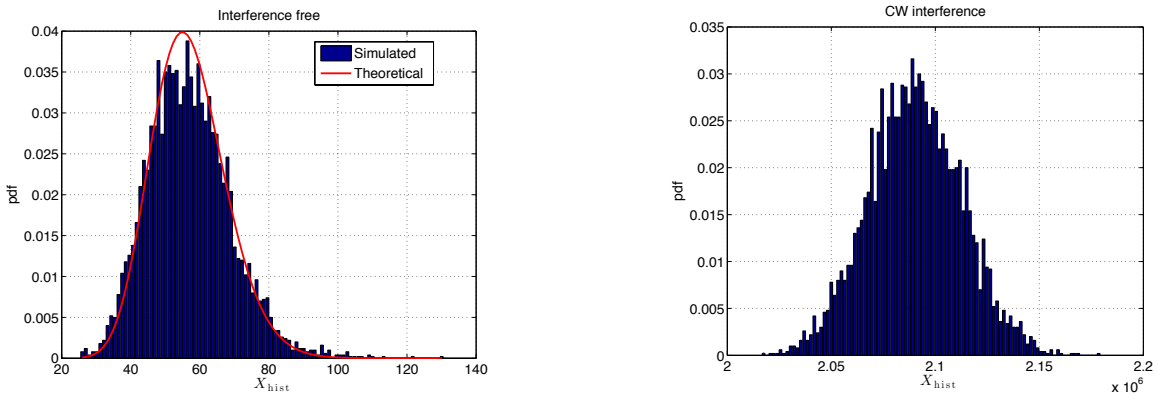

Figure 5.3: Statistical characterization of the chi-squared GoF test metric with a number of $N = 10^4$ samples and using $N_{\text{b}} = 50$ bins. (Left) Comparison between simulated (i.e. histogram) and theoretical pdf under $\mathcal{H}_0$, and (right) simulated pdf under $\mathcal{H}_1$ with a CW interference of INR = 20 dB.

$\mathcal{H}_1$ it departs from a central chi-squared distribution thus resulting in

$$\begin{aligned} \mathcal{H}_0 &: X_{\text{hist},m} \sim \chi^2_{N_b-1} \;\; m < v \\ \mathcal{H}_1 &: X_{\text{hist},m} \nsim \chi^2_{N_b-1} \;\; m \geq v \end{aligned} . \tag{5.3}$$

Figure 5.3 shows this behavior; that is, under $\mathcal{H}_0$ (see left plot) the histogram of the simulated data almost fits the theoretical χ^2 with $N_b - 1$ degrees of freedom. The right plot shows the histogram for the case when the simulated interference is present, departing from that obtained under $\mathcal{H}_0$.

Application of QCD: CUSUM for the Histogram metric (CUSUM-Hist)

So far, we have seen that the distribution of the histogram-based SLM under $\mathcal{H}_1$ is different to that under $\mathcal{H}_0$, but we have no knowledge about the former. Hence, since the distribution under $\mathcal{H}_1$ is unknown, the LLR cannot be completely defined, and then we are unable to apply the CUSUM algorithm directly to $X_{\text{hist},m}$. Notwithstanding, since $\mathcal{H}_0$ is known, we can use the O-CUSUM alternative (see Section 4.3.5). To do so we need to propose a metric ρ that has a negative mean before the change and a positive mean after the change. It is known that the mean of a χ^2 random variable is equal to the number of degrees of freedom of the χ^2. Therefore, $\text{E}_\infty[X_{\text{hist},m}] = N_b - 1 > 0$. Thus, we cannot directly apply $X_{\text{hist},m}$ as the function ρ to be used in the O-CUSUM, but we define the following modified function

$$\rho_{\text{hist}}(m) \doteq X_{\text{hist},m} - b, \tag{5.4}$$

with b a proper offset for which the mean of $\rho_{\text{hist}}(m)$ under $\mathcal{H}_0$ is negative, but it is positive under $\mathcal{H}_1$. In fact, the choice of the offset b should be large enough to ensure a negative mean before change and to provide a certain false alarm rate, always maintaining a positive mean after change. From (4.64) and using $\rho_{\text{hist}}(m)$ we are able to adjust the false alarm rate through the non-zero root ω_0 with (4.66), which turns out to be the non-zero root of

$$e^{\omega b} = (1 - 2\omega)^{-\frac{N_b-1}{2}} , \tag{5.5}$$

which can be solved numerically. Thus, the choice of the offset b will fix a value for ω_0 given by the equation above, then making use of the henceforth referred to as CUSUM-Hist decision

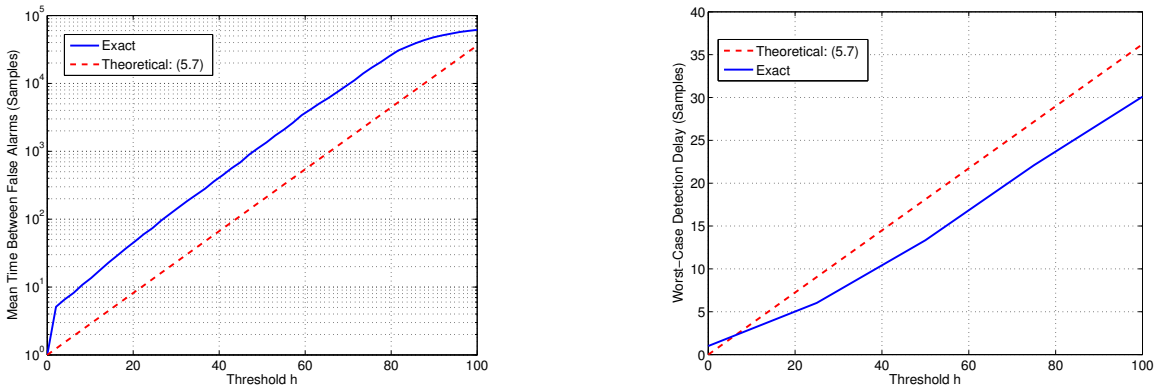

Figure 5.4: CUSUM-Hist performance with a fixed $N_b = 50$, $N = 10^4$, and $b = 55$. (Left) Mean time between false alarms, and (right) detection delay for a CW with INR $= -20$ dB.

rule, given by

$$g_m^{(\text{hist})} = \left(g_{m-1}^{(\text{hist})} + \rho_{\text{hist}}(m)\right)^+ \geq h, \tag{5.6}$$

we obtain the following performance

$$\mathfrak{N}_{\text{hist}} \geq e^{\omega_0 h} \text{ and } \mathfrak{D}_{\text{hist}} \leq \frac{h}{\text{E}_1\left[\rho_{\text{hist}}(m)\right]}, \tag{5.7}$$

where ω_0 stands for the non-zero solution of (5.5). These bounds are illustrated in Figure 5.4, which shows the CUSUM-Hist performance using $\rho_{\text{hist}}(m)$, and compares it with the theoretical bounds above. The left plot shows how $\mathfrak{N}_{\text{hist}}$ is greater than the lower bound given in (5.7). This observation allows us to set a threshold h guaranteeing certain desired false alarm rate. In addition, the right plot shows the detection delay measured in samples with respect to the set threshold, and shows similar values for the simulated results and theoretical ones, being the simulated delay below the upper bound given in (5.7).

5.2.2 Kurtosis detector

The kurtosis of a random variable X is defined as

$$R \doteq \frac{\xi_{X,4}}{\xi_{X,2}^2}, \tag{5.8}$$

where $\xi_{X,n}$ is the n-th central moment of the random variable X, which for a Gaussian random variable leads to

$$\xi_{X,n} = 1 \cdot 3 \cdots (n-1)\sigma^n, \tag{5.9}$$

with σ the standard deviation of X and n an even integer, since for n odd $\xi_{X,n} = 0$. Thus, for a Gaussian variable and applying (5.9), the kurtosis value is equal to 3 regardless of the value of σ. Equivalently, for a received GNSS signal without interference, the kurtosis value is equal to 3. In case the received signal is corrupted by some interference signal, the distribution of a received GNSS signal is likely to deviate from a Gaussian distribution, and the value of the kurtosis may then deviate from 3. Hence, the presence of interference can be inferred by evaluating whether the kurtosis deviates form 3. Actually, the kurtosis is measured from a random variable, so that its measured value is a random variable as well that can be characterized by its expected value

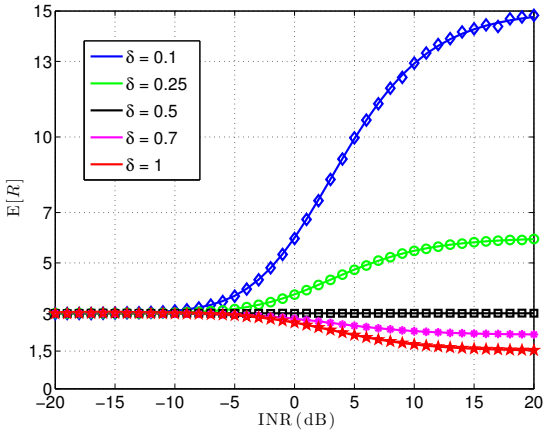

Figure 5.5: (Markers) Simulated and (lines) theoretical kurtosis values for different INR and duty cycles.

given in [De 07] as

$$\mathrm{E}[R] = 3 \cdot \frac{\left(1 + 2\mathrm{INR} + \frac{1}{2\delta}\mathrm{INR}^2\right)}{(1 + \mathrm{INR})^2}, \tag{5.10}$$

where a pulsed sinusoidal is assumed, INR stands for the Interference-to-Noise power Ratio (INR), and δ is the interference duty cycle. In the absence of RFI (i.e INR $\to 0$) $\mathrm{E}[R] = 3$, as expected.

When pulsed sinusoidal RFI is present, the kurtosis measure behaves as follows:

- In the CW limit ($\delta = 1$), $\mathrm{E}[R]$ ranges from 3 to 1.5 as the INR increases.
- As $\delta \to 0$ and INR increases $\mathrm{E}[R] > 3$.
- For $\delta = 0.5$, $\mathrm{E}[R] = 3$ as well, generating a blind spot at 0.5 duty cycle.

This behavior is depicted in Figure 5.5, showing how for low levels of interference (low INR) the kurtosis takes the theoretical value of 3, corresponding to the Gaussian distribution obtained in absence of RFI. As the presence of RFI increases, represented by the increasing of INR, the kurtosis deviates from the expected value thus allowing the detection of interferences. It is worth noting that depending on the duty cycle, the kurtosis value deviates below or above 3, ranging in the limit of CW ($d = 1$) from 3 to 1.5. It can also be seen the blind spot at 0.5 duty cycle, where the kurtosis value becomes 3 like in the absence of RFI in spite of a pulsed interference is present. These results prove the capability of the kurtosis value to detect the presence of pulsed interferences and the ability to extract information about the duty cycle, depending on whether the kurtosis deviates above or below the reference value of 3. To overcome the blind spot for duty cycles of 50%, the possibility of using statistics of higher order is investigated in [De 10].

Statistical characterization

Let $\widehat{R}_m$ be the N-samples estimation of the kurtosis value, thus [D'A86, Ch. 9] states that for large N (i.e. $N \gg 1000$) the kurtosis estimate of a Gaussian variable is another Gaussian variable with mean and variance respectively given by

$$\mu_{R,0} = \mathrm{E}_\infty\left[\widehat{R}_m\right] = 3 \cdot \frac{N-1}{N+1}, \tag{5.11}$$

$$\sigma^2_{R,0} = \mathrm{var}_\infty\left[\widehat{R}_m\right] = \frac{24}{N}. \tag{5.12}$$

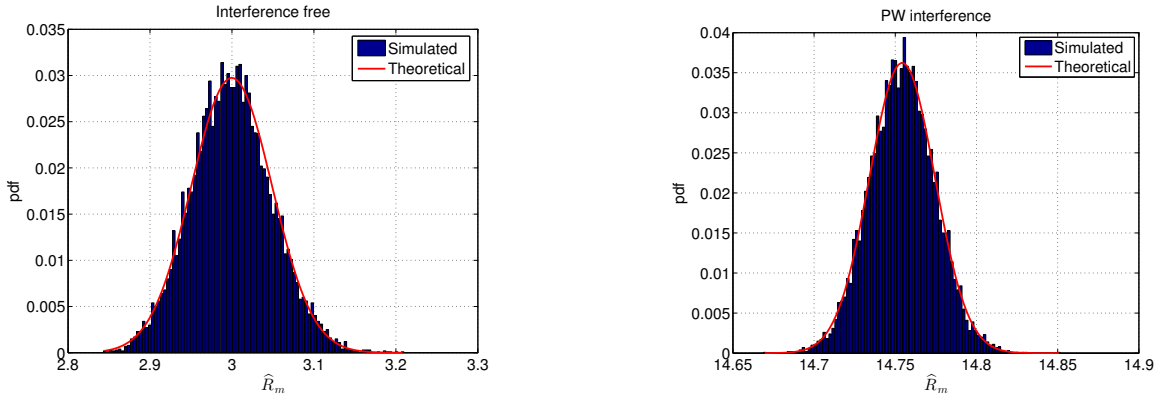

Figure 5.6: Statistical characterization of the kurtosis value with $N = 10^4$. Comparison between simulated (i.e. histogram) and theoretical pdf under (left) $\mathcal{H}_0$ and (right) under $\mathcal{H}_1$ with $\delta = 0.1$ and INR = 20 dB.

Moreover, in the presence of interference, the kurtosis estimates are still Gaussian distributed but suffering from a significant change in the mean, given as in (5.10) by

$$\mu_{R,1} = \mathrm{E}_1\left[\widehat{R}_m\right] = \mu_{R,0} \cdot \frac{1 + 2\mathrm{INR} + \frac{\mathrm{INR}^2}{2\delta}}{(1+\mathrm{INR})^2}. \tag{5.13}$$

Thereby, we can formulate the kurtosis-based detection in a QCD framework as

$$\widehat{R}_m \sim \begin{cases} \mathcal{H}_0 : \mathcal{N}\left(\mu_{R,0}, \sigma^2_{R,0}\right), & m < v \\ \mathcal{H}_1 : \mathcal{N}\left(\mu_{R,1}, \sigma^2_{R,0}\right), & m \geq v \end{cases}. \tag{5.14}$$

This is shown in Figure 5.6, where the distribution of the kurtosis is illustrated for both hypotheses and we can see how its value departs from 3 when an interference signal is present.

Application of QCD: CUSUM for the kurtosis metric (CUSUM-Kurt)

From (5.14) we have characterized the statistical behavior of the kurtosis, with $\mu_{R,0}$, $\sigma^2_{R,0}$ and $\mu_{R,1}$ known. It is worth saying that we have considered the same variance under both $\mathcal{H}_0$ and $\mathcal{H}_1$. Nevertheless, actually there is a slight variation of the variance of the kurtosis with the presence of interference. This variation is negligible compared to the change in the mean of the kurtosis, so that we can fairly assume that both variances before and after the change are equal, and then use the CUSUM algorithm as a Gaussian mean change detector. Doing so, we obtain the following LLR:

$$\mathrm{LLR}_{\mathrm{kurt}}(m) = \frac{\mu_{R,1} - \mu_{R,0}}{\sigma^2_{R,0}} \left(\widehat{R}_m - \frac{\mu_{R,1} + \mu_{R,0}}{2}\right), \tag{5.15}$$

with $\mu_{R,0}$, $\sigma^2_{R,0}$ and $\mu_{R,1}$ given by (5.11)–(5.13), and $\widehat{R}_m$ the N-sample kurtosis estimate at snapshot m. It is important to say that $\mu_{R,1}$ is known to depend on the INR and duty cycle of the interference. Hence, a way to proceed is to fix a certain value for $\mu_{R,1}$ according to the minimum INR that one expects to detect. Moreover, the duty cycle should be fixed to the value that produces the minimum change possible so that any larger change will easily be detected. Thereby, the kurtosis-based CUSUM decision rule, henceforth referred to as CUSUM-Kurt, is given by

$$g_{\mathrm{kurt},m} = \left(g_{\mathrm{kurt},m-1} + \mathrm{LLR}_{\mathrm{kurt}}(m)\right)^+ \geq h, \tag{5.16}$$

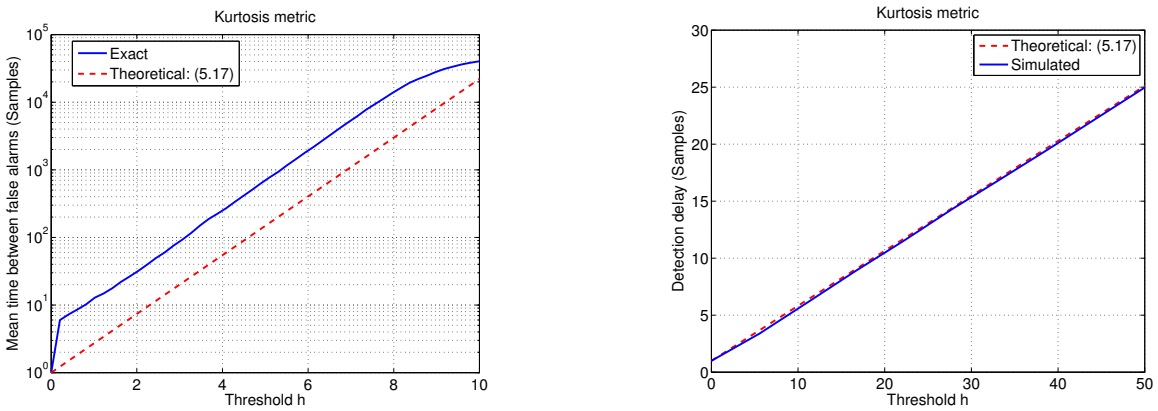

Figure 5.7: CUSUM-Kurt performance with $N = 10^4$. (Left) Mean time between false alarms in samples, and (right) detection delay for a PW with INR $= -10$ dB.

leading to the following performance in terms of false alarms and detection delay

$$\mathfrak{N}_{\text{kurt}} \geq e^h \text{ and } \mathfrak{D}_{\text{kurt}} \leq \frac{h}{d_{\text{kurt}}}, \tag{5.17}$$

with $d_{\text{kurt}} \doteq \mathrm{D}_{\text{KL}}(f_1, f_0) = (\mu_{R,1} - \mu_{R,0})^2/(2\sigma^2_{R,0})$ the KL divergence between f_1 and f_0 for the kurtosis value. These bounds are illustrated in Figure 5.7, which are compared with simulated results. The left plot shows that the simulated number of samples between false alarms is larger than the lower bound, which allows us to set a threshold h assuring certain desired false alarm rate. Moreover, the right plot confirms the match between simulated and theoretical results, even with the assumption of constant variance.

GNSS considerations

Once the kurtosis have been characterized, we can say that its mean value under the presence of interference will depend on the duty cycle (see Figure 5.5). After a deep review of pulsed interferences in GNSS, it is seen that we can expect duty cycles between 0.0014–0.33, so that realistic duty cycles in practice are always below 0.5. Thereby, the kurtosis in presence of PW will be larger than 3 in such a way that, for a fixed INR, the lower duty cycle, the larger the kurtosis value. With this behavior, $\mu_{R,1}$ should be selected as the mean obtained with the largest duty cycle, fixing a minimum change value that will allow to detect lower duty cycles, which is our case of interest. This is for PW interferences. On the other hand, we also have to detect CW interferences, leading to a kurtosis of 1.5. Therefore, two different CUSUMs are needed, one to detect an increase in the mean of the kurtosis, and another to detect a reduction. Moreover, in order to have all the parameters tuned, we need to fix the value of $\mu_{R,0}$ and $\sigma^2_{R,0}$. We know from (5.11), that these parameters only depend on the number of samples N that are used to calculate the kurtosis estimates at every snapshot. Thus, once the snapshot time and sampling rate are fixed, both $\mu_{R,0}$ and $\sigma^2_{R,0}$ can be fixed using the corresponding expressions.

5.2.3 Auto-Correlation Function (ACF) detector

Another statistical measure that can be used to infer the presence of interference is the Auto-Correlation Function (ACF) of the received GNSS signal. This measure provides valuable information on the underlying structure of the received signal, concretely the ACF provides information about the statistical relationship between the received samples. Let us denote, for

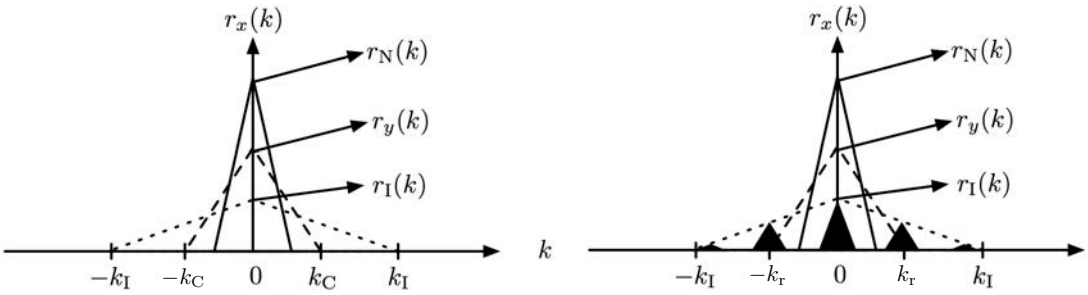

Figure 5.8: ACF of a received GNSS signal corrupted by noise and (left) affected by a CW interference or (right) a pulsed narrow-band interference.

the sake of notation simplicity, the discrete-time samples at the GNSS front-end output as $x(n)$ instead of $r_{\mathrm{GNSS}}(n)$. Hence, based on the theoretical definition of the discrete-time ACF given by $r_x(k) \doteq \mathrm{E}\left[x(n+k)x^*(n)\right]$, an estimate of the ACF can be obtained as

$$\widehat{r}_x(k) = \frac{1}{N}\sum_{n=1}^{N} x(n+k)x^*(n), \tag{5.18}$$

where N is the number of samples within the observation interval to be processed. The analysis of the ACF estimate takes into account the very well-known premise that in absence of any interference, the received samples of a GNSS receiver are dominated by the presence of thermal noise, whose samples exhibit a very narrow ACF. Ideally, this ACF should be a delta function. In practice, though, the presence of the front-end input filter in GNSS receivers will shape the ACF in such a way that a more or less narrow response (instead of a perfect delta) will be obtained. Taking into account this ACF as a reference, the presence of any other signal impinging onto the receiver will make additional shapes appear in the overall ACF of the received samples.

This effect is depicted in Figure 5.8, in which apart from the ACF of the noise, additional ACFs are visible because of the presence of the desired GNSS signal and narrow-band interference. The GNSS signal ACF is denoted by $r_y(k)$, whereas the interference ACF is denoted by $r_I(k)$. Moreover, the smaller the bandwidth of these additional signals, the wider the corresponding ACF, and thus the easier to detect their presence. This behavior certainly provides some hints to implement interference detection based on the observation of the ACF. The left plot of Figure 5.8 shows how the interference ACF ($r_{\mathrm{I}}(k)$) is extended between $-k_{\mathrm{I}}$ and k_{I}, being k_{I} the number of processed signal samples. In the case of the desired GNSS signal ($r_y(k)$), the ACF is extended between $\pm k_{\mathrm{C}}$, whose value is inversely proportional to the receiver filter bandwidth. Thus, with these time hints it is possible to extract some information about the relative bandwidth between the desired GNSS signal and the received interference signal. In the right plot of Figure 5.8 instead of having a peak extended over $\pm k_{\mathrm{I}}$, more than one peaks appear with a lower wide than in the CW case and repeating itself every k_{r} lags. This type of ACF is produced by cyclo-stationary random processes, as it is the case of pulsed interferences [Pro07]. This kind of ACF contains relatively large peaks at lags $k = \pm i k_{\mathrm{r}}, \forall i = 0, 1, 2, \ldots$, with k_{r} the periodicity period in samples. Again, these time hints can be used to detect and characterize pulsed interferences.

Statistical characterization

Based on the definition of the ACF in (5.18), we define the ACF metric as

$$X_{\text{acf},m} \doteq \inf\left\{k : \frac{|\widehat{r}_x(k;m)|}{\max_n\{|\widehat{r}_x(n;m)|\}} \geq \frac{1}{e}\right\}, \tag{5.19}$$

where $m = 1, 2, \ldots$ stands for the snapshot index, with each snapshot including N samples of the received signal, and k is the lag value. This expression is equivalent to finding the first lag for which the ACF maximum has decreased by a factor e. An important property of this definition is that $X_{\text{acf},m}$ does not depend on the noise power, nor on the number of snapshot samples, but on the bandwidth of the dominant signals. That is, when no interference is present, the received samples are dominated by thermal noise and then the width of the ACF depends on the bandwidth of the front-end filter. However, when some interference is present, the ACF width is dominated by the bandwidth of the interference, which is typically smaller than the front-end bandwidth, thus leading to a wider ACF. In general, the ACF width metric will remain constant in both hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, but it may slightly vary (one lag up or down) due to possible round-off errors when solving (5.19). Nevertheless, we can fairly model the ACF width metric as a Gaussian random variable with a small variance to account for the possible round-off errors. Hence, we can model the ACF width metric as

$$X_{\text{acf},m} \sim \begin{cases} \mathcal{N}\left(\mu_{\text{acf},0}, \sigma^2_{\text{acf},0}\right) & m < v \\ \mathcal{N}\left(\mu_{\text{acf},1}, \sigma^2_{\text{acf},0}\right) & m \geq v \end{cases}, \tag{5.20}$$

with $\sigma^2_{\text{acf},0} = 10^{-12}$, $\mu_{\text{acf},0}$ and $\mu_{\text{acf},1}$ the value of the ACF width before and after the change, respectively. As we have already said, the ACF width before and after the change is fixed by the front-end and interference bandwidth, respectively. Specifically, the ACF width is the inverse of the front-end bandwidth, and then, since the bandwidth for CW and PW interferences will typically be much smaller, the $\mu_{\text{acf},1}$ will be greater than $\mu_{\text{acf},0}$.

Application of QCD: CUSUM for the ACF metric (CUSUM-ACF)

As for the metrics investigated previously, using (5.20) we characterize the statistical behavior of the ACF width under the QCD framework, with $\mu_{\text{acf},0}$ known (inverse of the filter bandwidth), $\sigma^2_{\text{acf},0} = 10^{-12}$, and $\mu_{\text{acf},1}$ dependent upon the INR. Thus, a way to proceed is to fix a certain value for the mean after change according to the minimum INR that one expects to detect. In this way, a minimum change detection is set allowing the detection of any larger change caused by higher power interferences. Therefore, we can obtain the exact LLR for a Gaussian mean change given in (5.15), but with the ACF metric parameters. Thus, applying the CUSUM algorithms we have the henceforth referred to as CUSUM-ACF detection rule, leading to similar performance as for the kurtosis given in (5.17) but using the ACF parameters.

GNSS considerations

Among all the possible interferences in GNSS, the ACF width will be able to detect CW and PW interference, including chirp interferences. This is so because these are the types of interference that result in an apparent change in the ACF width, with respect to the free-interference case. For wide-band interferences, the width of the ACF will be fixed by the front-end filter bandwidth, and then the ACF width will not change with respect nominal conditions. After analyzing the ACF behavior (see Figure 5.8), we get some insight on how computing X_{acf}. A proper way to calculate the ACF is to only compute a moderate number of lags, given by the fixed maximum lag, $k_{\max}$, and separation between lags, Δk. Thus, a possible configuration may be

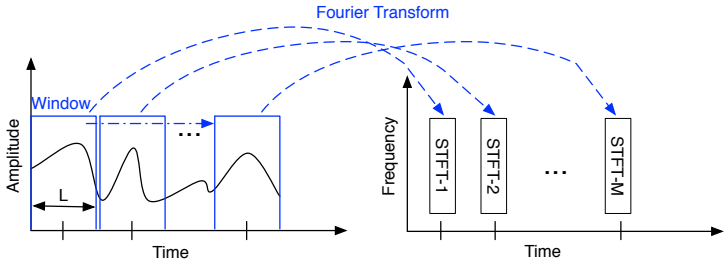

Figure 5.9: Spectrogram construction: The signal is divided in segments that are Fourier transformed in order to obtain the frequency description in a particular time.

- To fix the maximum lag as $T_{\mathrm{p}} < k_{\max} < 1/\mathrm{PRF}$, with T_{p} the pulse width and PRF the pulse repetition frequency.
- To fix the separation between lags as $\Delta k < T_{\mathrm{p}}$.

Both T_{p} and PRF should be fixed accordingly to the characteristics of typical pulsed interferences in GNSS. After a deep overview, we conclude that proper parameters may be to fix $k_{\max} = 1$ ms and $\Delta k = 20$ μs. With this configuration we would be able to detect almost any kind of pulsed interference, except those with very low pulse width. For this kind of interferences, we propose to leave them to be detectable with the kurtosis value (very relevant for these kind of interferences because it has very low duty cycle, and thus very large change in the kurtosis).

5.2.4 Spectrogram detector

Frequency analysis is probably the most widely adopted approach to detect the presence of a stationary interference signal, relying on classical spectral estimation techniques [Hay09]. These techniques perform well in the case of stationary signals. Unfortunately, this is not always the case in practice, since interference signals may appear and disappear within the observation interval, or they even may change their characteristics as a function of time (e.g. a swept or chirp interference source). This time-varying behavior, which is typically encountered in practice, leads most interference signals to be non-stationary in the sense that the ACF exhibits a time-variant dependence. In that case, interference detection must resort to generalized metrics beyond the traditional PSD analysis, the so-called TFA schemes. Different TFA schemes are available in the current literature [Coh89], but we focus here on the spectrogram. The reason is because it is not limited to specific types of interference sources, moreover it provides a versatile and affordable approach for interference detection and characterization in GNSS. Particularly, when compared to some other much more complex TFA methods such as the Wigner-Ville analysis.

The Short-Time Fourier Transform (STFT) is the most widely used method to study non-stationary signals. The concept behind it is simple and powerful, for instance, suppose a signal is present during a certain time and whose frequency takes three different values in three different interval times. If the whole content of the signal is Fourier analyzed, the conventional frequency analysis will show peaks at the frequencies corresponding to the three different values present in the signal. Apparently, these three frequencies will be reported as being present during the whole observation time, when actually they are not, since the conventional Fourier analysis is unable to determine the time instant at which every spectral component appears. The most straightforward approach to circumvent this limitation is to break up the total observation

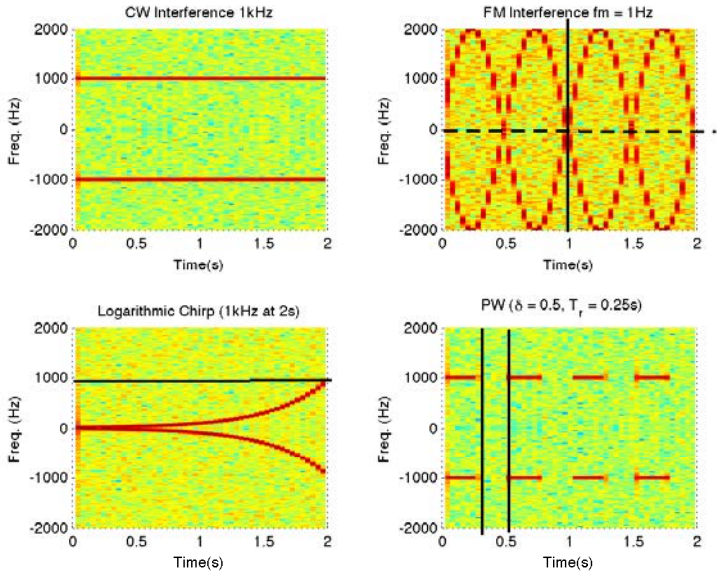

Figure 5.10: Spectrogram of different types of interferences.

length into several segments and Fourier analyze each interval separately. Upon examining the spectrum of each segment it will be possible to see in which time interval different frequencies occur. This is the basic idea behind the STFT (see Figure 5.9). The total set of spectra indicates how the frequency content of the signal under analysis varies in time. This procedure is often named *spectrogram* and its discrete version $S_x(m, f)$, at snapshot m and frequency f, is given by

$$S_x(m, f) \doteq \frac{1}{N} \left| \sum_{i=1}^{N} x\left(i + mN - N\right) e^{-j2\pi\left(\frac{if}{N}\right)} \right|^2 , \tag{5.21}$$

where $m = 1, 2, \ldots$ stands for the snapshot index, with each snapshot including N samples of the received signal $x(n)$. Thereby, with this expression we have an N-points estimate of the Power Spectral Density (PSD) every snapshot. Figure 5.10 shows the spectrogram for different interference sources with the aim of proving the capability of this TFA technique to capture the time-frequency evolution of non-stationary signals. In addition, the spectrogram representation allows extracting some interference features as in the case of a frequency modulated signal, the modulation frequency, or in the case of pulsed interferences, the duty cycle. In that sense, the spectrogram is a good candidate for interference detection and characterization, since it is able to detect a great variety of interference types and to extract information about them.

Statistical characterization

We see from (5.21) that under $\mathcal{H}_0$ the spectrogram at snapshot m and frequency f (i.e. $S_x(m, f)$) follows a central χ^2 distribution with two degrees of freedom, proportionality parameter related to the noise power, and non-central parameter related to the interference power. This is so because $S_x(m, f) = (1/N)(\mathrm{Re}\{y\}^2 + \mathrm{Im}\{y\}^2)$, with $\mathrm{Re}\{y\}$ and $\mathrm{Im}\{y\}$ being Gaussian random variables with variance equal to $N \cdot \sigma_w^2$ and mean equal to $\{0, i(n)\}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. This is taking into account the statistical characterization in (5.1). Thus, since both components

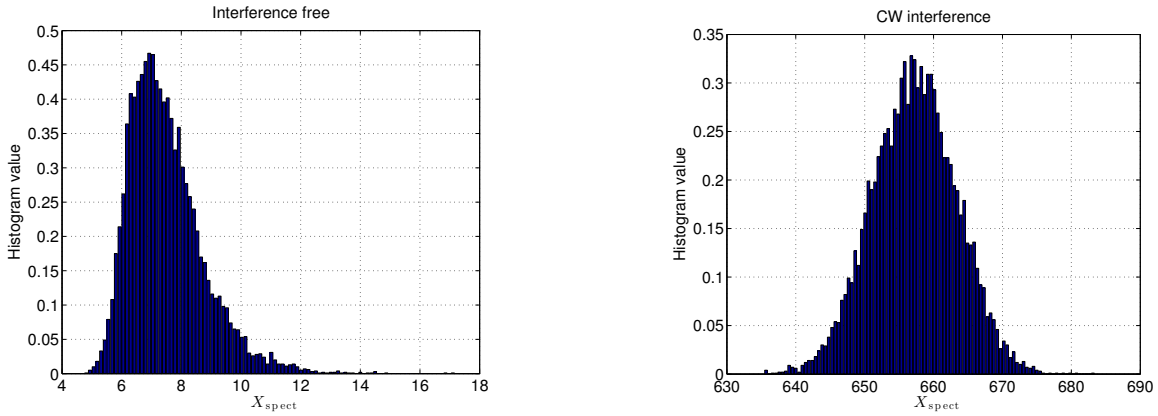

Figure 5.11: Statistical characterization of the spectrogram metric with $N = 10^4$. (Left) Experimental pdf under $\mathcal{H}_0$ and (right) under $\mathcal{H}_1$ with a CW interference with INR = 10 dB.

are quadratically added, we have

$$S_x(m,f) \sim \begin{cases} \mathcal{H}_0 : \sigma_w^2 \cdot \chi_2^2 & m < v \\ \mathcal{H}_1 : \sigma_w^2 \cdot \chi_2^2(\lambda) & m \geq v \end{cases}, \tag{5.22}$$

with $\chi_2^2(\lambda)$ the non-central chi-squared distribution with 2 degrees of freedom and non-centrality parameter $\lambda = 2 \cdot \text{INR}$. Now, in order to have a scalar metric to be fed into the CUSUM, we define

$$X_{\text{spect},m} \doteq \frac{\max_f \{S_x(m,f)\}}{\frac{1}{N-1}\sum_{i=1}^{N-1} S_x(m,f_i)}, \tag{5.23}$$

where f_i denotes the i-th frequency bin of the N-points power spectral density defined on (5.21), excluding the frequency bin of the maximum (i.e. $\max_f\{S_x(m,f)\}$). The term in the denominator of (5.23) acts as a normalization factor with respect to the noise power. In this way, we remove the noise dependence, and we get a metric that is independent of the noise power.

Under $\mathcal{H}_0$, the spectrogram is flat over all frequencies, and then the maximum value and the mean value will not depart so much one from the other. On the other hand, under $\mathcal{H}_1$, the spectrogram presents peaks in the frequency components of the interference. Hence, the maximum value will now depart considerably with respect to the mean value, and then the spectrogram metric will be much greater than under $\mathcal{H}_1$. Thereby, the spectrogram metric distribution depends on the distribution of the maximum of all frequency bins of the spectrogram (i.e. $\max_f\{S_x(m,f)\}$). The statistical characterization is shown in Figure 5.11, which presents the experimental pdf of the spectrogram metric in the absence (left plot) and presence (right plot) of interference. It is shown how besides the change in the mean, the distribution of the spectrogram also changes in presence of interference. These distributions do not fit the common Gaussian or chi-squared distribution seen so far, and they have to be analyzed under the framework of Extreme Value Theory (EVT). EVT shows that the cdf of the maximum of N iid random variables, with cdf F, has one of three possible functional forms depending on the tail of the parent distribution F. For the chi-squared distribution, which is the case of the spectrogram under $\mathcal{H}_0$, the cdf of the normalized maximum in (5.23) has a double exponential

form with mean and variance given by [Tur07]

$$\mu_0 = \ln(N) + \gamma \text{ and } \sigma_0^2 = \frac{\pi^2}{6} = 1.645, \tag{5.24}$$

where $\gamma \approx 0.5772$ the Euler constant. Thus, under $\mathcal{H}_0$, the mean value of the spectrogram metric depends upon the logarithm of the snapshot samples, whereas the variance is constant. Under the presence of interference, a closed-form expression for the distribution of X_{spect} is not available (they will depend upon the interference). However, we know that the mean will depart from the one obtained under $\mathcal{H}_0$, and also it will depend upon the INR. With this setting, we can think of a mean change detection problem with the spectrogram metric.

Application of QCD: CUSUM for the spectrogram metric (CUSUM-Spect)

Once the interference is present, the distribution depends on the maximum of a set of non-central chi-squared random variables, for which a tractable expression is difficult to find. In that case, the LLR cannot be completely defined, and then we are unable to apply the CUSUM algorithm directly to X_{spect}. An alternative would be to use the O-CUSUM variant, because the distribution under $\mathcal{H}_0$ is known. However, in order to use the O-CUSUM we have to obtain the characteristic function of the distribution under $\mathcal{H}_0$, which is not a straightforward task. In order to avoid such calculation, we propose to use a Gaussian mean change CUSUM to detect the change on the mean of the spectrogram metric. In this way, we know that the algorithm may not be optimal since the actual distributions are not truly Gaussian. However, if the mean change value is properly chosen, the algorithm will perform well using the same LLR as for the kurtosis, in (5.15), but using $X_{\text{spect},m}$. The mean and variance before change can be fixed as in (5.24), and the mean after change can be experimentally set depending on the minimum INR expected to be detected. Thus, making use of the CUSUM decision rule with the spectrogram metric parameters, henceforth referred to as CUSUM-Spect, we have

$$\mathfrak{N}_{\text{spect}} \geq e^h \text{ and } \mathfrak{D}_{\text{spect}} \leq \frac{h}{\mathrm{E}_1\left[X_{\text{spect},m}\right]}, \tag{5.25}$$

with $\mathrm{E}_1[X_{\text{spect},m}] = \mu_{\text{spect},1}$ the fixed mean after change.

GNSS considerations

Among all the possible interferences in GNSS the spectrogram metric will be able to detect CW and PW interferences, including chirp interferences as happened with ACF-based detection. This is so because these kind of interferences introduce some peaks on the spectrogram representation, which makes possible the detection of the interference. For the case of Gaussian wide-band interferences, the spectrogram will still be flat as in the absence of interference, and then the spectrogram metric will not significantly change with respect to nominal conditions. However, if the wide-band interference is a modulated CW or PW (e.g. AM-FM modulations, BPSK, ...), which is usually the case in practice, the spectrogram will present some peaks in the frequency components of the modulated signal, and then the interference can be detected. So far we have considered a flat spectrogram over all frequencies under $\mathcal{H}_0$. However, in practice this is only true for the range of frequencies filtered by the GNSS receiver front-end filter. Nevertheless, this is not a problem for the proposed metric. On the one hand, the maximum of the spectrogram will remain the same because it is not filtered, since it remains inside the pass-band of the filter. Moreover, in practice it is not feasible to calculate the STFT using all the samples in a snapshot due to the huge computational and memory requirements. Instead, it is usual to set an internal

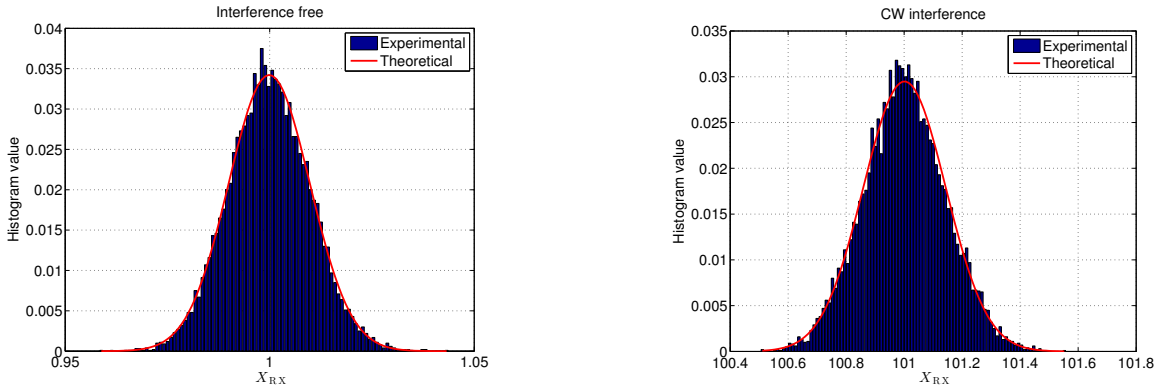

Figure 5.12: Statistical characterization for the power-based detection with $N = 10^4$ samples and noise power $\sigma_w^2 = 2$. Comparison between simulated (i.e. histogram) and theoretical pdf under (left) $\mathcal{H}_0$ and under (right) $\mathcal{H}_1$ with a CW interference with INR = 20 dB.

number of samples to sequentially calculate the STFT and then average the results within a snapshot period. Hence, due to the front-end filter shape (i.e. variation of the spectrogram mean value) and the internal processing of the STFT (i.e. averaging of the STFT over the snapshot) the real value for the mean and variance before the change will vary with respect to that given by expressions in (5.24). It is for this reason that it is convenient to calibrate the fixed values for those parameters with experimental results obtained with real life signals.

5.2.5 Received power detection

Another useful metric to infer the presence of interference is the power of the received signal. For simplicity, we will consider the total power of the received signal within a snapshot interval (i.e. N samples). This is equivalent to the integration of all the frequency bins of the spectrogram within a snapshot time. The operational principle is as follows: In the absence of interference, the estimated power should be around the noise power, whereas in the presence of some interference, the estimated power should significantly deviate from that noise power. Based on the above considerations, the received power estimation at snapshot m is defined as

$$\widehat{P}_{\mathrm{RX}}(m) \doteq \frac{1}{N}\sum_{i=1}^{N} |x(i+mN-N)|^2\,, \tag{5.26}$$

with $m = 1, 2, \ldots$ the snapshot index, with each snapshot including N samples of the received signal $x(n)$. So, with the monitoring of $\widehat{P}_{\mathrm{RX}}$ it is possible to infer the presence of interference and to extract information about its power. Particularly, we define the metric

$$X_{\mathrm{RX},m} \doteq \frac{\widehat{P}_{\mathrm{RX}}(m)}{2\sigma_w^2}, \tag{5.27}$$

with σ_w^2 the noise power, which may be estimated.

Statistical characterization

It can be easily shown that under $\mathcal{H}_0$ $|r_{\mathrm{GNSS}}(n)|^2$ follows a central χ^2 distribution with two degrees of freedom, while in the presence of interference it follows a non-central χ^2 with two degrees of freedom and non-central parameter related to the interference power. Therefore, from (5.27), $X_{\mathrm{RX},m}$ will be equal to one under $\mathcal{H}_0$, and depart form one otherwise. Thus, it can be

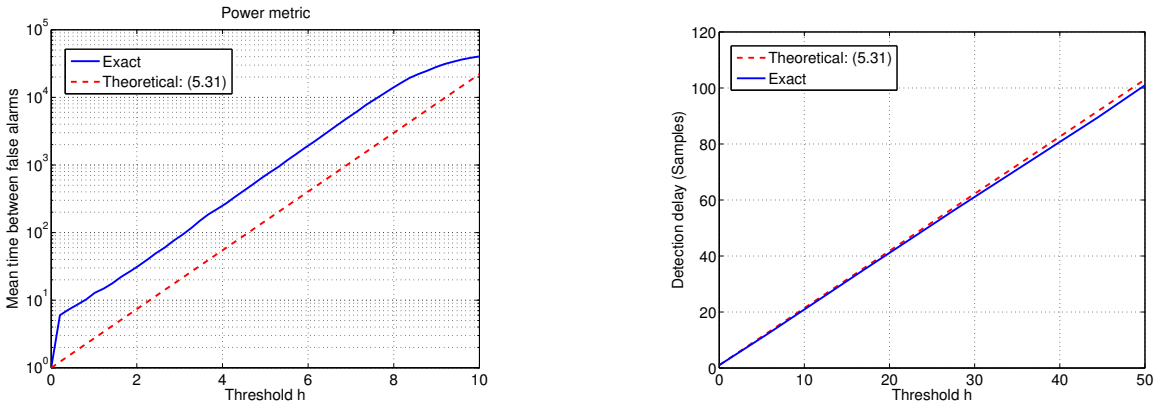

Figure 5.13: CUSUM-Power performance as a function of the detection threshold for $N = 10^4$ snapshot samples and power noise $\sigma_w^2 = 2$. (Left) Mean time between false alarms in samples, and (right) detection delay for a CW with INR $= -20$ dB.

modeled as a Gaussian distribution with mean and variance related by the noise and interference power in virtue of the CLT. That is,

$$X_{\mathrm{RX},m} \sim \begin{cases} \mathcal{H}_0 : \mathcal{N}\left(\mu_{\mathrm{RX},0}, \sigma^2_{\mathrm{RX},0}\right) & m < v \\ \mathcal{H}_1 : \mathcal{N}\left(\mu_{\mathrm{RX},1}, \sigma^2_{\mathrm{RX},1}\right) & m \geq v \end{cases}, \tag{5.28}$$

with expressions for the mean and variance, derived from the CLT, given by

$$\begin{aligned} \mu_{\mathrm{RX},0} &= \tfrac{\sigma_w^2}{2\sigma_w^2} \cdot \mathrm{E}\left[\chi_2^2\right] = 1, & \sigma^2_{\mathrm{RX},0} &= \tfrac{1}{N}\left(\tfrac{\sigma_w^2}{2\sigma_w^2}\right)^2 \cdot \mathrm{var}\left[\chi_2^2\right] = \tfrac{1}{N}, \\ \mu_{\mathrm{RX},1} &= \tfrac{\sigma_w^2}{2\sigma_w^2} \cdot \mathrm{E}\left[\chi_2^2(\lambda)\right] = (1 + \mathrm{INR}), & \sigma^2_{\mathrm{RX},1} &= \tfrac{1}{N}\left(\tfrac{\sigma_w^2}{2\sigma_w^2}\right)^2 \cdot \mathrm{var}\left[\chi_2^2(\lambda)\right] = \tfrac{1}{N}\left(1 + 2\mathrm{INR}\right), \end{aligned} \tag{5.29}$$

with INR the INR and $\lambda = 2 \cdot \mathrm{INR}$ the non-central parameter. This is shown in Figure 5.12, in which the distribution of the proposed metric is shown for the cases of absence and presence of interference. We can notice the Gaussian distribution for both hypotheses. On the other hand, we see the departure between the mean and variance under $\mathcal{H}_0$ (i.e. left plot) and $\mathcal{H}_1$ (i.e. right plot). Indeed, the mean under $\mathcal{H}_0$ should be close to one, whereas for the interference case the mean should be around the sum between the noise and interference power (see (5.29)).

Application of QCD: CUSUM power metric (CUSUM-Power)

In (5.28) we have statistically characterized the power estimate metric. It includes the mean and variance before and after the change, which are completely known from (5.29). Therefore, using the power estimation metric, $X_{\mathrm{RX},m}$, given by (5.26) we can fully characterize the LLR, and then use the CUSUM algorithm to detect a change in both the mean and variance, given by

$$\mathrm{LLR}_{\mathrm{RX}}(m) = \ln\left(\frac{\sigma_{\mathrm{RX},0}}{\sigma_{\mathrm{RX},1}}\right) + \frac{\left(X_{\mathrm{RX},m} - \mu_{\mathrm{RX},0}\right)^2}{2\sigma^2_{\mathrm{RX},0}} - \frac{\left(X_{\mathrm{RX},m} - \mu_{\mathrm{RX},1}\right)^2}{2\sigma^2_{\mathrm{RX},1}}. \tag{5.30}$$

Both $\sigma^2_{\mathrm{RX},1}$ and $\mu_{\mathrm{RX},1}$ depend on the INR of the interference. As before, a way to proceed is to fix a certain value for these parameters according to the minimum INR that one expects to detect. Thus, a minimum change detection is fixed allowing the detection of any larger change caused by larger INR. Thereby, we can make use of the CUSUM detection rule with threshold h, henceforth referred to as CUSUM-Power, leading to the following performance in terms of

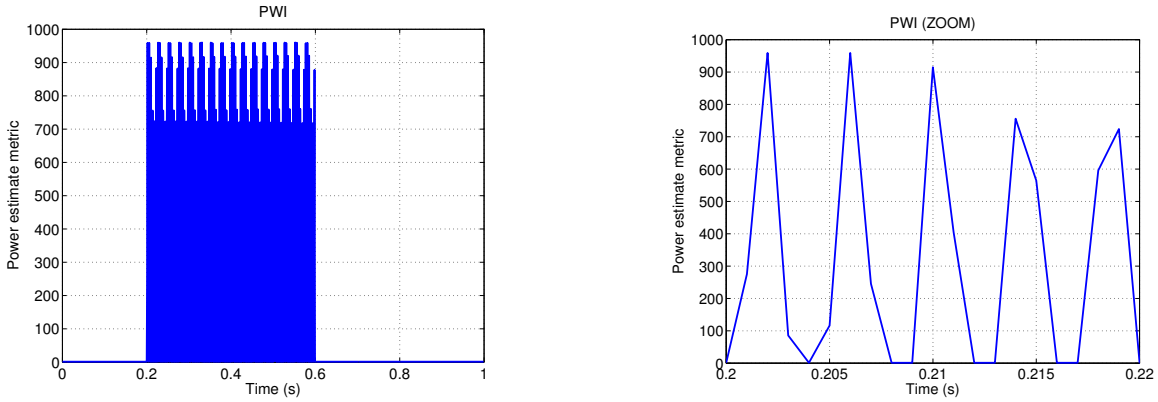

Figure 5.14: (Left) Behavior of the power-based metric, $X_{\mathrm{RX},m}$, under the presence of a PW interference, and (right) its zoomed version.

QCD:

$$\mathfrak{N}_{\mathrm{RX}} \geq e^{h} \text{ and } \mathfrak{D}_{\mathrm{RX}} \leq \frac{h}{d_{\mathrm{RX}}}, \tag{5.31}$$

with

$$d_{\mathrm{RX}} = \ln\left(\frac{\sigma_{\mathrm{RX},0}}{\sigma_{\mathrm{RX},0}}\right) + \frac{\left(\mu_{\mathrm{RX},1} - \mu_{\mathrm{RX},0}\right)^2}{2\sigma^2_{\mathrm{RX},0}} + \frac{\sigma^2_{\mathrm{RX},1}}{2\sigma^2_{\mathrm{RX},0}} - \frac{1}{2}. \tag{5.32}$$

The above bounds are depicted in Figure 5.13 and compared with simulated results.

GNSS considerations

Due to the high power of GNSS interferences with respect to the received signal power the change in the spectrogram-based metric, X_{RX}, due to the presence of interference will be large, so that interference may be promptly detected. This is particularly true for wide-band and frequency varying continuous interferences. Indeed, these kinds of interferences are the most important ones to be detected with TFA techniques, since they might not be visible for statistical analysis metrics. This is specially true for the Gaussian wide-band one that would not introduce significant variations in the histogram, since it is Gaussian, too. However, once the snapshot time is fixed (i.e. N), some pulsed interference may be perceived as a continuous one, or be perceived as the same interference but with a smaller duty cycle. This effect is not convenient for the CUSUM-Power, because the smaller the number of samples that are observed under the ON state of the pulsed interference, the smaller the estimated power, and thus, it will be more difficult to detect the interference. This is illustrated in the left plot of Figure 5.14, in which for an INR = 25 dB we should expect a mean around 316, but lower and larger values are also observed, reflecting the above behavior. The problem with this behavior is that it might be possible to obtain a power estimate in presence of interference that is lower than the fixed minimum change. This is shown in the right plot of Figure 5.14, in which a zoomed version of X_{spect} is shown. In contrast, this behavior is not a problem for the CUSUM-Kurt because smaller duty cycles cause larger changes in the kurtosis mean, and therefore, it will be easier to detect the presence of interference. With the above considerations, we conclude that although the CUSUM-Power is able to visualize PW interference, the suggested metric is not adequate for PW detection. So, we propose the use of the CUSUM-Power for the detection of Gaussian wide-band or frequency varying interferences, and the use of the CUSUM-Kurt for the detection of PW interferences.

5.3 Quickest Multipath Detection

Actually, multipath and in particular NLOS propagation are nowadays the main local effect limiting the performance of GNSS in urban environments. As a matter of fact, multipath mitigation has attracted the attention of many researchers during the past years, leading to a plethora of contributions in the existing literature [Bra01, Bhu07]. Nevertheless, as already stated, detection of these threats has often remained in background, when indeed it is often more important than mitigation. Among this line, substantial contributions for GNSS multipath detection have appeared, but most of them relying on external information [TM10, Bha09, Shy14] or adopting a classical detection framework. In this section, though, we take a leap forward and we focus on the framework of QCD, based on several post-correlation metrics. Specifically, we assume we operate at the output of a bank of L correlators with a given CI period, modeled as

$$\mathbf{y}(k) = \begin{cases} \mathcal{H}_0 : \boldsymbol{\alpha}_0(k) + \mathbf{w}(k) & k < v \\ \mathcal{H}_1 : \boldsymbol{\alpha}_0(k) + \boldsymbol{\alpha}_1(k) + \mathbf{w}(k) & k \geq v \end{cases}, \tag{5.33}$$

where $\mathbf{y}(k)$ stands for the $(L \times 1)$ vector with the L correlator outputs at the k-th CI period, and $\boldsymbol{\alpha}_j(k)$, with $j = \{0, 1\}$, is the vector collecting the L complex amplitudes corresponding to the LOS and multipath post-correlation outputs components, respectively. Finally, $\mathbf{w}(k)$ is the noise vector, whose components are the post-correlation complex Gaussian noise.

The statistical characterization of this model is complicated, so that an exact theoretical statistical characterization of the post-correlation metrics considered in this thesis becomes intractable in most cases. It is for this reason that we will consider a Gaussian approximation for the distribution of the considered SLM. We will see that despite the complexity of the above model, the analyzed SLM can be fairly approximated as Gaussian distributed by virtue of the CLT. This approximation will be evaluated with real GNSS signal analyses. Particularly, the analyzed data were gathered during 80 s by a moving vehicle in a dense urban environment. The vehicle was under benign condition the first 40 s, and then it changed to harsh conditions until the end of the data record. This discrimination between benign and harsh environments was done by the use of a truth trajectory. The detailed scenario will be introduced in Chapter 7. Before entering into details of each SLM, let us first introduce the general (Gaussian) model for the considered metrics; that is

$$X_{\mathrm{mp},k} \sim \begin{cases} \mathcal{H}_0 : \mathcal{N}(\mu_0, \sigma_0^2) & \text{if } k < v \\ \mathcal{H}_1 : \mathcal{N}(\mu_1, \sigma_1^2) & \text{if } k \geq v \end{cases}, \tag{5.34}$$

with $X_{\mathrm{mp},k}$ the samples of the post-correlation SLM, and $\{\mu_i, \sigma_i^2\}$, for $i = 0, 1$, the mean and variance under $\mathcal{H}_i$. Hence, the LLR for these metrics can be written as

$$\mathrm{LLR}_{\mathrm{mp}}(k) = \ln\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{(X_{\mathrm{mp},k} - \mu_0)^2}{2\sigma_0^2} - \frac{(X_{\mathrm{mp},k} - \mu_1)^2}{2\sigma_1^2}. \tag{5.35}$$

Thus, in general, a set of four different parameters need to be tuned for the CUSUM implementation. These are $\{\mu_0, \sigma_0^2, \mu_1, \sigma_1^2\}$, and they can be tuned as follows:

- μ_0: Under nominal conditions the mean of the metric is known or it can be calibrated

experimentally.

- σ_0^2: We can determine the variance under nominal conditions based on the relationship between the standard deviation and the population of a Gaussian distribution. In particular, we know that the maximum deviation is $\Delta_{\max} \approx 3\sigma_0$, so that

$$\sigma_0^2 \approx \left(\frac{\Delta_{\max}}{3}\right)^2. \tag{5.36}$$

  For instance, we know that due to noise and the dynamic of the user the DLL output can vary $\pm[0.02, 0.05]$ chips, leading to a variance of $[4.44 \cdot 10^{-5}, 2.78 \cdot 10^{-4}]$.

- μ_1: The mean under $\mathcal{H}_1$ is unknown since it will depend on the multipath conditions, which are actually unknown. However, it can be fixed as

$$\mu_1 = \mu_0 \pm \delta_{\min}, \tag{5.37}$$

  with $\delta_{\min}$ a proper value selected as the minimum detectable mean change due to multipath, which can be tuned experimentally.

- σ_1^2: Similarly as for σ_0^2, we can fix the variance in presence of multipath as the minimum detectable variability due to multipath, $\Delta_{\min}$, as

$$\sigma_1^2 \approx \left(\frac{\Delta_{\min}}{3}\right)^2. \tag{5.38}$$

  For instance, if we want to detect multipath leading to 20 m of error this leads to a variation in the DLL output of ± 0.07 chips (for GPS L1 C/A), which results in $\sigma_1^2 = 5.5 \cdot 10^{-4}$.

5.3.1 Carrier-to-Noise ratio (C/N_0) detector

One of the common observable in a satellite navigation receiver, apart from the PVT solution, is the so-called carrier-to-noise ratio (C/N_0) measured for each satellite (i.e. the ratio between the satellite signal carrier power and the noise power spectral density at the receiver). This observable can be used to detect the presence of multipath, since the received signal power is an average of the composite signal power of the direct and reflected signals. In static conditions, the C/N_0 is reasonably constant, but the presence of multipath that constructively and destructively combines with the LOS signal, or even its blockage, can cause fluctuations of the C/N_0. Different approaches are available in the literature in order to estimate the C/N_0 in a GNSS receiver [Fal11]. In fact, in a GNSS receiver, the carrier power C can be often related with the multipath parameters as [Ray99]

$$C = A\left(1 + \alpha^2 + 2\alpha\cos(\gamma)\right), \tag{5.39}$$

where A denotes a power factor related with the received carrier amplitude, α is another factor related with the Signal-to-Multipath (power) Ratio (SMR) and the time-delay of the multipath component, and γ is the reflected signal relative phase.

The behavior described by the above expression is depicted in the left plot of Figure 5.15, which shows the power gain in dBs for a fixed SMR and varying the multipath delay (upper plot) and also the case of fixed multipath delay and varying the SMR (lower plot). We can see

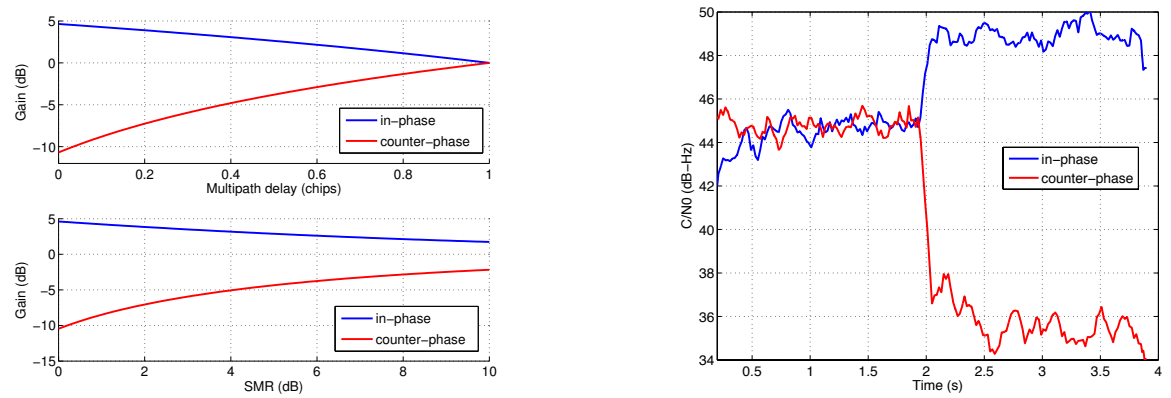

Figure 5.15: (Left) Multipath effects on the received carrier power (gain) as a function of the multipath delay for a fixed SMR = 3 dB (up); and as a function of the SMR for a fixed multipath delay of 0.3 chips (down). (Right) Effects of a multipath ray appearing at second 2 with multipath delay of 0.1 chips and SMR = 3 dB on the C/N_0 metric.

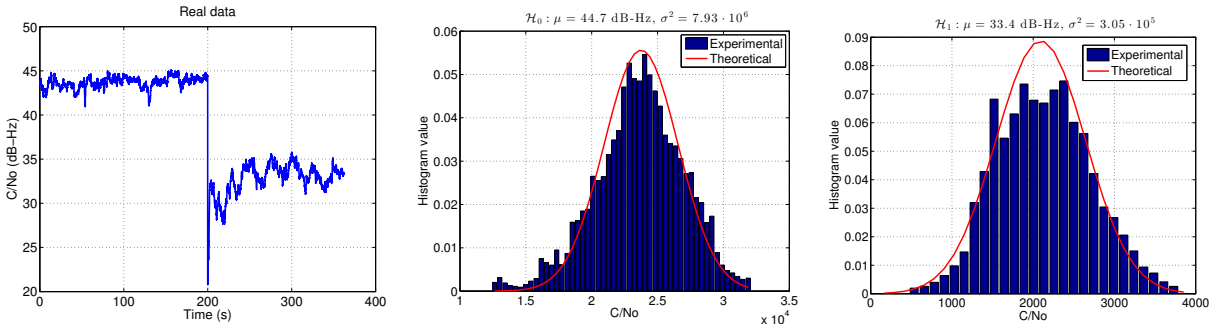

Figure 5.16: Statistical characterization of the C/N_0. (Left) Real data C/N_0 time behavior. (Middle) Statistical characterization under $\mathcal{H}_0$, corresponding to the first 200 s of the C/N_0 in the left plot; and (right) under $\mathcal{H}_1$, corresponding to the data after second 200 s of the C/N_0 in the left plot.

how the change on the received carrier power due to multipath increases as the multipath delay or the SMR decrease. Depending on the phase of the received multipath ray (γ), its contribution will be constructive or destructive, as it can be seen in the results for the in- and counter-phase case. For the former the contribution leads to an increase of the received power, whereas for the counter-phase case it becomes to a reduction of the received power. Hence, it can be concluded that observing the received carrier power, or equivalently the measured C/N_0, the presence of multipath can be inferred. This is illustrated in the right plot of Figure 5.15, which shows the temporal evolution of the C/N_0 when a multipath ray appears at second 2.

Statistical characterization

Since the C/N_0 estimates are based on the average of the prompt correlator, the Gaussianity of the C/N_0 follows in virtue of the CLT, and then the C/N_0-based multipath detection problem can be solved as a Gaussian mean change. This is shown in Figure 5.16, which illustrates the statistical characterization of the C/N_0 using the real data presented in the left plot. We see in the middle plot how the histogram of the C/N_0 values (in linear units) under $\mathcal{H}_0$ fits pretty well to a Gaussian pdf with mean and variance equal to 44.7 dB-Hz (in linear magnitude) and $7.93 \cdot 10^6$, respectively, as expected in nominal conditions. Under $\mathcal{H}_1$ (see right plot), we see that the main component of the histogram also fits pretty well to a Gaussian pdf with mean and variance equal to 33.4 dB-Hz (linear) and $3.05 \cdot 10^5$, respectively. We see however that in both cases there are some disagreements between theoretical an experimental results, particularly in the left tails. This is because the C/N_0 metric is also sensible to attenuation due to fading, thus

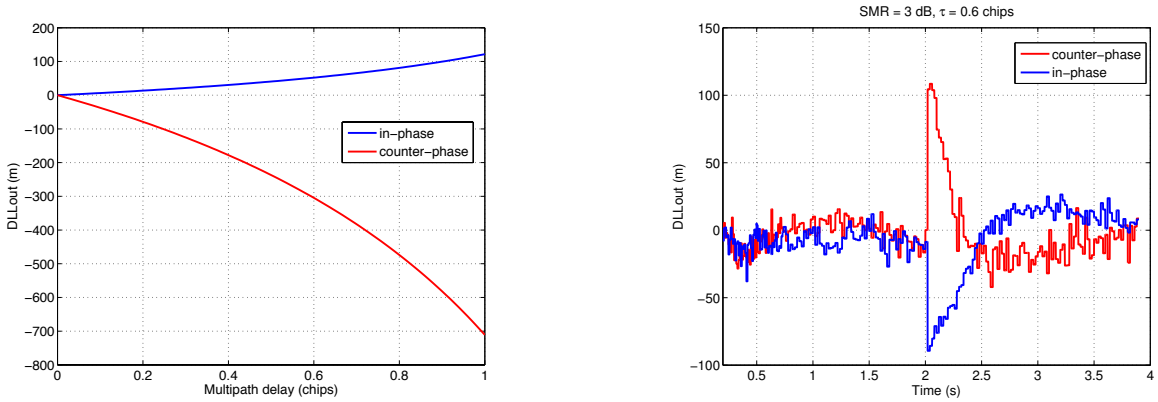

Figure 5.17: (Left) DLLout under the presence of a multipath ray with 3 dB of SMR and varying the multipath delay. (Right) Temporal behavior with a multipath ray appearing at second 2 with multipath delay of 0.6 chips and SMR = 3 dB.

introducing some outliers in the left tail of the Gaussian histogram.

Application of QCD: CUSUM-C/N_0

According to the above behavior we could say that the C/N_0 distribution in both hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ can roughly be approximated by a Gaussian distribution with different mean and variance. However, the variance after change will depend on the multipath parameters and then it will be unknown. In addition, the change on the mean is large enough to neglect the fact that there is a slight change in variance, too. Hence, we can formulate the problem of multipath detection as a C/N_0 mean change detection, so that the LLR to be used in the CUSUM takes the same form as that used for the kurtosis in (5.15), but with the C/N_0 parameters. The configuration of these parameters accounts for $\{\mu_0, \mu_1, \sigma_0^2\}$, which should be tuned as stated in the introduction of Section 5.3. Values for μ_0 ranges from 42–47 dB-Hz, whereas variations of around 2–3 dB should be taken into account to fix σ_0^2.

5.3.2 Code discriminator output (DLLout) detector

Various types of coherent and non-coherent discriminator functions are employed for code tracking [Kap05, Mis11, Par96a]. Without loss of generality, the following analysis is carried out for a non-coherent dot-product type of discriminator. For continuous tracking, the discriminator output is driven to zero by feeding the discriminator output to the code loop filter, and then using this filtered correction to update the code NCO. So, the discriminator output at each epoch will be a function of the multipath time-delay. Assuming the multipath delay is smaller than the chip width, the error in the DLL output (DLLout) due to multipath is dependent upon the time-delay and SMR. An illustration is given in the left plot of Figure 5.17, which shows the DLLout due to multipath with fixed SMR and varying the multipath delay. As shown, the DLLout is greater as the multipath delay increases. The positive or negative sign of the DLLout means that the NCO has to be delayed or advanced, respectively. This means that the presence of multipath/NLOS propagation may also be visible at the DLLout as the presence of some spikes, after which the value reverts to zero. This behavior is shown in the right plot of Figure 5.17, in which a multipath ray is simulated to appear at second 2. It is important to remark that the behavior of the DLLout is in contrast to the C/N_0 one; that is, the C/N_0 is greater as the multipath delay decreases and it experiences a shift in the mean instead of the

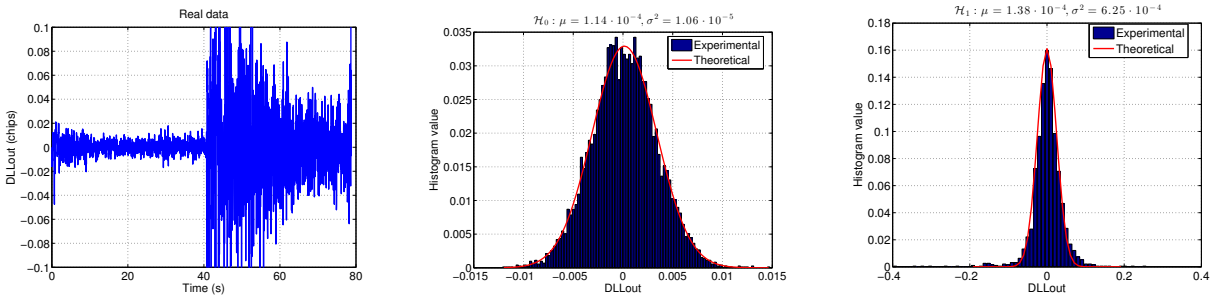

Figure 5.18: Statistical characterization the DLLout. (Left) Real data DLLout time behavior. (Middle) Statistical characterization under $\mathcal{H}_0$, corresponding to the first 40 s of the DLLout in the left plot; and (right) under $\mathcal{H}_1$, corresponding to the last 40 s of the DLLout in the left plot.

presence of spikes.

Statistical characterization

We know that under $\mathcal{H}_0$ the DLLout should be close to zero, in average, with just small variations due to noise and residual corrections needed to track the code dynamics (i.e. user movement). However, according to the above discussion and with real-life multipath coming in and out, it is clear that the variance of the DLLout is the parameter of interest implementing a multipath QCD technique. Similarly as the C/N_0, the DLLout can be approximated by a Gaussian random variable by virtue of the CLT. Thus, we can address the multipath detection problem as a Gaussian variance change on the DLLout samples. This is depicted in Figure 5.18, which shows the statistical characterization of the DLLoutbased on the real data in the left plot. The middle plot shows the histogram of the DLLout under $\mathcal{H}_0$, and it confirms the Gaussianity of the DLLout with a mean close to 0 and low variance corresponding to variations of 0.015 chips (i.e. 10^{-5}) due to the noise and code dynamics. In the right plot, we present the histogram under $\mathcal{H}_1$. We see how the histogram fits a Gaussian distribution too. In this case, the mean still is close to 0, but a much greater variance than under $\mathcal{H}_0$ is experienced.

Application of QCD: CUSUM-DLLout

We have seen that the DLLout follows a Gaussian distribution with known mean before and after change (i.e. ≈ 0) but unknown a-priori variance before and after the change. Using the configuration mentioned in Section 5.3 for the CUSUM parameters we are able to use the CUSUM algorithm for a Gaussian variance change. The LLR for such a change is given by

$$\mathrm{LLR}_{\mathrm{DLL}}(k) = \ln\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{\left(X_{\mathrm{DLL},k} - \mu_0\right)^2 \left(\sigma_1^2 - \sigma_0^2\right)}{2\sigma_0^2\sigma_1^2}, \tag{5.40}$$

with $x_{\mathrm{d}}(k)$ the DLL output at the k-th CI period. Then, applying this LLR into the CUSUM, we have the next performance

$$\mathfrak{N}_{\mathrm{DLL}} \geq e^h \text{ and } \mathfrak{D}_{\mathrm{DLL}} \leq \frac{h}{d_{\mathrm{DLL}}}, \tag{5.41}$$

with $d_{\mathrm{DLL}} = \ln(\sigma_0/\sigma_1) + (\sigma_1^2 - \sigma_0^2)/(2\sigma_0^2)$.

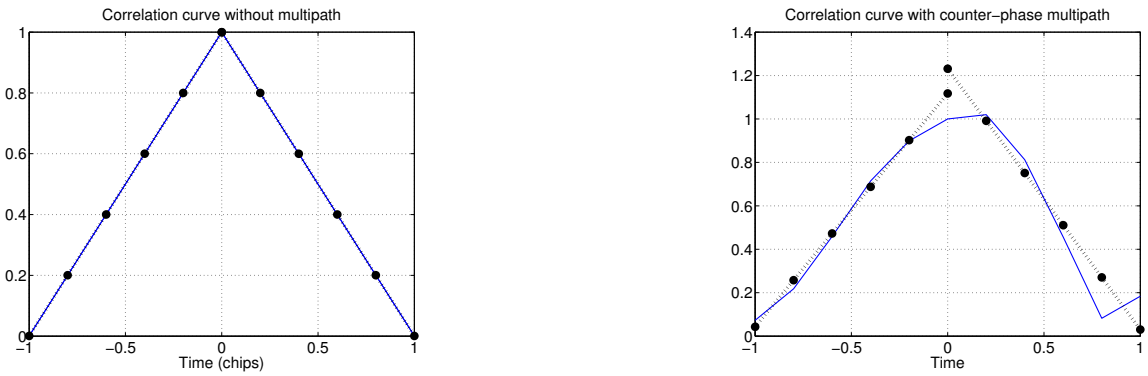

Figure 5.19: Correlation curves for (left) the case of free-multipath, and (right) counter-phase multipath; and LS fitting of the two side slopes (black).

5.3.3 Slope Asymmetry Metric (SAM) detector

Multipath rays always arrive later than the direct signal, thus distorting the correlation curve in a different manner on the early and late slopes. The shape of the correlation curve in nominal conditions is known, so the idea is to formulate a metric that measures how distant some features of the actual correlation are with respect to their nominal values. This idea is the basis of the Slope Asymmetry Metric (SAM), where a straight line is fitted to each side of the absolute value of the correlation, and subsequently the slopes are compared. The behavior is illustrated in Figure 5.19. In the left plot we see the correlation curve for a free-multipath case, and it is observed how the correlation curve is symmetrical. Thus, the curve fitting through LS provides the same side slopes. On the other hand, when multipath is present the correlation curve is shown to be asymmetrical and the curve fitting provides different side slopes (see right plot). Hence, this asymmetry metric can certainly allow us to detect the presence of NLOS/multipath. According to the previous behavior, a good metric to measure multipath from the correlation curve consists on comparing the left and right slopes of the measured correlation, computed within the tracking loop of a GNSS receiver. Ideally, both slopes should be equal, but sign reversed, and thus their sum should be close to zero, otherwise the slopes are not equal inducing the presence of multipath. The procedure to obtain the SAM is as follows [LS09]:

1. Normalize the correlation peaks with the square root of an estimate of the $\mathrm{C/N_0}$.

2. For each side of the correlation peak, find the LS estimate of the straight line that best fits the samples next to the maximum of the correlation. The maximum is not included in the set of correlation samples. The parameters of the line $a_\pm \cdot t + b_\pm$ are obtained as $[\widehat{a}_\pm, \widehat{b}_\pm]^T = \mathbf{M}_\pm^\dagger \mathbf{z}_\pm$, where $\mathbf{M}_i^\dagger$ is the Moore-Penrose inverse of

$$\mathbf{M}_- = \begin{bmatrix} -\tau_L & \cdots & -\tau_2 & -\tau_1 \\ 1 & \cdots & 1 & 1 \end{bmatrix} \text{ and } \mathbf{M}_+ = \begin{bmatrix} \tau_1 & \tau_2 & \cdots & \tau_L \\ 1 & 1 & \cdots & 1 \end{bmatrix}, \tag{5.42}$$

   and $\mathbf{z}_\pm$ is a $L \times 1$ vector with the L samples of each correlation curve side. This procedure is illustrated in Figure 5.20.

3. The SAM is defined as the sum of the estimated slopes of the left and right side of the correlation peak; that is $\mathrm{SAM} \doteq \widehat{a}_\mathrm{l} + \widehat{a}_\mathrm{r}$.

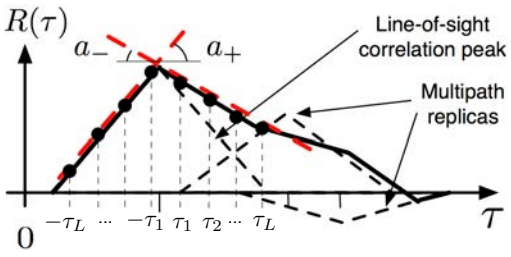

Figure 5.20: Procedure to calculate the SAM through the LS estimation of the side slopes of the correlation curve of the GNSS received signal.

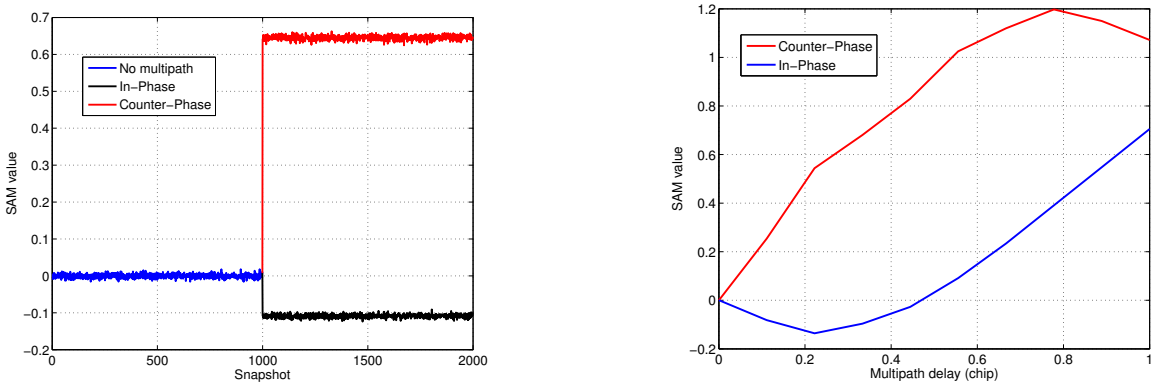

Figure 5.21: (Left) Temporal behavior of the SAM for a multipath ray with SMR = 10 dB and 0.3 chips of delay, appearing at snapshot 1000. (Right) SAM value behavior as a function of the multipath delay with SMR = 10 dB, using 11 correlation points with a separation of 0.2 chips. A $C/N_0 = 45$ dB-Hz is considered.

We show in Figure 5.21 the behavior of the SAM with the presence of multipath. The left plot shows the SAM value for the case of a multipath ray appearing at snapshot 1000. Both in- and counter-phase rays are shown. We see how just at the moment when multipath appears the SAM value deviates from the nominal value of zero. In general, results show that the SAM value increases as it does the multipath delay, with a behavior dependent on the multipath phase. From the above results we can conclude that the SAM can also be used to detect the presence of multipath, specially for large multipath delays. Actually, when multipath is present the SAM experiences two different effects:

1. Under LOS conditions, the mean of the SAM departs from 0.
2. Under NLOS conditions, the variance of the SAM fluctuates.

Indeed, in both LOS and NLOS conditions the mean and variance vary, but the mean change is predominant in LOS propagation (i.e. the deterministic component prevails), while the variance change is predominant in NLOS (i.e. random components due to multipath prevails). This behavior will be very useful to discriminate between LOS and NLOS multipath conditions.

Statistical characterization

The SAM is close to zero under $\mathcal{H}_0$ (indicating symmetry), whereas when multipath is present it departs from zero indicating asymmetry. This metric can also be approximated as a Gaussian random variable because it is calculated from an LS estimation of Gaussian distributed data. Actually, we can address the multipath detection as a change in both the mean and variance of a Gaussian distribution, as illustrated in Figure 5.22. The middle plot shows the histogram of the SAM values, shown in the left plot, under $\mathcal{H}_0$ (i.e. first 40 s). We see the Gaussianity of the

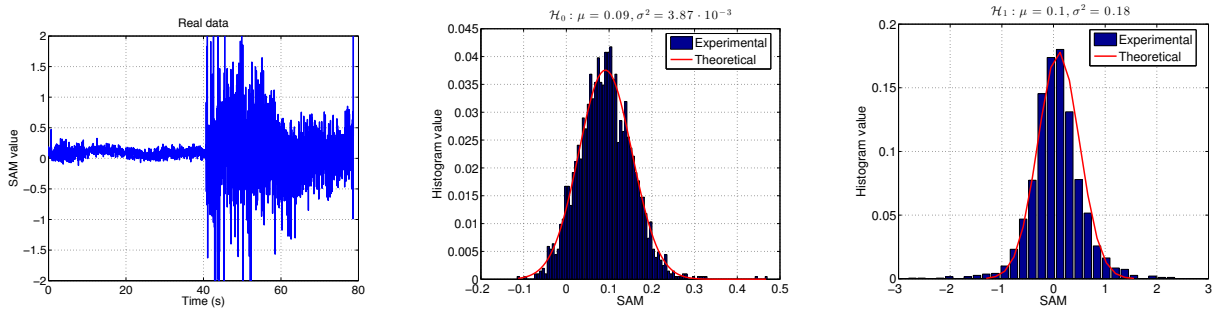

Figure 5.22: Statistical characterization of the SAM. (Left) Real data SAM time behavior. (Middle) Statistical characterization under $\mathcal{H}_0$, corresponding to the first 40 s of the SAM in the left plot; and (right) under $\mathcal{H}_1$, corresponding to the last 40 s of the SAM in the left plot.

SAM with a mean about 0.1. This value is due to the asymmetry introduced by the front-end filter and can be calibrated. On the other hand, in the right plot, we present the histogram under $\mathcal{H}_1$, corresponding to the last 40 s of the SAM presented in the left plot. We see how the histogram fits a Gaussian distribution quite well with mean also close to 0, but a much greater variance than under $\mathcal{H}_0$. Thereby, we can detect the presence of multipath as a Gaussian mean and variance change. Nevertheless, the SAM can also be used to discriminate between LOS and NLOS multipath. This is because under LOS a mean change of the SAM prevails, whereas under NLOS conditions a variance change prevails. In this sense, we can classify the presence of multipath (i.e. $\mathcal{H}_1$) into two different hypotheses, $\mathcal{H}_{\mathrm{L}}$ and $\mathcal{H}_{\mathrm{N}}$ for LOS and NLOS conditions, respectively, as

$$X_k \sim \begin{cases} \mathcal{H}_0 : \mathcal{N}\left(\mu_0, \sigma_0^2\right) & \text{if } k < v \\ \mathcal{H}_{\mathrm{L}} : \mathcal{N}\left(\mu_{\mathrm{L},1}, \sigma_0^2\right) & \text{if LOS multipath} \\ \mathcal{H}_{\mathrm{N}} : \mathcal{N}\left(\mu_0, \sigma_{\mathrm{N},1}^2\right) & \text{if NLOS multipath} \end{cases} . \tag{5.43}$$

To analyze this behavior, let us consider the following model for the received GNSS signal from the i-th satellite:

$$r_i(n) = \eta A s_i(n) + \sum_{l=1}^{N_i} A_{i,l} s_i\left(n - \tau_{i,l}\right) e^{j\psi_{i,l}} + w(n), \tag{5.44}$$

where $s_i(n)$ denotes the complex base-band signal arriving from the i-th satellite, including any time-delay and Doppler deviations, A is the signal amplitude, N_i is the number of reflected multipath rays for the i-th satellite, and $\{A_{i,l}, \tau_{i,l}, \psi_{i,l}\}$ the triple formed by the amplitude, delay (given in samples) and phase of each multipath replica, with $w(n)$ the noise at the receiver. The signal and multipath amplitudes are related by the SMR of each component, defined as $\mathrm{SMR}_{i,l} \doteq A/A_{i,l}$, and $\eta = \{0, 1\}$ for NLOS and LOS conditions, respectively.

CASE 1: SAM characterization under LOS conditions

Based on (5.43), we can discriminate between LOS and NLOS identifying whether the mean or variance change in the SAM is prevalent, respectively. In particular, for the LOS case we have the following model:

$$X_{\mathrm{L},k} \sim \begin{cases} \mathcal{H}_0 : \mathcal{N}\left(\mu_0, \sigma_0^2\right) & \text{if } k < v \\ \mathcal{H}_{\mathrm{L}} : \mathcal{N}\left(\mu_{\mathrm{L},1}, \sigma_0^2\right) & \text{if } k \geq v \end{cases}, \tag{5.45}$$

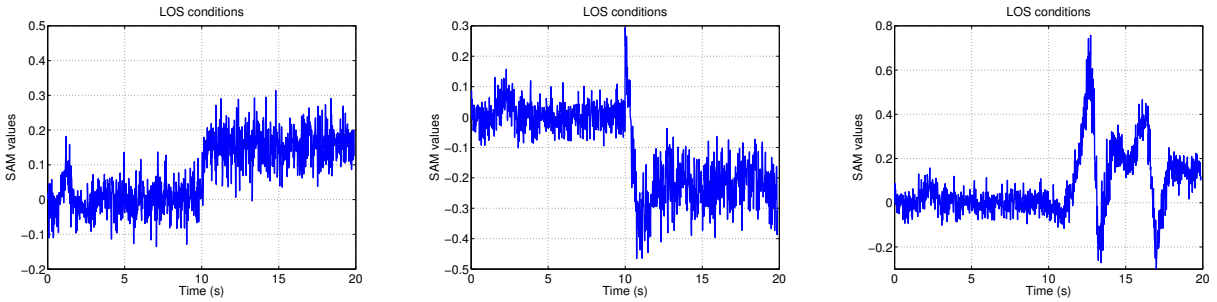

Figure 5.23: (Left) Simulated SAM for LOS multipath propagation with one ray with $\tau = 0.6$ chips and $\psi = 0$ rad; and (middle) with $\tau = 0.3$ chips and $\psi = \pi$ rad, both with SMR = 10 dB. (Right) Simulated SAM for LOS multipath with three rays with SMR $= \{5, 10, 7\}$ dB, $\tau = \{0.3, 0.7, 0.4\}$ chips, $\psi = \{0, \pi, \pi/2\}$ rad, and Doppler frequency $f_{\mathrm{d}} = 1.38 \cdot c \cdot \tau \cdot t_{\mathrm{ch}}$ mHz, and $t_{\mathrm{ch}} = 0.9775$ μs. All simulations have considered a $\mathrm{C/N_0} = 45$ dB-Hz.

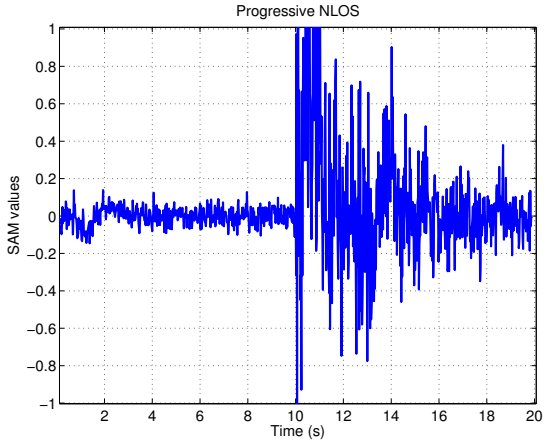

Figure 5.24: SAM for a NLOS multipath propagation simulating a moving receiver approaching the reflector with the following varying multipath parameters: SMR ranging from 20 to 5 dB, delay τ ranging from 0.7 to 0.3 chips, $\psi = 0$–π rad, and $\mathrm{C/N_0} = 45$ dB-Hz.

where $X_{\mathrm{L},k}$ and $\mu_{\mathrm{L},1}$ are the SAM metric measurements and mean when LOS multipath is present. In other words, we model LOS conditions as a mean change of the SAM. This is illustrated in the left plot of Figure 5.23, showing how the mean change prevails. Indeed, we see a mean change but the variance is similar. This result is obtained by simulating an static GNSS receiver with a multipath ray appearing at second 10. The left plot of Figure 5.23 shows no change in the variance. However, there may be situations where the variance may slightly vary when LOS multipath is present. This is shown in the middle plot of Figure 5.23, in which beyond to the mean change, the variance of the SAM slightly increases. Notwithstanding, the mean change is prevalent, so that (5.45) is still valid. Similar results are obtained in the right plot with a more realistic simulation including three multipath rays appearing at second 10 and simulating the movement of the satellite as in [Van92].

CASE 2: SAM characterization under NLOS conditions

Let us consider now the characterization of the SAM under NLOS conditions, given by the following model:

$$X_{\mathrm{N},k} \sim \begin{cases} \mathcal{H}_0 : \mathcal{N}(\mu_0, \sigma_0^2) & \text{if } k < v \\ \mathcal{H}_{\mathrm{N}} : \mathcal{N}(\mu_0, \sigma_{\mathrm{N},1}^2) & \text{if } k \geq v \end{cases}, \tag{5.46}$$

where $X_{\mathrm{N},k}$ and $\sigma_{\mathrm{N},1}^2$ stand for the SAM observations and variance when NLOS multipath is present. Thus, NLOS conditions are modeled as the presence of a change in variance of the SAM.

Figure 5.24 shows this behavior by simulating a moving receiver approaching the reflector with a NLOS multipath ray appearing at second 10. We see how the variance of the SAM increases abruptly just when the multipath appears, and then it decreases as the receiver is approaching the reflector. Regarding the mean of the SAM, it remains constant along the whole simulation. Similarly as for the LOS case, there may be situations where the mean of the SAM slightly varies under NLOS conditions, but the change in the variance will prevail.

Application of QCD: CUSUM-SAM

We see that the SAM follows a Gaussian distribution with known mean before change (i.e. it must be calibrated) but unknown a-priori variance before and after the change. However, these variances can be tuned as mentioned at the introduction of Section 5.3. Doing so, we can use the CUSUM algorithm for the case of a change in both the mean and variance of a Gaussian distribution. In this case, the LLR is given as in (5.35) but using the SAM parameters. This CUSUM gives raise to the same performance in terms of QCD as given for the CUSUM-Power in (5.31), but with the SAM parameters. Alternatively, we can exploit the properties of the SAM to discriminate between LOS and NLOS multipath. On the one hand, we model LOS conditions as a mean change of the SAM, so that the LLR for a Gaussian mean change applies (see (5.15)). On the other hand, NLOS conditions are modeled as the presence of a change in variance of the SAM, thus we can use the LLR for a Gaussian variance change (see (5.40)). In this way, based on the above results, we can use two different CUSUMs, working in parallel, one aimed at detecting a change in the mean of the SAM and another for the variance. Doing so we will detect multipath whenever one of the stopping times declares so, and we will able to discriminate between LOS or NLOS when the change in the mean or variance is declared, respectively. We will analyze this parallel stopping time under the framework of TCD in the next chapter.

5.4 Quickest Multi-Antenna Threat Detection

The SLM and detection techniques presented so far are based on a single-antenna GNSS receiver. In this section, though, we focus on metrics based on multi-antenna GNSS receivers. We will deal with both interference and multipath detection, and a QCD framework will be considered. Before entering into detail on each SLM let us first introduce the general model for multi-antenna GNSS receivers. We assume an antenna array composed of a Uniform Linear Array (ULA) with N_a antenna elements spaced at a distance d. For an incident signal with Direction-of-Arrival (DoA) θ, the array response or steering vector has the following form:

$$\mathbf{s}\left(\theta\right) = \left[1 \quad e^{j2\pi(d/\lambda)\cos(\theta)} \cdots e^{j(N_a-1)2\pi(d/\lambda)\cos(\theta)}\right]^T, \tag{5.47}$$

where λ is the GNSS signal wavelength. The noise at each antenna is modeled as an iid zero-mean complex Gaussian random process with variance σ_w^2, and uncorrelated both spatially and temporally. Next, we will analyze the different proposed SLM for interference and multipath detection in GNSS multi-antenna receivers.

5.4.1 Multi-Antenna interference detection: Eigenvalue ratio test

The detection problem for multi-antenna interference detection becomes

$$\mathbf{r}(n) = \begin{cases} \mathcal{H}_0 : \mathbf{w}(n) & n < v \\ \mathcal{H}_1 : i(n) \cdot \mathbf{s}(\theta_t) + \mathbf{w}(n) & n \geq v \end{cases}, \tag{5.48}$$

where $\mathbf{r}(n) \doteq [r_1(n), \cdots, r_{N_a}(n)]^T$ contains the discrete-time base-band complex samples of the N_a antennas at time n, $i(n)$ denotes an interference signal with DoA θ_t and $\mathbf{w}(n)$ is the noise vector disturbing the received samples. The spatial correlation matrix of this signal, $\mathbf{R}_{rr} \doteq \mathrm{E}\left[\mathbf{r}\mathbf{r}^H\right]$, will be used herein to obtain a SLM, since it shows a clearly different behavior in the presence of interference, as inferred from (5.48). Specifically, we will exploit the idea of the eigenvalue ratio test used in cognitive radio for spectrum sensing [Nad11]. That is, under $\mathcal{H}_0$, all the eigenvalues of $\mathbf{R}_{rr}$ are equal to σ_w^2. However, under $\mathcal{H}_1$ the eigenvalues are different, with a maximum value given by

$$\lambda_{\max}|_{\mathcal{H}_1} = P_{\mathrm{int}} N_{\mathrm{a}} + \sigma_w^2, \tag{5.49}$$

where P_{int} denotes the interference power; and minimum eigenvalue $\lambda_{\min} = \sigma_w^2$. It is then clear that

$$\frac{\lambda_{\max}}{\lambda_{\min}} = \begin{cases} \mathcal{H}_0 : & 1 \\ \mathcal{H}_1 : & 1 + N_{\mathrm{a}}\mathrm{INR} \end{cases}, \tag{5.50}$$

with $\mathrm{INR} \doteq P_{\mathrm{int}}/\sigma_w^2$ the received INR. This behavior, together with results in [Nad11], suggest that the following metric could be fed to the CUSUM:

$$X_{\mathrm{i},k} \doteq \frac{\widehat{\lambda}_{\max}(k)}{\widehat{\lambda}_{\min}(k)}, \tag{5.51}$$

where the subscript $_{\mathrm{i}}$ indicates observations for interference detection. To obtain these observations, the receiver estimates the spatial correlation matrix based on a set of N_{sp} samples of $\mathbf{r}(n)$ as

$$\widehat{\mathbf{R}}_r(k) \doteq \frac{1}{N_{\mathrm{sp}}} \sum_{n=0}^{N_{\mathrm{sp}}-1} \mathbf{r}(n - kN_{\mathrm{sp}})\mathbf{r}^H(n - kN_{\mathrm{sp}}); \tag{5.52}$$

and then computes the maximum and minimum eigenvalue of each $\widehat{\mathbf{R}}_r(k)$, denoted by $\widehat{\lambda}_{\max}(k)$ and $\widehat{\lambda}_{\min}(k)$, respectively.

In order to fed $X_{\mathrm{i},k}$ into the CUSUM algorithm, we first need to know their statistical distribution. A useful result here is the Gaussian approximation for the distribution of extreme eigenvalues, for even low N_{sp} [Wei10], whereby under $\mathcal{H}_0$ we have $X_{\mathrm{i},k} \sim \mathcal{N}(A_{\mathrm{i},0}, \sigma_{\mathrm{i},0}^2)$ with

$$A_{\mathrm{i},0} \approx \frac{a_1 - 1.771 b_1}{a_K - 1.771 b_K} \text{ and } \sigma_{\mathrm{i},0}^2 \approx \frac{b_1^2}{N_{sp} b_K^2}, \tag{5.53}$$

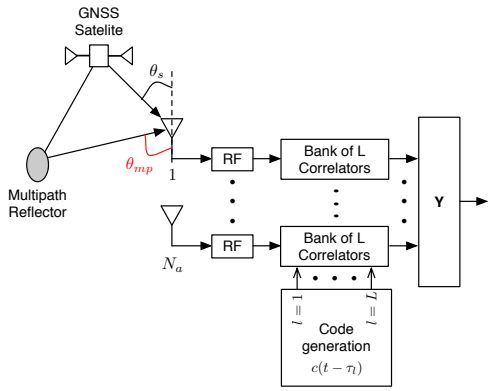

Figure 5.25: Model adopted under the presence of multipath in the used multi-correlator receiver architecture.

and $a_{\frac{1}{K}}$, $b_{\frac{1}{K}}$ given by

$$a_{\frac{1}{K}} \doteq \left(\sqrt{N_a} \pm \sqrt{N_{sp}}\right)^2 \text{ and } b_{\frac{1}{K}} \doteq \left(\sqrt{N_a} \pm \sqrt{N_{sp}}\right)\left(\sqrt{\frac{1}{N_a}} \pm \sqrt{\frac{1}{N_{sp}}}\right)^{1/3}. \tag{5.54}$$

Therefore, we can formulate the change detection problem as:

$$X_{\mathrm{i},k} \sim \begin{cases} \mathcal{H}_0 : \mathcal{N}\left(A_{\mathrm{i},0}, \sigma^2_{\mathrm{i},0}\right) & k < v \\ \mathcal{H}_1 : \mathcal{N}\left(A_{\mathrm{i},1}, \sigma^2_{\mathrm{i},1}\right) & k \geq v \end{cases}, \tag{5.55}$$

where $A_{\mathrm{i},1} \approx 1 + N_{\mathrm{a}}\mathrm{INR}$ and, variance given as in [Wei10] by $\sigma^2_{\mathrm{i},1} = (N_{\mathrm{a}}A_{\mathrm{i},1} + 1)^2 + 1/N_{\mathrm{sp}}$. This leads to a mean and variance change problem. A way to proceed in this case is to fix a certain value for $A_{\mathrm{i},1}$ according to the minimum INR that one expects to detect. In this way, a minimum change detection is set allowing the detection of any larger change caused by higher power interferences. More details on the eigenvalue ratio test for interference detection in GNSS see [Ege14a].

5.4.2 Multi-Antenna multipath detection

Figure 5.25 shows the situation when a GNSS signal arrives with DoA θ_{s} at the antenna array and a multipath ray is present with DoA θ_{mp}. Let us consider we operate at the output of a bank of correlators with a CI time equal to one code period (e.g. 1 ms in the case of GPS L1 C/A signals). We also consider a total of L correlation outputs per array arm, which is possible due to the multi-correlation architecture in Figure 5.25. With this scheme, the following $(N_{\mathrm{a}} \times L)$ received data matrix can be generated:

$$\mathbf{Y}_k \doteq \left[\mathbf{y}\left(1;k\right), \cdots, \mathbf{y}\left(L;k\right)\right], \tag{5.56}$$

where $\mathbf{y}\left(l;k\right)$ stands for the $(N_{\mathrm{a}} \times 1)$ vector with the l-th correlator output at different antennas at the k-th CI period. Depending on the presence or absence of multipath, the two following

hypotheses can be written, respectively:

$$\mathbf{y}\left(l;k\right)=\begin{cases}\mathcal{H}_0:\alpha_0(l)\cdot\mathbf{s}(\theta_{\mathrm{s}})+\mathbf{n}(l) & k<v\\ \mathcal{H}_1:\alpha_0(l)\cdot\mathbf{s}(\theta_{\mathrm{s}})+\alpha_1(l)\cdot\mathbf{s}(\theta_{\mathrm{mp}})+\mathbf{n}(l) & k\geq v\end{cases}, \tag{5.57}$$

where the complex amplitude α_j, with $j=\{0,1\}$, corresponds to the LOS and multipath post-correlation outputs components, respectively. Finally, $\mathbf{n}(l;k)$ is the noise vector whose components are the post-correlation zero-mean complex Gaussian noise with variance σ^2. Without loss of generality only one multipath ray is assumed. The correlator output samples are taken on either side of the prompt correlator. Thus, the first $(L-1)/2$ samples correspond to early correlators, whereas the last $(L-1)/2$ correlator samples correspond to the late correlators. Moreover, we will assume that the correlation points are uniformly distributed in the range of ±1 chip period. The single-correlator receiver architecture case, with a single correlator output corresponding to the prompt correlator sample will be also considered. In such a case, $\mathbf{y}(l;k)$ is denoted by $\mathbf{y}(k)$.

Single-Correlation multi-antenna (SC-MA) detector

Let us first investigate the Single-Correlation Multi-Antenna (SC-MA) metric. In order to obtain the required observations for the CUSUM algorithm, we will first obtain $\mathbf{z}(m)\doteq\mathbf{y}(m)\odot\mathbf{s}^*(\theta_s)$, with $\odot$ the Schur-Hadamard product. By doing so, we remove the effects of the desired signal and obtain a reference signal that provides us information on the presence of multipath. This can be observed by taking a look at the resulting signal model:

$$\mathbf{z}(k)=\begin{cases}\mathcal{H}_0:\alpha_0\mathbf{1}+\mathbf{n}(k) & k<v\\ \mathcal{H}_1:\alpha_0\mathbf{1}+\alpha_1\left[\mathbf{s}\left(\theta_t\right)\odot\mathbf{s}^*\left(\theta_s\right)\right]+\mathbf{n}(k) & k\geq v\end{cases}, \tag{5.58}$$

with $\mathbf{1}$ a $(N_a\times1)$ all-ones vector. Based on $\mathbf{z}(k)$, the following scalar observations are proposed to implement the CUSUM algorithm:

$$X_{\mathrm{mp},k}=\frac{1}{N_{\mathrm{a}}-1}\sum_{n=2}^{N_{\mathrm{a}}}\left(\left[\mathbf{z}(k)\right]_n-\left[\mathbf{z}(k)\right]_1\right) \tag{5.59}$$

where $\left[\mathbf{z}(k)\right]_j$ indicates the j-th component of $\mathbf{z}(k)$. This expression denotes the average of the differences between antenna pairs, once the LOS steering vector has been removed. In the absence of multipath the result is expected to be negligible, since $\left[\mathbf{z}(k)\right]_n$ is the same for all antennas. However, in the presence of multipath it contains the different contribution of multipath in different antennas, and thus multipath becomes detectable. To implement the CUSUM algorithm based on the above observations, considering (5.58), we can assume that $X_{\mathrm{mp},k}$ is Gaussian distributed with mean and variance under $\mathcal{H}_0$ respectively given by

$$A_{\mathrm{mp},0}=0\text{ and }\sigma^2_{\mathrm{mp},0}=\frac{2\sigma^2}{N_a-1}. \tag{5.60}$$

On the other hand, when a deterministic multipath ray is present, we have

$$A_{\mathrm{mp},1}=\frac{\alpha_1}{N_a-1}\cdot\sum_{n=1}^{Na-1}\left(e^{j\omega(\theta_t,\theta_s)n}-1\right) \tag{5.61}$$

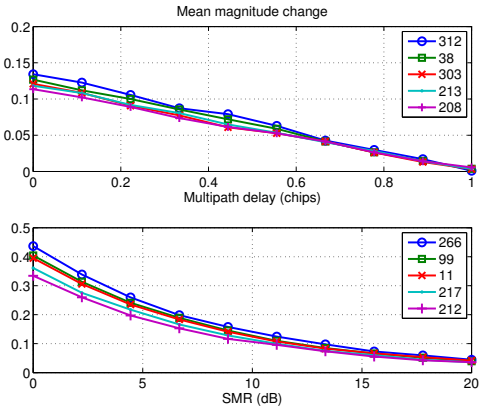

Figure 5.26: (Up) Sensitivity of $|A_{\text{mp},1}|$ as a function of multipath delay at SMR = 10 dB; and (down) SMR at 0.1 chips delay, for $\theta_t = 20^\circ$ and $\theta_s = 95^\circ$. A $\text{C/N}_0 = 45$ dB-Hz has been considered.

with $\omega(\theta_t, \theta_s) \doteq 2\pi \frac{d}{\lambda} (\cos(\theta_t) - \cos(\theta_s))$. Hence, we can formulate the change detection problem as

$$X_{\text{mp},k} \sim \begin{cases} \mathcal{H}_0 : \mathcal{N}\left(0, \sigma^2_{\text{mp},0}\right) & k < v \\ \mathcal{H}_1 : \mathcal{N}\left(A_{\text{mp},1}, \sigma^2_{\text{mp},0}\right) & v \geq v \end{cases}. \tag{5.62}$$

While $\sigma^2_{\text{mp},0}$ can easily be obtained in practice, $A_{\text{mp},1}$ depends on the amplitude and DoA of the multipath ray, and thus it remains unknown. The solution here is to replace the unknown parameter by its MLE (i.e. GLR) as suggested in [Ege14a] and [Ege14b]. The difficulties on fixing $A_{\text{mp},1}$ are illustrated in Figure 5.26. The upper and lower plots show the value of $A_{\text{mp},1}$ for different code-delays of the multipath ray and different SMRs. Moreover, five randomly chosen carrier phases have been considered on the amplitude of the multipath ray α_1. The results evidence the complexity in selecting a minimum change magnitude. This is because depending on the multipath phase and its DoA with respect to the LOS, the change magnitude will take different values even for a fixed multipath delay and SMR.

Multi-Correlation multi-antenna (MC-MA) detector

In order to derive a metric for the Multi-Correlation multi-antenna (MC-MA) case, we will consider the covariance matrix of $\mathbf{Y}_k$, which can be estimated as [Clo11]

$$\widehat{\mathbf{W}} \doteq \text{Cov}\left[\mathbf{Y}_k\right] = \widehat{\mathbf{R}}_{yy} - \frac{\mathbf{r}_{yc}\mathbf{r}^H_{yc}}{P_c}, \tag{5.63}$$

with the following definitions,

$$\widehat{\mathbf{R}}_{yy} \doteq \frac{1}{L}\mathbf{Y}\mathbf{Y}^H; \quad \mathbf{r}_{yc} \doteq \frac{1}{L}\mathbf{Y}\mathbf{r}^H_{cc}, \tag{5.64}$$

$$P_c \doteq \frac{1}{L}\mathbf{r}_{cc}\mathbf{r}^H_{cc}, \tag{5.65}$$

and $\mathbf{r}_{cc} \doteq [R_{cc}(\tau_1), \cdots, R_{cc}(\tau_L)]$, where $R_{cc}(\tau_i)$ is the ACF of the GNSS code evaluated at the time delay τ_i. For the sake of clarity the index k has been omitted, keeping in mind that each data matrix corresponds to a given CI interval. From [Clo11] it is known that $\widehat{\mathbf{W}}$ is a diagonal matrix under $\mathcal{H}_0$, whereas under $\mathcal{H}_1$, it departs from diagonal due to the presence of correlated signals among antenna elements. Thus, the following metric is proposed with the aim

of measuring to which extent $\widehat{\mathbf{W}}$ departs from a diagonal matrix:

$$\rho_d(k) = 2LN_a \ln \left(\frac{\frac{1}{N_a}\sum_{i=1}^{N_a} \widehat{\lambda}_i(k)}{\left[\prod_{i=1}^{N_a} \widehat{\lambda}_i(k)\right]^{1/N_a}} \right), \tag{5.66}$$

where $\widehat{\lambda}_i(k)$ is the i-th eigenvalue associated to the estimated covariance matrix $\widehat{\mathbf{W}}$ at CI time k. This metric follows a chi-squared distribution with $N_a^2 - 1$ degrees of freedom under $\mathcal{H}_0$ [Clo11], thus

$$\begin{aligned} \mathcal{H}_0 &: \mu_0 = N_a^2 - 1, \\ \mathcal{H}_1 &: \mu_1 > N_a^2 - 1, \end{aligned} \tag{5.67}$$

with μ_i for $i = \{0, 1\}$ the mean of $\rho_d(k)$ under hypothesis $\mathcal{H}_i$. Nevertheless, the distribution under $\mathcal{H}_1$ is unknown.

Hence, the use of the O-CUSUM is proposed by using the following modified metric:

$$\rho_{\text{mc}}(k) \doteq \rho_d(k) - p. \tag{5.68}$$

In this way, by selecting a proper p, the mean of $\rho_{\text{mc}}(k)$ before change will be negative, but it will become positive after the change. As in previous cases the choice of the offset p should be large enough to assure a negative mean before change and provide a certain false alarm rate. But at the same time, p should be small enough to maintain a positive mean after change for the test metric. From (4.64) and using ρ_{mc} we are able to adjust the mean time between false alarm through the non-zero root, s_0, of the following equation:

$$\mathrm{E}_\infty \left[e^{s \cdot \rho_d(k)} \right] e^{-s \cdot p} = 1, \tag{5.69}$$

with E_∞ the expectation operator when hypothesis $\mathcal{H}_0$ is true. We also know that under $\mathcal{H}_0$, $\rho_d(k)$ is a chi-squared variable with $N_a^2 - 1$ degrees of freedom. Then, the characteristic function of $\rho_d(k)$, under $\mathcal{H}_0$, is given by

$$\mathrm{E}_\infty \left[e^{s \cdot \rho_d(k)} \right] = (1 - 2s)^{-\frac{N_a^2 - 1}{2}}, \tag{5.70}$$

and thus s_0 turns out to be the non-zero root of

$$e^{-s \cdot p} = (1 - 2s)^{\frac{N_a^2 - 1}{2}}, \tag{5.71}$$

which can be solved numerically. For more details and simulations see [Ege14b].

5.5 Conclusions

In this chapter we have moved forward into the design of local threat detectors. This has been motivated by the widespread use of GNSS-based critical applications in terrestrial environments, where local threats abound. Thus integrity analyses on the received GNSS signal must be taken into consideration, thus leading to the so-called SLI. This new level of integrity is needed for the safely use of critical applications in terrestrial environments, driving to the definition of two requirements. These are, as shown in Figure 5.1, (i) the use of signal observable metrics able to

detect any misleading behavior of the received signal, and (ii) to perform a timely detection so that the user can be alerted as soon as the threat appears. The discussion of this chapter has addressed these two requirements by providing a complete set of detection metrics to detect any type of interference and multipath propagation. Furthermore, QCD has been adopted in order to promptly detect these local threats. Once several SLM have been presented and analyzed, we have at our disposal a complete set of SLM to detect any kind of local threats, and we have therefore fulfilled the first requirement for SLI. At this point, we have moved a step ahead by showing the application of QCD into the above mentioned SLM. In particular, the same flow as for optimal detectors design have been carried out; that is, we have first shown the theoretical statistical characterization of these metrics, followed by the ultimate design of the corresponding CUSUM algorithm. Both the statistical characterization and performance of the CUSUM algorithms have been corroborated.

After the theoretical and simulation analysis, we can conclude that most of the detectors can be designed as Gaussian mean, variance or mean and variance change detectors. This is the case for the kurtosis, power, C/N_0, DLLout, and SAM metrics. Other metrics, such as the spectrogram, have a completely unknown statistical characterization. Notwithstanding, it has been shown that they can be fed into any of the above mentioned Gaussian changes. The only exception that is not be fed into the CUSUM is the histogram SLM. This is because its distribution under nominal conditions is known, but the one under faulty conditions is not. Thus, we have relied on the O-CUSUM alternative to implement the histogram-based detection. As a general conclusion of this chapter, the use of the proposed detectors allows the provision of SLI to GNSS users, making the user aware when the GNSS signal quality and trust are being threatened by interference or multipath. This is of potential interest in practical applications involving GNSS signal quality monitoring, such as SCA and LCA. Doing so, we have bridged the gap between QCD and threat detection in GNSS. Notwithstanding, for integrity purposes a bounded delay given by the TTA is desirable. For this reason, the TCD framework should be considered instead. Moreover, for a fully integrity service, the connection between signal quality monitoring and integrity algorithms must be investigated. These two points are the ultimate goal of this dissertation and they will be considered in Chapter 6.

Chapter 6

Improving GNSS Integrity Through Signal-Level RAIM

6.1 Introduction

Up to now we have studied RAIM algorithms and its design for critical applications. Nevertheless, RAIM (as well as SBAS) schemes were designed for civil aviation applications, so that many limitations can be found in this design when moving into terrestrial environments. Our proposed approach to overcome these limitations is based on applying SCD to the very first stages of the GNSS receiver. This approach is in line with the substantial recent interest in critical applications in which an established maximum tolerable delay m_{d} is desirable. Detections declared after this tolerable delay are actually considered as unbounded. In these applications, which is also the case of GNSS integrity, the TCD criterion introduced in Chapter 4 is very appropriate. Based on these considerations, the need of a new family of integrity schemes based on SCD, and particularly on TCD, is apparent. This will highlighted in this chapter using numerical simulations to compare current RAIM techniques with novel SCD solutions. The first attempts connecting integrity algorithms and SCD dates back to the work of Nikiforov [Nik96, Nik97a, You98, Nik02]. These references introduced the concept of sequential RAIM and they took into account the probability minimizing criterion considered in TCD. Nevertheless, at that time the framework of TCD was not know as it is known nowadays, so the use of the CUSUM algorithm was considered. Moreover, Nikiforov did not consider the application of signal-level integrity.

In this chapter we will focus on the implementation of both RAIM and SLI algorithms. To do so, we firstly show the integration of the multipath detectors, presented in Chapter 5, with the PVT and integrity module (PVT+I), including the RAIM algorithm implementation. This is done in order to provide the additional SLI, leading to what we call PVT+I^2 module. The structure of this chapter is as follows: In Section 6.2 we treat the current implementation of RAIM-based PVT integrity in GNSS. We will also consider the implementation of RAIM based on SCD solutions, the so-called sequential RAIM algorithms. This will confirm that the TCD framework is the best suited for the integrity monitoring problem, and this is why we will consider it for the rest of the chapter. Next, Section 6.3 deals with the implementation of the sig-RAIM algorithm. This implementation is based on the philosophy of RAIM algorithms but

considering SLI. This level of integrity is given by using the SLM investigated in Chapter 5 to detect local threats in a TCD framework. We will show with numerical simulations the goodness of fit of our theoretical findings and simulation results. In addition, the simulation will show the better performance of the proposed solution with respect to other available solutions in the literature. Finally, Section 6.4 will consider the integration of both PVT+I and sig-RAIM algorithms, and then Section 6.5 concludes the chapter. Most of the work in this chapter is based on our own publications in [ER15d, ER17c, ER17d].

6.2 PVT Integrity (PVT+I)

This section is devoted at introducing the current provision of PVT integrity in GNSS based on RAIM techniques. In the following we will use the terminology of integrity monitoring requirements, that is, AL, TTA, Integrity risk ($\mathcal{P}_{\text{risk}}$), and false alarm probability ($\mathcal{P}_{\text{fa}}$) previously defined in Section 2.5. Using this terminology, it is important to recall the role of RAIM techniques in GNSS receivers in order to further link them with the TCD framework to be presented in this dissertation. This is shown in the following.

6.2.1 RAIM implementation

As explained in Section 3.5, RAIM algorithms calculate a scalar test statistic that provides information about the positioning error. This statistic is based on the redundancy of observables coming from different satellites in view and/or the consistency of the position solution. Traditional RAIM approaches are snapshot-based, meaning that they compute the test statistic using a block of pseudoranges at the present time instant. At each epoch the test statistic is evaluated, without information of past samples, to decide whether there is an integrity fault within the snapshot time (i.e. $\mathcal{H}_1$) or not (i.e. $\mathcal{H}_0$). Based on the presented formulation for detection theory in Chapter 4, the underlying problem can be modeled as

$$Z_n \sim \begin{cases} \mathcal{H}_0: & f_0(Z_n) = \chi^2_k \\ \mathcal{H}_1: & f_1(Z_n, \lambda) = \chi^2_k(\lambda) \end{cases}, \tag{6.1}$$

where Z_n denotes the test statistic at snapshot n, f_i, with $i = \{0, 1\}$, is the pdf of Z_n under $\mathcal{H}_i$ and λ is the change parameter introduced by a faulty measurement in the pdf of the test statistic. The equality follows if we use the SSE test statistic given in (3.20). For the implementation of RAIM, the following parameters need to be considered:

- *Design Parameters ($\widetilde{\beta}$, $\widetilde{\alpha}$)*: The RAIM algorithm is designed to meet a maximum permitted integrity risk $\widetilde{\beta}$ and probability of false alarm $\widetilde{\alpha}$ such that $\mathcal{P}_{\text{risk}}(\lambda) \leq \widetilde{\beta}$ and $\mathcal{P}_{\text{fa}}(h) \leq \widetilde{\alpha}$, respectively.
- *Failure Detection Threshold (h)*: RAIM detects a failure whenever the test statistic exceeds a threshold (i.e. $Z_n \geq h$). The threshold is chosen so that $\mathcal{P}_{\text{fa}}(h) \leq \widetilde{\alpha}$.
- *Minimum Detectable Change Parameter (λ_{det})*: When a threat is present, the minimum detectable change parameter is the value of λ such that $\mathcal{P}_{\text{risk}}(\lambda) = \widetilde{\beta}$.

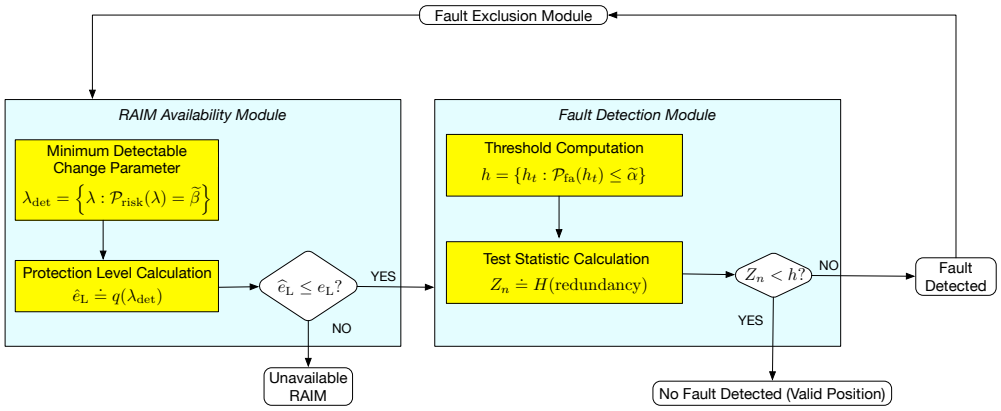

Figure 6.1: RAIM algorithm operation flowchart.

- *PL ($\widehat{e}_{\text{L}}$)*: The PL is calculated as the positioning error that would generate a change parameter equal to λ_{det}, that is

$$\hat{e}_{\text{L}} \doteq q(\lambda_{\text{det}}), \tag{6.2}$$

  with $q(\cdot)$ a function that maps change parameters into position errors.

Based on the above parameters, the RAIM algorithm implementation at the GNSS receiver consists of two modules (see Figure 6.1):

1. The **RAIM availability check**, which is in charge of checking if the RAIM algorithm can provide the established requirements in terms of $\widetilde{\alpha}$, $\widetilde{\beta}$ and TTA.

2. The **FDE module**, in case the RAIM algorithm is available, is in charge of detecting whether an integrity fault is present. In that case it tries to exclude the faulty measurement from the PVT computation.

Thus, at each epoch, $\widehat{e}_{\text{L}}$ is calculated depending on the requirements of integrity risk $\widetilde{\beta}$ and false alarm rate $\widetilde{\alpha}$. If the computed PL exceeds the AL (i.e. $\widehat{e}_{\text{L}} > e_{\text{L}}$), RAIM is considered not available because it cannot monitor integrity with the desired requirements. In other words, the estimated position is not trustworthy. If the PL is equal to or lower than the AL, the RAIM is considered to be available and it proceeds to check whether the estimated position is faulty or not. The test statistic is computed, and if it is higher than the fixed threshold h (given by the false alarm rate $\widetilde{\alpha}$), a fault is detected, and the position should not be used. Otherwise, the position is valid. In case a fault is detected, the RAIM algorithm can optionally perform FDE, which tries to eliminate the erroneous measurements.

In the following sections, we analyze and compare different methods for RAIM, including traditional snapshot RAIM and some new sequential algorithms that we have proposed by exploiting the SCD framework introduced in Chapter 4. It is shown that sequential algorithms can give better availability performance than currently used snapshot methods. One of the reasons is that currently existing RAIM algorithms are not designed in an optimal way, in the sense that they are not designed taking into account the traditional design flow of optimal detectors presented in Chapter 4. Instead, they are rather heuristic methods based on some metric relying on some consistency check. It is for this reason that we consider different solutions based on the presented optimal criteria studied in Chapter 4. In particular, we will consider the

traditional snapshot RAIM as a benchmark for comparison, and sequential solutions based on the CUSUM, NP and FMA stopping times. This is done with the aim of comparing with the QCD, classical and TCD, respectively.

6.2.2 Snapshot RAIM

Snapshot RAIM is based on the SSE χ^2-statistic, denoted here by $Z_n = \mathbf{e}_n^T\mathbf{e}_n$, with stopping time given by

$$T_{\mathrm{s}} \doteq \inf\left\{n \geq 1 : \mathbf{e}_n^T\mathbf{e}_n \geq h\right\}, \tag{6.3}$$

where $\mathbf{e}_n = \mathbf{y}_n - \widehat{\mathbf{y}}_n$ is the residual vector computed by (3.19) at snapshot n, and h is a given threshold. We consider the traditional pseudorange error model introduced in Section 3.4.4 assuming Gaussian distributed pseudoranges. A mean and variance equal to the true range and the UERE, respectively, is assumed under nominal conditions. On the other hand, an additional bias is assumed under the presence of a threat. Let $\mathbb{P}_0$ and $\mathbb{P}_1$ be the probability measures under nominal and faulty conditions, respectively. Thus, the PFA of the test $Z_n = \mathbf{e}_n^T\mathbf{e}_n \geq h$ is $\mathrm{P_F} \doteq \mathbb{P}_0\left(Z_n \geq h\right)$ and the PMD is $\mathrm{P_M} \doteq \mathbb{P}_1\left(Z_n < h\right)$. We know that under $\mathcal{H}_0$ the SSE is distributed according to the χ^2 distribution with $r = N_{\mathrm{sat}} - 4$ degrees of freedom, where N_{sat} is the number of satellites in view. Therefore, $\mathrm{P_F} = 1 - \Gamma(h/\sigma^2_{\mathrm{UERE}}, r)$, where $\Gamma(h, r)$ is the cdf of the central chi-squared distribution with r degrees of freedom, evaluated at h. This is the case for a single sample of the SSE test statistic. Thus, in order to guarantee a given level of PFAP, given by $\widetilde{\alpha}$, we have

$$h = \sigma^2_{\mathrm{UERE}}\Gamma^{-1}\left(\left(1-\widetilde{\alpha}\right)^{\delta}, r\right), \tag{6.4}$$

with δ the number of SSE samples we want to guarantee a PFAP equal to $\widetilde{\alpha}$.

Let us consider now the case when there is an integrity threat, so that a fault occurs in pseudorange l introducing a bias b_l. In this case, the expectation of the residual vector is given by $\mathrm{E}_1\left[\mathbf{e}\right] = (\mathbf{I} - \mathbf{P_H})\boldsymbol{\eta}$, with $\boldsymbol{\eta} = [0, \ldots, 0, b_l, 0, \ldots, 0]^T$ a vector with all zeros except in the l-th component, $\mathbf{P_H} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ and $\mathbf{H}$ the geometry matrix. Hence, under $\mathcal{H}_1$, the SSE statistic is distributed according to the non-central chi-squared distribution with r degrees of freedom and the non-centrality parameter $\lambda = (b_l^2\pi_{ll})/\sigma^2_{\mathrm{UERE}}$, where $\pi_{ll} = [\mathbf{I} - \mathbf{P_H}]_{l,l}$. Therefore, the PMD is given by $\mathrm{P_M} = \Gamma(h/\sigma^2_{\mathrm{UERE}}, r; \lambda)$, with $\Gamma(h, r; \lambda)$ the cdf of the non-central chi-squared distribution with r degrees of freedom and non-centrality parameter λ. Now, considering the TTA, the worst-case PUD (or integrity risk) is given by

$$\mathcal{P}_{\mathrm{risk}} \doteq \mathrm{P_{ud}} = \sup_{v\geq 1}\mathbb{P}_v\left(T_{\mathrm{s}} - v + 1 > m\right) = \mathbb{P}_1\left(\bigcap_{i=1}^{m}\left\{\mathbf{e}_i^T\mathbf{e}_i < h\right\}\right) = \left[\Gamma\left(\frac{h}{\sigma^2_{\mathrm{UERE}}}, r; \lambda\right)\right]^m, \tag{6.5}$$

with m the number of snapshots within the TTA. We see that the PUD is a function of the fault magnitude b_l. In practice, this magnitude is completely unknown but, we can compute the worst-case fault magnitude b_l^* for each visible satellite. This is done by taking into account the satellites geometry and AL as

$$|b_l|^* = \frac{\mathrm{HAL}}{\sqrt{a_{1l}^2 + a_{2l}^2}}, \tag{6.6}$$

with $a_{ll} = [\mathbf{A}]_{l,l}$ and $\mathbf{A} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$.

6.2.3 Sequential RAIM

Among sequential solutions, the ones relying on the TCD framework are actually the most promising ones. To better illustrate this statement, let us introduce the intuition that make us believe that the TCD framework is convenient for integrity monitoring. In general, integrity norms and standards establish values for TTA and AL. They also limit by $\{\widetilde{\alpha}, \widetilde{\beta}\}$ the probability of false alarm, $\mathcal{P}_{\text{fa}}$, and integrity risk, $\mathcal{P}_{\text{risk}}$, respectively, resulting from the test $Z_n \geq h$. With these requirements, integrity algorithms fix the detection threshold h so that $\mathcal{P}_{\text{fa}}(h) \leq \widetilde{\alpha}$, and then they monitor the integrity risk, given by its definition as

$$\mathcal{P}_{\text{risk}} \doteq \Pr\left(\{Z_n < h\} \bigcap \{n > v + t_{\text{TTA}}\} \,\middle|\, \mathcal{H}_1 \right), \tag{6.7}$$

where v stands for the unknown time in which a fault appears, and t_{TTA} is the TTA measured in samples. In order to monitor the integrity risk, the flow in Figure 6.1 is applied by RAIM algorithms, which are characterized by the test $Z_n \geq h$. Whenever the integrity algorithm is available, an alarm will be raised based on this test. In this sense, it would be interesting to find the test statistic Z_n that minimizes the integrity risk given by (6.7). Doing so, λ_{det} would be reduced, and then the PL would be reduced as well, increasing the availability of the integrity algorithm. To this end, it is worth noting that when using some other test $G_n \geq h$ instead of $Z_n \geq h$, and $t_{\text{TTA}} = m$, the probability measure in (6.7) could be written as the PUD used in TCD problems, $\text{P}_{\text{ud}}(T)$ with $T = \inf\{n \geq 1 : G_n \geq h\}$, given by (4.68). Hence, the problem of minimizing the integrity risk is closely connected to the TCD problem.

Let us define the following hypotheses based on the statistical model for the pseudorange under nominal and faulty conditions, respectively:

$$\mathbf{y}_n \sim \begin{cases} \mathcal{H}_0 : \mathcal{N}\left(\mathbf{H}\Delta\mathbf{u}, \sigma^2_{\text{UERE}}\mathbf{I}\right) & n < v \\ \mathcal{H}_1 : \mathcal{N}\left(\mathbf{H}\Delta\mathbf{u} + \boldsymbol{\eta}, \sigma^2_{\text{UERE}}\mathbf{I}\right) & v \geq v \end{cases}, \tag{6.8}$$

where $\boldsymbol{\eta}$ is the *informative* parameter and $\Delta\mathbf{u}$ is the *nuisance* parameter denoting the unknown position of the user. We are interested in detecting a change from 0 to $\boldsymbol{\eta}$, while considering $\Delta\mathbf{u}$ as unknown. Following [Nik02], a proper design is to consider a *mini-max algorithm* consisting on finding the least favorable value of $\Delta\mathbf{u}$ for which the KL divergence is minimum, and in computing the LLR for this value. Doing so, we take the worst-case detection delay and PUD among all possible values of $\Delta\mathbf{u}$. For the above defined hypotheses, this LLR becomes

$$\text{LLR}^*(n) = \frac{|b_l|\, e_{l,n}}{\sigma^2_{\text{UERE}}} - \frac{b_l^2 \pi_{ll}}{2\sigma^2_{\text{UERE}}}, \tag{6.9}$$

where $e_{l,n} \doteq [\mathbf{e}_n]_l$ is the l-th element of the residual vector $\mathbf{e}_n$ at snapshot n. Since the sign of the bias b_l is unknown, we should repeat the selected stopping time twice. Next, we consider three different two-sided stopping times, namely, the CUSUM-, FMA-, and NP-based stopping times. The first one was already considered in [Nik02] by Nikiforov, and compared with the snapshot RAIM. The two last solutions are proposed by us and submitted to [ER17c].

CUSUM-Based RAIM

Based on the above discussion, the two-sided CUSUM-based stopping time is given by

$$T_{\mathrm{c}}^{l} \doteq \min\left\{T_{\mathrm{c}}^{l+}, T_{\mathrm{c}}^{l-}\right\}, \quad T_{\mathrm{c}}^{l\pm} \doteq \inf\left\{n \geq 1 : g_n^{\pm} \geq h\right\} \text{ and } g_n^{\pm} = \left(g_{n-1}^{\pm} \pm \mathrm{LLR}^*(n)\right)^+. \tag{6.10}$$

The number l of faulty satellite is also unknown, for this reason, we should repeat the stopping time above for each visible satellite, hence, the sequential fault detection algorithm is given by a collection of N_{sat} parallel two-sided CUSUMs; that is,

$$T_{\mathrm{c}} \doteq \min\left\{T_{\mathrm{c}}^{1}, \ldots, T_{\mathrm{c}}^{N_{\mathrm{sat}}}\right\}. \tag{6.11}$$

It follows from the above discussion that this algorithm has $N_{\mathrm{sat}}+1$ tuning parameters, namely, the threshold h and the magnitudes b_1,..., $b_{N_{\mathrm{sat}}}$. Regarding the threshold h, from (4.102)–(4.103) it can be chosen, so that $\mathrm{P}_{\mathrm{fap}}(h) \leq \widetilde{\alpha}$, as

$$h = -\ln\left(\frac{\widetilde{\alpha}}{m_\alpha}\right), \tag{6.12}$$

with m_α the number of samples (snapshots n) we want to guarantee that $\mathrm{P}_{\mathrm{fap}}(h) \leq \widetilde{\alpha}$. Regarding the bias magnitudes b_i for $i = 1, \ldots, N_{\mathrm{sat}}$ they can be fixed accordingly to the HAL as in (6.6), thus leading, again from (4.102)–(4.103), to the following bound for the worst-case PUD:

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{c}}, m\right) \leq \Phi\left(\frac{h - m \cdot \mu_{\mathrm{s}}}{\sigma_{\mathrm{s}}\sqrt{m}}\right), \tag{6.13}$$

where $\Phi(x)$ stands for the standard Gaussian cdf, and

$$\mu_{\mathrm{s}} = \frac{|b_l^*|^2 \pi_{ll}}{2 \cdot \sigma_{\mathrm{UERE}}^2}; \quad \sigma_{\mathrm{s}}^2 = 2 \cdot \mu_{\mathrm{s}}. \tag{6.14}$$

Without loss of generality, we have considered that only the l-th GNSS measurement is contaminated and that the bias v_l is positive.

Proposed FMA-Based RAIM

Similarly, we define the two-sided FMA-based stopping time as

$$T_{\mathrm{f}}^{l} \doteq \min\left\{T_{\mathrm{f}}^{l+}, T_{\mathrm{f}}^{l-}\right\}; \quad T_{\mathrm{f}}^{l\pm} \doteq \inf\left\{n \geq 1 : \pm \sum_{i=n-m+1}^{m} \mathrm{LLR}^*(i) \geq h\right\}, \tag{6.15}$$

and then, for N_{sat} satellites in view, we have

$$T_{\mathrm{f}} \doteq \min\left\{T_{\mathrm{f}}^{1}, \ldots, T_{\mathrm{f}}^{N_{\mathrm{sat}}}\right\}. \tag{6.16}$$

Now, taking into account the bounds for a FMA stopping time, given in Theorem 8 and results in Corollary 1, we can fix the detection threshold h from (4.97), so that $\mathrm{P}_{\mathrm{fap}}(h) \leq \widetilde{\alpha}$, as

$$h = \sqrt{m \cdot \sigma_{\mathrm{s}}^2} \cdot \Phi^{-1}\left[(1-\widetilde{\alpha})^{1/m_\alpha}\right] - m \cdot \mu_{\mathrm{s}}. \tag{6.17}$$

For the worst-case PUD, a similar bound as for the CUSUM-based RAIM is satisfied. Details of how these results can be derived are found in Section 6.3.2.

Proposed NP-Based RAIM

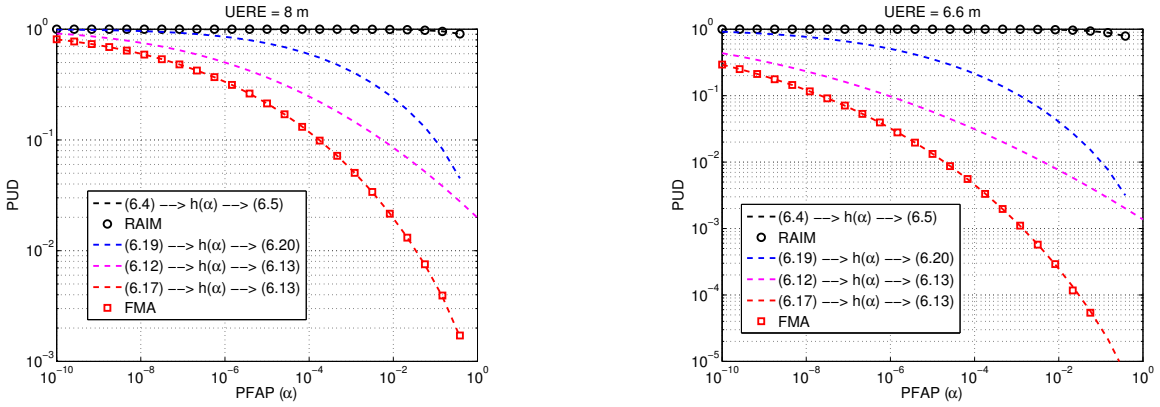

Figure 6.2: Comparison between simulated (markers) and theoretical (dashed line) ROC for the traditional RAIM (black), optimal FSS (blue), CUSUM (magenta), and FMA (red). (Left) Big and (right) small UERE values.

Finally, let us consider the stopping time based on the optimal NP test in the sense of the optimal TCD criterion; that is

$$T_{\mathrm{np}}^{l} \doteq \min\left\{T_{\mathrm{np}}^{l+}, T_{\mathrm{np}}^{l-}\right\}; \quad T_{\mathrm{np}}^{l\pm} \doteq N \inf\left\{n \geq 1 : \pm \sum_{i=(n-1)N+1}^{nN} \mathrm{LLR}^{*}(i) \geq h\right\}, \tag{6.18}$$

with N the optimal sample size of the NP test for the TCD problem given asymptotically in Theorem 10 by $N \sim \lceil m/2 \rceil$ (see (4.138)). Now, taking into account the results for the optimal LRT in the TCD problem, given in Proposition 9, the detection threshold h can be fixed from (4.132) as

$$h = \sqrt{N \cdot \sigma_{\mathrm{s}}^{2}} \cdot \Phi^{-1}\left[(1-\widetilde{\alpha})^{1/\lceil m_{\alpha}/N \rceil}\right] - N \cdot \mu_{\mathrm{s}}, \tag{6.19}$$

thus leading from (4.133) to

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{np}}\right) \sim \Phi\left(\frac{h - N \cdot \mu_{\mathrm{s}}}{\sigma_{\mathrm{s}}\sqrt{N}}\right) \text{ as } \widetilde{\alpha} \to 0. \tag{6.20}$$

6.2.4 Performance assessment

The goal of this section is to compare different RAIM schemes in terms of the TCD framework, and to evaluate their availability. To do so, we will consider HAL = 40 m (ICAO requirement) and that only 6 satellites are visible in the following experiments.

TCD criterion evaluation

Let us consider TTA = 10 s and two different values for the UERE; that is $\sigma_{\mathrm{UERE}} = \{8, 6.6\}$ m, corresponding to single- and dual-frequency receivers, respectively [Nik02]. We also fix $\pi_{ll} = 0.35$ and $a = \sqrt{a_{1l}^{2} + a_{2l}^{2}} = 2$, so that from (6.6) we have a worst-case bias error in the pseudorange of the l-th satellite given by $|b_{l}^{*}| = 20$ m. The results for the PUD as a function of the fixed PFAP, henceforth referred to as ROC, are illustrated in Figure 6.2 for the above considered parameters. The left and right plot respectively illustrate the ROC for $\sigma_{\mathrm{UERE}} = \{8, 6.6\}$ m. We include both simulated (markers) and theoretical (dashed lines) results. We only include simulated results for the snapshot RAIM and FMA stopping time given by (6.3) and (6.15), respectively, which are the benchmarks for comparison. Regarding the theoretical results, we select the detection thresholds from (6.4), (6.12), (6.17) and (6.19); for the snapshot RAIM,

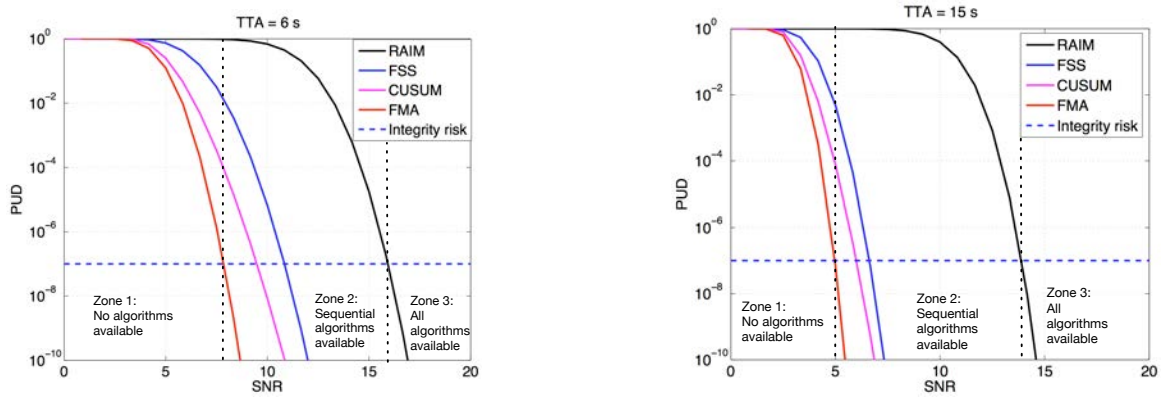

Figure 6.3: Comparison between the snapshot RAIM (black) and their competitors (i.e. optimal FSS, CUSUM and FMA) in blue, magenta and red, respectively, with different values of the magnitude of change to noise ratio $|b_l| / \sigma^2_{\text{UERE}}$ (SNR). (Left) PUD as a function of the SNR for a TTA of 6 s and (right) 15 s. A required integrity risk of 10^{-7} is indicated by the dashed line, and the detection threshold has been selected to guarantee the required PFAP of 10^{-5}/h. Moreover, we have considered $\pi_{ll} = 0.5$ and $\sigma^2_{\text{UERE}} = 12.5$ m.

CUSUM-, FMA-, and optimal FSS-RAIM, respectively. This is done for the different values of fixed PFAP, α, indicated in the x-axis. Then, we compute the PUD value indicated in the y-axis given by (6.5), (6.13) and (6.20); for the snapshot RAIM, CUSUM- and FMA-, and FSS-RAIM, respectively.

Results depicted in Figure 6.2 show a perfect fit between theoretical and simulated results for the snapshot and FMA RAIM. More important, we see that for a given value of α, all the considered sequential schemes improve the performance, in terms of the TCD criterion, with respect to the traditional RAIM scheme. In particular, the FMA solution provide the best results in terms of PUD for a given value of α. This shows the appropriateness of the sequential schemes in general, and of the FMA solution in particular, for the modernization of RAIM algorithms. These results are confirmed for both considered values of UERE, giving the FMA solution an improvement of around two and five order of magnitude for $\sigma_{\text{UERE}} = \{8, 6.6\}$ m, respectively. This huge improvement obtained by the use of sequential RAIM might lead to an increment of the availability of the RAIM algorithm in urban environments, thus allowing the safely use of critical applications in these environments. It is important to note that this is the main limitation of current RAIM algorithms in urban environments. So, the use of sequential RAIM is useful to overcome this issue, as it will be shown next.

Availability assessment

A direct consequence of the results in Figure 6.2 is an improvement of the availability of the RAIM algorithm, or equivalently, to an improvement of the minimum detectable magnitude of the fault-to-noise ratio (SNR) for a given value of PUD (i.e. integrity risk). To see so, rather than a ROC plot, it is more interesting to analyze the PUD as a function of the SNR for a fixed value of PFAP. In this line, we consider the TTA $= \{6, 15\}$ s, accounting for precision approach and in terminal en-route flight operations, respectively, (see Table 2.1). In these cases a required integrity risk of $\mathcal{P}_{\text{risk}} = 10^{-7}$, shown by the blue dashed line, is considered. In this case, we only show the theoretical results for SNR $= |b_l^*| / \sigma_{\text{UERE}} \in [0, 20]$. The same theoretical expressions as before are used, but fixing the detection threshold using a fixed PFAP of $\widetilde{\alpha} = 10^{-5}$/h, considering a sampling rate of 0.6 s. The left and right plots of Figure 6.3 shows the obtained results for TTA $= \{6, 15\}$ s, respectively. From the results we see a minimum detectable SNR, within a given TTA $= \{6, 15\}$ s, of SNR $= \{8, 5\}$. Therefore, the RAIM would be available

within the interval SNR $\in [0, 8]$ and SNR $\in [0, 5]$ (see Zone 1 in Figure 6.3) for TTA $= \{6, 15\}$ s, respectively. This is so when using the FMA solution, thus if SNR $\geq \{8, 5\}$ the RAIM would be available if and only if the FMA solution is applied. It is within the interval SNR $\in [8, 10.83]$ and SNR $\in [5, 6.67]$ when the RAIM would start being available (see Zone 2) when using any of the considered sequential schemes. After that, the RAIM would still be available with the sequential schemes. Regarding the traditional RAIM, it would not be available until SNR $\geq \{15.83, 13.8\}$ for TTA $= \{6, 15\}$ s, respectively, (see Zone 3 in Figure 6.3). This behavior clearly highlights the huge improvements of using sequential RAIM, gaining around 4 dB of SNR for a given value of required integrity risk. This is a notable improvement that confirms the capabilities of sequential schemes to overcome the limitations of current RAIM schemes in urban environments, in terms of availability. Finally, let us now focus on the sequential schemes. With the above hints we see that the FMA solution is the most sensitive solution, showing the lowest detectable SNR for a given requirement of integrity risk. This is so for both considered TTAs, showing an improvement of around 1.3 and 0.6 dB with respect to the FSS and CUSUM solution, respectively.

6.3 Signal-Level Integrity (Sig-RAIM)

We will focus here on the SLI algorithm, henceforth termed sig-RAIM, which may play a prominent role for the design of future integrity algorithms in terrestrial environments. The key point is that sig-RAIM can compute metrics for signal quality monitoring from features of the received signal that are measurable within the GNSS receiver without need of external information. In the following, we show the application of TCD to sig-RAIM, so that a prompt detection is possible, and we will show it through the example of multipath detection.

6.3.1 Implementation

Following the same general idea of RAIM algorithms, let $\{X_n\}_{n\geq 1}$ be independent and sequential observations of some SLM providing information about the GNSS signal. Let v be the time instant at which a signal integrity threat appears, and let $\mathbb{P}_v$ be the probability measure associated with the following behavior

$$X_n \sim \begin{cases} f_0(X_n) & \text{if } n < v \text{ or } n \geq v + m \\ f_1(X_n, \boldsymbol{\lambda}) & \text{if } v \leq n < v + m \end{cases}, \tag{6.21}$$

where $\boldsymbol{\lambda}$ denotes those parameters that define the change in distribution due to the appearance of a signal integrity threat and $m = t_{\text{TTA}}$, with t_{TTA} given in samples by the corresponding regulations and standards. In this case, the signal metric X_n is designed with the goal of detecting local degrading effects and it is based on the properties of the GNSS signal. This is crucial because through the signal-level approach we avoid the use of redundant measurements coming from different satellites, so that metrics can effectively be obtained in terrestrial environments, where the redundancy of conventional RAIM is often not available. It is important to note that the model in (6.21) is equivalent to that presented in (4.67) for the TCD problem. In this case, though, $\boldsymbol{\lambda}$ is unknown, and so is the statistical characterization of X_n. Thus, as done in RAIM, we use a reference parameter ($\widetilde{\boldsymbol{\lambda}}$), including all unknown parameters, which should be fixed as

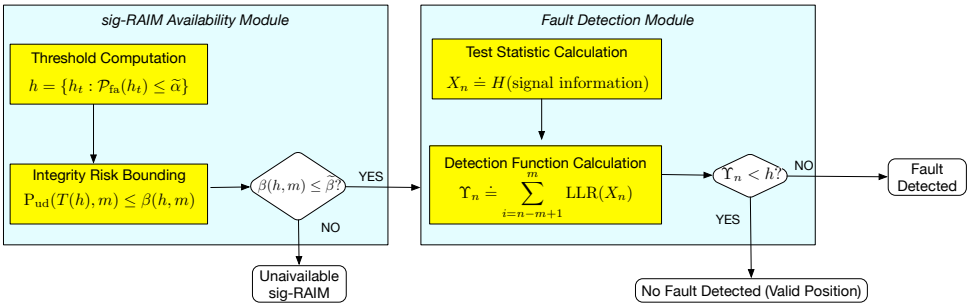

Figure 6.4: Signal integrity (sig-RAIM) algorithm flowchart.

the minimum change parameter we want to detect, given by

$$\widetilde{\boldsymbol{\lambda}} \doteq q_{\rm s}(e_{\rm L}), \tag{6.22}$$

with $e_{\rm L}$ the AL and $q_{\rm s}(\cdot)$ mapping positioning errors to change parameters in the SLM. The reference parameter $\widetilde{\boldsymbol{\lambda}}$ can be regarded as the minimum detectable change parameter $\lambda_{\rm det}$ used in RAIM algorithms to compute the PL. But here, $\widetilde{\boldsymbol{\lambda}}$ is obtained from the maximum value of the error that can be tolerated, $e_{\rm L}$.

Now we would like to fix the integrity risk as $\mathcal{P}_{\rm risk} \leq \widetilde{\beta}$, subject to the false alarm constrain $\mathcal{P}_{\rm fa} \leq \widetilde{\alpha}$ within the interval of time t_α. So we seek a test statistic G_n that minimizes the integrity risk $\mathcal{P}_{\rm risk}$, given as in (6.7) by

$$\mathcal{P}_{\rm risk} \doteq \mathbb{P}_1\left(\{G_n < h\} \bigcap \{n > v + t_{\rm TTA}\}\right). \tag{6.23}$$

That is, to minimize the probability that a threat is present but it is not detected within the TTA. Since v is unknown we should consider the following worst-case probability:

$$\sup_{v}\{\mathbb{P}_1(\{G_n < h\} \cap \{n > v + t_{\rm TTA}\})\}. \tag{6.24}$$

Doing so, the above minimization can be formulated as the optimal criterion for TCD, given in (4.70), by using stopping times notation and noting that

$$\mathcal{P}_{\rm risk} = \mathrm{P}_{\rm ud}\left(T, m\right) \text{ and } \mathcal{P}_{\rm fa} = \mathrm{P}_{\rm fap}\left(T, m_\alpha\right), \tag{6.25}$$

where $T \doteq \inf\{n : G_n \geq h\}$ stands for the stopping time used to raise an alert, and $m_\alpha = t_\alpha$, with t_α given in samples by norms and standards. Thereby, we seek the optimal stopping time T in the sense of criterion (4.70). Notwithstanding, we have already mentioned that there is not an optimal solution in the literature for finite $m > 1$, and then the problem is still open. It is for this reason that we propose the use of the FMA stopping time, defined in (4.90), to implement the sig-RAIM algorithm. Henceforth, we use the following notation: $G_n = \Upsilon_n$, $T = T_{\rm FMA}(h)$, $\mathcal{P}_{\rm risk} = \mathrm{P}_{\rm ud}\left(T_{\rm FMA}(h), m\right)$ and $\mathcal{P}_{\rm fa} = \mathrm{P}_{\rm fap}\left(T_{\rm FMA}(h), m_\alpha\right)$.

With the above framework, the implementation of the sig-RAIM at the GNSS receiver also consists of two modules, but slightly different from those for RAIM: (i) the sig-RAIM availability check; and (ii) the sig-RAIM FDE module (see Figure 6.4). At each epoch we monitor the integrity risk and if it exceeds the desired level $\widetilde{\beta}$, the sig-RAIM is not available because it cannot monitor integrity with the desired requirements in terms of t_α, $\widetilde{\alpha}$, $t_{\rm TTA}$ and $\widetilde{\beta}$. In this

case, SLI cannot be guaranteed and thus the signal coming from the analyzed satellite is declared untrustworthy. On the other hand, if the integrity risk is equal to or less than the desired level $\widetilde{\beta}$, the sig-RAIM is available and it proceeds to check whether the signal information is faulty or not. The signal metric X_n is calculated and used as input for the detection statistic Υ_n, which is compared to the detection threshold h (given by $\widetilde{\alpha}$). If the result is higher than h, a fault is declared, and the information of the analyzed satellite is excluded from the positioning solution computation. Otherwise, the signal can be used in the positioning algorithm. Note that here we monitor the integrity risk by its calculation and comparison to the given requirement $\widetilde{\beta}$. This is in contrast to traditional RAIM, in which the PL is computed and compared to the AL. The computation of the PL is based on the minimum detectable change parameter λ_{det}, which is analogous to the reference parameter $\widetilde{\boldsymbol{\lambda}}$ used in sig-RAIM. But here, $\widetilde{\boldsymbol{\lambda}}$ is obtained from the AL, and it is used to completely define the statistical characterization of X_n.

The definition of the reference parameter $\widetilde{\boldsymbol{\lambda}}$ is useful because the LLR can be completely defined and then the results in Corollary 1 can be used to fix the detection threshold h accordingly to $\widetilde{\alpha}$. Doing so, we can compute the bound on the integrity risk, so that if the bound is below $\widetilde{\beta}$ the algorithm can be declared available. So, even the way the integrity risk is monitored in RAIM and sig-RAIM is different, the same philosophy is used in both schemes. It is worth pointing out, though, that further work is needed to fully connect sig-RAIM and integrity algorithms. For instance, a key factor is the definition of the function $q_{\text{s}}(\cdot)$. This issue and others must be considered for a full integrity service, but they remain out of the scope of this thesis. The presented application of TCD to sig-RAIM is mainly intended to give the mathematical framework aimed at minimizing the integrity risk for signal quality monitoring. It is also important to clarify that the application of TCD to integrity is for the sake of exemplification; the presented theoretical results in Chapter 4, given by Theorem 8 and Corollary 1, are not restricted to GNSS, but they could be used in any TCD problem.

6.3.2 Multipath SLM performance bounds

In this section, we focus on the application of TCD to sig-RAIM, so that a prompt detection is possible. Specifically, we will focus on multipath detection, and then the multipath detection metrics considered in this dissertation will be analyzed in the following. When using the TCD framework, the pdf of $\Upsilon_m \doteq \sum_{i=1}^{m} \text{LLR}(i)$ is needed for performance evaluation. Indeed, this distribution will depend upon the distribution of the SLM. Hence, the first step to apply TCD is to statistically characterize the SLMs. As shown in Chapter 5, the distribution of the considered metrics (i.e. C/N_0, DLLout and SAM) can be fairly modeled as Gaussian under both the absence and presence of multipath (i.e. $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively), according to the CLT. Thereby, we model the multipath detection problem as a change on the parameters of a Gaussian distribution:

$$X_n \sim \begin{cases} \mathcal{H}_0 : \mathcal{N}(\mu_0, \sigma_0^2) & \text{if } n < v \text{ or } n \geq v + m \\ \mathcal{H}_1 : \mathcal{N}(\mu_1, \sigma_1^2) & \text{if } v \leq n < v + m \end{cases}, \tag{6.26}$$

where $\mathcal{N}(\mu_i, \sigma_i^2)$, with $i = \{0, 1\}$, is the Gaussian pdf with mean μ_i and variance σ_i^2, given by the mean and variance of X_n under $\mathcal{H}_i$. X_n contains the SLM samples, and m is fixed accordingly to the given TTA.

Now, once the model to be used has been presented, the next step is to calculate the LLR

of the proposed metrics, which in general, for the model in (6.26) is given by

$$\mathrm{LLR}(n) = \ln\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{(X_n - \mu_0)^2}{2\sigma_0^2} - \frac{(X_n - \mu_1)^2}{2\sigma_1^2}. \tag{6.27}$$

With this expression, and the statistical model of each metric, the cdf of Υ_m can be obtained, and then the results in Corollary 1 can be applied. We start with the simpler cases of having a change in either the mean or variance of a Gaussian distribution, corresponding to the $\mathrm{C/N_0}$ and DLLout metrics, respectively. Next, we present the most general case of having both a change in the mean and variance of a Gaussian distribution, corresponding to the SAM. Indeed, these cases represent a wide-range of practical TCD problems in which the Gaussian distribution appears, so that the presented results here are not restricted to GNSS. Moreover, it is also worth clarifying that the theoretical results in Section 4.4.4 are not restricted to Gaussian distributions, but they are valid for the general TCD theory. Finally, we further investigate the problem of discriminating between LOS and NLOS multipath conditions.

$\mathrm{C/N_0}$ metric (Gaussian mean change)

Let us start with the statistical characterization of the LLR for the $\mathrm{C/N_0}$ metric. To do so it is interesting to remind its statistical model, already discussed in Section 5.3.1, given by

$$X_{\mathrm{c},n} \sim \begin{cases} \mathcal{H}_0: & \mathcal{N}\left(\mu_{\mathrm{c},0}, \sigma_{\mathrm{c}}^2\right) \\ \mathcal{H}_1: & \mathcal{N}\left(\mu_{\mathrm{c},1}, \sigma_{\mathrm{c}}^2\right) \end{cases}, \tag{6.28}$$

with $X_{\mathrm{c},n}$ the $\mathrm{C/N_0}$ samples, $\mu_{\mathrm{c},0}$ and σ_{c}^2 the known mean and variance of $X_{\mathrm{c},n}$ under $\mathcal{H}_0$, and $\mu_{\mathrm{c},1}$ the unknown change parameter, denoting the mean of $X_{\mathrm{c},n}$ under $\mathcal{H}_1$. For the sake of notation simplicity, we have omitted the time conditions in each hypothesis (i.e. if $n < v$ or $n > v + m, \ldots$) corresponding to the TCD model. Thereby, we model the appearance of multipath as a Gaussian mean change, and then the following result is obtained.

Corollary 5. Let $\widetilde{\mu}_{\mathrm{c},1}$ and $\mu_{\mathrm{c},1}$ be the minimum detectable change parameter, as in (6.22), and the actual change parameter, respectively, and let $\mu_{\mathrm{c},0}$ and σ_{c} be known. Therefore the LLR for the $\mathrm{C/N_0}$ metric is given by

$$\mathrm{LLR}_{\mathrm{c}}(n) = Y_n = \frac{\widetilde{\mu}_{\mathrm{c},1} - \mu_{\mathrm{c},0}}{\sigma_{\mathrm{c}}^2}\left(X_{\mathrm{c},n} - \frac{\widetilde{\mu}_{\mathrm{c},1} + \mu_{\mathrm{c},0}}{2}\right), \tag{6.29}$$

with mean and variance equal to

$$\begin{aligned} \mu_{y,0} &= -\frac{(\widetilde{\mu}_{\mathrm{c},1} - \mu_{\mathrm{c},0})^2}{2\sigma_{\mathrm{c}}^2}, \quad \sigma_y^2 = -2\mu_{y,0}, \\ \mu_{y,1} &= \frac{\widetilde{\mu}_{\mathrm{c},1} - \mu_{\mathrm{c},0}}{\sigma_{\mathrm{c}}^2}\left(\mu_{\mathrm{c},1} - \frac{\widetilde{\mu}_{\mathrm{c},1} + \mu_{\mathrm{c},0}}{2}\right). \end{aligned} \tag{6.30}$$

Now, let $T_{\mathrm{C/N_0}}(h_{\mathrm{c}})$ be the FMA stopping time in (4.90) for the $\mathrm{C/N_0}$ metric with LLR in (6.29) and threshold h_{c} so that $\mathrm{P_{fap}}(T_{\mathrm{C/N_0}}(h_{\mathrm{c}}), m_\alpha) \leq \widetilde{\alpha}$; and let $\Phi(x)$ denote the cdf of the standard

normal distribution. Hence

$$h_{\mathrm{c}}\left(\widetilde{\alpha}\right) = \sqrt{m\sigma_y^2}\Phi^{-1}\left[\left(1-\tilde{\alpha}\right)^{1/m_\alpha}\right] + m\mu_{y,0}, \tag{6.31}$$

$$\mathrm{P}_{\mathrm{ud}}(T_{\mathrm{C/N_0}}\left(h_{\mathrm{c}}\right), m) \leq \beta_{\mathrm{C}}(h_{\mathrm{c}}, m) = \Phi\left(\frac{h_{\mathrm{c}} - m\mu_{y,1}}{\sqrt{m\sigma_y^2}}\right), \tag{6.32}$$

$$\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{C/N_0}}\left(h_{\mathrm{c}}\right), m_\alpha) \leq 1 - \left[\Phi\left(\frac{h_{\mathrm{c}} - m\mu_{y,0}}{\sqrt{m\sigma_y^2}}\right)\right]^{m_\alpha}. \tag{6.33}$$

Proof. The change in the $\mathrm{C/N_0}$ metric is modeled as a Gaussian mean change and thus, substituting $\mu_0 = \mu_{\mathrm{c},0}$, $\mu_1 = \widetilde{\mu}_{\mathrm{c},1}$ and $\sigma_0 = \sigma_1 = \sigma_{\mathrm{c}}$ into (6.27), (6.29) follows after simple calculus. Thereby, it is trivial to see, from (6.29) and (6.28), that the LLR for the $\mathrm{C/N_0}$ is Gaussian distributed in both hypotheses, with mean and variance as in (6.30). Hence, $\Upsilon_m^{(\mathrm{c})} = \sum_{i=1}^{m} \mathrm{LLR_c}(i)$ is Gaussian as well, but with mean and variance scaled by a factor m, and (6.31)–(6.33) thus follows by direct application of Theorem 8 and Corollary 1. □

DLLout metric (Gaussian variance change)

Now, we analyze the characterization of the LLR for the DLLout metric, modeled as in Section 5.3.2 by

$$X_{\mathrm{d},n} \sim \begin{cases} \mathcal{H}_0: & \mathcal{N}\left(0, \sigma_{\mathrm{d},0}^2\right) \\ \mathcal{H}_1: & \mathcal{N}\left(0, \sigma_{\mathrm{d},1}^2\right) \end{cases}, \tag{6.34}$$

with $X_{\mathrm{d},n}$ the DLLout samples, $\sigma_{\mathrm{d},0}^2$ the known variance under $\mathcal{H}_0$, and $\sigma_{\mathrm{d},1}^2$ the unknown variance under $\mathcal{H}_1$. Thereby, we model the appearance of multipath as a change in the variance of the DLLout samples, obtaining the following result.

Corollary 6. Let $\widetilde{\sigma}_{\mathrm{d},1}^2$ and $\sigma_{\mathrm{d},1}^2$ be the minimum detectable change parameter, as in (6.22), and the actual change parameter, respectively, and let $\sigma_{\mathrm{d},0}$ be known, then the LLR for the DLLout is given by

$$\mathrm{LLR_d}(n) = aX_{\mathrm{d},n}^2 + c, \tag{6.35}$$

with

$$a = \frac{\widetilde{\sigma}_{\mathrm{d},1}^2 - \sigma_{\mathrm{d},0}^2}{2\sigma_{\mathrm{d},0}^2\widetilde{\sigma}_{\mathrm{d},1}^2}; \quad c = \ln\left(\frac{\sigma_{\mathrm{d},0}}{\widetilde{\sigma}_{\mathrm{d},1}}\right). \tag{6.36}$$

Now, let $T_{\mathrm{DLL}}(h_{\mathrm{d}})$ be the FMA stopping time for the DLLout metric with LLR in (6.35) and threshold h_{d} so that $\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{DLL}}(h_{\mathrm{d}}), m_\alpha) \leq \widetilde{\alpha}$, and let $\Gamma_m(x)$ denote the cdf of the chi-squared distribution with m degrees of freedom, and $k_i = \sigma_{\mathrm{d},i}^2 a$, with $i = \{0, 1\}$. Hence,

$$h_{\mathrm{d}}\left(\widetilde{\alpha}\right) = k_0\Gamma_m^{-1}\left(\left(1-\tilde{\alpha}\right)^{1/m_\alpha}\right) + mc, \tag{6.37}$$

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{DLL}}(h_{\mathrm{d}}), m\right) \leq \beta_{\mathrm{D}}(h_{\mathrm{d}}, m) = \Gamma_m\left(\frac{h_{\mathrm{d}} - mc}{k_1}\right), \tag{6.38}$$

$$\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{DLL}}(h_{\mathrm{d}}), m_\alpha\right) \leq 1 - \left[\Gamma_m\left(\frac{h_{\mathrm{d}} - mc}{k_0}\right)\right]^{m_\alpha}. \tag{6.39}$$

Proof. The change in the DLLout metric is modeled as a Gaussian variance change and thus, substituting $\sigma_0 = \sigma_{\mathrm{d},0}$, $\sigma_1 = \widetilde{\sigma}_{\mathrm{d},1}$ and $\mu_0 = \mu_1 = 0$ into (6.27), (6.35) follows. Thereby, under

$\mathcal{H}_1$, and denoting $\Upsilon_m^{(\mathrm{d})} \doteq \sum_{i=1}^{m} \mathrm{LLR}_{\mathrm{d}}(i)$, we have

$$\Upsilon_m^{(\mathrm{d})}|\mathcal{H}_1 = a\sum_{n=1}^{m}\sigma_{\mathrm{d},1}^2\left(\frac{X_{\mathrm{d},n}}{\sigma_{\mathrm{d},1}}\right)^2 + m\cdot c = k_1\sum_{n=1}^{m}X_{\mathrm{n}}^2 + m\cdot c, \tag{6.40}$$

with $X_{\mathrm{n}} \sim \mathcal{N}(0,1)$ a standard Gaussian random variable. A similar result is obtained under $\mathcal{H}_0$, and then we can write $\Upsilon_m^{(\mathrm{d})}|\mathcal{H}_i = k_i\widetilde{X} + mc$, with $i = \{0,1\}$ and $\widetilde{X}$ a chi-squared random variable with m degrees of freedom. Hence,

$$\mathcal{H}_i : \frac{\Upsilon_m^{(\mathrm{d})} - m\cdot c}{k_i} \sim \chi_m^2, \tag{6.41}$$

where χ_m^2 stands for the chi-squared pdf with m degrees of freedom, and (6.37)–(6.39) thus follow by direct application of Theorem 8 and Corollary 1. □

SAM metric (General Gaussian Change)

Previously, we have analyzed the characterization of the LLR for the particular cases of having a change in either the mean or variance of a Gaussian distribution. In this section, we analyze the characterization of the LLR for the most general case of having a change in both the mean and variance of a Gaussian distribution. This is the case of the SAM-based detection, which can be modeled as

$$X_{\mathrm{s},n} \sim \begin{cases} \mathcal{H}_0 : & \mathcal{N}\left(\mu_{\mathrm{s},0}, \sigma_{\mathrm{s},0}^2\right) \\ \mathcal{H}_1 : & \mathcal{N}\left(\mu_{\mathrm{s},1}, \sigma_{\mathrm{s},1}^2\right) \end{cases}, \tag{6.42}$$

with $X_{\mathrm{s},n}$ the SAM samples, $\{\mu_{\mathrm{s},0}, \sigma_{\mathrm{s},0}^2\}$ the known mean and variance of $X_{\mathrm{s},n}$ under $\mathcal{H}_0$, respectively, and $\{\mu_{\mathrm{s},1}, \sigma_{\mathrm{s},1}^2\}$ the unknown change parameters, denoting the mean and variance of $X_{\mathrm{s},n}$ under $\mathcal{H}_1$, respectively. Thereby, we model the appearance of multipath as a change in the mean and variance of the SAM samples, and then, from (6.27) and after some straightforward manipulations, we can write the LLR of the SAM metric as

$$\mathrm{LLR}_{\mathrm{s}}(n) = a_{\mathrm{s}}X_{\mathrm{s},n}^2 + b_{\mathrm{s}}X_{\mathrm{s},n} + c_{\mathrm{s}}, \tag{6.43}$$

with

$$a_{\mathrm{s}} = \frac{\widetilde{\sigma}_{\mathrm{s},1}^2 - \sigma_{\mathrm{s},0}^2}{2\sigma_{\mathrm{s},0}^2\widetilde{\sigma}_{\mathrm{s},1}^2}; \quad b_{\mathrm{s}} = \frac{\sigma_{\mathrm{s},0}^2\widetilde{\mu}_{\mathrm{s},1} - \widetilde{\sigma}_{\mathrm{s},1}^2\mu_{\mathrm{s},0}}{\sigma_{\mathrm{s},0}^2\widetilde{\sigma}_{\mathrm{s},1}^2}; \quad c_{\mathrm{s}} = \ln\left(\frac{\sigma_{\mathrm{s},0}}{\widetilde{\sigma}_{\mathrm{s},1}}\right) + \frac{\widetilde{\sigma}_{\mathrm{s},1}^2\mu_{\mathrm{s},0}^2 - \sigma_{\mathrm{s},0}^2\widetilde{\mu}_{\mathrm{s},1}^2}{2\sigma_{\mathrm{s},0}^2\widetilde{\sigma}_{\mathrm{s},1}^2}; \tag{6.44}$$

where $\widetilde{\mu}_{\mathrm{s},1}$ and $\widetilde{\sigma}_{\mathrm{s},1}$ are the minimum detectable change parameters.

In this case we cannot find the distribution of $\mathrm{LLR}_{\mathrm{s}}(n)$ in an straightforward way as for the previous cases. Here, in order to find the pdf of $\Upsilon_m^{(\mathrm{s})} \doteq \sum_{i=1}^{m}\mathrm{LLR}_{\mathrm{s}}(i)$, we make use of the so-called Edgeworth series approximation and EVT, which provide a very tight closed-form expression for the bounds to the FMA stopping time of the SAM metric, $T_{\mathrm{SAM}}(h_{\mathrm{s}})$ [ER17b]. For the sake of notation clarity, let us write $\Upsilon_m^{(\mathrm{s})}$ as the random variable Z_{s} (i.e. $Z_{\mathrm{s}} = \Upsilon_m^{(\mathrm{s})}$). Thereby, we can state the following result.

Corollary 7. Let $\mu_{\mathrm{s},i}$ and $\sigma_{\mathrm{s},i}^2$, with $i = \{0,1\}$, be the actual mean and variance of the SAM metric under hypothesis $\mathcal{H}_i$, and let a_{s}, b_{s} and c_{s} be defined as in (6.44). Now, let $T_{\mathrm{SAM}}(h_{\mathrm{s}})$ be the FMA stopping time for the SAM metric with LLR in (6.43) and threshold h_{s} so that

$\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{SAM}}(h_{\mathrm{s}}), m_\alpha) \leq \widetilde{\alpha}$, and let $\phi(x) \doteq \mathcal{N}(0,1)$ be the standard normal pdf. Hence, we have

$$h_{\mathrm{s}}\left(\widetilde{\alpha}\right) = \delta - \frac{\ln\left(-\ln\left(1-\widetilde{\alpha}\right)\right)}{\gamma}, \tag{6.45}$$

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{SAM}}\left(h_{\mathrm{s}}\right), m\right) \leq \beta_{\mathrm{S}}(h_{\mathrm{s}}, m) = F_{\mathrm{s},1}\left(h_{\mathrm{s}}\right), \tag{6.46}$$

$$\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{SAM}}\left(h_{\mathrm{s}}\right), m_\alpha\right) \leq 1 - \exp\left(-e^{-\gamma(h_{\mathrm{s}}-\delta)}\right), \tag{6.47}$$

with

$$\delta = F_{\mathrm{s},0}^{-1}\left(1 - \frac{1}{m_\alpha}\right); \quad \gamma = m_\alpha f_{\mathrm{s},0}\left(\delta\right), \tag{6.48}$$

and

$$\begin{aligned} F_{\mathrm{s},i}(z) &\approx \Phi(\tilde{z}_i) - \sigma_{z_{\mathrm{s}},i}\phi(\tilde{z}_i)\sum_{k\in\mathcal{A}} C_{k,\mathcal{H}_i} H_{k-1}(\tilde{z}_i), \\ f_{\mathrm{s},0}(z) &\approx \phi(\tilde{z}_0)\left[1 + \sum_{k\in\mathcal{A}} C_{k,\mathcal{H}_0} H_k(\tilde{z}_0)\right], \end{aligned} \tag{6.49}$$

where $\mathcal{A} = \{3,4,6\}$, $C_{k,\mathcal{H}_i}$, with $i = \{0,1\}$, are the coefficients C_k (expressions can be found in Appendix 6.A) under $\mathcal{H}_i$, $H_k(z)$ is the Hermite polynomial of degree k evaluated at z and $\tilde{z}_i = (z - \mu_{z_{\mathrm{s}},i})/\sigma_{z_{\mathrm{s}},i}$, with

$$\begin{aligned} \mu_{z_{\mathrm{s}},i} &= m\left[a_{\mathrm{s}}\left(\sigma_{\mathrm{s},i}^2 + \mu_{\mathrm{s},i}^2\right) + b_{\mathrm{s}}\mu_{\mathrm{s},i} + c_{\mathrm{s}}\right], \\ \sigma_{z_{\mathrm{s}},i}^2 &= m\left[\sigma_{\mathrm{s},i}^2\left[2a_{\mathrm{s}}\left(a_{\mathrm{s}}\sigma_{\mathrm{s},i}^2 + 2a_{\mathrm{s}}\mu_{\mathrm{s},i}^2 + b_{\mathrm{s}}\mu_{\mathrm{s},i}\right) + b_{\mathrm{s}}^2\right]\right]. \end{aligned} \tag{6.50}$$

Proof. The proof is given in [ER17b] and it is included in Appendix 6.A. □

SAM for LOS/NLOS discrimination

We have analyzed the detection of multipath as a change in the mean and variance of the SAM. Nonetheless, the discrimination between LOS and NLOS was not considered. This discrimination can be very helpful in order to apply multipath mitigation techniques or to decide to apply techniques capable of exploiting the NLOS multipath such as [Gen15]. Recently, several contributions for distinguishing NLOS from LOS multipath have been proposed, but relying on the use of additional hardware [Gro10, Hsu15]. We already have shown in Chapter 5 the capability of the SAM for LOS and NLOS multipath discrimination. Based on this fact, we propose here the use of two FMA stopping times, working in parallel, one to detect a change in the mean of the SAM and another to detect a change in the variance. This is beneficial, as we will show, to obtain simple expressions for theoretical performance bounds. Otherwise, above shown, a closed-form expression is not available, leading to approximations. This is a novel contribution since the proposed solution works at the signal processing level, without using additional aid, and adopts a TCD framework easily implementable in mass-market GNSS receivers. This work was presented by us in [ER16]. Indeed, we have already shown in Section 5.3.3 the statistical characterization of the SAM under both LOS and NLOS conditions. Based on this characterization, we present here the proposed stopping time to detect multipath and to be able to discriminate between LOS and NLOS.

Let us first recall the statistical characterization of the SAM under both LOS and NLOS conditions. From (5.45) the LOS conditions are modeled as a mean change of the SAM. So, we

have the following LLR

$$\mathrm{LLR_M}(n) = \frac{\mu_{\mathrm{L},1} - \mu_{\mathrm{s},0}}{\sigma^2_{\mathrm{s},0}} \left(X_{\mathrm{L},k} - \frac{\mu_{\mathrm{L},1} + \mu_{\mathrm{s},0}}{2} \right), \tag{6.51}$$

where $X_{\mathrm{L},k}$ and $\mu_{\mathrm{L},1}$ are the SAM metric observations and mean when LOS multipath is present and $\{\mu_{\mathrm{s},0}, \sigma^2_{\mathrm{s},0}\}$ are the mean and variance of the SAM under $\mathcal{H}_0$. On the other hand, under NLOS situation, the SAM experiences a change in the variance, and then we have

$$\mathrm{LLR_V}(n) = \gamma X^2_{\mathrm{N},k} + \delta, \tag{6.52}$$

with

$$\gamma = \frac{\sigma^2_{\mathrm{N},1} - \sigma^2_{\mathrm{s},0}}{2\sigma^2_{\mathrm{s},0}\sigma^2_{\mathrm{N},1}} \text{ and } \delta = \ln\left(\frac{\sigma_{\mathrm{s},0}}{\sigma_{1,\mathrm{N}}}\right), \tag{6.53}$$

where $X_{\mathrm{N},k}$ and $\sigma^2_{\mathrm{N},1}$ stand for the SAM observations and variance when NLOS multipath is present, respectively. Based on the previous results, we propose the use of two different stopping times working in parallel, one for detecting a change in the mean of the SAM and another for the variance. Thereby, we will detect multipath whenever one of the stopping times declares so, and we will be able to discriminate between LOS or NLOS when the change in the mean or variance is declared, respectively. To do so, we define

$$T_{\mathrm{P}}(h_{\mathrm{M}}, h_{\mathrm{V}}) \doteq \inf\left\{ n \geq m : \{\Upsilon_{\mathrm{M},n} \geq h_{\mathrm{M}}\} \bigcap \{\Upsilon_{\mathrm{V},n} \geq h_{\mathrm{V}}\} \right\} = \min\left\{T_{\mathrm{M}}(h_{\mathrm{M}}), T_{\mathrm{V}}(h_{\mathrm{V}})\right\}, \tag{6.54}$$

where T_{M} (h_{M}) and T_{V} (h_{V}) are the stopping times (detection thresholds) for detecting a change in the mean and variance, respectively, defined as

$$T_{\mathrm{M}}(h_{\mathrm{M}}) \doteq \inf\left\{n \geq m : \Upsilon_{\mathrm{M},n} \geq h_{\mathrm{M}}\right\} \text{ and } T_{\mathrm{V}}(h_{\mathrm{V}}) \doteq \inf\left\{n \geq m : \Upsilon_{\mathrm{V},n} \geq h_{\mathrm{V}}\right\}, \tag{6.55}$$

with

$$\Upsilon_{\mathrm{M},n} \doteq \sum_{i=n-m+1}^{n} \mathrm{LLR_M}(i); \quad \Upsilon_{\mathrm{V},n} \doteq \sum_{i=n-m+1}^{n} \mathrm{LLR_V}(i), \tag{6.56}$$

and $\{\mathrm{LLR_M}, \mathrm{LLR_V}\}$ given by (6.51) and (6.52), respectively. Apart from the interest in discriminating between LOS/NLOS, the use of parallel stopping times is also motivated by the difficulties in finding closed-form bounds for PUD and PFAP. Notwithstanding, the use of the parallel stopping time extremely reduces these difficulties, as shown in the following corollary.

Corollary 8. Let $\{\mu_{\mathrm{s},0}, \sigma^2_{\mathrm{s},0}\}$ be the mean and variance of the SAM under $\mathcal{H}_0$, and let $\mu_{\mathrm{L},1}$ and $\sigma^2_{\mathrm{N},1}$ be the mean and variance of the SAM under LOS and NLOS conditions, respectively. Furthermore, let $T_{\mathrm{P}}(h_{\mathrm{M}}, h_{\mathrm{V}})$ be the parallel stopping time in (6.54) and detection thresholds $\{h_{\mathrm{M}}, h_{\mathrm{V}}\}$. Hence, we have

$$\mathrm{P_{fap}}\left(T_{\mathrm{P}}\left(h_{\mathrm{M}}, h_{\mathrm{V}}\right), m_\alpha\right) \leq \alpha_{\mathrm{p}}\left(h_{\mathrm{M}}, h_{\mathrm{V}}, m_\alpha\right) = \max\left\{\alpha_{\mathrm{M}}\left(h_{\mathrm{M}}, m_\alpha\right), \alpha_{\mathrm{V}}\left(h_{\mathrm{V}}, m_\alpha\right)\right\}, \tag{6.57}$$

with

$$\alpha_{\mathrm{M}}\left(h, m_{\alpha}\right)=1-\left[\Phi\left(\frac{h+m \cdot \mu_{\mathrm{M}}}{\sqrt{m \cdot \sigma_{\mathrm{M}}^{2}}}\right)\right]^{m_{\alpha}}, \tag{6.58}$$

$$\alpha_{\mathrm{V}}\left(h, m_{\alpha}\right)=1-\left[\Gamma_{m}\left(\frac{h-m \cdot \delta}{k_{0}}\right)\right]^{m_{\alpha}}, \tag{6.59}$$

where $\Phi(x)$ denotes the standard normal cdf evaluated at x, $\Gamma_m(x)$ stands for the cdf of the chi-squared distribution with m degrees of freedom, $k_0 = \sigma_{\mathrm{s},0}^2\gamma$, with γ and δ given in (6.53); and

$$\mu_{\mathrm{M}}=\frac{\left(\mu_{\mathrm{L},1}-\mu_{\mathrm{s},0}\right)^{2}}{2 \cdot \sigma_{\mathrm{s},0}^{2}}, \quad \sigma_{\mathrm{M}}^{2}=2 \cdot \mu_{\mathrm{M}}. \tag{6.60}$$

Moreover, under LOS conditions we have

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{P}}\left(h_{\mathrm{M}}, h_{\mathrm{V}}\right), m\right) \leq \beta_{\mathrm{p,L}}\left(h_{\mathrm{M}}, h_{\mathrm{V}}, m\right)=\min\left\{\beta_{\mathrm{M}}\left(h_{\mathrm{M}}, m\right), \beta_{\mathrm{V,L}}\left(h_{\mathrm{V}}, m\right)\right\}, \tag{6.61}$$

whereas under NLOS we have

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{P}}\left(h_{\mathrm{M}}, h_{\mathrm{V}}\right), m\right) \leq \beta_{\mathrm{p,N}}\left(h_{\mathrm{M}}, h_{\mathrm{V}}, m\right)=\min\left\{\beta_{\mathrm{M,N}}\left(h_{\mathrm{M}}, m\right), \beta_{\mathrm{V}}\left(h_{\mathrm{V}}, m\right)\right\}, \tag{6.62}$$

with

$$\beta_{\mathrm{M}}\left(h, m\right)=\Phi\left(\frac{h-m\mu_{\mathrm{M}}}{\sqrt{m\sigma_{\mathrm{M}}^{2}}}\right), \tag{6.63}$$

$$\beta_{\mathrm{V}}\left(h, m\right)=\Gamma_{m}\left(\frac{h-m\delta}{k_{1}}\right), \tag{6.64}$$

$$\beta_{\mathrm{M,N}}\left(h, m\right)=\Phi\left(\frac{h+m\mu_{\mathrm{M}}}{\sqrt{m}\widetilde{\sigma}}\right), \tag{6.65}$$

$$\beta_{\mathrm{V,L}}\left(h, m\right)=\Gamma_{m}\left(\frac{h-m\delta}{k_{0}}; \lambda\right), \tag{6.66}$$

and where $\Gamma_m(x;\lambda)$ is the cdf of the non-central chi-squared distribution with non-central parameter $\lambda = m(\mu_{\mathrm{L},1}^2/\sigma_{\mathrm{s},0})$, $k_1 = \sigma_{\mathrm{N},1}^2\gamma$ and $\widetilde{\sigma}^2 = \sigma_{\mathrm{M}}^2(\sigma_{\mathrm{N},1}^2/\sigma_{\mathrm{s},0}^2)$.

Proof. Let us first consider the bounds for T_{M} and T_{V}. As a Gaussian mean change detection, for the LOS case we have the bounds in (6.63) and (6.58) for T_{M} (see Corollary 5), whereas under NLOS conditions the PUD of T_{M} is bounded by (6.65). On the other hand, under NLOS situation, as a Gaussian variance change detection, we have the bounds in (6.64) and (6.59) for T_{V} (see Corollary 6), while under LOS conditions the PUD of T_{V} is bounded by (6.66). Thereby, from the definition of P_{ud} and $\mathrm{P}_{\mathrm{fap}}$ as well as T_{P} in (6.54), we have

$$\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{P}}\right) \doteq \sup_{v>m} \mathbb{P}_{v}\left(T_{\mathrm{P}} \geq v+m-1 \,|\, T_{\mathrm{P}} \geq v\right) \leq \min\left\{\mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{M}}\right), \mathrm{P}_{\mathrm{ud}}\left(T_{\mathrm{V}}\right)\right\}, \tag{6.67}$$

where the inequality follows because despite T_{M} detected faster than T_{V} on average, there might be realizations in which T_{V} detects faster, or viceversa. Similarly, for $\mathrm{P}_{\mathrm{fap}}$ we have

$$\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{P}}\right) \doteq \sup_{l \geq m} \mathbb{P}_{\infty}\left(l \leq T_{\mathrm{P}}<l+m_{\alpha}\right) \geq \max\left\{\mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{M}}\right), \mathrm{P}_{\mathrm{fap}}\left(T_{\mathrm{V}}\right)\right\}. \tag{6.68}$$

From (6.68) we have a lower bound, and then (6.57) will be only true when the change in the mean or variance is much greater than the other, which is our case for the SAM, and then (6.68) turns out to be an equality, and the proof of the corollary thus follows. □

C/N$_0$		DLLout		SAM	
$\mu_{c,0}$	$10^{4.4}$	$\sigma^2_{d,0}$	$1.11 \cdot 10^{-5}$	$\mu_{s,0}$	0.1
$\mu_{c,1}$	$10^{3.7}$			$\sigma^2_{s,0}$	$1.14 \cdot 10^{-3}$
σ^2_c	$2.5 \cdot 10^5$	$\sigma^2_{d,1}$	$5.44 \cdot 10^{-4}$	$\mu_{s,1}$	0.2
				$\sigma^2_{s,1}$	$2.03 \cdot 10^{-3}$
m	6				
m_α	60			300	

Table 6.1: Simulation parameters for the different presented multipath SLM.

6.3.3 Performance assessment

This section introduces numerical simulations to assess the goodness of the presented theoretical results in Section 6.3, based on different study cases.

Evaluation of the probability minimization criterion

Here, we compare the FMA stopping time with those stopping times currently available in the literature of TCD. This comparison is done with both simulated and available theoretical results. The simulated results include Monte-Carlo simulations (10^6 runs) of the exact worst-case PUD, $\mathrm{P_{ud}}(T(h), m)$, as a function of the exact worst-case PFAP, $\mathrm{P_{fap}}(T(h), m_\alpha)$, henceforth referred to as the ROC. Regarding the theoretical results, on the one hand they include those obtained in Section 6.3.2 for the FMA stopping time of the different metrics used in this dissertation, stated in Corollaries 5, 6 and 7. On the other hand, they also include the bounds for the CUSUM and WLC. For the PUD of these two latter methods we have an upper bound in the same form as those in the previous mentioned corollaries (see Section 4.4.4). For the PFAP, we use the upper bound given by (4.103), which holds for both the CUSUM and WLC. For the representation of the theoretical results shown in Figures 6.5 and 6.6, the procedure is as follows:

1. For the FMA bound (dashed line), the detection threshold is fixed as given in the corollaries for the corresponding metrics (i.e. (6.31), (6.37) and (6.45)) using $\widetilde{\alpha} = \mathrm{P_{fap}}$, with $\mathrm{P_{fap}}$ the PFAP value indicated on the x-axis; and

2. then the obtained threshold is substituted into the formula for the bound $\beta(h, m)$ in the corollary pertaining to the corresponding metric (i.e. (6.32), (6.38), (6.46)), giving the value depicted in the y-axis.

Similarly, for the CUSUM and WLC bounds (solid line), the same formulas for $\beta(h, m)$ are used, but the detection threshold is computed with (4.103). For instance, for the C/N$_0$, results in Corollary 5 apply, and hence (6.31) is used to obtain the threshold $h_c(\mathrm{P_{fap}})$, which is further applied to (6.32) leading to the value depicted on the y-axis. This procedure is indicated in the figure's legend (see left plot of Figure 6.5) as (6.31)$\rightarrow$ $h_c(\mathrm{P_{fap}})$ $\rightarrow$(6.32) for the FMA bound, and as (4.103)$\rightarrow$ $h_c(\mathrm{P_{fap}})$ $\rightarrow$(6.32) for the WLC and CUSUM bound. Figure 6.5 shows the results corresponding to the C/N$_0$ (left) and SAM (right) metrics. In this case, the obtained theoretical results in Corollaries 5 and 7 apply, respectively for the C/N$_0$ and SAM, so that {(6.31),(6.45)} and {(6.32),(6.46)} are used to obtain the threshold and PUD bound, respectively. This is done by considering the parameters shown in Table 6.1. For the sake of comparison, we assume that the actual change parameters have been used to formulate the LLR (i.e. $\widetilde{\boldsymbol{\lambda}} = \boldsymbol{\lambda}$).

It can be concluded from Figure 6.5 that the FMA stopping time performs better (for both the C/N$_0$ and SAM metrics), in the sense of the optimality criterion in (4.70), than the other

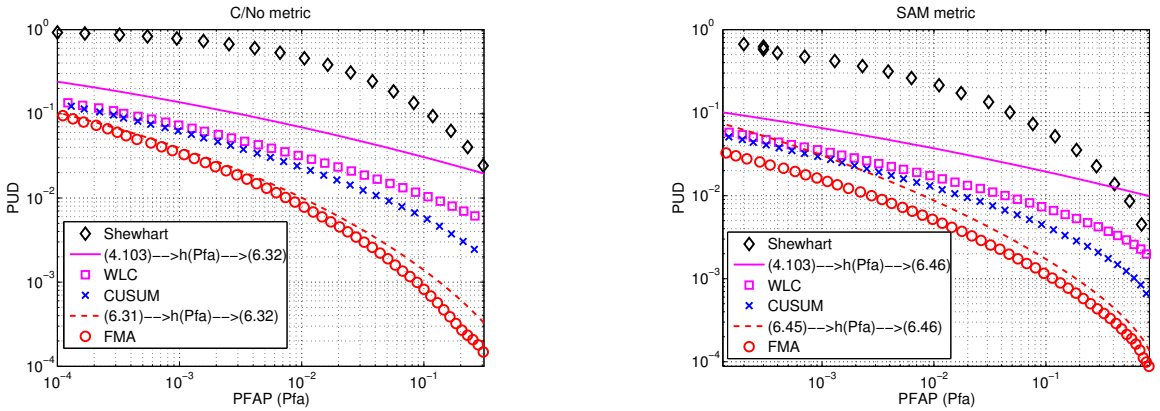

Figure 6.5: (Left) Simulated ROC for the C/N_0 and (right) SAM FMA stopping times and their competitors (i.e. CUSUM, WLC and Shewhart), with the theoretical results given in (6.31)–(6.32) and (6.45)–(6.46), respectively, and those obtained with (4.103) for the bound of the PFAP.

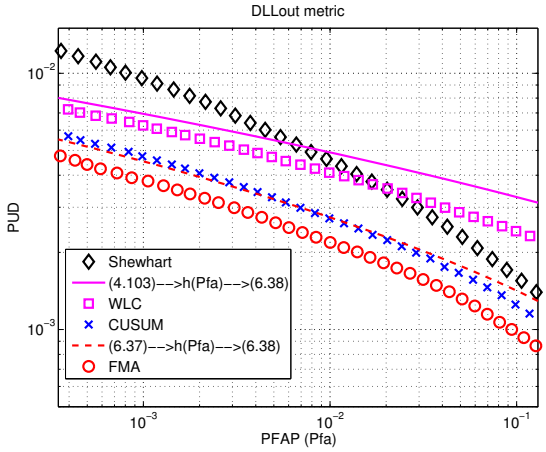

Figure 6.6: Simulated ROC for the DLLout FMA stopping time and its competitors, with the theoretical results given in (6.37)–(6.38) and those obtained with (4.103) for the bound of the probability of false alarms.

considered stopping times. Figure 6.6 shows similar results but for the case of the DLLout metric and using the corresponding parameters shown in Table 6.1. In this case, the theoretical results in Corollary 6 apply; and hence for the FMA test, (6.37) is used to compute the threshold and (6.38) to calculate the value depicted on the y-axis. For the CUSUM and WLC bounds, the only difference is that the threshold is obtained with (4.103). In this case we also see an improvement of the FMA stopping time with respect to the CUSUM and WLC. Moreover, it is important pointing out that the bounds will also impact the real performance of the methods because the threshold h is in practice fixed using the bounds, with the aim of fixing certain desirable performance. Thus, the availability of tight bounds is important not only for a theoretical study but also to set the threshold in practice and provide a level of performance that is close to the desired one. We see in Figure 6.5 and Figure 6.6 how the improvement of the FMA bounds with respect to those available in the literature for other stopping times is quite significant, providing between half- to two-orders-of-magnitude improvement. This is something that, as we will see next, greatly contributes on improving the availability of the sig-RAIM. Finally, it is important to remark that the Shewhart stopping time is not giving the best results nether for the DLLout nor for the C/N_0 and SAM, making evident the loss of its optimality properties for $m > 1$.

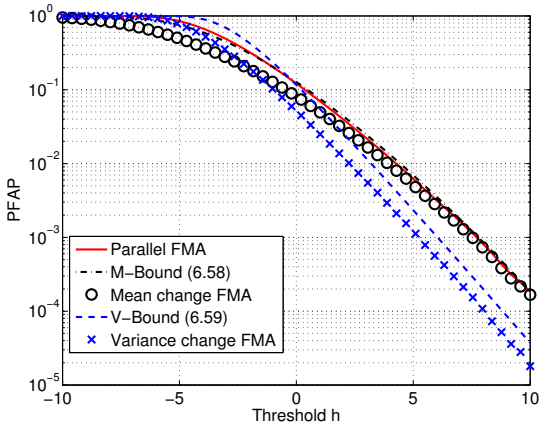

Figure 6.7: Numerical simulations of the PFAP, $\mathrm{P}_{\mathrm{fap}}(T)$, for the parallel, mean and variance change stopping times, with their respective bounds in (6.58) and (6.59). The following parameters have been used: $m = 6$, $m_\alpha = 60$, $\mu_0 = 0$, $\mu_{\mathrm{L},1} = 0.15$, $\sigma_0^2 = 5 \cdot 10^{-3}$.

Availability assessment

Next, we show the behavior of the C/N$_0$and DLLout metrics in terms of availability of the sig-RAIM. As before, the simulation parameters for each metric are presented in Table 6.1, but here we will consider different values for the parameters under $\mathcal{H}_1$, stated below. Let us start with the C/N$_0$ metric assuming that we have a tolerable error equivalent to a mean change in the C/N$_0$ of 7dB; thus we fix the change parameter as $\widetilde{\mu}_{\mathrm{c},1} = 10^{3.7}$, but the actual change parameter is $\mu_{\mathrm{c},1} = 10^{3.4}$. Therefore, for the case of using the FMA stopping time, fixing the detection threshold from (6.31) to $h_{\mathrm{c}} = 2.92$ so that $\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{C/N_0}}(h_{\mathrm{c}}), m_\alpha) \leq \widetilde{\alpha} = 10^{-1}$, and substituting the previous values in (6.30) and (6.32), the integrity risk is bounded as $\mathcal{P}_{\mathrm{risk}} \doteq \mathrm{P}_{\mathrm{ud}}(T_{\mathrm{C/N_0}}(h_{\mathrm{c}}), m) \leq \beta_{\mathrm{C}}(h_{\mathrm{c}}, m) = 6.97 \cdot 10^{-4}$. For the CUSUM or WLC stopping time we fix the threshold $\widetilde{h}_{\mathrm{c}}$ from (4.103), which for $\widetilde{\alpha} = 10^{-1}$ gives $\widetilde{h}_{\mathrm{c}} = 6.40$, and thus from (6.32) we get $\mathcal{P}_{\mathrm{risk}} \doteq \mathrm{P}_{\mathrm{ud}}(T_{\mathrm{WLC}}(\tilde{h}_{\mathrm{c}}), m) \leq \beta_{\mathrm{C}}(\widetilde{h}_{\mathrm{c}}, m) = 4.56 \cdot 10^{-3}$. Now, suppose the maximum allowed integrity risk is $\widetilde{\beta} = 10^{-2}$; then since $\mathcal{P}_{\mathrm{risk}} \leq \beta_{\mathrm{C}}(h_{\mathrm{c}}) < \beta_{\mathrm{C}}(\tilde{h}_{\mathrm{c}}) < \widetilde{\beta}$ the sig-RAIM will be available in the case of using any of the analyzed stopping times. On the other hand, we suppose we need $\widetilde{\alpha} = 10^{-2}$ so from (6.31) and (4.103) we get $h_{\mathrm{c}} = 3.59$ for the FMA and $\widetilde{h}_{\mathrm{c}} = 8.70$ for the CUSUM and WLC, respectively. Thus, from (6.32) we have that $\beta_{\mathrm{C}}(h_{\mathrm{c}}) = 1.02 \cdot 10^{-3}$ and $\beta_{\mathrm{C}}(\widetilde{h}_{\mathrm{c}}) = 1.33 \cdot 10^{-2}$. Hence, in this case, the sig-RAIM will be available only if the FMA stopping time is used. Otherwise, it will not be available since $\beta_{\mathrm{C}}(\widetilde{h}_{\mathrm{c}}) > \widetilde{\beta}$, showing the improvements of the FMA in terms of the sig-RAIM availability.

Now, for the DLLout, imagine that a change is present with $\sigma_{\mathrm{d},1}^2 = 5.44 \cdot 10^{-4}$, but the maximum tolerable error in the measured range within the GNSS receiver for each satellite is equal to 14.65 m. For a GPS signal, 14.65 m of error is equivalent to a variation of ± 0.05 chips, which converted to DLLout variance as explained in Section 5.3 gives a minimum detectable change parameter of $\widetilde{\sigma}_{\mathrm{d},1}^2 = 2.78 \cdot 10^{-4}$. Assuming we want $\mathrm{P}_{\mathrm{fap}}(T(h), m_\alpha) \leq \widetilde{\alpha} = 10^{-2}$, from (6.37), we have $h_{\mathrm{d}} = 3.14$ for the FMA and, from (4.103), $\widetilde{h}_{\mathrm{d}} = 8.70$ for the CUSUM and WLC. Thereby, if we fix the actual parameter as $\sigma_{\mathrm{d},1}^2 = \widetilde{\sigma}_{\mathrm{d},1}^2$, we get from (6.38) that $\beta_{\mathrm{D}}(h_{\mathrm{d}}, m) = 1.70 \cdot 10^{-2}$ and $\beta_{\mathrm{D}}(\widetilde{h}_{\mathrm{d}}, m) = 4.25 \cdot 10^{-2}$. If we wish $\widetilde{\beta} = 10^{-2}$, then since the bounds are above $\widetilde{\beta}$, the sig-RAIM is not available.

LOS/NLOS discrimination: performance bounds

Here, we assess the goodness of the theoretical bounds, proposed in Section 5.3.3, with numerical

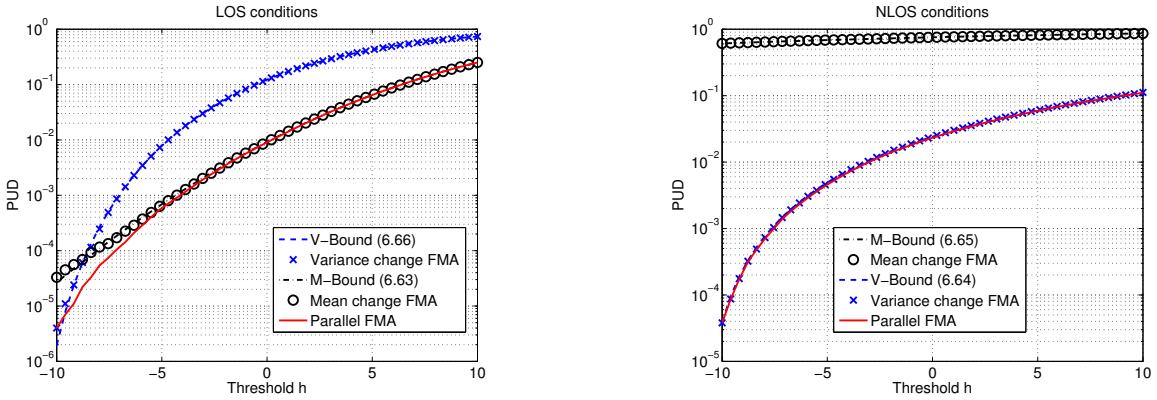

Figure 6.8: Numerical simulations of the PUD, $\mathrm{P}_{\mathrm{ud}}(T)$, for the considered stopping times, with their respective bounds in (6.63) and (6.66), with $m = 6$, $m_\alpha = 60$, $\mu_0 = 0$, $\mu_{\mathrm{L},1} = 0.15$, $\sigma_0^2 = 5 \cdot 10^{-3}$ and $\sigma_{\mathrm{N},1}^2 = 17 \cdot \sigma_0^2$. (Left) Under LOS and (right) NLOS conditions, with their respective bounds in (6.64) and (6.65).

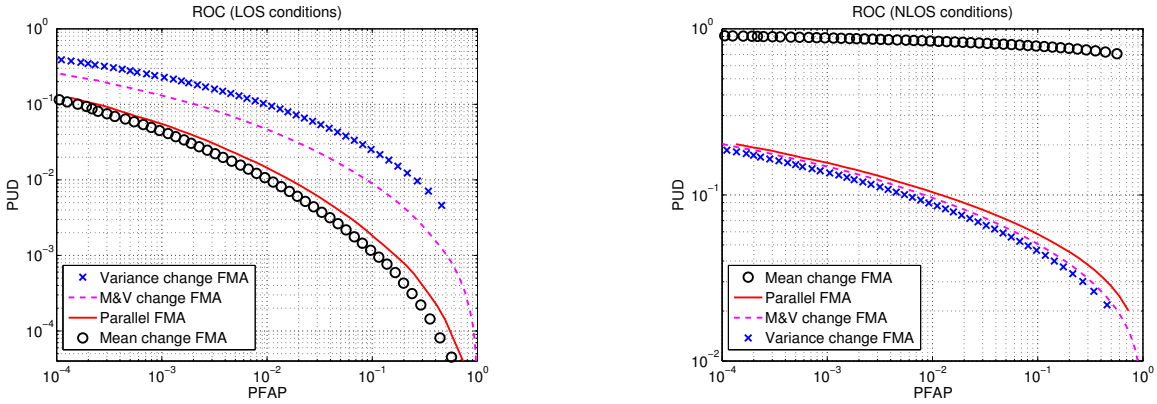

Figure 6.9: (Left) ROC under LOS and (right) NLOS conditions for the analyzed stopping times with $m = 6$, $m_\alpha = 60$, $\mu_0 = 0$, $\mu_{\mathrm{L},1} = 0.15$, $\sigma_0^2 = 4 \cdot 10^{-3}$ and $\sigma_{\mathrm{N},1}^2 = 12 \cdot \sigma_0^2$.

simulations of the worst-case PUD, $\mathrm{P}_{\mathrm{ud}}(T)$, and PFAP with duration m_α, $\mathrm{P}_{\mathrm{fap}}(T)$. We fix $h_{\mathrm{M}} = h_{\mathrm{V}} = h$ so that we can compare T_{M}, T_{V} and T_{P} with a common threshold. Figure 6.7 shows the PFAP as a function of the detection threshold h. This probability is compared with the bounds for T_{M}, T_{V} and T_{P} given by (6.58), (6.59) and (6.57), respectively. We see how the proposed bound for the PFAP is fulfilled with all the stopping times. Indeed, we see how for small values of threshold h, $\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{V}})$ is greater than $\mathrm{P}_{\mathrm{fap}}(T_{\mathrm{M}})$, and then the upper-bound for the parallel stopping time T_{P} is given by the bound for T_{V} in (6.59). On the other hand, for big h, T_{M} gives worse PFAP, so that the bound for T_{P} is given by the bound for T_{M} in (6.58). Similar results are obtained for the PUD, under LOS conditions, in the left plot of Figure 6.8; that is, for small h, T_{V} gives better PUD, and then $\mathrm{P}_{\mathrm{ud}}(T_{\mathrm{P}})$ is upper-bounded by the bound for T_{V} in (6.66). For big h, T_{M} provides the lowest PUD, and then the bound for T_{P} is given by the bound for T_{M} in (6.63). Finally, the right plot of Figure 6.8 shows the PUD for the case of NLOS conditions. We see how in this case the T_{V} always gives the best results, in terms of unbounded detection, and then the bound for the parallel stopping time is given by the bound for T_{V} in (6.64).

LOS/NLOS discrimination: ROC Comparison

Now, we compare T_{P}, in terms of ROC, with T_{FMA}, T_{M} and T_{V}. To compute the ROC, we fix

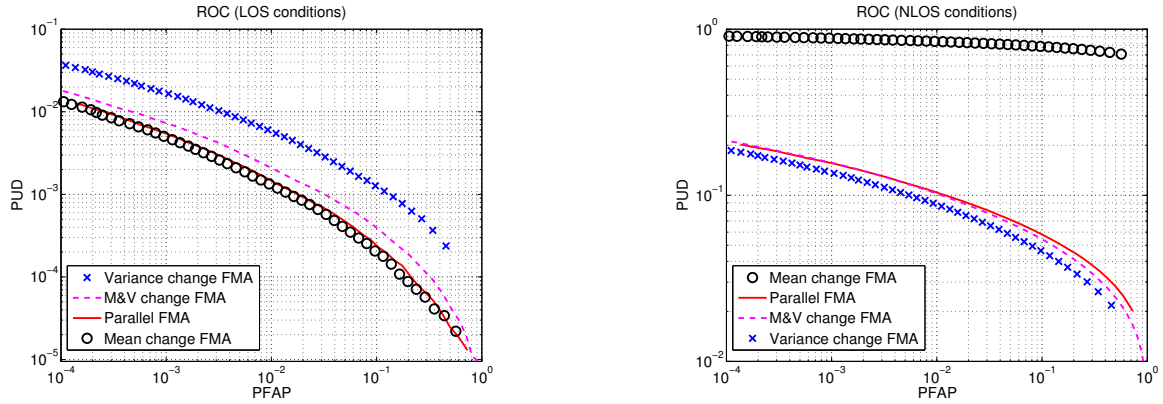

Figure 6.10: (Left) ROC under LOS and (right) NLOS conditions for the analyzed stopping times with $m = 6$, $m_\alpha = 60$, $\mu_0 = 0$, $\mu_{\mathrm{L},1} = 0.2$, $\sigma_0^2 = 4 \cdot 10^{-3}$, $\sigma_{\mathrm{N},1}^2 = 12 \cdot \sigma_0^2$ and $\sigma_{\mathrm{L},1}^2 = 2 \cdot \sigma_0^2$.

h_{M} and h_{V} from (6.58) and (6.52), respectively, as

$$\begin{aligned} h_{\mathrm{M}} &= \sqrt{m\sigma_{\mathrm{M}}^2}\Phi\left((1-\widetilde{\alpha})^{1/m_\alpha}\right) - m\mu_{\mathrm{M}}, \\ h_{\mathrm{V}} &= k_0\Gamma_m\left((1-\widetilde{\alpha})^{1/m_\alpha}\right) + m\delta, \end{aligned} \tag{6.69}$$

with $\widetilde{\alpha} \in [10^{-4}, 1]$. Then, we numerically obtain $\mathrm{P_{ud}}(T)$ and $\mathrm{P_{fap}}(T)$ with 10^6 Monte-Carlo runs. The left and right plot of Figure 6.9 show the ROC for each analyzed stopping time for LOS and NLOS conditions, respectively. We see in the left plot how for the LOS case T_{P} gives better performance than T_{SAM}. This is because there is not change in variance, and then the model used for T_{SAM} is not actually true, hindering the detection. Meanwhile, T_{P} approaches T_{M}, which provides the best results in terms of ROC. Furthermore, we see in the right plot of Figure 6.9 that for the NLOS case, the T_{SAM} slightly performs better than the parallel stopping time. In this case, the difference in terms of performance between T_{M} and T_{V} is huge, and then T_{P} is badly affected by the poor performance of T_{M}, whereas for T_{SAM} this difference makes it to approach T_{V}, which produce the best results in terms of ROC. The left plot of Figure 6.10 consider those cases where the variance under LOS conditions vary. We see that T_{P} still outperforms T_{SAM}, but slightly less than previously. This is so because we have included a change in the variance and then the T_{SAM} is benefited, but not enough as for outperforming T_{P}. On the other hand, in the right hand plot of Figure 6.10, we see that for the NLOS case both T_{P} and T_{SAM} provides similar performance. It is worth noting that the parameters used for these simulations are intended to provide qualitative results. In practice, the cases where the variance vary under LOS conditions is linked with a change in the mean greater than that simulated. Hence, in practice, under LOS conditions the difference between T_{P} and T_{SAM} will be greater, and then T_{P} will still perform better even with larger changes in the variance.

6.4 PVT Integrity with Signal-Level Integrity (PVT+I^2)

Once the application of SCD for signal-level threat detection has been presented, we will show how to integrate this approach into current GNSS receivers with PVT integrity capabilities, thus leading to what we call the PVT+I^2 module. At this point, it is worth distinguishing the following modules:

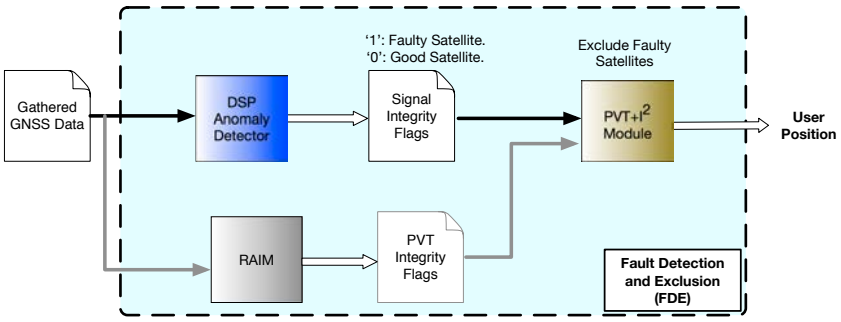

Figure 6.11: Diagram for the implementation of the FDE mechanism based on the integration of PVT and SLI (i.e. PVT+I^2 module)

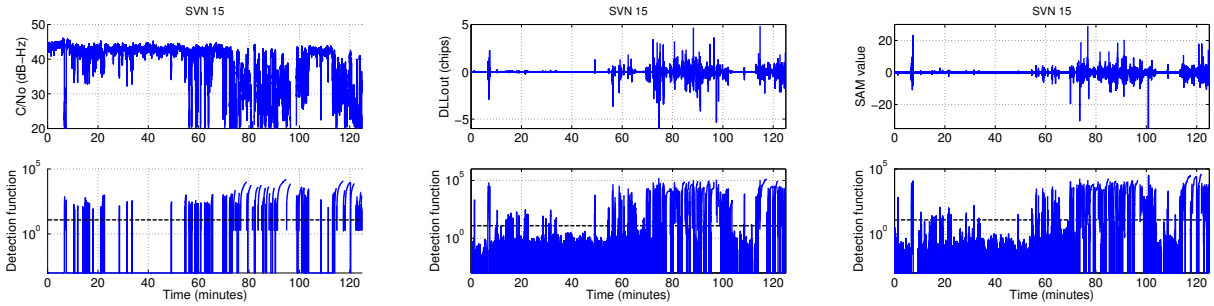

Figure 6.12: (Left) C/N_0, (middle) DLLout, and (right) SAM; and (down) detection function (e.g. CUSUM) time-evolution for the satellite with Satellite Vehicle Number (SVN) 15.

- The **PVT module**, which is focused on computing the navigation solution;
- The **PVT+I module**, which is focused on the computation of the navigation solution with PVT integrity. This level of integrity is obtained by traditional techniques such as RAIM or augmentation systems (e.g. SBAS);
- The **PVT+I^2 module**, proposed in this thesis, which is focused on the computation of the navigation solution but implementing either signal or PVT integrity or both of them.

A schematic representation is shown in Figure 6.11. First of all, SCD algorithms are implemented into the so-called "Digital Signal Processing (DSP) anomaly detector". This module implements the detectors described so far, and it generates validity flags indicating whether the satellites in view are faulty or not. Flags are activated whenever the output of the corresponding detection function (i.e. CUSUM, FMA, ...) exceeds the detection threshold. Finally, the PVT+I^2 module uses these SLI flags in order to exclude the faulty satellites from the computation of the navigation solution. Moreover, the PVT+I^2 module may also use the PVT-level integrity flags generated by PVT integrity techniques. For the results to be presented in this thesis, we use the classical RAIM algorithm (see Section 3.5). Next, we describe the FDE, with special interest in the generation of SLI flags.

6.4.1 Signal-Level integrity flags

We will show through a practical example how the signal-level integrity flags are generated. We show herein the results of processing some collected GNSS samples in urban environments, details on the collected data may be found in the following chapter. The output of the multipath SLM as well as the corresponding detection function values are shown in Figure 6.12, for the

SVN	12	15	17	24	28
C/N_0	**27%**	68%	29%	84%	68%
DLLout	**30%**	70%	**49%**	88%	75%
SAM	41%	71%	**47%**	89%	75%

Table 6.2: Percentage of time that the most available satellites are used.

case of the satellite with Satellite Vehicle Number (SVN) 15. As illustrated, this satellite is in nominal conditions most of the time, except from minute 70 to 100 where harsh propagation conditions are experienced. This is shown by the the detection function repeatedly exceeding the detection threshold (dashed black line) during this time period. For the sake of clarity, we have fixed the detection threshold in order to guarantee a time between false alarms of 1 h. Specifically, from (4.52) we have a logarithmic relationship with the threshold (i.e. $h = \ln(N_{\text{fa}})$), being $N_{\text{fa}} \doteq t_{\text{fa}}/T_{\text{sp}}$ the metric samples between false alarms, with t_{fa} the time between false alarms and T_{sp} the sampling time of the metric. For the presented results in Figure 6.12 we have used $T_{\text{sp}} = 20$ ms, so for $t_{\text{fa}} = 1$ h we have $h = \ln(3600/20 \cdot 10^{-3}) = 12.1$. During those time periods in which the detection function values are above the detection threshold, the SLI flag for the different metrics is activated, indicating that the SVN 15 should be excluded from the navigation solution computation.

Now, based on the detection function and the corresponding flags, we can compute the availability of each GPS satellite in view as the amount of time that each metric is not reporting faulty measurements. Table 6.2 shows the availability in percentage for the most available satellites in view within the analyzed data. We can see how the results for the DLLout and the SAM are quite similar, indicating that they declare a faulty measurement roughly at the same time. In contrast, the detector based on the C/N_0 metric exhibits more restrictive results, thus leaving less useful satellites for the subsequent position calculation. This effect is due to the fact that the C/N_0 values may fluctuate due to other effects rather than multipath, such as shadowing or fading that may become less perceptible to the rest of metrics (i.e. the DLLout and SAM). For instance, we see for SVN 17 how the DLLout and SAM exhibit a similar availability of 49% and 47%, respectively, whereas for the C/N_0 it becomes 29%. However, we can see that there are some exceptions where the DLLout and SAM do not coincide that much. For instance, for SVN 12, availability for the C/N_0 and DLLout is quite similar (i.e. 27% and 30%, respectively), but for the SAM it exhibits a higher value of 41%.

6.4.2 Flag fusion rules

In order to circumvent potential disagreements between different detection metrics, and to obtain only one flag per satellite in view (as required by the PVT module), we present two heuristic flag fusion rules:

- **Restrictive flag**: This option is the default one and it involves declaring a faulty measurement whenever any of the multipath detection flags declares a threat. That is,

$$T_{\text{r}}(i) \doteq T_{\text{c}}(i) | T_{\text{d}}(i) | T_{\text{s}}(i), \tag{6.70}$$

  with $T_{\text{c}}(i)$, $T_{\text{d}}(i)$ and $T_{\text{s}}(i)$ the C/N_0, DLLout and SAM flags of the ith satellite, respectively, at a given epoch; and $|$ the logical OR operator. This flag is named restrictive

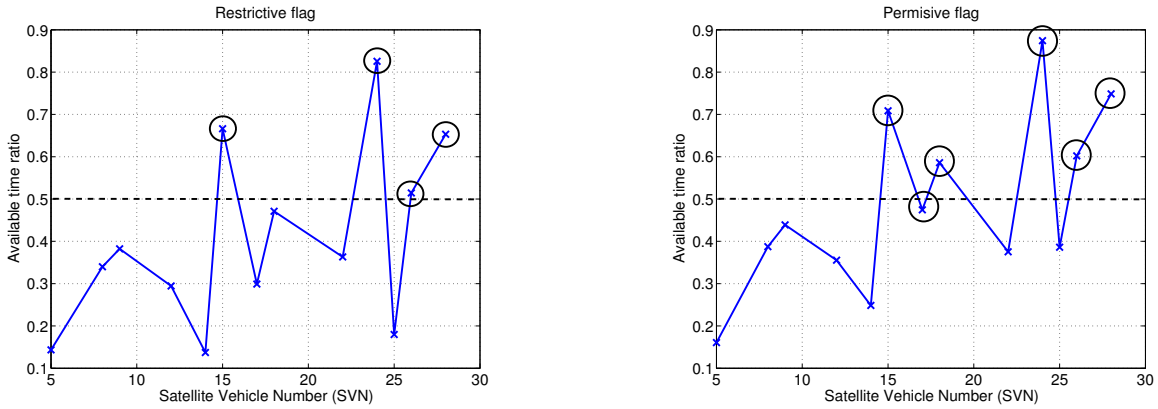

Figure 6.13: Percentage of time that the different satellites, indicated by their SVN, are decided to be used. (Left) restrictive and (right) permissive flags.

because it is the more sensitive to the presence of faults.

- **Permissive flag**: This option is proposed in order to compensate the restrictiveness of the C/N_0 metric and then avoid detections due to other effects different from multipath like shadowing or attenuation. The idea is to declare a faulty measurement when at least two multipath detection flags declare a fault. Thereby, we also provide robustness against false alarms because a false alarm will be declared only when two metrics produce a false alarm at the same time, which is an unlikely situation. The expression for the permissive flag of the ith satellite at a given snapshot is as follows:

$$T_{\mathrm{p}}(i) \doteq T_{\mathrm{c}}(i) + T_{\mathrm{d}}(i) + T_{\mathrm{s}}(i) > 1. \tag{6.71}$$

  In this case the flag is named permissive because it is not until two metrics declare a threat that the flag is activated.

Figure 6.13 shows the availability of every GPS satellite in view for the two fusion rules described above. The two plots show how for the restrictive flag the availability of GPS satellites is smaller than for the permissive flag. Indeed, we see that for the restrictive case, only three satellites are available more than the 50% of the time, whereas for the permissive one, up to five of them are close to 50% of time. In the sequel, we will use both flags configurations into the integrity PVT+I^2 algorithm and see which one produce better results. Anyway, this results and those presented in Table 6.2 show a very low availability of "healthy" GPS satellites, being likely that we only have the minimum number of satellites needed to obtain the position solution (i.e. 4 satellites) or even less and then losing the continuity of the GPS. This evidence the necessity of using multi-constellation receivers when implementing SLI, so that the availability of satellites is increased, thus improving the continuity and accuracy of the position solution.

6.5 Conclusions

In this chapter we have studied a new era of integrity schemes based on SCD. This has been motivated by the urgent need of a new level of integrity for critical applications, namely the the signal-level integrity. This new level of integrity requires the prompt detection of local degrading effects in order to assure the safely use of these critical applications in terrestrial environments.

In particular, we have investigated how to improve current PVT integrity algorithms through the use of SLI. Regarding the PVT integrity module (i.e. PVT+I), we have analyzed the implementation of traditional sequential RAIM. This has been important to recall the role of RAIM in GNSS receivers in order to further link them with the TCD framework. Once this role has been presented, we have investigated different decision rules based on different SCD solutions. Indeed, we have compared them under the setting of RAIM with the use of numerical simulations. From the obtained results we can conclude that the best solutions, in terms of availability, are those based on solutions of the SCD problem. Specifically, the best solution is the one given by the use of the FMA stopping time, the one proposed in this dissertation. The formulation of the integrity problem under the framework of TCD relieves the second requirement introduced in Chapter 5 for the modernization of integrity algorithms. That is, a timely detection is approached, understanding timely detection as the detection that minimizes the probability of unbounded detections (i.e. delayed more than a given time).

It is for the above discussion that the use of the so-called sig-RAIM may play a prominent role for the design of future integrity algorithms in terrestrial environments. The key point is that sig-RAIM can compute metrics to monitor the quality of the signal from features of the received signal, measurable within the GNSS receiver, without need of external information. This is beneficial for the availability of RAIM schemes, which is the main drawback of traditional RAIM in terrestrial environments. This increment in the availability is also driven by the use of the FMA stopping time, which has been shown to give the best results in terms of integrity risk for a fixed false alarm probability. These results are translated into either an improvement of the availability or equivalently an improvement on the minimum detectable change (i.e. sensitivity). Finally, sig-RAIM provides the additional level of SLI needed to detect local effects, which currently is not available in traditional RAIM schemes. In order to take advantage of these properties, so that integrity is improved in general, and in urban environments in particular, this module has to be combined with both PVT and PVT+I modules. This combination leads to the so-called PVT+I^2 algorithm, which is the one providing the additional SLI in real working conditions. So far in this dissertation, we have provided all the theoretical and simulation results needed to define and design the PVT+I^2 algorithm. At this point, we need to test this algorithm in a real-life scenario in order to validate the proposed concept of SLI. This proof of concept is given in Chapter 7, which is aimed at providing experimental results obtained from a real GNSS receiver and post-processing the gathered data.

Appendix 6.A Edgeworth Series Expansion and EVT Approximations

We focus here on the proof of Corollary 7 based on the materials we published in [ER17b]. We study the development of the Edgeworth series expansion and we apply the EVT in order to obtain approximations for the bounds of the SAM stopping time, T_{SAM}. We refer the interested reader to [ER17b] for the evaluation of these approximation by numerical simulations. It is important to say, that although these approximations are given for the TCD problem when dealing with multipath detection based on the SAM, they are valid for any detection problem involving inhomogeneous quadratic tests. These tests involve the sum of a dependent non-central chi-squared with a Gaussian random variable, and they arise in many signal processing-related problems in bio-medicine [Ber07], finance [Sch11] or engineering [ER15d], just to mention a few. Nevertheless, the use of inhomogeneous quadratic forms has typically remained in the realm of estimation or optimization theory, where some parameters need to be estimated or where the forms are part of some optimization cost function [Stk11]. To the best of the author's knowledge, little attention has been paid so far to inhomogeneous quadratic tests in detection problems, where the statistical characterization of these forms becomes of paramount importance.

Motivated by this observation, the problem addressed in [ER17b] was that of finding a closed-form approximation to the cdf of inhomogeneous quadratic forms, in order to easily compute the error probabilities in detection problems. The main goal is to obtain a compact and analytical formulation that can easily be parameterized in time-varying working conditions. The approach considered herein is based on the use of the Edgeworth series expansion, which provides an analytical expression for the pdf under analysis based on its constituent moments [Ken48, p. 169]. In contrast to other methods, such as the saddle-point approximation, the advantage of the Edgeworth expansion is that it can easily be integrated to give an analytical expression for the cdf, and does not require the cumulant generating function to be known in closed-form [BN79]. In the following we give some preliminaries in order to provided some stepping-stone results needed to prove Corollary 7. Then, closed-form approximations for the PUD and PFAP are provided, making use of the Edgeworth series and EVT, respectively. Finally, the use of these closed-form approximations to the bounds of T_{SAM} is provided, reaching the ultimate goal of proving Corollary 7.

6.A.1 Preliminaries

Let $\{Y_n\}_{n\geq 1}$ be a sequence of iid random variables whose inner structure is given by the following inhomogeneous quadratic form,

$$Y_n \doteq aX_n^2 + bX_n + c \tag{6.72}$$

for some constants $\{a, b, c\}$ and X_n some random variable. Because of the presence of a quadratic and a linear term in (6.72), the distribution of Y_n is not straightforward. This problem is further aggravated when we actually intend to find the distribution for the sum of $\{Y_n\}_{n\geq 1}$, denoted herein by the random variable Z as follows,

$$Z \doteq \sum_{n=1}^{m} Y_n \tag{6.73}$$

The distribution of Z, denoted by f_Z, involves m times the convolution of the pdf of Y_n. This poses insurmountable obstacles for the derivation of a closed-form expression for f_Z, and therefore for the corresponding cdf, F_Z. In order to circumvent this limitation, we propose the use of some approximations that provide a tight match to the original distribution, while providing a mathematically tractable closed-form expression. We will briefly recall here the CLT, which becomes a simple reference benchmark for the approximations to be proposed later on, as well as some indications on the statistical moments of Z to be used as well.

Theorem 11 (CLT for the density of Z). Let Z be the sum of m independent random variables $Y_1, Y_2, \ldots, Y_m$, with mean $\mu_Z = \mu_1 + \mu_2 + \ldots + \mu_m$ and variance $\sigma_Z^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_m^2$. Then,

$$f_Z(z) \xrightarrow[m\to\infty]{} \phi(\tilde{z}) \doteq \frac{1}{\sqrt{2\pi}} e^{-\tilde{z}^2/2} \tag{6.74}$$

$$F_Z(z) \xrightarrow[m\to\infty]{} \Phi(\tilde{z}) \doteq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\tilde{z}} e^{-\lambda^2/2} d\lambda \tag{6.75}$$

where $\tilde{z} \doteq (z - \mu_Z)/\sigma_Z$, $\phi(\tilde{z})$ the standard Gaussian pdf and $\Phi(\tilde{z})$ the standard Gaussian cdf [Pap02].

Lemma 3 (Moments of Z). Let Z be the sum of m independent random variables $Y_1, Y_2, \ldots Y_m$. The k-th order moment of Z denoted by $\xi_Z^k \doteq \mathrm{E}\left[Z^k\right]$ can be computed using the multi-nomial theorem [Bol68] as

$$\xi_Z^k \doteq \mathrm{E}\left[Z^k\right] = \sum_{l_1+l_2+\ldots+l_m=k} \frac{k!}{l_1! l_2! \cdots l_m!} \xi_{Y_1}^{l_1} \xi_{Y_2}^{l_2} \cdots \xi_{Y_m}^{l_m} \tag{6.76}$$

for all sequences $\{l_n\}_{n=1}^m \in \mathbb{Z}$ such that their sum is equal to k, and where ξ_Y^l stands for the l-th order moment of the inhomogeneous quadratic form in (6.72).

Remark 7. The computation of moments up to order four will be used later in this work, so it is convenient to provide here the particular case of (6.76) for $k = \{1, 2, 3, 4\}$. After some cumbersome but straightforward manipulations, we have:

$$\xi_Z = m\xi_Y \tag{6.77}$$

$$\xi_Z^2 = m\left[\xi_Y^2 + (m-1)(\xi_Y)^2\right] \tag{6.78}$$

$$\xi_Z^3 = m\Big[\xi_Y^3 + (m-1)\left[3\xi_Y^2\xi_Y + (m-2)(\xi_Y)^3\right]\Big] \tag{6.79}$$

$$\begin{aligned}\xi_Z^4 = m\Big[&\xi_Y^4 + (m-1)\big[4\xi_Y^3\xi_Y + 3\xi_Y^2 \\ &+ (m-2)\big[6\xi_Y^2(\xi_Y)^2 + (m-3)(\xi_Y)^4\big]\big]\Big].\end{aligned} \tag{6.80}$$

Lemma 4 (Moments of Y_n). Let $Y_\mathrm{s} \doteq a_\mathrm{s} X_{\mathrm{s},n}^2 + b_\mathrm{s} X_{\mathrm{s},n} + c_\mathrm{s}$ for some constants $\{a, b, c\}$ and some random variable X_n. Then the moments of Y_s are given by

$$\xi_Y^k \doteq \mathrm{E}[Y_\mathrm{s}^k] = \sum_{i=0}^{k} A(i), \tag{6.81}$$

with

$$A(i) \doteq \sum_{j=0}^{i} \binom{k}{i} \binom{i}{j} a_\mathrm{s}^{k-i} b_\mathrm{s}^{i-j} c_\mathrm{s}^{j} \xi_{X,2k-i-j}, \tag{6.82}$$

where $\xi_X^k \doteq \mathrm{E}[X_{\mathrm{s},n}^k]$ is the k-th order moment of $X_{\mathrm{s},n}$ and $\binom{l}{i} \doteq k!/(i!(k-i)!)$ is the binomial coefficient.

Proof. Let us rewrite Y_n as $Y_\mathrm{s} = u + v$, with $u \doteq a_\mathrm{s} X_{\mathrm{s},n}^2$ and $v \doteq b_\mathrm{s} X_{\mathrm{s},n} + c_\mathrm{s}$. Applying the binomial expansion we have,

$$Y_\mathrm{s}^k = \sum_{i=0}^{k} \binom{k}{i} u^{k-i} v^i = \sum_{i=0}^{k} \binom{k}{i} a_\mathrm{s}^{k-i} X_{\mathrm{s},n}^{2(k-i)} \left(b_\mathrm{s} X_{\mathrm{s},n} + c_\mathrm{s}\right)^i . \tag{6.83}$$

Proceeding in the same way for the right-hand side of (6.83),

$$\begin{aligned} Y_\mathrm{s}^k &= \sum_{i=0}^{k} \binom{k}{i} a_\mathrm{s}^{k-i} X_{\mathrm{s},n}^{2(k-i)} \left(\sum_{j=0}^{i} \binom{i}{j} (b_\mathrm{s} X_{\mathrm{s},n})^{i-j} c_\mathrm{s}^j \right) \\ &= \sum_{i=0}^{k} \sum_{j=0}^{i} \binom{k}{i} \binom{i}{j} a_\mathrm{s}^{k-i} b_\mathrm{s}^{i-j} c_\mathrm{s}^j X_{\mathrm{s},n}^{2k-i-j} \end{aligned} \tag{6.84}$$

which leads to (6.81) once the expectation is applied. □

6.A.2 Inhomogeneous quadratic tests in TCD problems

We will first introduce how inhomogeneous quadratic forms arise in the TCD problem under study and later on we will introduce the detection performance metrics to be evaluated and the distributions whose closed-form expression is sought.

Test statistic

We will focus on the FMA stopping time given by (4.90), whose statistic results in

$$\Upsilon_n \doteq \sum_{k=n-m+1}^{n} \mathrm{LLR}_k, \tag{6.85}$$

where $\mathrm{LLR}_k \doteq f_{X,1}(X_k)/f_{X,0}(X_k)$ is the LLR of the sample X_k. The LLR turns out to be an inhomogeneous quadratic form when evaluated for the signal model in (6.26). That is,

$$\mathrm{LLR}_k = aX_k^2 + bX_k + c \tag{6.86}$$

where the constants $\{a, b, c\}$ are given by

$$a = \frac{\sigma_{\mathrm{s},1}^2 - \sigma_{\mathrm{s},0}^2}{2\sigma_{\mathrm{s},0}^2 \sigma_{\mathrm{s},1}^2}, \quad b = \frac{\sigma_{\mathrm{s},0}^2 \mu_{\mathrm{s},1} - \sigma_{\mathrm{s},1}^2 \mu_{\mathrm{s},0}}{\sigma_{\mathrm{s},0}^2 \sigma_{\mathrm{s},1}^2}, \text{ and } c = \ln\left(\frac{\sigma_{\mathrm{s},0}}{\sigma_{\mathrm{s},1}}\right) + \frac{\sigma_{\mathrm{s},1}^2 \mu_{\mathrm{s},0}^2 - \sigma_{\mathrm{s},0}^2 \mu_{\mathrm{s},1}^2}{2\sigma_{\mathrm{s},0}^2 \sigma_{\mathrm{s},1}^2}. \tag{6.87}$$

The detection metric in (6.85) is actually the accumulation of m inhomogeneous quadratic forms, and therefore it can be modeled by the random variable Z in (6.73). While the exact density of (6.85) is unknown, the statistical moments can be derived using Lemma 3 and 4.

Detection performance

The detection performance is measured in terms of the PUD and PFAP, which are bounded by

$$\beta_m(h) \doteq \mathbb{P}_1(\Upsilon_n < h) \text{ and } \alpha_{m_\alpha}(h) \doteq 1 - [\mathbb{P}_\infty(\Upsilon_n < h)]^{m_\alpha}, \tag{6.88}$$

respectively. Due to the one-to-one relationship between (6.85) and (6.73), we can reformulate the above bounds in as

$$\beta_m(h) = F_{Z,1}(h) \tag{6.89}$$

$$\alpha_{m_\alpha}(h) = 1 - [F_{Z,0}(h)]^{m_\alpha} \tag{6.90}$$

with $F_{Z,0}$ and $F_{Z,1}$ the distribution of Z in the absence and in the presence of a transient change, respectively. These distributions have no closed-form expression, but tight approximations can adopted instead. For instance, using the Edgeworth series expansion to be presented next in Section 6.A.3. This approach works well for $\beta_m(h)$ in (6.89), since it directly depends on the cdf of Z, for which the Edgeworth expansion can readily be derived using the moments of Z in Lemma 3. However, some difficulties are found for $\alpha_{m_\alpha}(h)$ in (6.90) due to the presence of the m_α-th power on the cdf of Z. In that case, the approximation errors incurred by the Edgeworth series tend to be amplified, thus potentially violating the upper bound inequality in (4.92). We will address this issue by adopting an alternative closed-form approximation using results from EVT, as described next in Section 6.A.4.

6.A.3 Closed-Form approximation for $\beta_m(h)$, the upper bound on $\mathrm{P_{ud}}$

For a sufficiently large m, the density of Z in (6.73) can be assumed to be Gaussian in virtue of the CLT. This certainly relaxes the complexity of the problem at hand, and provides an acceptable match with the target density. However, the CLT approximation is often too loose for small m or when focusing on the tails of the resulting distribution, as it occurs when dealing with error probabilities (e.g. $\mathrm{P_{ud}}$ and $\mathrm{P_{fap}}$). A tighter approximation can be obtained through the following theorem [Cra46, p. 223]:

Theorem 12 (Gram-Charlier type-A expansion). The error between the target density f_Z and the CLT approximation can be modeled by a series expansion as follows:

$$\epsilon(\tilde{z}) \doteq f_Z(z) - \phi(\tilde{z}) = \phi(\tilde{z}) \sum_{p=3}^{\infty} C_p H_p(\tilde{z}) \tag{6.91}$$

where $H_p(\tilde{z})$ is the Hermite polynomial of degree p and C_p the projection of the target density onto $H_p(\tilde{z})$,

$$C_p \doteq \frac{1}{p!} \int_{-\infty}^{\infty} H_p(\tilde{z})\, f_Z(\tilde{z}) d\tilde{z}. \tag{6.92}$$

Corollary 9 (Gram-Charlier type-A expansion for f_Z). Using the result in (6.91), the density of Z can be represented through the following series expansion,

$$f_Z(z) = \phi(\tilde{z}) \left[1 + \sum_{p=3}^{\infty} C_p H_p(\tilde{z}) \right]. \tag{6.93}$$

A useful result to extend the previous result for f_Z to an expansion for F_Z is given by the following lemma.

Lemma 5. Let $H_p(\tilde{z})$ be the Hermite polynomial of degree p, then

$$\left(\frac{\mathrm{d}}{\mathrm{d}\tilde{z}}\right)^p \phi(\tilde{z}) = (-1)^p H_p(\tilde{z})\phi(\tilde{z}). \tag{6.94}$$

Proof. First, note that $\phi(\tilde{z}) \doteq e^{-\tilde{z}^2/2}/\sqrt{2\pi}$, and then

$$e^{-\tilde{z}^2/2} = \sqrt{2\pi}\phi(\tilde{z}) \text{ and } e^{\tilde{z}^2/2} = \left(\sqrt{2\pi}\phi(\tilde{z})\right)^{-1}. \tag{6.95}$$

On the other hand, the Hermite polynomials are defined as

$$H_p(\tilde{z}) \doteq (-1)^p e^{\tilde{z}^2/2}\left(\frac{d}{d\tilde{z}}\right)^p e^{-\tilde{z}^2/2}. \tag{6.96}$$

Hence, substituting (6.95) into (6.96) we obtain

$$H_p(\tilde{z}) = (-1)^p(\phi(\tilde{z}))^{-1}\left(\frac{d}{d\tilde{z}}\right)^p \phi(\tilde{z}), \tag{6.97}$$

which leads to (6.94). □

With this result we can prove the following corollary.

Corollary 10 (Gram-Charlier type-A expansion for F_Z). Integrating the result in (6.91), the distribution of Z can be represented through the following series expansion,

$$F_Z(z) = \Phi(\tilde{z}) - \sigma_Z\phi(\tilde{z})\sum_{p=3}^{\infty} C_p H_{p-1}(\tilde{z}). \tag{6.98}$$

Proof. The proof of Corollary 9 follows straight away from the Taylor series expansion of $f_Z(\tilde{z})$ and the orthogonal properties of the Hermite polynomials [Cra46]. However, some further manipulations are required to proof Corollary 10. Starting from the distribution definition, we have that

$$F_Z(z) \doteq \int_{-\infty}^{z} f_Z(u)du \approx \Phi(\tilde{z}) + \sum_{p=3}^{\infty} C_p \int_{-\infty}^{z} \phi(\tilde{u})H_p(\tilde{u})du, \tag{6.99}$$

where $\tilde{u} \doteq (u - \mu_Z)/\sigma_Z$ and the first term follows by the definition of the standard Gaussian distribution. The integral is solved by integrating (6.94),

$$\int_{-\infty}^{z} \phi(\tilde{u})H_p(\tilde{u})\,\mathrm{d}u = \sigma_Z \int_{-\infty}^{z} (-1)^p \left(\frac{\mathrm{d}}{\mathrm{d}\tilde{u}}\right)^p \phi(\tilde{u})\,\mathrm{d}\tilde{u} = \sigma_Z(-1)^p\left[\left(\frac{\mathrm{d}}{\mathrm{d}\tilde{u}}\right)^{p-1}\phi(\tilde{u})\right]_{-\infty}^{z}, \tag{6.100}$$

where the first equality follows by applying a change of variable. Applying (6.94) we have

$$\int_{-\infty}^{z} \phi(\tilde{u})H_p(\tilde{u})du = \sigma_Z(-1)^{2p-1}H_{p-1}(\tilde{z})\phi(\tilde{z}) = -\sigma_Z H_{p-1}(\tilde{z})\phi(\tilde{z}), \tag{6.101}$$

and then (6.98) follows by substituting this result into (6.99). □

While the results in (6.93)-(6.98) provide a closed-form approximation for $f_Z(z)$ and $F_Z(z)$, it is well-known that the Gram-Charlier approximation may suffer from some instabilities and

convergence issues [Ken48]. In particular, the terms of the infinite series in (6.91) do not monotonically decrease with increasing the order p, thus making the truncation of the asymptotic series a not trivial task. Notwithstanding, these issues can be circumvented by rearranging the error terms so as to provide a series expansion with guaranteed convergence [Cra46]. This rearrangement of terms leads to the so-called Edgeworth series expansion, which consists on grouping the error terms with similar order. This is the case, for instance, of terms $p = 3$, $p = \{4, 6\}$ and $p = \{5, 7, 9\}$. Using this observation, we are now in position to provide a closed-form approximation for the upper bound on $\mathrm{P_{ud}}$ in (6.89).

Proposition 10 (Edgeworth approximation for $\mathrm{P_{ud}}$). Using the result in (6.98), the upper bound on $\mathrm{P_{ud}}$ in (6.89) can be approximated as follows

$$\beta_m(h) \approx \beta_{\mathrm{EDG},m}(h) = \Phi(\tilde{h}_1) - \sigma_{Z,1}\phi(\tilde{h}_1) \sum_{p\in\mathcal{A}} C_{p,1} H_{p-1}(\tilde{h}_1) \tag{6.102}$$

where $\mathcal{A} = \{3, 4, 6\}$ and $C_{p,1}$ are the Hermite coefficients computed using $f_{Z,1}(z)$ under $\mathcal{H}_1$, and given by

$$\begin{aligned} C_{3,1} &= \frac{\xi_{Z,1}^3 - 3\xi_{Z,1}\xi_{Z,1}^2 + 2(\xi_{Z,1})^3}{6\sigma_{Z,1}^3}, \\ C_{4,1} &= \frac{\xi_{Z,1}^4 - 4\xi_{Z,1}\xi_{Z,1}^3 + 6(\xi_{Z,1})^2\xi_{Z,1}^2 - 3(\xi_{Z,1})^4}{24\sigma_{Z,1}^4} - \frac{1}{8}, \\ C_{6,1} &= 10C_{3,1}^2, \end{aligned} \tag{6.103}$$

where $\xi_{Z,1}^k$ is the k-th order moment of Z, which can be evaluated using the results in Lemma 3 under $\mathcal{H}_1$. Finally, $\sigma_{Z,1}$ is the standard deviation of Z that can be obtained as $\sigma_{Z,1} = \sqrt{\xi_{Z,1}^2 - (\xi_{Z,1})^2}$, and $\tilde{h}_1 \doteq (h - \xi_{Z,1})/\sigma_{Z,1}$.

6.A.4 Closed-Form approximations for $\alpha_{m_\alpha}(h)$, the upper bound on $\mathrm{P_{fap}}$

Edgeworth series approximation

A closed-form approximation for the upper bound of $\mathrm{P_{fap}}$ in (6.90) can similarly be obtained substituting $F_{Z,0}$ by its Edgeworth series expansion, as already done in (6.102) for $\mathrm{P_{ud}}$.

Proposition 11 (Edgeworth approximation for $\mathrm{P_{fap}}$). Using the result in (6.98), the upper bound on $\mathrm{P_{fap}}$ in (6.90) can be approximated as follows

$$\alpha_{m_\alpha}(h) \approx \alpha_{\mathrm{EDG},m_\alpha}(h) = 1 - \left[\Phi(\tilde{h}_0) - \sigma_Z\phi(\tilde{h}_0) \sum_{p\in\mathcal{A}} C_{p,0} H_{p-1}(\tilde{h}_0)\right]^{m_\alpha} \tag{6.104}$$

where $\mathcal{A} = \{3, 4, 6\}$, $\tilde{h}_0 \doteq (h - \xi_{Z,0})/\sigma_{Z,0}$ and $C_{p,0}$ are the Hermite coefficients computed using $f_{Z,0}(z)$ under $\mathcal{H}_0$, and given by (6.103) replacing $\xi_{Z,1}$ by $\xi_{Z,0}$.

EVT approximation

Although the Edgeworth series provides a better approximation for the tails of the density of Z than the CLT, there is still some mismatch between the tail of the approximation and the true density. This mismatch is negligible for the case of approximating $F_{Z,1}$ in (6.89), but it is not when approximating the m_α-th power of $F_{Z,0}$ in (6.90). The approximation inaccuracies become amplified and the upper bound inequality for the PFAP is not guaranteed to be preserved

anymore. In order to circumvent this issue we will formulate an alternative approximation for $\mathrm{P}_{\mathrm{fap}}$ making use of results from extreme value theory (EVT). EVT has historically been linked to the statistical problem of flood frequency analysis, where predicting such extreme events is of paramount importance. However, EVT is also widely adopted today in applications dealing with finance, insurance or engineering [Gum12]. In the problem at hand, we can understand the upper bound in (6.90) as the probability that none of the m_α trials of Z under $\mathcal{H}_0$ exceeds the threshold h. This is equivalent to state that the maximum of these m_α trials does not exceed it either. Following this rationale we make use of the following theorem.

Theorem 13 (EVT distribution). Let $U \doteq \max_{1\leq i\leq N}\{Z_i\}$ with $\{Z_i\}_{i=1}^N$ iid samples whose density f_Z exhibits exponentially decreasing tails. Then the distribution of U becomes

$$F_U(u) \overset{N\to\infty}{=} \exp\left(-e^{-\gamma(x-\delta)}\right), \tag{6.105}$$

with

$$\delta = F_Z^{-1}\left(1-\frac{1}{N}\right) \text{ and } \gamma = N f_Z(\delta). \tag{6.106}$$

Proof. See [Gum12, p. 166]. □

Using the result above we can provide an alternative approximation for the upper bound on $\mathrm{P}_{\mathrm{fap}}$ as follows,

Proposition 12 (EVT approximation for $\mathrm{P}_{\mathrm{fap}}$). Using the result in Theorem 1, the upper bound on $\mathrm{P}_{\mathrm{fap}}$ in (6.90) can be approximated as follows

$$\alpha_{m_\alpha}(h) \approx \alpha_{\mathrm{EVT},m_\alpha}(h) = 1-\exp\left(-e^{-\gamma_{m_\alpha}(\tilde{h}_0-\delta_{m_\alpha})}\right), \tag{6.107}$$

where

$$\delta_{m_\alpha} = F_{Z,0}^{-1}\left(1-\frac{1}{m_\alpha}\right), \tag{6.108}$$

$$\gamma_{m_\alpha} = m_\alpha f_{Z,0}(\delta_{m_\alpha}). \tag{6.109}$$

Proof. We can rewrite the term $[F_{Z,0}(h)]^{m_\alpha}$ in (6.90) as

$$[F_{Z,0}(h)]^{m_\alpha} = \mathbb{P}_\infty\left(\bigcap_{i=1}^{m_\alpha} Z_i < h\right) = \mathbb{P}_\infty\left(M_{m_\alpha} < h\right), \tag{6.110}$$

with $M_{m_\alpha} \doteq \max_{1\leq i\leq m_\alpha}\{Z_i\}$. That is, we can obtain $[F_{Z,0}(h)]^{m_\alpha}$ as the probability that m_α iid signal samples of Z are below the value h, which is equivalent to the probability that the maximum of all m_α samples is below h. Thereby, we can make use of EVT to obtain $[F_{Z,0}(h)]^{m_\alpha}$. Since Z is the sum of dependent non-central chi-squared and Gaussian random variables, the density of Z has an exponentially decreasing tail and Theorem 13 is applicable to (6.110). On the other hand, since the density and distribution of Z are unknown, we apply the corresponding Edgeworth expansion in order to use Theorem 13, and the proof of Corollary 12 thus follows. □

Before concluding this section it is worth noting that to obtain δ_{m_α} in (6.108) we need to evaluate the inverse distribution $F_{Z,0}^{-1}$. Even though there is no closed-form for this inverse, it can easily be found by numerically solving $F_{Z,0}(\delta) = 1-(1/m_\alpha)$.

Chapter 7

Experimental Results with Real GNSS Signals

7.1 Introduction

In Chapter 6 we have dealt with the integration of SLI with RAIM techniques, making use of the SLMs investigated in Chapter 5. This was done in order to provide the additional SLI needed in terrestrial critical applications, leading to what we call PVT+I^2 module. Throughout the above mentioned chapters we have provided simulation results showing the capability of the proposed SLMs to detect interference and multipath in GNSS. We also have evaluated the theoretical findings in these chapters, and compared the considered detection schemes in Chapter 4. Nevertheless, so far, we have not presented results using real data measurements. This is the main goal of this chapter, that is to provide realistic results of the proposed techniques in order to show the capability of our suggested sig-RAIM algorithm to operate in real working conditions. Real signals considered in this work were gathered in the framework of the "Integrity Receivers" project (iGNSSrx) funded by the European Commission, being the UAB part of the consortium. Thanks to the use of these real data, recommendations are provided herein to set and tune the algorithms in practice. The results will show that the proposed algorithms for threat detection can successfully be tuned for different scenarios, and that satisfactory results in terms of accuracy, integrity and availability are obtained, thus paving the way for the implementation of a new era of realistic integrity monitoring algorithms. It is worth saying that the presented results considered a QCD framework because it was the framework used in the iGNSSrx project. The use of the TCD should be considered as a future line.

It is important to clarify here the procedure carried out to obtain the results presented in this chapter. One of the first task carried out by the UAB in the iGNSSrx project was to experimentally characterize the effects of the different analyzed threats into the proposed SLMs. That is to say, what happens to the SLM X when threat Y is present. Based on the observed effects and the metric outputs of those results, the second goal of our task in the iGNSSrx project, and the goal of the present chapter, was to use this information to design and validate sequential algorithms capable of automatically detecting these threats. The structure of this chapter is the following: First of all, in Section 7.2 we will introduce the experimental set-up used to obtain the real signal gathered in the iGNSSrx project, as well as the procedure carried

out to obtain the results presented in this chapter. Next, Section 7.3 and Section 7.4 show the experimental results intended to validate the use of the presented SLMs in Chapter 5 to detect interference and multipath, respectively. Moreover, the statistical characterization and behavior of the metrics under the absence and presence of local threats will be validated with real signal results. Finally, Section 7.5 deals with the evaluation of the impact of the proposed sig-RAIM in the navigation performance. Then, Section 7.6 concludes the chapter and give some ideas for future work in line with the introduced SLI in this thesis. The discussion presented in this chapter is based on the work carried out in the iGNSSrx, and most of them was published in [ER15c, ER15d].

7.2 Experimental Set-Up

In this section we introduce the experimental set-up used to obtain the presented results by processing real-life captured signal from both GPS and GLONASS in the framework of the iGNSSrx project. Particularly, two different measurement campaigns were carried out:

1. The first one comprised real signals captured with a GNSS receiver at the Joint Research Center (JRC) of the European Commision, in Ispra (Italy). This data is intended to analyze the interference SLMs with real signals;

2. The second campaign comprised real signals captured within the framework of the iGNSSrx project in urban environments. This kind of data is intended to analyze the multipath SLMs.

Next, we will explain with further detail both data campaigns. First, though, it is worth commenting that the GNSS receiver used in both data collection campaign was a multi-constellation single-antenna receiver able to capture GPS and GLONASS L1 signals with the following features: RF bandwidth of 4 MHz and 10 MHz, sampling rate of 10.28 MHz and 21.576 MHz, intermediate frequency of 2.57 MHz and 5.394 MHz, respectively for GPS and GLONASS, and 8 quantification bits.

7.2.1 Interference data set

This section describes the interference data collection carried out at the JRC between the 25-th and the 27-th of June 2014 in the European Microwave Signature Laboratory (EMSL). The tested jamming scenarios were aimed at covering those interferences consisting on non-navigation-like signals with the potential of disrupting or degrading the performance of GNSS receivers, comprising both unintentional and the intentional interferences (see Chapter 3). The set-up used in the JRC for the jamming tests is shown in Figure 7.1. An antenna was place on the roof of the EMSL building as the source of GNSS signals. A signal generator was used to produce the interferences. Then, the GNSS and interference signals were combined to feed the GNSS receiver. Moreover, the combined signal was split to also feed an spectrum analyzer, used to monitor the test. The combined signal was also filtered with a GNSS band-pass filter to emulate that the interference would have been filtered by the antenna.

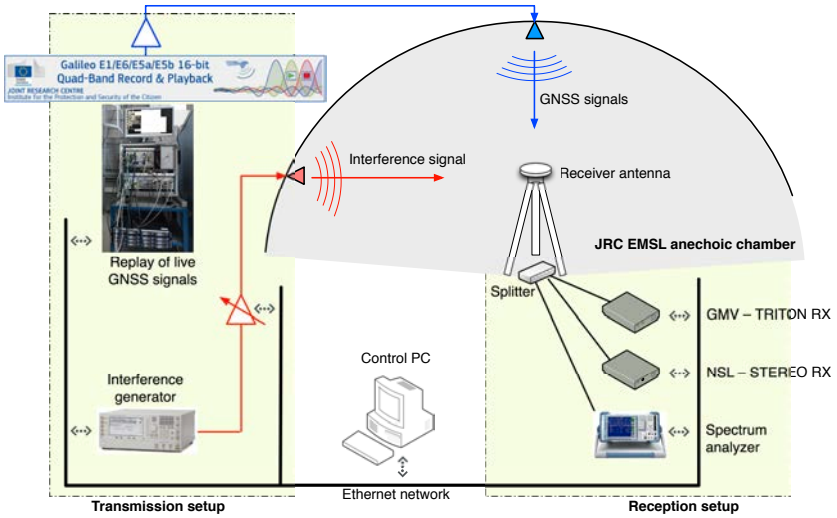

Figure 7.1: Experimental set-up for the interference analysis data collection campaign carried out at the EMSL of the JRC.

Figure 7.2: Truth trajectory of the analyzed scenario with data captured in London's (UK) downtown.

7.2.2 Multipath data set

In order to test the performance of the integrated PVT+I^2 approach, a test on a realistic urban scenario is considered. In particular, GNSS samples were gathered by the above mentioned GNSS receiver on board of a vehicle that was moving around London's (UK) downtown (see trajectory in Figure 7.2). Since the trajectory is in an urban area, the scenario covers both static (due to traffic) and dynamic cases with a maximum velocity of 50 km/h. In order to discriminate between nominal and faulty conditions, we first need to analyze the positioning errors that have been incurred by the GNSS receiver with the aim of identifying the pieces of data suffering from low and large errors. The positioning error can be calculated by means of a truth trajectory available thanks to the developed platform in the iGNSSrx project. We will consider to be large errors those that are greater than some error bound. This error bound must be selected in such a way that errors exceeding this bound are likely produced by the presence of multipath or some disturbing effect. On the other hand, low errors are those not exceeding this bound and then they should not be due to the presence of multipath, just to nominal conditions. This will allow us to analyze the behavior of the SLMs with real signals without multipath (i.e. piece of data with low positioning error) and with presence of multipath (i.e. data with large error), and to confirm the behavior of these metrics. These errors are shown in Figure 7.3, which illustrates the mean positioning error at different time instants of the collected data the 24-th of April 2014, which is the scenario analyzed herein. The mean error is calculated computing

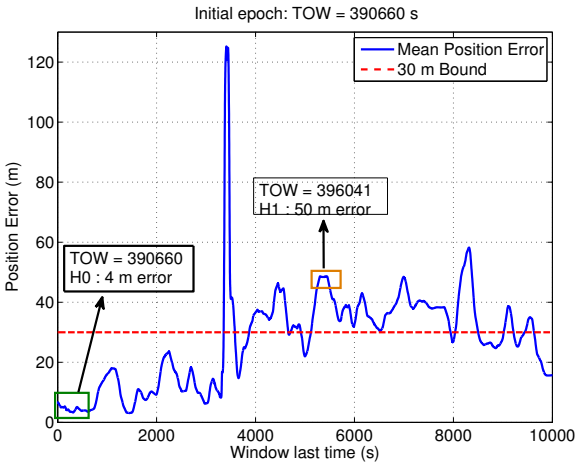

Figure 7.3: Positioning errors for the urban data collection analyzed.

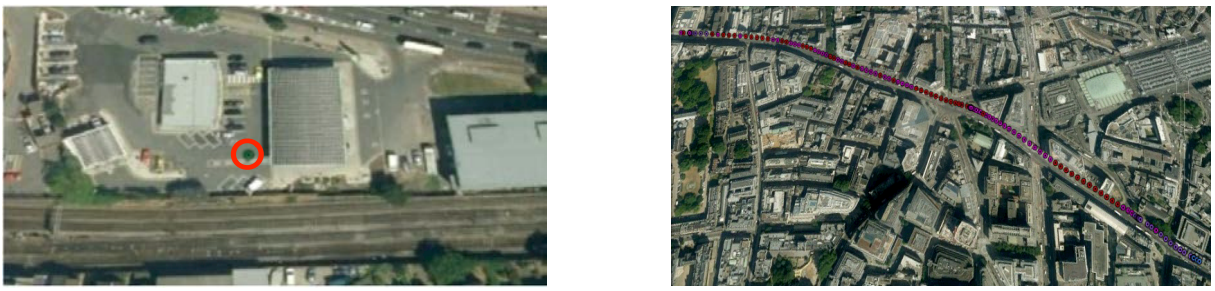

Figure 7.4: Truth trajectory for the processed pieces of data. (Left) TOW0 and (right) TOW1.

the average over 100 positions provided by the receiver. This averaged position is subtracted to the corresponding mean position of the truth trajectory, thus giving the mean positioning error. In fact, the receiver provides a position every second, then we obtain the mean error averaging 100 seconds.

Figure 7.3 also shows a dashed red line denoting the selected error bound of 30 m error, which is the selected boundary between the positioning errors treated as nominal errors or as integrity faults. We choose this error bound because we know that the first part of the collected data was captured in a free-multipath environment (i.e. before second 4000), whereas the rest of the data was captured in an environment likely to be affected by multipath. Thus, the 30 m bound is a good choice since it will discriminate between free-multipath and multipath environments. In order to select some time instants to analyze a piece of data with nominal conditions and another with integrity faults, we make the following choice:

- **Nominal Conditions ($\mathcal{H}_0$)**: For the data under $\mathcal{H}_0$ we take a window of 200 s centered at the time instant denoted by TOW0 = 390660 s. As we can see in the figure, the positioning error at this TOW0 is 4 m and then it is supposed to be under nominal conditions (i.e. without multipath).

- **Harsh Conditions($\mathcal{H}_1$)**: For the data under $\mathcal{H}_1$ we take a window of 138 s with starting point at 100 seconds before TOW1 = 396041 s (i.e. we have 100 seconds of data before TOW1 and 38 seconds after). In the figure we can see how the positioning error at TOW1 is 50 m and therefore, it is supposed to be under multipath conditions.

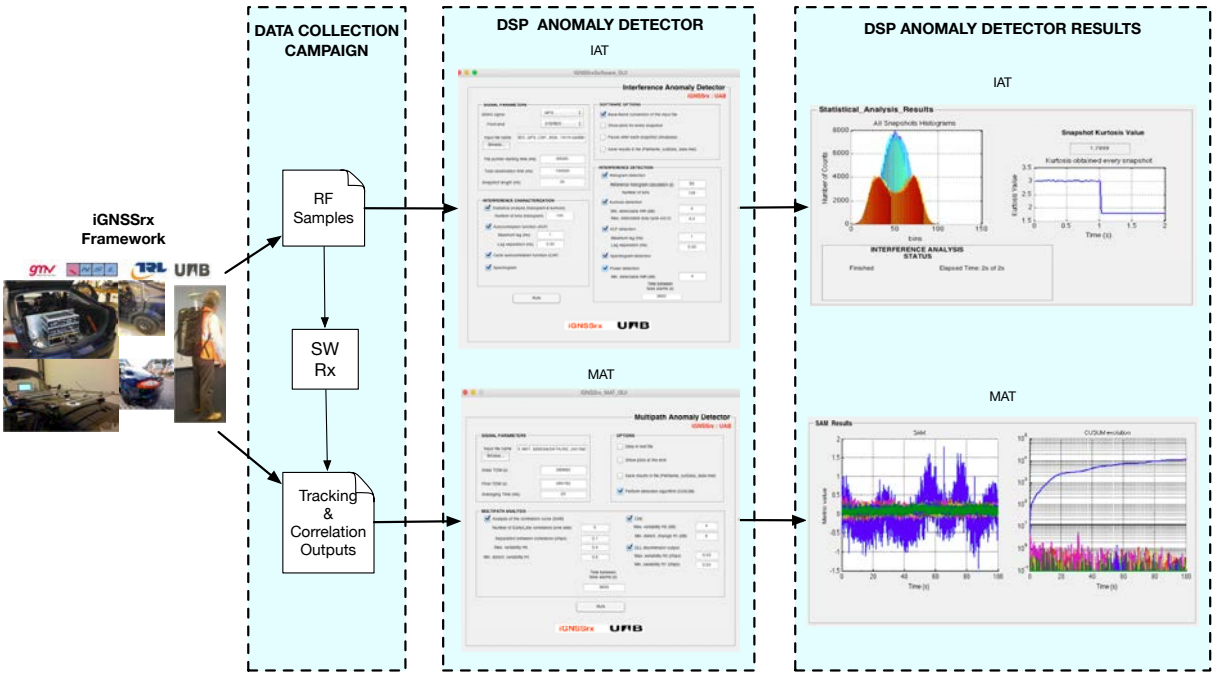

Figure 7.5: Process carried out to obtain the results presented in this chapter, based on the collected data and further DSP anomaly detector processing.

Figure 7.4 shows the truth trajectory that corresponds to the data in nominal conditions (i.e. TOW0) and harsh conditions (i.e. TOW1):

- **TOW0**: It can be seen how the receiver is static all the time. The place where the receiver is placed is a gas station, which as shown the picture is a clear space with not too many obstacles blocking the visibility with the satellites or generating several reflections.
- **TOW1**: In this case we can see how the receiver is in movement during the 138 seconds that last the data. This movement is inside the city within narrow streets plenty of buildings with different heights and different obstacles that can block the visibility of some satellite and produce reflections.

7.2.3 DSP anomaly detector

In order to analyze the data collection campaigns carried out during the iGNSSrx project, a MATLAB-based software was developed by UAB in order to implement the proposed SLM and the corresponding detectors. This software was designed to read the data gathered within the iGNSSrx project, process it in order to obtain the SLMs and to provide the functionality of automatically detecting anomalies in the form of interference or multipath. This software is the so-called DSP anomaly detector, which is composed by two sub-systems, namely the Interference Analysis Tool (IAT) and the Multipath Analysis Tool (MAT). These tools process the samples at the front-end filter and the correlation/tracking outputs, respectively, provided by the GNSS receiver used in the project. These samples are processed in order to detect anomalies in the received signals. Particularly, the IAT carries out different types of interference analyses on the input data samples. These analyses are those studied in Section 5.2. The IAT was designed to be started either by using a Graphical User Interface (GUI) or a configuration file, in order to set the parameters of the tool. Once the detectors have been properly configured, the IAT runs the

specified tests and provides the corresponding results. Similarly, the MAT is composed of a set of independent routines, each one implementing one of the different multipath SLM presented in Section 5.3. Contrary to the IAT, the input samples for the MAT come from the different stages of a multi-correlator GNSS receiver; that is, the MAT uses the outputs of the tracking stage of a GNSS receiver. As for the IAT, the MAT can be configured through a GUI or manually editing the configuration file. Once the configuration is set, the MAT runs the specified tests and provides the corresponding results. A scheme of the process to obtain the presented results in this chapter is illustrated in Figure 7.5. The data gathered in the data collection campaigns is stored in form of the signal samples gathered by the GNSS receiver, as well as in form of the tracking and correlation outputs. These two kind of data is used as input for the DSP anomaly detector, composed of the IAT, which processes the RF samples, and the MAT, which processes the tracking and correlation outputs. Finally, the DSP anomaly detector provides the results of processing the collected data in form of the graphical representation of the selected SLMs as well as output files. In the following, we show the obtained results after processing the gathered data in both interference and multipath analysis data campaigns.

7.3 Experimental Results for Interference Detection

This section includes results of the interference data collection with real signals carried out at the JRC EMSL. It is worth mentioning that for the jamming data collection, the AGC of the front-end was disabled. Nevertheless, the dynamic range was adjusted in such a way that it allows us to obtain a Gaussian distribution for the received signal samples under nominal conditions. The GNSS receiver and set-up above mentioned is used. We divide the results according to the different interference signals, which include the most prevalent and dangerous types of interferences that we can find in GNSS (i.e. CW, PW and Gaussian wide-band interferences). Actually, chirp interferences are also common and were also captured at the JRC. However, we do not present here the results because they can be inferred from that obtained with continuous and pulsed wave interferences. Then, for each one we show the results of the interference detection algorithms presented in Chapter 5. In fact, we show the more relevant metrics for each type of interference. The presented results include the statistical characterization (i.e. histogram) of the methods for both presence and absence of interference. Moreover, they include the evolution of the SLM. This is obtained by processing 2 minutes of the real signals data, whose first minute includes the absence of interference case, and the last minute includes the presence of interference case. The snapshot time is fixed to 20 ms, which is a common used value in practice by receivers and it is a convenient snapshot time in order to get a sufficient large value for the snapshot samples assuring the assumption of large sample size made in Chapter 5. It is also worth pointing out that the results presented in this section corresponds to the GPS captured data. So, in each snapshot we process $N \approx 2 \cdot 10^5$ samples in order to provide each SLM.

7.3.1 CW interference

Next, we show the results after processing an in-band CW. This interference can be detected by any of the proposed methods in Chapter 5 for interference detection, but we only show here the histogram- and ACF-based detection. Specifically, the interference starts being detected when it takes an INR = 4 dB. However, here, we show the results for the case of INR = 24 dB. This

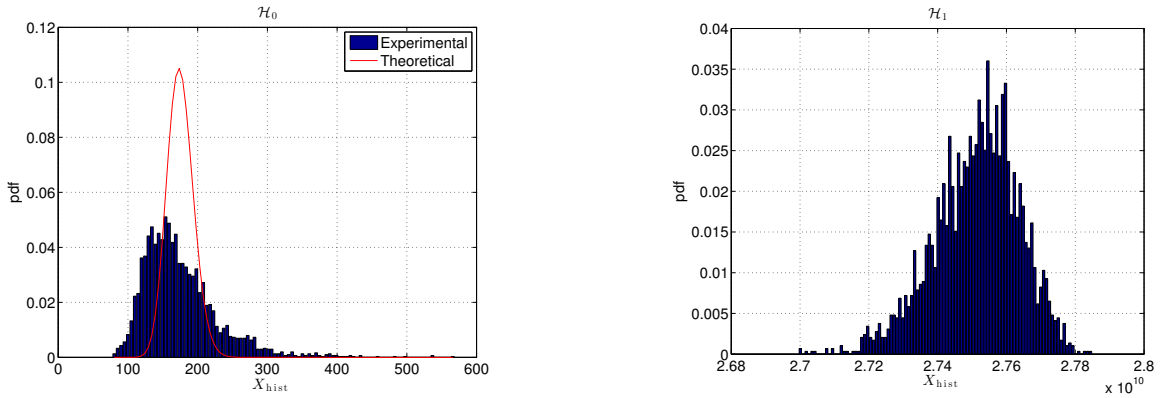

Figure 7.6: Statistical characterization of the Histogram-based SLM. (Left) Characterization under $\mathcal{H}_0$ and (right) under the presence of the real CW interference with INR = 24 dB.

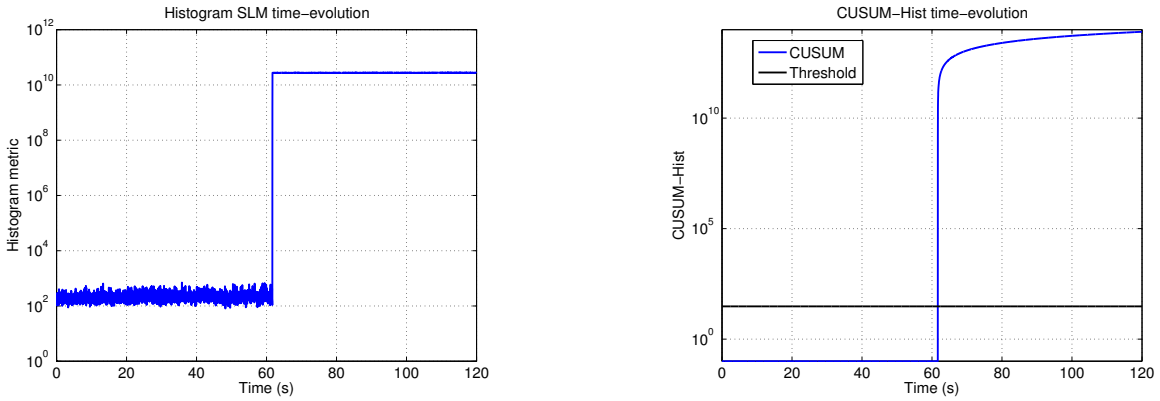

Figure 7.7: CW interference detection based on the histogram SLM. (Left) Metric and (right) CUSUM time-evolution.

is so because in this way we can clearly see the change in the metrics when the interference appears. Moreover, this value fits well with the range of values we can encounter in practice, since the interference power is much greater than the noise level.

Histogram-based detection

Figure 7.6 shows the statistical characterization of the histogram metric for both absence and presence of the CW interference. We can see how the histogram in both cases is quite different. Thus, the histogram metric exhibits a large change that allows the detection of the interference. The results are obtained by using 100 bins to compute the histogram (i.e. $N_b = 100$). In the left plot we can see how the experimental distribution of the histogram metric does not fit the theoretical chi-squared distribution. The fitted chi-squared distribution is a chi-squared distribution with degrees of freedom equal to the mean of the metric, which in this case was around 200. This fitted pdf corresponds to the theoretical one since the mean of a chi-square distribution is equal to the degrees of freedom. The fact that the experimental and theoretical characterizations do not fit may be due to the fact that the distribution of the received signal in absence of interference is not actually a zero-mean Gaussian distribution. After analyzing different sets of data we realized that the mean of the metric for different snapshot times is around $2 \cdot N_b$. Then, as for the histogram-based detection is based on the O-CUSUM, we can still use the proposed method but taking into account that the mean before the change is equal to $\mu_0 = 2 \cdot N_b$ instead of being equal to $N_b - 1$.

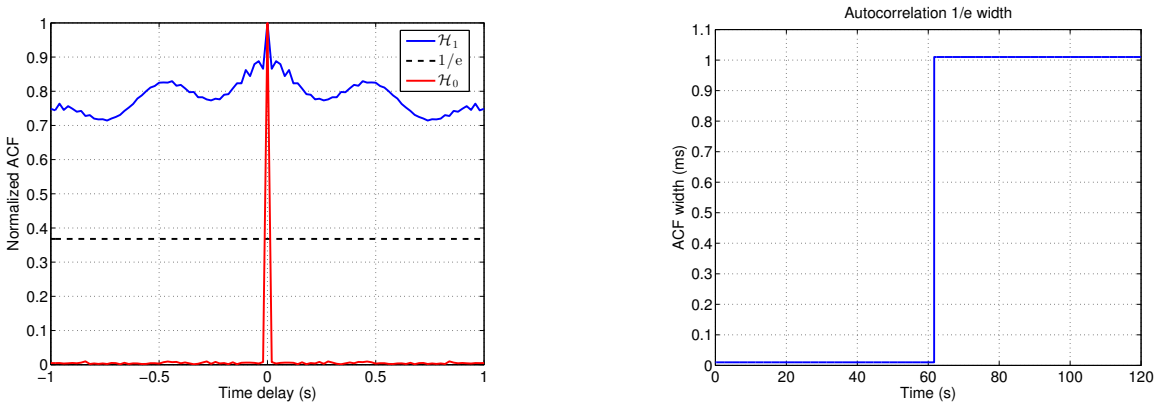

Figure 7.8: (Left) ACF width under $\mathcal{H}_0$ and under the presence of the CW; and (right) ACF width time-evolution.

Figure 7.7 shows the histogram-based detection metric and the CUSUM time-evolution. In the left plot we see how the SLM changes just when the interference appears (i.e. second 60). The value after change will depend on the INR since the increment of the INR leads to an increment on the effects of the interference, and subsequently, an increment of the detection metric. Finally, in the right plot, we can see the CUSUM time-evolution. In the absence of interference, it is close to 0, whereas just at the change time, it increases and crosses the threshold, clearly indicating the presence of interference. These results are obtained tuning the algorithm as follows:

- We set the degrees of freedom to $2 \cdot N_b$, with $N_b = 100$ bins. Then, we have a mean before the change equal to $\mu_0 = 2 \cdot N_b$. Moreover, we use an offset $b = 5 \cdot \mu_0$, which is large enough to maintain a negative mean before change and it is short enough to maintain a positive mean after change (see Figure 7.6).
- We fix the threshold $h = \ln(N_{\text{fa}}/\omega_0)$, with N_{fa} the number of desired metric samples between false alarms and ω_0 the non-zero root of (5.5). Finally, we use the Hist-CUSUM decision rule in (5.6).

ACF-based detection

Figure 7.8 shows the ACF behavior under $\mathcal{H}_0$ and $\mathcal{H}_1$ in the left plot, whereas in the right plot shows the evolution of the ACF width as a function of time. For the ACF calculation we use $k_{\max} = 1$ ms and $\Delta k = 20$ μs. In the left plot, we see how the ACF under $\mathcal{H}_0$ (red curve) decays abruptly when the time delay departs from 0. Indeed, the ACF falls below the $1/e$ factor just for the first computed lag. On the other hand, under $\mathcal{H}_1$ we see that the ACF remains above the factor $1/e$ for all the computed range (i.e. $\pm k_{\max}$). Hence, the ACF width exhibits a change in the presence of interference that allows us to detect it, as it is depicted in the right plot of Figure 7.8. We see how the ACF width takes an abrupt change just when the interference appears (i.e. 60-th second). The value obtained under $\mathcal{H}_0$ is equal to the lag resolution (i.e. 20 μs), whereas when the interference appears it takes a value of 1 ms. This is so because, as shown in the left plot, the ACF does not decay by a factor $1/e$ anywhere within the computed range. Therefore, as the computed range extends from $-k_{\max}$ to $k_{\max}$, the obtained ACF width is equal to $k_{\max}$. Here we do not show the CUSUM behavior because it is similar to that obtained for the histogram case (see Figure 7.7). The tuning of the CUSUM-ACF becomes the following:

- We set the mean and variance before change to $\mu_0 = \Delta k$ and $\sigma_0^2 = 10^{-12}$, respectively,

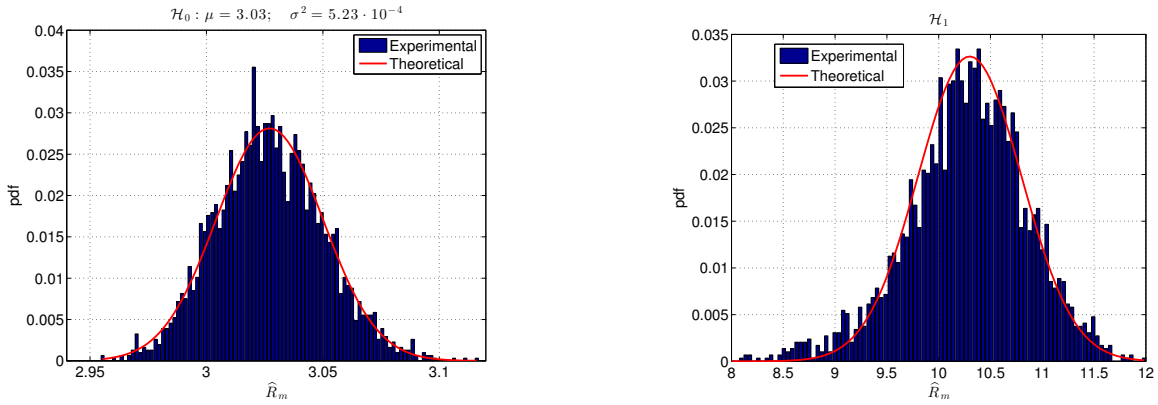

Figure 7.9: Statistical characterization of the kurtosis SLM. (Left) Characterization under $\mathcal{H}_0$ and (right) under the presence of the real PW1 interference.

with Δk the separation between obtained lags in the ACF calculation by (5.18).

- We set the mean after change to $\mu_1 = 5 \cdot \mu_0$ and fix the threshold to $h = 1$. Finally, we use the LLR for the Gaussian mean change, with the SLM in (5.19), and then we use the CUSUM decision rule.

7.3.2 PW interference

In the following, we show the results of the pulsed sinusoidal interferences. We present two different interferences each one with a different duty cycle, in order to see the corresponding effects into the detection metrics. In fact, we have chosen, among the generated interferences, the lowest (i.e. 0.98%) and greatest (i.e. 30%) duty cycle, which correspond to a small and large pulse width, respectively. We will denote the two interferences as PW1 and PW2, respectively. Moreover, we show the results for INR = 14 dB. We present the results for the kurtosis- and spectrogram-based detection, which are the most relevant metrics to detect PW interferences.

Kurtosis-Based detection

Figure 7.9 shows the statistical characterization of the kurtosis SLM for PW1. In the left plot we observe how the histogram under $\mathcal{H}_0$ takes a Gaussian shape, with mean 3 and variance approximately equal to $5 \cdot 10^{-4}$, whereas using the theoretical expression in (5.12), with $N = 2 \cdot 10^4$, we obtain a value of $1.2 \cdot 10^{-4}$. This is so because, actually, the value of N in (5.12) refers to the number of independent samples used to calculate the kurtosis estimate. Then, as in our case the signal samples are filtered by a base-band bandwidth of 2 MHz, not all the samples in a snapshot period are independent, since they will be correlated by the filter shape. In order to obtain the proper variance value we would have to define the number of independent samples as $N = N_{\text{sp}}(B/F_{\text{s}})$, with $N_{\text{sp}} = F_{\text{s}}T_{\text{sp}}$, where F_{s} and T_{sp} are the sampling rate and snapshot time, respectively. In our case, we obtain a factor 0.2 multiplying the snapshot samples. Substituting this effective number of samples into the variance expression we obtain the factor 5 that makes fit both experimental and theoretical results, as shown in the figure. On the other hand, in the right plot of Figure 7.9 we see how the histogram presents a change in the mean under $\mathcal{H}_1$. Moreover, we see how the histogram has a Gaussian shape, but presenting larger tails. This is so because the kurtosis oscillates between two different values due to the variation of the effective duty cycle from different snapshots, so that outliers increasing the tails of the Gaussian distribution appear.

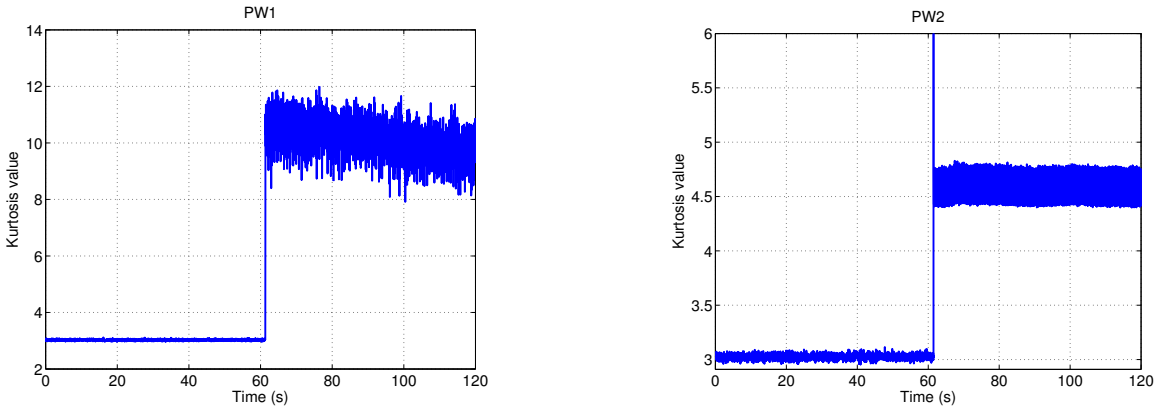

Figure 7.10: (Left) Kurtosis SLM time-evolution for the PW1 and (right) PW2.

Figure 7.10 shows the time-evolution of the kurtosis SLM for PW1 and PW2. We see how in both cases the metric fluctuates between two different values after the change (see last 60 seconds). This is due to the effect mentioned before on the duty cycle variation. In average, the kurtosis value after the change is around 10 and 5 for the PW1 and PW2, respectively. These values are close to the values obtained using the theoretical expression in (5.13) taking into account the effective INR (i.e. multiplied by the duty cycle). With these values we see that the lower the duty cycle, the larger the kurtosis value, confirming the behavior presented in Chapter 5. In view of this behavior, we can configure the CUSUM-Kurt as follows:

- We set the mean and variance before change to $\mu_0 = 3$ and $\sigma_0^2 = 5 \cdot (24/N_{\text{sp}})$, respectively. We set the mean after change to $\mu_1 = 4$, which is the theoretical value for a PW with INR = 14 dB and duty cycle of 30%, and fix the threshold $h = \ln(N_{\text{fa}})$. Indeed, the previous value of μ_1 is for the case of a PW, for a CW interference, we should use another CUSUM tuned with the proper mean after change in that case.
- Finally, we use the LLR for the Gaussian mean change in (5.15), and then we use the CUSUM-Kurt decision rule.

Spectrogram-Based detection

Here, we show the results of the spectrogram metric for the PW interference. Since the distributions in both hypotheses are unknown, the most important thing here is to see the large change in the mean of the metric, so that the use of a mean change CUSUM is coherent. It is for this reason that the statistical characterization of this metric is not shown. Among all the analyzed data, we see that the mean of the spectrogram metric in the absence of interference is around 9, whereas the variance value is about 0.2. On the other hand, we see how the metric under $\mathcal{H}_1$ departs from that under $\mathcal{H}_0$. In this case, the mean change value depends on the INR as well as the duty cycle, since depending on the ON state time of the pulsed interference the peak appearing in the spectrogram will be more or less large, therefore producing a greater or smaller value of the metric after change. This is shown in Figure 7.11, which presents the evolution in time of the spectrogram metric for the PW1 and PW2 (i.e. short and large duty cycle). We see how the metric varies between two different values after the change, as for the kurtosis case. After analyzing different data, we propose the following configuration for the Spect-CUSUM:

- We set the mean and variance before change to $\mu_0 = 9$ and $\sigma_0^2 = 0.2$, respectively; and we set the mean after change to $\mu_1 = 50$ and fix the threshold $h = \ln(N_{\text{fa}})$.

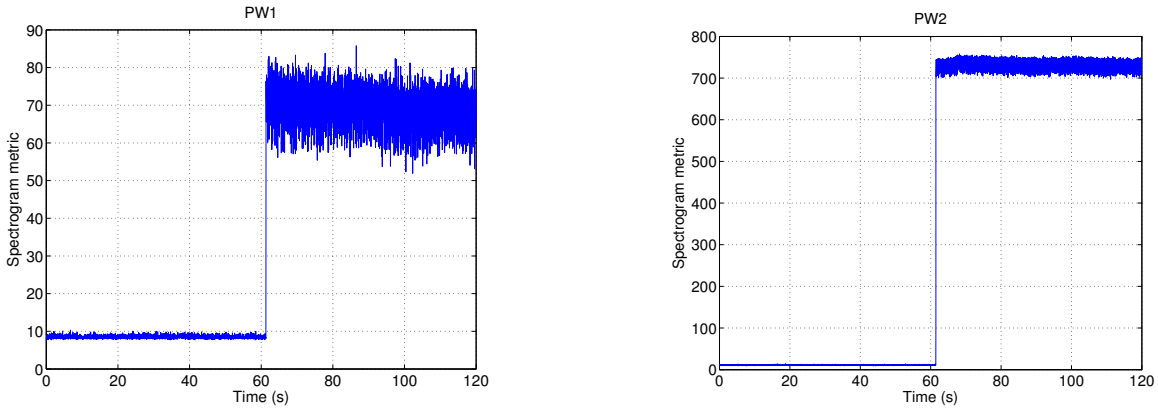

Figure 7.11: (Left) Spectrogram SLM time-evolution for the PW1 and (right) PW2.

- Finally, we use the LLR for the Gaussian mean change, with the spectrogram SLM defined in (5.23), and then we use the CUSUM decision rule.

7.3.3 Gaussian wide-band interference

This section is dedicated to present the results of a wide-band interference corresponding to a Gaussian signal of 50 MHz centered in-band (i.e. at 1.575420 GHz). We process all the INR available in the captured data, which starts with the absence of interference and then gradually increases in steps of 5 dB (-21, . . . , 24 dB). The interference introduces slight variations into all the SLM, except for the ACF width and spectrogram metrics for which there is not variation. However these variations are clearly detectable by the histogram- and received power-based detection, which are shown in this section. Since the interference is Gaussian, it is striking that it is detectable by the histogram-based metric, as the Gaussianity of the received signal samples is maintained even with interference, and then the statistical analysis metrics should not vary. However, for the histogram metric, the presence of interference is translated into an increase of the variance of the received signal samples (i.e. increase of the noise power). Therefore, if we know the noise power under nominal conditions (e.g. we estimate it), the presence of the wide-band interference spreads the histogram of the received samples, and then the histogram-based metric will depart from the ideal conditions value. The same happens with the received power metric, if we know the noise power under ideal conditions, we are able to discriminate an increase of this power due to the presence of interference.

Histogram-Based detection

Figure 7.12 shows the histogram metric and its CUSUM time evolution. In the left plot we see how as the INR increases (5 dB every minute) the histogram metric does so. Specifically, the metric takes a slight variation around 120 seconds, which is due to the interference with INR $= -11$ dB. However, the change is not substantial until about 180 seconds (i.e. INR $= -6$ dB). It is for this reason that the CUSUM algorithm starts detecting the interference around 180 s (see right plot of Figure 7.12). Nevertheless, the CUSUM is changing of decision among the time the interference with INR $= -6$ dB is present. This is so because the change is not large enough to maintain a mean after change positive, and then there are some snapshots where the metric is small enough to make the CUSUM decides in favor of $\mathcal{H}_0$. Hence, it is not until about 240 s (i.e. INR $= -1$ dB) when the CUSUM detects the interference among the entire interval.

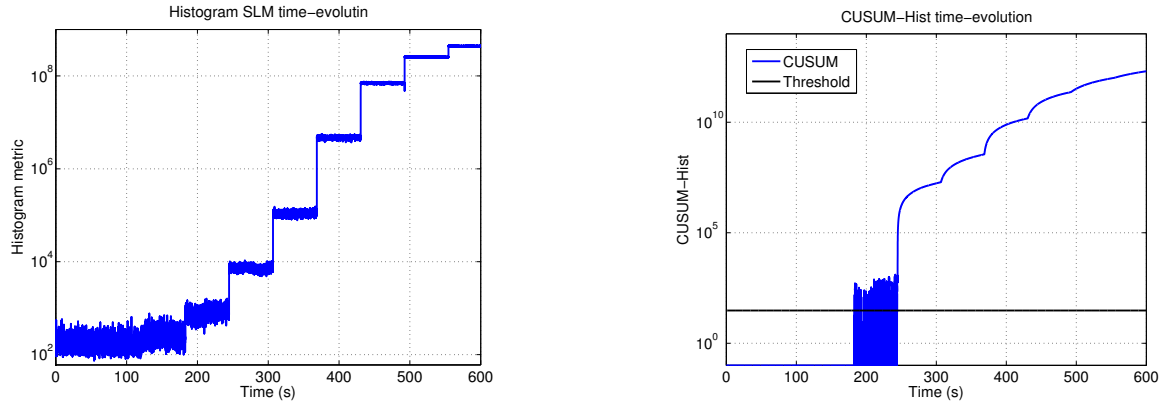

Figure 7.12: Histogram-Based SLM. (Left) Metric and (right) CUSUM time-evolution under the presence of a Gaussian wide-band interference.

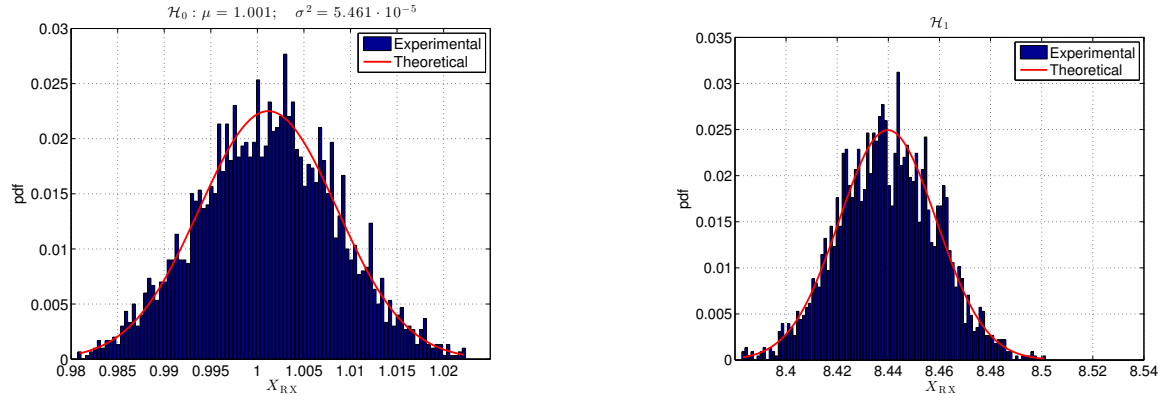

Figure 7.13: Statistical characterization of the power SLM for the wide-band interference. (Left) Characterization under $\mathcal{H}_0$ and (right) under the presence of the wide-band interference with INR = 24 dB.

The same configuration as the above stated for the CUSUM-Hist is used.

Power-Based detection

Figure 7.13 illustrates the statistical characterization of the power metric. As for the other methods, we show in the left plot the histogram under $\mathcal{H}_0$, and in the right plot the one when the wide-band interference is present with INR = 24 dB. We see how under $\mathcal{H}_0$, the expressions for the mean and variance before the change in (5.29) are approximately equal to the experimental results (i.e. $\mu_0 \approx 1$, $\sigma_0^2 = 4.4 \cdot 10^{-5}$). On the other hand, we see in the right plot how the histogram significantly changes in the presence of interference, and then it becomes detectable, as we expected. In this case, though, the relationship between the theoretical and experimental results for the mean and variance does not match exactly. Still, the distribution is Gaussian, as indicated in the figure. The reason is that the wide-band interference is filtered with a lower bandwidth, and then the total power of the interference after filtering is reduced. Figure 7.14 shows the evolution in time of the power metric and its CUSUM. In the left plot, we see how the change on the metric increases as the INR of the interference does so. In particular, with a minimum change on the metric fixed to 3, equivalent to an INR of 3 dB, the interference is detected around second 360 (i.e. INR = 4 dB). As for the histogram case, the power metric starts changing before second 360, and this can be detected by decreasing the minimum fixed change. However, detecting an interference with INR = 4 dB is enough for practical situations, since interferences will often appear with powers much greater than just 4 dB above the GNSS received power levels. Therefore, the minimum change fixed to 3 is a proper value to detect interferences

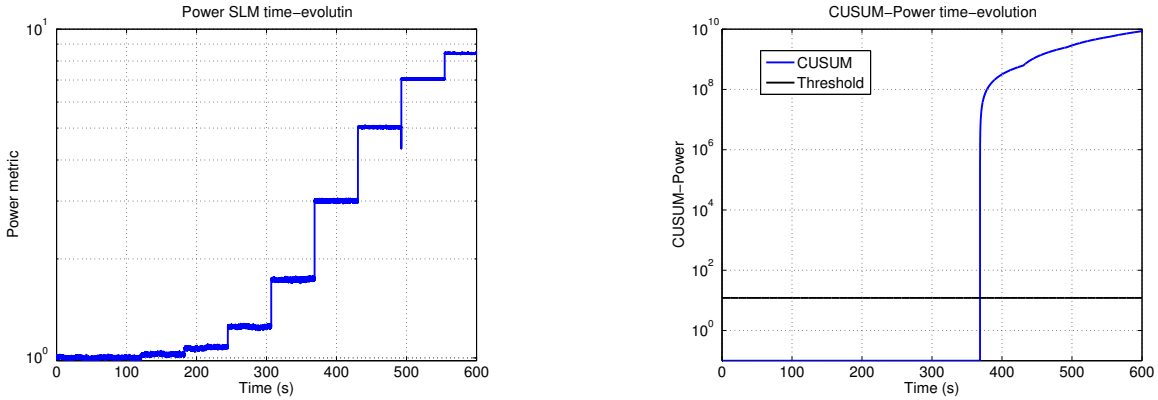

Figure 7.14: Wide-Band interference detection with the power-based SLM. (Left) Power metric and (right) CUSUM time-evolution.

SLM	Histogram	Kurtosis	Power
Detection rule type	O-CUSUM	Mean-Change	Mean& Variance-Change
Target interferences	All	All except wide-band	CW and wide-band
Tuning Parameters	None	$\mu_{\text{CW,PW}} = 2.2, 4$	$\mu_1 = 3$ $\sigma_1^2 = 2 \cdot 10^{-4}$
Threshold	$h = \ln\left(\frac{N_{\text{fa}}}{\omega_0}\right)$	$h = \ln(N_{\text{fa}})$	
Performance	$\mathfrak{D} \leq \frac{h}{\mu_1}$	$\mathfrak{D} \leq \frac{2h \cdot \sigma_0^2}{(\mu_1 - \mu_0)^2}$	$\mathfrak{D} \leq \frac{h}{\ln(\lvert\sigma_0/\sigma_1\rvert) + (\mu_1 - \mu_0)^2/(2\sigma_0^2) - 0.5}$

Table 7.1: Properties of the suggested interference detection algorithms.

with low INR and it provides a good level of false alarms. This is because decreasing the minimum change level would detect lower INR, but at the expense of increasing the number of false alarms. Hence, we can configure the CUSUM-Power as follows:

- We set the mean and variance before change to $\mu_0 = 1$ and $\sigma_0^2 = 5 \cdot (1/N_{\text{sp}})$, respectively; and we set the mean and variance after change to $\mu_1 = 3$ and $\sigma_1^2 = 5 \cdot (5/N_{\text{sp}})$ and fix the threshold to $h = \ln(N_{\text{fa}})$.
- Finally, we use the LLR for the Gaussian mean and variance change in (5.30), and then we use the CUSUM decision rule.

7.3.4 Summary of the selected SLM

Based on the results previously presented and those obtained in the iGNSSrx project after analyzing all the data collection campaign, we can summarize the main features of the proposed SLMs in terms of their suitability to certain scenarios, the type of sequential algorithm they are fed, and their main configuration parameters. A summary for the interference SLM is provided in Table 7.1, for the histogram, kurtosis and power SLMs, which are the selected metrics to provide a complete set able to detect all types of interferences. The ACF-width and TFA metrics are discarded due to their higher computational load, and the fact that they do not provide additional information that may be relevant for detection purposes. Regarding the histogram, it is based on the Bias-CUSUM, for which we only need to know the statistical characterization of the interference-free scenario (i.e. $\mathcal{H}_0$). This has the advantage of reducing the number of parameters to be tuned, but at the same time, it departs from the optimal

approach (e.g. CUSUM) and thus incurs in a larger detection delay. Instead, the mean-change (MC) and mean&variance-change (MVC) CUSUM implemented for the kurtosis and power SLM, respectively, do provide the quickest detection by minimizing the detection delay. For the power SLM, this is particularly true for CW and wide-band interference, since these are the interferences for which the Power-CUSUM has been specifically designed (i.e. using the mean and variance values in Table 7.1). For the kurtosis, wide-band interferences (particularly those occupying the whole front-end bandwidth) are the ones that cannot be detected because they are often perceived just as an increased noise level. Regarding the noise, it should be mentioned that all the proposed techniques except for the kurtosis, do depend on the noise power, but this has already been taken into account when processing the data measurements by sequentially estimating the noise power.

The CUSUM parameters are also indicated in Table 7.1. As we have already said, for the histogram we do not need to fix a minimum change magnitude. For the kurtosis, after analyzing all the results, a proper mean after change is 2.2 and 4 for CW and PW interferences, respectively. Those values belong to a minimum INR detectable of 4 dB for the CW case and of 10 dB and duty cycle equal to 0.25 for the PW case. On the other hand, for the power metric, we suggest a mean and variance after change of 3 and $2 \cdot 10^{-4}$, respectively, which correspond to a minimum detectable INR of 3 dB. Table 7.1 also shows the selected threshold as a function of the fixed metric samples between false alarms N_{fa}. Moreover, we also provide the corresponding bound for the detection delay, which depends on N_{fa} and the CUSUM parameters. For the histogram case, the threshold depends on both N_{fa} and the non-zero root ω_0, which is calculated from (5.5). On the other hand, the threshold for both the kurtosis and power metric depends on the N_{fa}, only. Since the non-zero root ω_0 is smaller than one, for the same N_{fa}, the threshold of the CUSUM-Hist will be larger than that obtained for the CUSUM for the kurtosis and power metric. For the same reason we can say that the CUSUM algorithm for the kurtosis and power metric will be quicker than the histogram one. In addition, if we analyze the detection delay for the kurtosis and the power metric, we see that the denominator of the power metric delay is greater than the one for the kurtosis. Therefore, the detection delay for the power metric is shorter than the kurtosis metric one.

7.4 Experimental Results for Multipath Detection

Next we show the results of processing the urban data collection campaign gathered on 24-th of April 2014. It is important to note that the results presented herein are intended to label or classify different data sets; that is, to differentiate between data sets captured under nominal and faulty conditions. We do so because we have already presented the statistical characterization of the SLM for multipath detection with real signal. It is for this reason that we show here the characterization process of the multipath threats. Thereby, we illustrate the method carried out to identify different scenarios dealing with the presence and absence of multipath.

7.4.1 Behavior of the SLM under nominal conditions

We will focus here on the results of the data that have a low error on the position, and therefore they are supposed to correspond to signals received under nominal conditions. As we have

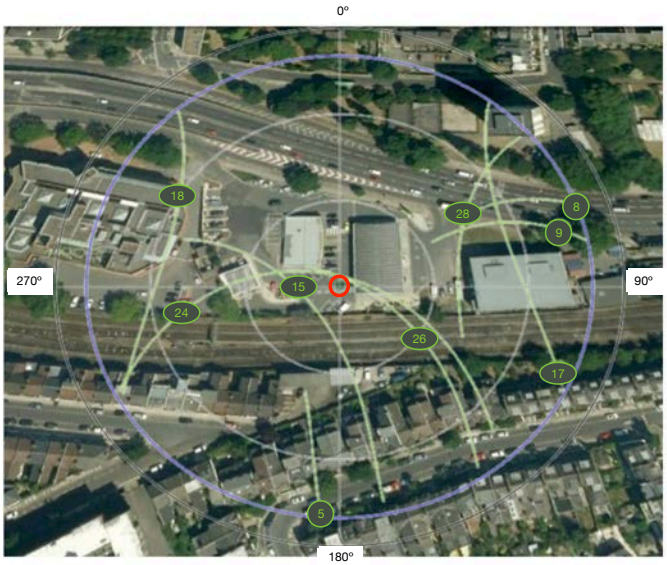

Figure 7.15: Sky plot with the configuration of the GNSS satellites in view at TOW0.

already commented, the piece of data to be analyzed belongs to the truth trajectory in the left plot of Figure 7.4, which is a static position located on a gas station with a relatively clear vision to the sky. This kind of environment often leads to nominal reception conditions. Figure 7.15 shows the sky plot for the TOW0 case. The sky plot shows the position of the different satellites in view on the region where the data have been captured at some time. The figure shows the position of the satellites in term of the azimuth and elevation angle with respect to the position of the receiver. The azimuth angle is indicated with the angle in the circumference (e.g. satellite 15 take an azimuth angle of 270°, satellite 5 is close to 180°). On the other hand, the elevation angle is indicated by the distance to the center (i.e. the closest to the center the greatest elevation angle). Sky plots presented in this dissertation has been obtained from [Tri14]. As we have already mentioned, here we process 200 s of data, and we analyze the sky plot at the analyzed position at the given time TOW0. The sky plot will allow us to confirm the different hypothesis about effects on the SLM generated by the MAT. As a general comment we can say that the results show how there are some satellites likely to be completely blocked by some obstacle (e.g. SVN 8), others are received under NLOS conditions (e.g. SVN 5), others are under LOS multipath condition (e.g. SVN 9 and 17). The rest of satellites are under nominal conditions (i.e. SVN 15, 18, 24, 26 and 28). Next we show the proposed multipath SLM for SVN 5 and 15, given by the MAT after processing the piece of data at TOW0.

SVN 5

Tracking measurements for SVN 5 are not available until around second 40 since the receiver is not able to track the signal of SVN 5. Figure 7.16 shows a behavior of the SLM that correspond to the expected behavior in the presence of multipath; that is, C/N_0 below the nominal value of 45 dB-Hz, DLLout taking values departing from the nominal value of 0, and SAM values also departing from 0. This behavior can be explained by evaluating the sky plot in Figure 7.15. We see that the elevation of SVN 5 is near to 10°, which is the limit at which a receiver would typically lose the view with the satellite, even without any obstacle blocking the LOS. In our case, the receiver do not lose the view with the satellite, but as shown in Figure 7.17 we have short houses in the path between the satellite and the receiver. This is why the received signal

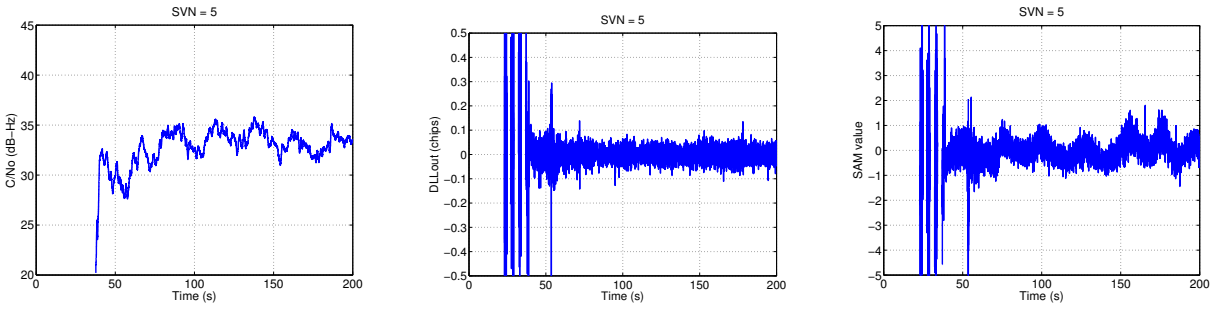

Figure 7.16: SVN 5: Multipath SLMs given by the MAT after processing the data captured at time TOW0. (Left) C/N_0, (middle) DLLout and (right) SAM value.

Figure 7.17: Line of view between SVN 5 and receiver.

may be received only by its multipath component (i.e. NLOS reception), producing the low mean C/N_0 and departure from zero of the DLLout and SAM. Finally, the fact that the results are not available at the beginning of the record, and that they appear at second 40 might be due to a low elevation of the satellite, thus loosing the view. Thereafter, the multipath condition seems to vary with time, which is the responsible of the oscillations in C/N_0 and SAM.

SVN 15

Figure 7.18 shows the results for SVN 15, summarized below:

- **C/N_0**: The mean C/N_0 value is around 44-45 dB-Hz, which reflects the nominal conditions reception. Nevertheless, we can also see small variations appearing in the form of spikes (i.e. green circles in C/N_0 plots on Figure 7.18) or as mean changes (i.e. red circles). These variations may suggest the presence of multipath. However, as these variations are small (i.e. lower than 2 dB) we can conclude that the impact of this multipath in the positioning error is small. This is why the final impact is nearly negligible in terms of positioning error.

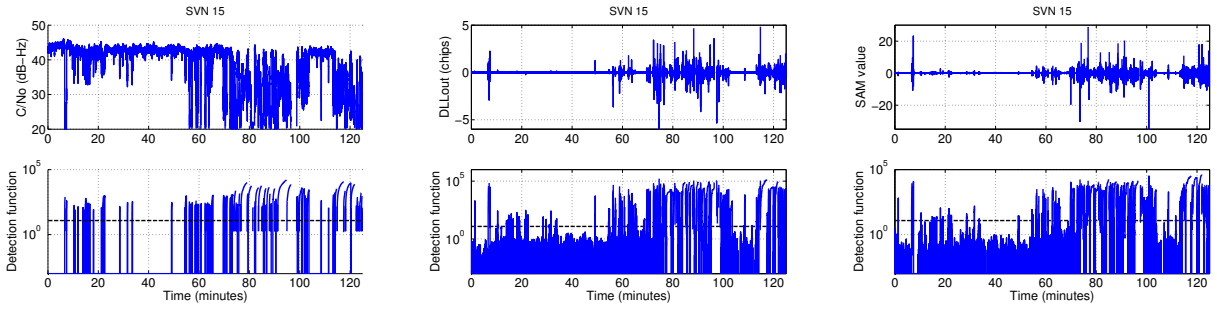

Figure 7.18: SVN 15: Multipath SLMs given by the MAT after processing the data captured at time TOW0. (Left) C/N_0, (middle) DLLout and (right) SAM value.

- **DLLout**: As illustrated in the figure, SVN 15 is tracked under nominal conditions, and then the DLLout is close to zero. The spike of 0.2 chips appearing around 50 seconds is due to the software receiver used to generate the tracking outputs, because it always appears.

- **SAM**: Since we have favorable propagation conditions, the SAM results coincide with the expected ideal values, with a mean around 0.1 (due to the filter shape). However, we can also notice some variations of the variance (i.e. square in the SAM plot). These variations do not appear in the C/N_0 values, this might be so because the multipath delay is so large that is not visible in the C/N_0 value, which is relevant for short delays, but it is visible in the SAM values, which is relevant for large delays.

As we can see in the sky plot of Figure 7.15, SVN 15 has direct line of sight with the receiver. In addition, it has a high elevation angle (i.e. 76°) and then it is not likely to suffer from multipath, since the possible obstacles near the receiver are not tall enough to reflect the signals coming from the satellites. If some multipath is present, it must come from some other tall building far away from the receiver, and thus the multipath rays will arrive with a large delay and low amplitude. It is for this reason that the results show a nominal behavior but in some moments they show a behavior like in presence of a large delay multipath.

7.4.2 Behavior of the SLM under harsh conditions

The piece of data to be analyzed here belongs to the truth trajectory showed in the right plot of Figure 7.4, which is located in an urban street with tall buildings and narrow streets. This kind of environment is a clear candidate to produce multipath. Moreover as the receiver is moving across the street, the multipath conditions will change depending on the visibility of each satellite with the receiver and the surroundings characteristic. Next, we show the different multipath SLM for the satellites SVN 15 and 24, as provided by the MAT after processing the piece of data at TOW1. Figure 7.19 shows the MAT results for the SVN 15 and 24. In the upper plot we see the results for SVN 15, and we see that it experiences different behaviors; for instance, from 0 to 40 s we see a behavior corresponding to multipath propagation (i.e. low C/N_0 mean, low DLLout and SAM variations). For 40–90 s the tracking loop is lost, and finally for 90–120 s the metrics behave under nominal conditions. Regarding SVN 24 (see lower plot), in general, we can notice that the results present a nominal behavior, with the C/N_0 metric having a mean value around 43 dB-Hz, and the DLLout and SAM metrics being close to zero all the time. For the latter, just small variations are observed due to the thermal noise. However, in the range of 50-90 s, we can see how the C/N_0 presents an abrupt change on its mean value. This change might be caused by a short delay multipath, which would explain the fact that this variation only appears in the C/N_0 and not in the DLLout and SAM metrics, since the latter are sensitive to large multipath delays and then the metrics may not be able to note the effects of the short delay multipath.

Next we provide a more detailed analysis of these results. In particular, we will present the sky plots on different time instants in order to see the visibility condition of the different satellites in view. We will analyze 3 different time instants, which provide information about three different relevant situations on the received signals:

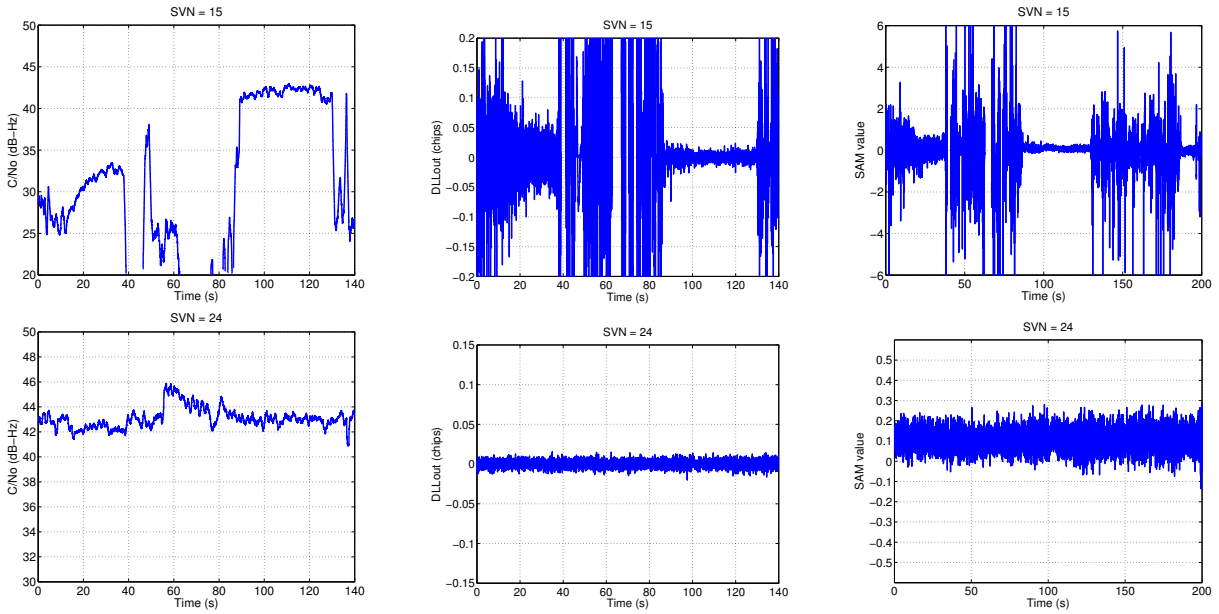

Figure 7.19: (Up) SVN 15 and (down) SVN 24: Multipath SLMs given by the MAT after processing the data captured at time TOW1. (Left) C/N_0, (middle) DLLout and (right) SAM value.

- **P1**: Most of the satellites, including SVN 15, are tracked with the presence of multipath, except SVN 24 that presents a nominal behavior. This situation belongs to the 0-50 s period, in Figure 7.19. We choose the second 20, corresponding to TOW = 395961 s, to represent a time instant in this period and taking the sky plot.

- **P2**: Only SVN 17 and 24 are tracked (so SVN 15 is not tracked), with some symptoms of being affected by multipath. This is the case of the period from 50 to 90 s in Figure 7.19. In this case we choose the second 70 corresponding to TOW = 396011 s in order to get the sky plot of this situation.

- **P3**: Finally, a different situation is when all the satellites are tracked except the SVN 17, which happens at 90-120 s. In this case we use the second 111 in this period, which corresponds to TOW = 396052 s.

Next, we further analyze the behavior of the metrics in each time instant, showing the corresponding sky plot.

CASE 1: TOW = 395961 (P1)

In this case we are able to track all satellites except SVN 18 and 26. Moreover, except SVN 24 the rest of satellites present some kind of multipath. Figure 7.20 shows the sky plot at the time instant TOW = 395961 s, and the following can be concluded for the two considered satellites:

- **SVN 15**: This satellite has an azimuth angle with some building between the satellites and receiver. However, in this case the elevation of the satellites is large enough to make possible the signal to exceed the buildings, but with some reflected ray. This is coherent with the results in Figure 7.19, which shows the presence of some multipath condition for SVN 15 but nominal behavior in general.

- **SVN 24**: As depicted in Figure 7.20, this is the satellite with the largest altitude (i.e. 78°) and without obstacles in the LOS, and this is the reason why an ideal behavior is

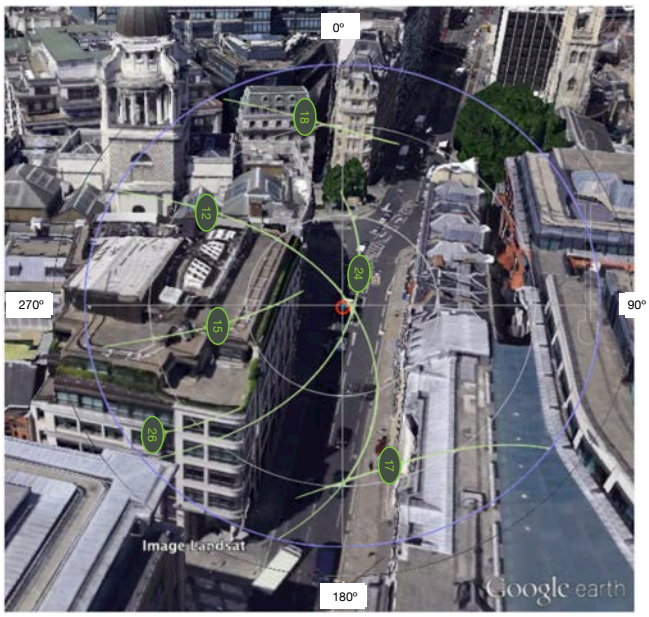

Figure 7.20: P1: Sky plot with the distribution of the satellites in view at TOW = 395961 s.

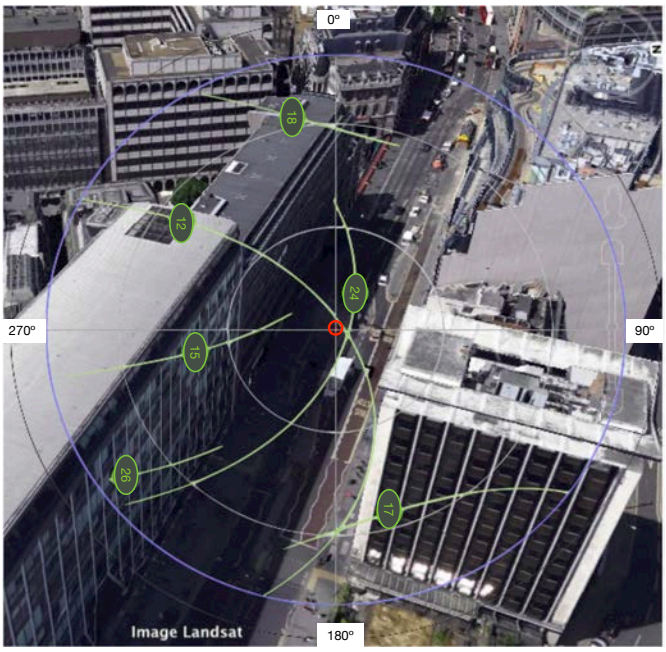

Figure 7.21: P2: Sky plot with the distribution of the satellites in view at TOW = 396011.

observed in its SLMs.

CASE 2: TOW = 396011 (P2)

We have already mentioned that in this case we are only able to track satellites SVN 17 and 24. Figure 7.21 shows the sky plot at the time instant corresponding to TOW = 396011 s. With the satellite and obstacles distribution displayed in the figure we can state the following:

- **SVN 24**: Although we obtain nominal behaviors for the DLLout and SAM, we observe small variations in the C/N_0. These variations, despite being low, suggest the presence of multipath. Checking the sky plot, and taking into account the large elevation of SVN 24, such multipath may be due to the quite narrowness of the street. Moreover, the street

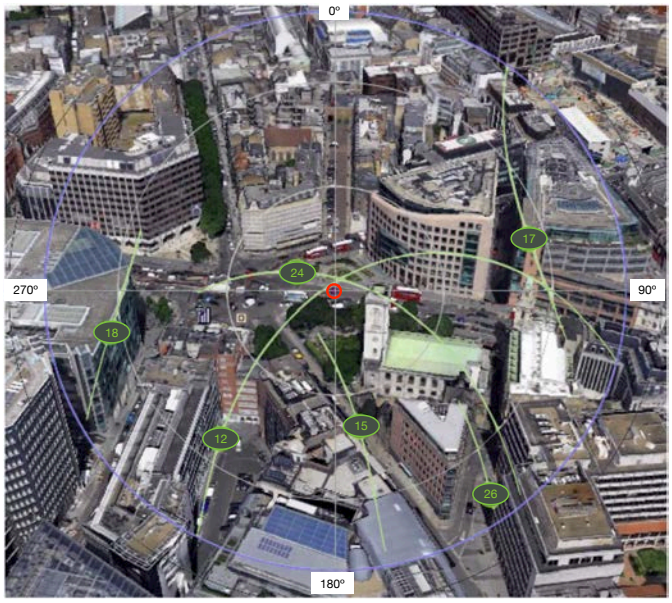

Figure 7.22: P3: Sky plot with the distribution of the satellites in view at TOW = 396052 s.

presents relatively tall buildings on both sides, which makes the street to seem like a corridor or deep urban canyon. This kind of situation usually produces several reflections with short delays. This is why the effects are only visible on the C/N_0, which is relevant for short multipath delay. These effects are not strong because of the big elevation of the satellite, which produces a strong LOS component, fixing a large SMR.

- **SVN 15**: It is not being tracked at this time instant since it is completely blocked by the building on the left hand side of the street. This is coherent because the building in this case is taller than in the previous time instant, and then it completely blocks the signal.

CASE 3: TOW = 396052 (P3)

In this case, as we have already commented, we are able to track all the satellites in view except SVN 17, which loses track at this time instant. Moreover, the tracked satellites exhibit nominal conditions in the sense they take a mean C/N_0 value around 40-44 dB-Hz, without oscillations not even spikes on the DLLout and SAM. With the sky plot displayed in Figure 7.22 we can see how in general it shows a clear environment in the sense that there are not a lot of critical obstacles (e.g. tall buildings, narrow streets, foliage,...) in the surroundings of the receiver. This environment leads to the following discussion:

- **SVN 15 and 24**: We can see how these satellites have no obstacle in the line of sight with the satellites. Moreover, they have a high enough elevation to avoid any reflection of some building of the surroundings and then receive the signals under nominal conditions. Maybe there is some reflection, but as far from the receiver that it arrives with negligible amplitude.

7.4.3 Summary of the selected SLM

Based on the results presented above and those obtained in the iGNSSrx project after analyzing all the data collection campaign, we can summarize the main features of the proposed SLMs.

SLM	C/N_0	DLLout	SAM
Detection type	MC	VC	MVC
Target delay	Low	High	
Comment	Nominal conditions estimation	Spikes presence instead of magnitude change	It can take nominal values even with multipath (oscillations
Limits	$\Delta\mu = \pm 6$ dB $\Delta_{\max} = \pm 4$ dB	$\Delta_{\max} = \pm 0.03$ chips $\Delta_{\min} = \pm 0.05$ chips	$\mu_1^{\pm} = 0.3/-0.1$ $\Delta_{\max} = \pm 0.3$ $\Delta_{\min} = \pm 0.5$
KL divergence	$d_1 = \frac{(\mu_1-\mu_0)^2}{2\sigma_0^2}$	$d_1 = \ln\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{\sigma_1^2}{2\sigma_0^2} - 0.5$	$d_1 = \ln\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{\left(\sigma_1^2-\sigma_0^2\right)\cdot(\mu_1-\mu_0)^2}{2\sigma_0^2\sigma_1^2} - 0.5$

Table 7.2: Properties of the suggested multipath detection algorithms.

A summary for the multipath SLMs is given in Table 7.2, in which they are classified in two groups, namely Low and High multipath delay, depending on the relevance of the analysis for the two delays cases. The C/N_0 corresponds to the low delay and the DLLout and SAM to the high delay. These two groups of techniques are complementary, and in practice they will have to be used jointly to ensure that we truly cover any possible multipath case. As we did for the interference case, Table 7.2 indicates the different type of changes experienced by the SLMs (i.e. MC, VC and MVC for the C/N_0, DLLout and SAM, respectively). The parameters that have been set before and after change are also indicated. These values have been chosen according to the experimental analyses of the proposed techniques with the data collection campaign. This is done in order to represent as accurately as possible the conditions before and after change (i.e. ideal and harsh conditions) that may be found in a representative scenario. Those values determine the performance in terms of detection delay, which also depends on the fixed false alarm rate. For the latter, and since we are using the CUSUM algorithm, we have the logarithmic relationship with the threshold (i.e. $h = \ln(N_{\text{fa}})$), being N_{fa} the metric samples between false alarms.

On the other hand, for the detection delay we have the known expression $\mathfrak{D} \leq h/d_1$, with d_1 the KL divergence given in the table for every metric. We can say that in general the quickest detection is obtained with the SAM metric, since it uses a MVC CUSUM, so that the KL divergence is greater than for the two other cases. Indeed, we see that this divergence for the SAM metric is equal to the sum of the one for the C/N_0 and DLLout metrics. For the other two cases (i.e. C/N_0 and DLLout) the delay will depend on the change magnitude. However, the changes on the mean of the C/N_0 are usually larger than the changes on the variance of the DLLout, therefore the KL divergence for the former case will usually be greater and then give a quicker detection than by using the DLLout metric. This will be true for the cases of low delay, where the C/N_0 is relevant. Nevertheless, if the multipath has a large delay, the effects on the C/N_0 will be lower than for the DLLout case, and then the DLLout will produce the quickest detection (with respect the C/N_0). Hence, making use of different analyses simultaneously we maximize the insight that can be obtained during a fault event. In essence, as we have seen, some interference and multipath SLM may fail in some circumstances, but it is unlikely that all of them will fail at the same time.

7.5 Integrity Analysis

This section evaluates the effects of the proposed PVT+I^2 into the navigation performance. In particular, we analyze the integration of the DSP module with the GNSS navigation and PVT integrity algorithm. The approach taken herein works offline and it is executed sequentially:

1. The raw GNSS data captured in the data campaigns is processed with the DSP anomaly detector. This results in an output file with validity flags corresponding to the signal integrity on the processed epochs.

2. The PVT navigation and integrity algorithm (PVT+I) is executed in order to generate the validity flags at PVT level. Optionally, the PVT+I^2 module can use the SLI flags file in step 1 as an input in order to exclude those measurement declared as faulty from the computation of the navigation solution. We will consider the first option as the benchmark for comparison and it will be referred to as nominal solution.

In order to assess the performance of this methodology, a test including the urban scenario studied in 7.2 is presented. At this point, it is important to say that we use here an offline approach in order to evaluate the PVT+I^2 performance. However, in a practical implementation, the output flags from the DSP module will feed the navigation algorithm online (i.e. real-time). So, the analysis performed in this section has illustrative purposes, and it is aimed at (i) evaluating the impact of the measurement rejection in the navigation performance; and (ii) at analyzing the real-time capability of the proposed techniques.

7.5.1 Navigation analysis

The first step to evaluate the performance of the DSP anomaly detector is to check whether the measurements flagged as faulty at this stage do correspond to potential outliers, while at the same time, that healthy measurements are consequently flagged as valid. The first topic is covered with the concept of PMD, while the second topic uses the PFA. The threshold value used to determine whether a pseudo-range error can be considered to be healthy or to be an outlier is 20 m (in absolute value). Last but not least, we will evaluate the impact of the whole FDE into the navigation performance, using the GPS+GLONASS constellation.

CASE 1: Original configuration
In the light of the obtained results in [ER15d] in terms of PMD and PFA, we can conclude that the DSP anomaly detector FDE is capable of detecting most of the outliers (around 90%), which should imply an improvement of the accuracy in the navigation solution. On the other hand, the obtained PFA is quite high, and measurements whose error is actually low are discarded (around 39% and 50% for the permissive and restrictive configurations, respectively). Fortunately, the fact that we use the dual GPS+GLONASS constellation, which increases the measurements availability, is expected to improve the accuracy of the position solution. It is worth mentioning that these probabilities were measured with fixing a threshold of 20 m for declaring an outlier. So, we should visualize this high value as an indicator that our detectors are detecting pseudorange errors lower than 20 m, which is a good thing, rather to think that the detectors are falsely detecting nominal measurements. The improvements on performance is validated in Figure 7.23, which represents the Horizontal Positioning Error (HPE) performance of the PVT+I^2

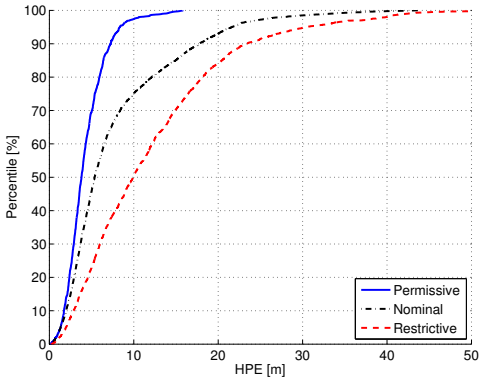

Figure 7.23: HPE cumulative distribution for the permissive and restrictive flags. Standard configuration with a sampling time for the metrics of 20 ms and the following parameters: $\mu_0 = \{10^{4.5}, 0, 0.1\}$, $\mu_1 = \{10^{2.9}, 0.4\}$ for the C/N$_0$ and SAM, respectively, $\sigma_0^2 = \{5.77 \cdot 10^7, 1.78 \cdot 10^{-4}, 1.78 \cdot 10^{-2}\}$ for the C/N$_0$, DLLout and SAM, respectively, and $\sigma_1^2 = \{5.5 \cdot 10^{-4}, 4 \cdot 10^{-2}\}$ for the DLLout and SAM, respectively.

algorithm, using both signal and PVT integrity, with each set of flags introduced in Section 6.4 (i.e. permissive and restrictive). This is compared to the nominal solution, in which only the RAIM is implemented to provide PVT integrity flags. These results were obtained by using the so-called standard configuration, which is provided after analyzing all the data in the urban data collection campaign of the iGNSSrx project, and it was aimed at detecting pseudorange errors above 20 m. We see in Figure 7.23 how the permissive configuration improves the performance of the navigation solution quite significantly (around 10 m at the 90-th percentile or 30 m in the 100-th percentile). Notwithstanding, the restrictive configuration leads to a worse performance than the nominal one. This is due to the high PFA in the restrictive configuration, which produces a large amount of discarded measurements. Thereby, the number of satellites to compute the navigation solution is low, thus reducing accuracy. For the permissive configuration, the PFA is also high but it is low enough to maintain sufficient measurements in order to improve the accuracy.

CASE 2: Modified configuration

We considered an outlier those measurements that present a pseudorange error greater than 20 m. However, the used configuration for the CUSUM parameters might not exactly correspond to an error of 20 m. This is due to the non-straightforward relationship between the magnitude of the SLMs variations, and the error they incur at positioning level. A first attempt to improve the performance in terms of PFA of the FDE is to modify the configuration of the CUSUM parameters in a more permissive way (i.e. declaring fault measurements those errors close to 20 m). Doing so, we consider two different configurations:

1. Maintaining the previous configuration (i.e. LowMD), but using a sampling time of 1 s (then the CUSUM parameters vary); and

2. modifying the CUSUM parameters in order to obtain lower PFA (i.e. LowFAR), using also a sampling time of 1 s

For the details on the specific values used for the CUSUM parameters in both configurations see [ER15d]. The results in terms of probabilities of these configurations are as follows: the LowMD results are similar to the obtained with the previous configuration, whereas for the LowFAR we obtain better results in terms of PFA (17%) but at the expense of increasing the

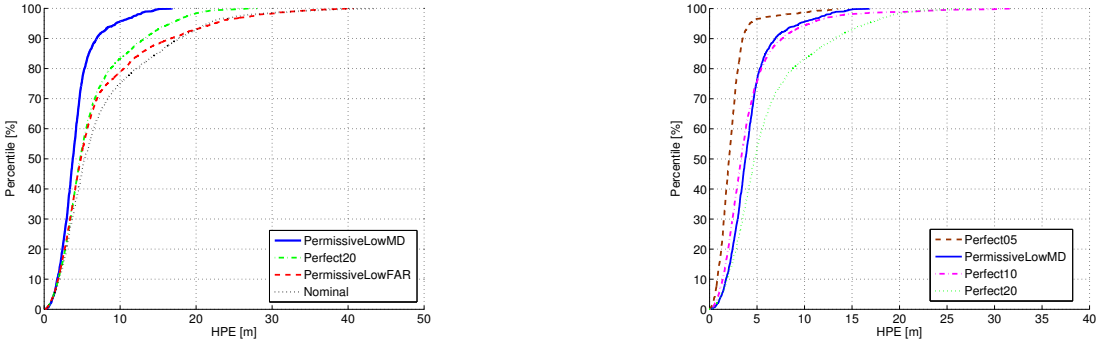

Figure 7.24: (Left) HPE cumulative distribution for the modified configurations; and (right) including different perfect flags.

Multipath Metric	SAM	C/N_0	DLLout
Processing time for a data set with 10 s observation time	0.1652 s	77.5 ms	65 ms

Table 7.3: Analysis of the real-time processing capability of the considered multipath detectors in this dissertation.

PMD (28%). The results in terms of HPE are presented in the left plot of Figure 7.24, in which we can notice that the LowMD configuration is the one providing the best performance. Hence, we can conclude that it is better to detect as many threats as possible (i.e. low PMD) instead of having a few false alarms at the expense of increasing the missed detections. This is in line with the expected results, considering the dual GPS+GLONASS constellation, which increases the measurement availability, and then the false alarms impact on the accuracy is not as critical as not detecting the actual faults.

Furthermore, we also include in the left plot of Figure 7.24 the curve corresponding to the perfect flags for measurement errors of 20 m. The idea of perfect flag is that we activate the flag in those cases where the actual pseudorange error is greater or equal to the desired threshold for the error to be detected, in this case 20 m. We see how the LowFAR curve fits the perfect curve until almost the 70th percentile, and then it is always below the perfect curve and approaching the nominal one. This is because the 70% of the time the LowFAR configuration is linked with a measurement error of 20 m, but the rest of the time it is linked with a higher error. On the other hand, we see how the LowMD configuration outperforms the perfect curve. This is due to the fact that this configuration is linked with a measurement error lower than 20 m, and then it improves the accuracy, with respect to the perfect flag, since we are discarding measurements with lower errors. This is reflected in the right plot of Figure 7.24, which shows how the LowMD configuration is close to the perfect10 (i.e. measurement error threshold of 10 m) curve and it is always below the perfect5 (i.e. 5 m error). Therefore, we can conclude that the LowMD configuration is linked with a measurement error threshold between 5 m and 10 m (closer to 10 m than 5 m), as we expected when configuring the algorithms. Therefore, the real probabilities will vary with respect to those presented previously.

7.5.2 Real-Time processing capability

Finally, we show some results on the time that the proposed algorithms take to process 10 s of signal. The results are obtained using a MacBook Pro computer with an Intel core i7 processor

@ 2.2 GHz, using 1 core. Results are presented in Table 7.3, which shows the time needed to process the data of only one satellite. Therefore, the final results will depend on the number of satellites in view, however even in the hypothetical case we have all the GPS constellation in view (i.e. 24 satellites), taking into account the times in the table, real-time processing would be possible. This is because here we only present the time needed to implement the CUSUM algorithm to the obtained metrics. These metrics are supposed to be already available within the GNSS receiver, and based on real-time calculations. For the SAM metric we need a multi-correlator receiver, so if the multi-correlation is real-time, the SAM metric calculation will be so, as Table 7.3 shows. We see that the times for the C/N_0 and DLLout metrics are quite lower than for the SAM, the reason for this is that for the SAM we need to apply bit synchronization in order to properly average the correlation curves in order to calculate the SAM. Moreover, there is the process itself of calculating the SAM, whereas for the C/N_0 and the DLLout, we only have to apply the CUSUM algorithm because these metrics are already available within the GNSS receiver. Finally, we see how the application of the CUSUM, which is reflected with the C/N_0 and DLLout metrics, takes around 70 ms. This short time makes clear the feasibility of using the CUSUM algorithm into real-time integrity monitoring applications. Note that the CUSUM-C/N_0takes a slightly greater time than the one for the DLLout. This is because for the C/N_0 we have to sequentially estimate the C/N_0 mean under ideal conditions, whereas for the DLLout this is not needed.

7.6 Conclusions

This chapter has been dedicated to provide realistic results of the proposed SLMs. This has been intended to show the capability of the proposed sig-RAIM algorithm to operate in real-working conditions. This is obtained by processing GNSS signal data gathered in the framework of the iGNSSrx project. This data has been used to show both the behavior of the SLM under the absence and presence of local threats (i.e. interference and multipath) and the improvement of the accuracy and integrity by using the proposed PVT+I^2, with respect to current methods (i.e. RAIM). After processing the data we can confirm the behavior of the SLMs, previously investigated in Chapter 5. Moreover, we have provided a summary of the main properties of the considered threat detectors in terms of their suitability to certain scenarios, the type of algorithm they implement, and their main configuration parameters. This lead to the consideration of the histogram, kurtosis and power metrics as the proper set of SLM able to detect any kind of interference in GNSS. Regarding the multipath SLM, the C/N_0 is considered for low multipath delay, whereas the DLLout and SAM are considered for high multipath delay. Finally, the integration of the sig-RAIM and PVT+I algorithms (i.e. PVT+I^2 module) have been analyzed. Particularly, we have analyzed the impact of the PVT+I^2 FDE algorithm into the navigation solution. The analysis with the gathered data in the iGNSSrx project has shown that the proposed configuration for the algorithms give a good performance in terms of accuracy (i.e. provides better navigation results than the nominal configuration with only the PVT+I algorithm) and integrity (i.e. rises warnings related with the SLI that are not raised with the PVT+I implementing traditional RAIM). However, further work is needed in order to further improve the results and to further link the presented sig-RAIM with current integrity algorithms:

- Regarding the tuning of the CUSUM parameters, in order to obtain a relation between

faults in the SLM and pseudo-range error (e.g. configuration of parameters to declare a fault an error greater than 20 m), and then improve the results in terms of false alarms, integrity risk and availability.

- Improvements on the availability of the satellites may be beneficial for large HPE. Two possible alternatives are:
  - To fix the minimum number of satellites to use in the navigation algorithm.
  - To weight the satellite measurements instead of discarding them.
- The presented results adopted a QCD, however as shown in this thesis the TCD is more convenient for the integrity problem. So, the implementation of the sig-RAIM with the framework of TCD should be considered, expecting an improvement in terms of integrity risk and availability.

Chapter 8

Conclusions and Future Work

This thesis has contributed to the topic of GNSS integrity monitoring by exploiting the unparalleled performance of statistical change detectors. Figure 8.1 illustrates the roadmap of the present dissertation, in which integrity and local threat detection are the starting points of the subsequent discussions. The topics covered in this thesis and fields in which contributions have been provided are highlighted for the sake of clarity. Let us first consider the topic of detection theory. This field has been classified into traditional hypothesis testing and SCD. This is a very important topic because it allows us to identify the SCD framework as the appropriate one for integrity monitoring purposes; that is, the one targeting a timely detection. Thereby, the goal of Chapter 4 has been to provide a comprehensive overview of detection theory, with particular interest in SCD. The main contribution has been the provision of a novel solution and performance bounds for TCD problems. Moreover, for the first time, we have compared both classical hypothesis testing and SCD within the same framework, showing that TCD is better for integrity purposes than any other detection scheme. This has constituted the first part of this dissertation.

Once the theoretical analysis of detection theory has been addressed, the next step has been to move forward into the design of local threat detectors. This was done with the aim of providing the additional level of integrity, namely SLI, needed for the safely use of GNSS-based critical applications. Two basic requirements are needed to provide SLI. The first one is to use signal observable metrics capable of detecting any misleading behavior of the received GNSS signal. The second requirement is to detect these threats as soon as possible so that the user can be promptly alerted. Chapter 5 has shown how to address these requirements by providing a complete set of SLM able to detect any type of interference and multipath propagation. In addition, the adoption of QCD has been considered in order to promptly detect any local threat. This has been done with the aim of providing a novel framework for signal quality monitoring, so that integrity and multipath can be promptly detected. Notwithstanding, for integrity purposes, the TCD criterion should be adopted instead. In particular, the suggested FMA stopping time and the provided novel bounds have been shown to be superior than any other available solution for TCD, and then for integrity. It is for this reason that Chapter 6 has dealt with the practical implementation of RAIM algorithms based on TCD and its integration with local threat detection, leading to the proposed sig-RAIM algorithm. Finally, Chapter 7 has validated the proposed algorithms in this dissertation with real GNSS signals captured within the framework of the iGNSSrx project.

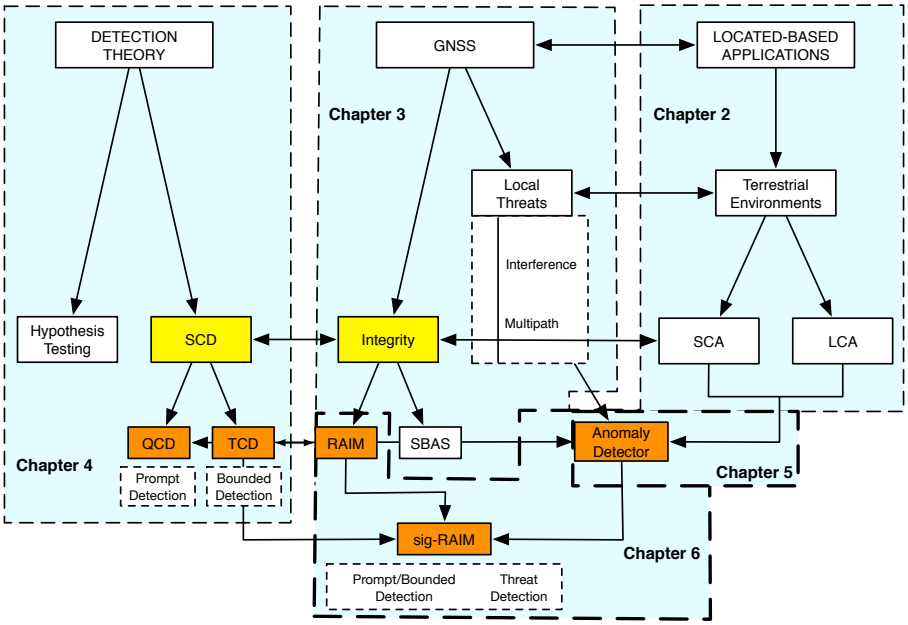

Figure 8.1: Schematic overview of the topics covered within the present dissertation.

8.1 Conclusions

Next, we summarize the contributions and conclusions for each of the topics covered in this thesis.

Detection theory

The mathematical theory of sequential detection has been investigated in Chapter 4. The main goal has been to provide a mathematical framework for the design of optimal sequential detectors that would later on be applied to integrity algorithms. After introducing the background of sequential detection based on Wald's work, we have brought together disparate results of the theory of QCD so that a formal and unified treatment is provided for a general reader. This part of the thesis has been intended to introduce and stimulate the use of QCD within the GNSS community. The main property of QCD problems is that they deal with a change in distribution of infinite duration. In contrast, there is the branch of TCD that deals with a change of finite duration. QCD has extensively been studied in the last decades, providing a very attractive optimal solution, namely the CUSUM algorithm. Unfortunately, in TCD a bounded delay is desirable, for which the CUSUM is not that attractive. This has motivated us to investigate the optimal properties of TCD problems. This has been intended to gain insight on the general problem of TCD, so that an optimal solution can be sought to be applied to integrity techniques. In this line, we have provided relevant contributions to find the optimal solution, leading us to the proposal of the FMA solution, given by (4.90), to approach the problem of TCD. This solution is of potential interest in GNSS integrity, and this is why we have investigated the statistical properties of this solution. This has led us to the provision of very tight novel bounds, stated in Theorem 8. These bounds are important to evaluate the statistical performance of the FMA solution, thus allowing the design of integrity algorithms. Numerical evaluations have shown that the FMA solution is the best solution so far proposed for the TCD problem, which is the most important contribution of this chapter.

Finally, another important contribution has been the comparison of both classical and sta-

tistical change detection, within the same framework. This is an important result because these fields have often been considered to be disparate, when actually we have shown that they are closely linked. Indeed, this connection has been found in the block sample size of the traditional LRT detector, used in classical detection. This block sample size has always been left as a degree of freedom to be fixed by the user. Nevertheless, this size might be fixed targeting one of the optimal criteria in SCD. This approach has led to the sample sizes in (4.122) and (4.138) for QCD and TCD, respectively. Once the optimal size for the LRT have been found, the resulting stopping time has been compared with the CUSUM and FMA solutions. The theoretical findings in theorems 9 and 10 have respectively led to corollaries 3 and 4, stating that the LRT loses its optimality when dealing with QCD and TCD, respectively. These findings are completely general and its novelty has allowed for the first time to fill the gap between the general theory of classical and sequential change detection.

Signal-Level techniques

Integrity algorithms were initially designed for civil aviation applications. So far, they have assumed that local effects have a controlled influence on the received GNSS signal. However, this is not the case in terrestrial environments, where analyses on the received signal should be considered. We have referred to this kind of integrity as SLI. Actually, as illustrated in Figure 5.1, in order to improve current integrity, two requirements have to be considered; that is, (i) the use of SLM to monitor the presence of any local threat, and (ii) the use of SCD to detect these threats as soon as they appear. The provision of these two requirements have been investigated in Chapter 5. In this dissertation, the first requirement has been approached by providing different observable SLMs that react under the presence of either interference or multipath. In particular, pre- and post-correlation metrics have been proposed to monitor interference and multipath, respectively. The provided set of SLMs has been considered to detect any possible interference source and multipath propagation in GNSS. Once the SLMs have been defined and its behavior characterized, the problem of interference and multipath detection, based on these metrics, have been considered. A QCD framework has been adopted, which is the main contribution of Chapter 5.

Following the design flow of optimal detectors, we have firstly shown the statistical characterization of the SLMs, followed by the ultimate design of the corresponding CUSUM algorithm. We can conclude that most of the detectors can be designed as Gaussian mean, variance, or mean and variance changes. Some metrics have a completely unknown statistical characterization, but they can be fairly fed to any of the above mentioned Gaussian changes. The only exception that has not been fed into the CUSUM is the histogram SLM. This is because its distribution under nominal conditions is known, but the one under harsh conditions is not. Thus, allowing the use of the O-CUSUM to implement the histogram-based detection. Doing so, we gain robustness with respect to the case of feeding the histogram metric into any of the considered Gaussian changes. Finally, we have also considered the problem of LOS and NLOS discrimination and multi-antenna threat detection. For the former we have evaluated the behavior of the SAM under both LOS and NLOS multipath conditions. From the obtained results we have seen that a mean change is prevalent under LOS, whereas a variance change is prevalent under NLOS conditions. Both the LOS and NLOS discrimination and multi-antenna threat detection are interesting contributions because they work at the signal processing level, without using additional aid like maps or sensors, and they adopt a SCD framework easily implementable in mass-market GNSS receivers. It is true that multi-antenna receivers are not prevalent nowa-

days, but they are expected to be extensively widespread in the coming years. Simulation results have been presented validating the capabilities of the proposed metrics to detect interference and multipath.

Integrity algorithm implementation

The adoption of a new generation of integrity techniques is needed to the further development and safely adoption of SCA and LCA, which are currently emerging. This is due to the predominance of local threats in terrestrial environments, where critical applications are deployed. In this situation there is no choice but to resort on sequential change threat detection. Notwithstanding, in order to provide the signal-level integrity needed in these applications, these detectors have to be integrated with the framework of RAIM algorithms. This challenging task has been carried out in Chapter 6. RAIM schemes have traditionally been designed as heuristic methods based on the SSE metric. Despite they work quite well for aviation operations, reality is different when turning into terrestrial environments. Theoretical results in this dissertation have shown that a proper mathematical framework is available to optimally design integrity algorithms. So, we have at our disposal the tools to switch the design of integrity algorithms from heuristic- to optimally-based, thus moving a step ahead into the modernization of integrity techniques. Contrary to these traditional schemes, the framework of TCD has been considered in this thesis to approach the problem of integrity in GNSS. Simulations have shown the superior performance of the FMA solution and TCD framework with respect to other available solutions. This is one of the most important contribution of this dissertation, showing the appropriateness of the novel TCD framework to implement RAIM algorithms in an attractive mathematical way.

The main operational functioning and implementation of traditional RAIM have been illustrated in Figure 6.1. This has been very important to recall the role of RAIM techniques in GNSS in order to further link them with TCD and threat detection. This has relieved the second requirement, indicated in Figure 5.1, targeted in this thesis to provide SLI. The first requirement, namely the provision of signal quality monitoring, have been approached by linking RAIM algorithms with threat detectors in a TCD framework. This link has defined what we have called the sig-RAIM algorithm, illustrated in Figure 6.4. For the use of this algorithm, the application of the theoretical results in Theorem 8 and Corollary 1 to the considered SLMs is needed. Doing so, we have provided the resulting bounds for the multipath SLMs, stated in corollaries 5–7. Numerical results have been provided to confirm the superiority of our proposed solution (i.e. FMA stopping time) with respect to other solutions available in the literature of TCD. Furthermore, we have also proposed the use of two parallel FMA stopping times to approach the discrimination between LOS and NLOS multipath conditions. In summary, the key point of sig-RAIM is that it can compute SLMs, without needing external information either coming from additional hardware, neither from redundant measurements from different satellites. This is beneficial for the availability of RAIM algorithms, which is the main drawback of traditional RAIM in terrestrial environments. This increment in the availability is also driven by the use of the FMA stopping time, which has been shown to give the best results in terms of integrity risk, leading to an improvement of the availability of RAIM.

Experimental results

The proposed sig-RAIM has been validated with real GNSS signal gathered within the framework of the iGNSSrx project funded by the European Commission. To this end, a software anomaly detector has been developed to process the data captured in the two data collection campaigns

carried out in the project. These collection campaigns have served to validate the behavior of the proposed SLMs and their capability to detect any kind of interference and multipath behavior in realistic situations. This is an important practical issue, since it has allowed us to provide a summary of the suitability to certain type of threats and their main configuration parameters to to be used in real-working conditions. This is given in Table 7.1 and Table 7.2, for interference and multipath SLM, respectively. As a summary the histogram, kurtosis and power metrics have been considered as the proper set of SLMs able to detect any kind of interference in GNSS. In particular, the histogram metric is able to detect all of the possible interferences. Nevertheless, it does not uses an optimal detection approach. It is for this reason that the kurtosis is considered to detect all metrics except wide-band, for which the power metric is devoted. Then, the kurtosis- and power-based detection are the optimal ones (in the sense they use the CUSUM), and the histogram SLM has been proposed as backup in those cases in which the others may fail. Regarding the multipath SLMs, the C/N_0 is considered for low multipath delay, whereas the DLLout and SAM are considered for high multipath delay.

Finally, one of the main contribution of this dissertation, the so-called PVT+I^2 module, has been evaluated with real GNSS signals. After the analysis of the impact of the PVT+I^2 module into the navigation solution, we have validated the proposed configuration, giving a good performance in terms of accuracy and integrity. Actually, results in figures 7.23 and 7.24 have shown that the use of sig-RAIM provides better navigation performance than the traditional solution, which only implements RAIM. Results have shown an improvement of around 10 m at the 90-th percentile (i.e. RAIM gives around 20 m of error, whereas sig-RAIM around 10 m) or as large as 30 m in the 100-th percentile. Moreover, the use of sig-RAIM raises warnings at signal level (e.g. when multipath is present) that are not raised with traditional RAIM. So, even though further work is needed to completely link SLI and RAIM algorithms, we have proved that the concept of SLI is very attractive to improve the current integrity practical implementation. In addition, the proposed implementation has been proved to fulfill real-time processing conditions, thus easily implementable in mass-market receivers. Last but not least, it is worth pointing out that the proposed implementation has departed from traditional approaches. Heuristic approaches based on consistency checks have been traditional sought, whereas our approach has followed a formal and novel mathematical framework very appropriate for the problem at hand.

8.2 Future Work

Next, we summarize the future research lines for each of the topics covered in this thesis. In Chapter 4 the following theoretical extension could be addressed:

- Contrary to QCD, the field of TCD is still immature with scarce literature about optimal aspects. So, the problem is still open and a direct future line of research is to seek the optimal solution to the general problem of TCD. Among this line we propose two alternatives:
  - To formulate the problem in a way that can be solved by means of the optimal stopping theory. This has already targeted in this dissertation, but without success. The issue with the proposed formulation in (4.72) is that it is not a Markovian problem, and it leads to a mathematical intractable problem. However, the problem

indeed takes an m-sample Markovian representation. So, at this point, one path may be to extend the results of Markov optimal stopping theory for an m-sample Markov process. Doing so a treatable solution for (4.72) can be sought. Another option among this alternative may be to re-formulate the problem in (4.70) in a treatable way.

- Unfortunately, the previous alternative may be quite complicated because a lot of experience is needed to do so. As an example, it is worth to say that it took more than 30 years since the CUSUM was first proposed by Page in the 1950s until Moustakides proved its optimality in the 1980s using an optimal stopping theory approach. For this reason an asymptotic optimal proof may be more feasible. The idea would be to prove that the FMA solution is asymptotically optimal in the TCD sense. To do so, the insights gained in Section 4.5 may be useful to prove that the provided performance bounds in this dissertation for the FMA stopping time are the best one can do for the TCD problem.

In Chapter 5 the following research lines can be considered:

- Alternative SLM can be investigated, as for instance:
  - Additional simple metrics may be considered, as for example the use of metrics based on the FLL/PLL.
  - The use of dual-polarized antennas may be considered for multipath detection and mitigation.
  - Spoofing SLMs should also be considered.

Finally, further work in line of Chapter 6 may be:

- The relation between the change introduced in the SLM with the resultant error in pseudorange must be investigated to completely link RAIM algorithms with the proposed TCD framework. That is to say, to propose a configuration of the detection algorithms in order to declare a fault those faults producing an error greater than the given AL. To this end, the relation defined in (6.22) must be found. Doing so, the relation between the AL and the minimum detectable change parameter used to define the LLR can be calculated, thus completely filling the gap between the RAIM and sig-RAIM implementation.
- Alternative fusion rules to generate the SLI flags can be investigated. It would be interesting to investigate fusion rules based on distributed detection; that is, to consider the information coming from different SLMs as they were information coming from different decentralized sensors. Then, optimal ways to combine the information coming from different SLMs can be sought.
- Improvements on the availability of the satellites may be beneficial to improve accuracy. Two possible alternatives are:
  - To fix the minimum number of satellites to use in the navigation algorithm.
  - To weight the satellite measurements instead of discarding them.

Bibliography

[Ami16] Moeness G. Amin, Pau Closas, Ali Broumandan, John L. Volakis, "Vulnerabilities, threats, and authentication in satellite-based navigation systems [scanning the issue]", *Proceedings of the IEEE*, Vol. 104, nº 6, pags. 1169–1173, 2016.

[ARS06] ARS Traffic & Transport Technology, "Accuracy and reliability of distance and position measurements by GNSS receivers", Tech. rep., 2006.

[AS09] Miguel Azaola-Saenz, Joaquín Cosmen-Schortmann, "Autonomous integrity: An error isotropy-based approach for multiple fault conditions", *Inside GNSS*, Vol. 1, nº JAN/FEB, pags. 28–36, 2009.

[Axe12] Erik Axell, Geert Leus, Erik Larsson, H. Vincent Poor, "Spectrum sensing for cognitive radio : State-of-the-Art and recent advances", *IEEE Signal Processing Magazine*, Vol. 29, nº 3, pags. 101–116, 2012.

[Bak00] Bacem Bakhache, Igor Nikiforov, "Reliable detection of faults in measurement systems", *International Journal of Adaptive Control and Signal Processing*, Vol. 14, nº 7, pags. 683–700, 2000.

[Bal09] Asghar Tabatabaei Balaei, Andrew G. Dempster, "A statistical inference technique for GPS interference detection", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 45, nº 4, pags. 1499–1511, 2009.

[Ban85] Stephen Bancroft, "An algebraic solution of the GPS equations", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-21, nº 1, pags. 56–59, 1985.

[Bas93] Michele Basseville, Igor Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice Hall Englewood Cliffs, 1993.

[BBC13] BBC, "What will be driving in 2050?", *BBC News*, Retrieved from : `http://www.bbc.com/future/story/20131108-what-will-we-be-driving-in-2050`, Accesed: 18-6-2017, 2013.

[Bee05] T.W. Beech, Miguel ngel Martinez-Olague, Joaquin Cosmen-Schortmann, "Integrity: A key enabler for liability critical applications", *Proceedings of the 61st Annual Meeting of The Institute of Navigation (ION)*, pags. 1–10, 2005.

[Ber07] Pietro Berkes, Laurenz Wiskott, "Analysis and interpretation of quadratic models of receptive fields", *Nature protocols*, Vol. 2, nº 2, pags. 400–407, 2007.

[Bha09] Umar Iqbal Bhatti, Washington Y. Ochieng, "Detecting multiple failures in GPS/INS integrated system: A novel architecture for integrity monitoring", *Journal of Global Positioning Systems*, Vol. 8, nº 1, pags. 26–42, 2009.

[Bhu07] Mohammad Zahidul H. Bhuiyan, Elena Simona Lohan, Markku Renfors, "Code tracking algorithms for mitigating multipath effects in fading channels for satellite-based positioning", *EURASIP Journal on Advances in Signal Processing*, Vol. 2008, nº 1, pags. 1–17, 2007.

[Bis05] Richard Bishop, "Intelligent vehicle R&D: A review and contrast of programs worldwide and emerging trends", *Annales Des Télécommunications, Springer*, Vol. 60, nº 3, pags. 228–263, 2005.

[Bla74] Richard Blahut, "Hypothesis testing and information theory", *IEEE Transactions on Information Theory*, Vol. IT-20, nº 4, pags. 405–417, 1974.

[Bla12] Juan Blanch, Todd Walter, Per Enge, Young Lee, Boris Pervan, Markus Rippl, Alexandre Spletter, "Advanced RAIM user algorithm description: Integrity support message processing, fault detection, exclusion, and protection level calculation", *Proceedings of the Global Navigation Satellite Systems Conference of the Institute of Navigation (ION GNSS)*, pags. 2828–2849, 2012.

[BMW13] BMW, "Active Cruise Control with Stop & Go", *BMW webpage*, Retrieved from : `http://www.bmw.com/com/en/insights/technology/connecteddrive/2013/driver%7B_%7Dassistance/intelligent%7B_%7Ddriving.html`, Accesed: 18-6-2017, 2013.

[BN79] Ole Barndorff-Nielsen, David R. Cox, "Edgeworth and saddle-point approximations with statistical applications", *Journal of the Royal Statistical Society. Series B (Methodological)*, pags. 279–312, 1979.

[Boj79] Tomasz Bojdecki, "Probability maximizing approach to optimal stopping and its application to a disorder problem", *Stochastics*, Vol. 3, pags. 61–71, 1979.

[Bol68] D. W. Bolton, "The multinomial theorem", *The Mathematical Gazette*, pags. 336–342, 1968.

[Bra01] Michael S. Braasch, "Performance comparison of multipath mitigating receiver architectures", *Proceedings of the IEEE Aerospace Conference*, pags. 1309–1315, 2001.

[Bro86] Grover Brown, Patrick Y.C. Hwang, "GPS failure detection by autonomous means within the cockpit", *Proceedings of the 42nd Annual Meeting of the Institute of Navigation*, 1986.

[Bro97] Grover Brown, "Solution of the two-failure GPS RAIM problem under worst-case bias conditions: Parity space approach", *Navigation*, Vol. 44, nº 4, pags. 425–431, 1997.

[Bro01] Karel A. Brookhuis, Dick de Waard, Wiel H. Janssen, "Behavioural impacts of advanced driver assistance systemsan overview", *European Journal of Transport and Infrastructure Research*, Vol. 1, nº 3, pags. 245–253, 2001.

[Bro12] Ali Broumandan, Ali Jafarnia-Jahromi, Vahid Dehghanian, John Nielsen, Gérard Lachapelle, "GNSS spoofing detection in handheld receivers based on signal spatial correlation", *Proceedings of the IEEE Position Location and Navigation Symposium (PLANS)*, pags. 479–487, 2012.

[Bro16a] Danilo Bronzi, Yu Zou, Federica Villa, Simone Tisa, Alberto Tosi, Franco Zappa, "Automotive three-dimensional vision through a single-photon counting SPAD camera", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, nº 3, pags. 782–795, 2016.

[Bro16b] Ali Broumandan, Ali Jafarnia-Jahromi, Saeed Daneshmand, Gérard Lachapelle, "Overview of spatial processing approaches for GNSS structural interference detection and mitigation", *Proceedings of the IEEE*, Vol. 104, nº 6, pags. 1246–1257, 2016.

[Cal01] Vincent Calmettes, Frédéric Pradeilles, Michel Bousquet, "Study and comparison of interference mitigation techniques for GPS receiver", *Proceedings of the 14th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2001)*, pags. 957–968, 2001.

[Cho71] Yuan Shih Chow, Herbert Robbins, David Siegmund, *Great Expectations: The Theory of Optimal Stopping*, Houghton Mifflin Boston, 1971.

[Clo11] Pau Closas, Carles Fernandez-Prades, "A statistical multipath detector for antenna array based GNSS receivers", *IEEE Transactions on Wireless Communications*, Vol. 10, nº 3, pags. 916–929, 2011.

[CNN16] CNN, "Getting lost near the Kremlin? Russia could be GPS spoofing", *CNN tech*, Retrieved from : `http://money.cnn.com/2016/12/02/technology/kremlin-gps-signals/index.html`, Accesed: 18-6-2017, 2016.

[Coh89] Leon Cohen, "Time-Frequency Distributions-A Review", *Proceedings of the IEEE*, Vol. 77, nº 7, pags. 941–981, 1989.

[Col09] Toll Collect, "Truck toll in Germany. User information", Tech. rep., 2009.

[Cov06] Thomas M. Cover, Joy A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 2nd ed., 2006.

[Cra46] Harald Cramér, *Mathematical Methods of Statistics.*, Princeton University Press, 1946.

[CS08] Joaquin Cosmen-Schortmann, Miguel Azaola-Sáenz, Miguel Ángel Martínez-Olagüe, Manuel Toledo-López, "Integrity in urban and road environments and its use in liability critical applications", *Proceedings of the IEEE Position Location and Navigation Symposium (PLANS)*, pags. 972–983, 2008.

[D'A86] Ralph B. D'Agostino, Michael A. Stephens, *Goodness-of-Fit Techniques*, CRC Press, 1986.

[De 07] Roger D. De Roo, Sidharth Misra, Christopher S. Ruf, "Sensitivity of the kurtosis statistic as a detector of pulsed sinusoidal RFI", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 45, pags. 1938–1946, 2007.

[De 10] Roger D. De Roo, Sidharth Misra, "A moment ratio RFI detection algorithm that can detect pulsed sinusoids of any duty cycle", *IEEE Geoscience and Remote Sensing Letters*, Vol. 7, nº 3, pags. 606–610, 2010.

[Des04] Sameet Mangesh Deshpande, *Study of interference effects on GPS signal acquisition*, PhD Thesis, University of Calgary, 2004.

[Do15] Van Long Do, Lionel Fillatre, Igor Nikiforov, "Sequential detection of transient changes in stochastic-dynamical systems", *Journal de la Société Française de Statistique*, Vol. 156, nº 4, pags. 101–113, 2015.

[Ege14a] Daniel Egea, José A. López-Salcedo, Gonzalo Seco-Granados, "Interference and multipath sequential tests for signal integrity in multi-antenna GNSS receivers", *Proceedings of the IEEE Sensor Array and Multichanel Signal Processing Workshop*, pags. 117–120, 2014.

[Ege14b] Daniel Egea, Gonzalo Seco-Granados, José A. López-Salcedo, "Single- and multi-correlator sequential tests for signal integrity in multi-antenna GNSS receivers", *Proceedings of the International Conference on Localization and GNSS (ICL-GNSS)*, pags. 117–120, 2014.

[ER15a] Daniel Egea-Roca, Gonzalo Seco-Granados, José A . López-Salcedo, "Quickest detection framework for signal integrity monitoring in low-cost GNSS receivers", *Proceedings of the IEEE Vehicular Technology Conference (VTC Fall)*, pags. 1–5, 2015.

[ER15b] Daniel Egea-Roca, Gonzalo Seco-Granados, José A. López-Salcedo, "On the use of quickest detection theory for signal integrity monitoring in single-antenna GNSS receivers", *Proceedings of the International Conference on Localization and GNSS (ICL-GNSS)*, pags. 1–6, 2015.

[ER15c] Daniel Egea-Roca, Gonzalo Seco-Granados, José A. López-Salcedo, Enrique Domínguez, Enrique Aguado, David Lowe, Denis Naberezhnykh, Fabio Dovis, Ignacio Fernández-Hernández, Juan Pablo Boyero, "Signal-Level integrity and metrics based on the application of quickest detection theory to interference detection", *Proceedings of the 28th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+)*, pags. 3136–3147, 2015.

[ER15d] Daniel Egea-Roca, Gonzalo Seco-Granados, José A. López-Salcedo, Carlos Moriana, Maciej Jerzy Paśnikowski, Enrique Domínguez, Enrique Aguado, David Lowe, Denis Naberezhnykh, Fabio Dovis, Ignacio Fernández-Hernández, Juan Pablo Boyero, "Signal-Level integrity and metrics based on the application of quickest detection theory to multipath detection", *Proceedings of the 28th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+)*, pags. 3136–3147, 2015.

[ER16] Daniel Egea-Roca, Gonzalo Seco-Granados, José A . López-Salcedo, "Transient change detection for LOS and NLOS discrimination at GNSS signal processing level", *Proceedings of the International Conference on Localization and GNSS (ICL-GNSS)*, pags. 1–6, 2016.

[ER17a] Daniel Egea-Roca, Gonzalo Seco-Granados, José A . López-Salcedo, "Comprehensive overview of quickest detection theory and its application to GNSS threat detection", *Gyroscopy and Navigation*, Vol. 8, nº 1, pags. 1–14, 2017.

[ER17b] Daniel Egea-Roca, Gonzalo Seco-Granados, José A. López-Salcedo, "Inhomogeneous quadratic tests in transient signal detection: Closed-Form upper bounds and application in GNSS", *submitted to Digital Signal Processing, Elsevier (Accepted with changes).* arXiv preprint arXiv:1608.00615., pags. 1–11, 2017.

[ER17c] Daniel Egea-Roca, Gonzalo Seco-Granados, José A. López-Salcedo, "Sequential change detection for next-generation RAIM algorithms", *Procceedings of the International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+), Accepted*, 2017.

[ER17d] Daniel Egea-Roca, Gonzalo Seco-Granados, José A. López-Salcedo, H. Vincent Poor, "Performance bounds for finite moving average change detection: Application to global navigation satellite systems", *IEEE Transactions on Signal Processing (Accepted with major changes).* arXiv preprint arXiv:1706.05552., pags. 1–12, 2017.

[ER17e] Daniel Egea-Roca, Gonzalo Seco-Granados, José A. López-Salcedo, H. Vincent Poor, "When Statistical Change Detection Meets Classical Detection", *IEEE Transactions on Signal Processing (To be submitted)*, 2017.

[Esa67] James D. Esary, Frank Proschan, David W. Walkup, "Association of random variables, with applications", *The Annals of Mathematical Statistics*, Vol. 38, nº 5, pags. 1466–1474, 1967.

[ESA17] ESA, "ESA website", *ESA*, Retrieved from : `http://www.esa.int/ESA`, Accesed: 18-6-2017, 2017.

[Exp06] Expert Group 9, "Report of Expert Group 9 working to support the European Comission on the work on Directive 2004/52/EC: Specification of the EFC application based on satellite technologies", Tech. rep., EG9, 2006.

[FAA16] FAA, "Satellite Navigation-Ground Based Augmentation System (GBAS)", Tech. rep., FAA website, 2016.

[Fal11] Emanuela Falletti, Marco Pini, Letizia Lo Presti, "Low complexity carrier-to-noise ratio estimators for GNSS digital receivers", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 47, nº 1, pags. 420–437, 2011.

[Fer16] Nunzia Giorgia Ferrara, Maciej Jerzy Paśnikowski, Susana María Sánchez-Naranjo, Fabio González, Raul Ramos-Pollán, Gonzalo Seco-Granados, José A. López-Salcedo, Daniel Egea-Roca, Marc Solé, Manuel Toledo-López, Elena Simona Lohan, "Combined architecture enhancing multi-dimensional signal quality in GNSS receivers", *Inside GNSS Working Papers*, pags. 54–62, 2016.

[For08] Borje Forssell, *Radionavigation systems*, GNSS Technology and Applications, 2008.

[GC16] David Gómez-Casco, José A. López-Salcedo, Gonzalo Seco-Granados, "Generalized integration techniques for high-sensitivity GNSS receivers affected by oscillator phase noise", *Proceedings of Statistical Signal Processing Workshop (SSP)*, pags. 1–5, 2016.

[Gen15] Christian Genter, Robert Pöhlmann, Markus Ulmschneider, Thomas Jost, Armin Dammann, "Multipath assisted positioning for pedestrians", *Proceedings of the annual meeting of the Institute of Navigation (ION GNSS+)*, pags. 2079–2086, 2015.

[GPS17] GPS.gov, "GPS.gov website", *GPS*, Retrieved from: `http://www.gps.gov/systems/gps`, Accesed: 18-6-2017, 2017.

[Gro10] Paul D. Groves, Ziyi Jiang, Benjamin Skelton, Paul A. Cross, Lawrence Lau, Yacine Adane, Izzet Kale, "Novel multipath mitigation methods using a dual-polarization antenna", *Proceedings of the annual meeting of the Institute of Navigation (ION GNSS+)*, pags. 140–151, 2010.

[GSA16] GSA, "First EGNOS LPV-200 approach implemented at Charles de Gaulle Airport", Tech. rep., European GSA website available at `http://www.gsa.europa.eu/`, 2016.

[gua15] The guardian, "Self-driving cars are coming, and the technology promises to save lives", *The guardian webpage*, Retrieved from : `https://www.theguardian.com/technology/2015/dec/17/self-driving-cars-safety-future-interactive`, Accesed: 18-6-2017, 2015.

[Gué12] Blaise Kévin Guépié, Lionel Fillatre, Igor Nikiforov, "Sequential detection of transient changes", *Sequential Analysis*, Vol. 31, nº 4, pags. 528–547, 2012.

[Gum12] Emil Julius Gumbel, *Statistics of extremes*, Courier Corporation, 2012.

[Haj09] Ali Hajbabaie, Rahim F. Benekohal, Madhav V. Chitturi, Ming-Heng Wang, Juan C. Medina, "Comparison of effects of automated speed enforcement and police presence on speeding in work zones", *Proceedings of the Transportation Research Board 88th Annual Meeting*, pags. 1–13, 2009.

[Han98] Chunming Han, Peter Willett, Biao Chen, Douglas Abraham, "A detection optimal min-max test for transient signals", *IEEE Transactions on Information Theory*, Vol. 44, nº 2, pags. 866–869, 1998.

[Han99] Chunming Han, Peter Willett, Douglas A. Abraham, "Some methods to evaluate the performance of Page's test as used to detect transient signals", *IEEE Transactions on Signal Processing*, Vol. 47, nº 8, pags. 2112–2127, 1999.

[Hay09] Monson H. Hayes, *Statistical Digital Signal Processing and Modeling*, John Wiley & Sons, 2009.

[Hea05] Philip Headley, "ESC as a baseline for active safety", *Proceedings of 19th Enhanced Safety of Vehicles (ESV)*, pags. 1–5, 2005.

[Hsu15] Li-Ta Hsu, Shunsuke Kamijo, "NLOS exclusion using consistency check and city building model in deep urban canyons", *Proceedings of the annual meeting of the Institute of Navigation (ION GNSS+)*, pags. 2390–2396, 2015.

[Hum08] Todd E. Humphreys, Brent M. Ledvina, Mark L. Psiaki, Brady W. Hanlon, Paul M. Kintner, "Assessing the spoofing threat: Development of a portable GPS civilian spoofer", *Proceedings of the 21st International Technical Meeting of the Satellite Division of The Institute of Navigation (ION)*, pags. 2314–2325, 2008.

[Hut97] L. C. Hutwagner, E. K. Maloney, N. H. Bean, L. Slutsker, S. M. Martin, "Using laboratory-based surveillance data for prevention: An algorithm for detecting salmonella outbreaks", *Emerging Infectious Diseases*, Vol. 3, nº 3, pags. 395–400, 1997.

[ICA02] ICAO, "Global air navigation plan for CNS/ATM systems", Tech. rep., 2nd Edition, Doc 9750 AN/963, 2002.

[ICA06] ICAO, "Annex 10 (Aeronautical Telecommunications): To the convention on international civil aviation, Volume I -Radio navigation aids, international standards and recommended practices (SARPs)", Tech. rep., 6th Edition, 2006.

[IWG14] SBAS IWG, "SBAS interoperability working group, global SBAS status-presentation", Tech. rep., 2014.

[Kal87] R. M. Kalafus, "Receiver autonomous integrity monitoring of GPS", *Project Memorandum DOTTSC-FAA-FA-736-1, US DOT Transportation System Center, Cambridge, MA*, 1987.

[Kap05] Elliott Kaplan, Christopher Hegarty, *Understanding GPS: Principles and Applications, Second Edition*, Artech House, 2nd ed., 2005.

[Kar07] John Karl, *Celestial Navigation in the GPS Age*, Paradise Cay Publications, 2007.

[Kay98a] Steven M Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*, Prentice Hall Upper Saddle River, NJ, USA, 1998.

[Kay98b] Steven M Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*, Prentice Hall Upper Saddle River, NJ, USA, 1998.

[Ken48] Maurice George Kendall, *The Advanced Theory of Statistics*, Charles Griffin and Co., 4th ed., 1948.

[Kni02] John C. Knight, "Safety critical systems: Challenges and directions", *Proceedings of 24th International Conference on Software Engineering*, pags. 547–550, 2002.

[Lai98] Tze Leung Lai, "Information bounds and quick detection of parameter changes in stochastic systems", *IEEE Transactions on Information Theory*, Vol. 44, nº 7, pags. 2917–2929, 1998.

[Lee86] Young C. Lee, "Analysis of range and position comparison methods as a means to provide GPS integrity in the user receiver", *Proceedings of the 42nd Annual Meeting of the Institute of Navigation*, 1986.

[Lee95] Nigel Lee, Stuart C. Schwartz, "Robust transient signal detection using the oversampled Gabor representation", *IEEE Transactions on Signal Processing*, Vol. 43, nº 6, pags. 1498–1502, 1995.

[Lee04] Hyung Keun Lee, Jang Gyu Lee, Gyu In Jee, "GPS multipath detection based on sequence of successive-time double-differences", *IEEE Signal Processing Letters*, Vol. 11, nº 3, pags. 316–319, 2004.

[Lev08] Bernard C. Levy, *Principles of Signal Detection and Parameter Estimation*, Springer, 2008.

[Li13] Di Li, Lifeng Lai, Shuguang Cui, "Quickest change detection and identification across a sensor array", *Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pags. 145–148, 2013.

[Lin06] Anders Lindgren, Fang Chen, "State of the art analysis: An overview of advanced driver assistance systems (ADAS) and possible human factors issues", *Human factors and economics aspects of safety*, pags. 38–50, 2006.

[Lin07] Jonas Lindstrom, Dennis Akos, Oscar Isoz, Marcus Junered, "GNSS interference detection and localization using a network of low cost front-end modules", *Proceedings of the 20th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+)*, 2007.

[Lor71] Gary Lorden, "Procedures for reacting to a change in distribution", *The Annals of Mathematical Statistics*, Vol. 42, nº 6, pags. 1897–1908, 1971.

[LS09] José A. López-Salcedo, Juan Manuel Parro-Jimenez, Gonzalo Seco-Granados, "Multipath detection metrics and attenuation analysis using a GPS snapshot receiver in harsh environments", *Proceedings of the 3rd European Conference on Antennas and Propagation. EuCAP*, pags. 3692–3696, 2009.

[LS16a] Sergi Locubiche-Serra, Gonzalo Seco-Granados, José A. López-Salcedo, "Doubly-adaptive autoregressive Kalman filter for GNSS carrier tracking under scintillation conditions", *Proceedings of International Conference on Localization and GNSS (ICL-GNSS)*, pags. 1–6, 2016.

[LS16b] Vicente Lucas-Sabola, Gonzalo Seco-Granados, José A. López-Salcedo, José A. García-Molina, Massimo Crisci, "Cloud GNSS receivers: New advanced applications made possible", *Proceedings of International Conference on Localization and GNSS (ICL-GNSS)*, pags. 1–6, 2016.

[M6t03] M6toll, "Ways to Pay available", *M6toll Webpage*, Retrieved from : `http://www.m6toll.co.uk/ways-to-pay/#!lightbox1599Item[lightbox1599]/0/`, Accesed: 19-6-2017, 2003.

[Med08] Wietske Medema, Brian S. Mcintosh, Paul J . Jeffrey, "From premise to practice : A critical assessment of integrated water resources management and adaptive management approaches in the water sector", *Ecology And Society*, Vol. 13, nº 2, pags. 29, 2008.

[Mei10] Yajun Mei, "Efficient scalable schemes for monitoring a large number of data streams", *Biometrika*, Vol. 97, nº 2, pags. 419–433, 2010.

[Mei11] Yajun Mei, "Quickest detection in censoring sensor networks", *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pags. 2148–2152, 2011.

[Mer01] Stelios P. Mertikas, "Automatic and online detection of small but persistent shifts in GPS station coordinates by statistical process control", *GPS Solutions*, Vol. 5, nº 1, pags. 39–50, 2001.

[Mis11] Pratap Misra, Per Enge, *Global positioning system: signals, measurements, and performance*, Ganga-Jamuna Press, 2nd ed., 2011.

[Møg12] Andreas Møgelmose, Mohan Manubhai Trivedi, Thomas B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, nº 4, pags. 1484–1497, 2012.

[Mou86] George V. Moustakides, "Optimal stopping times for detecting changes in distributions", *The Annals of Statistics*, Vol. 14, nº 4, pags. 1379–1387, 1986.

[Mou14] George V. Moustakides, "Multiple optimality properties of the Shewhart test", *Sequential Analysis*, Vol. 33, pags. 318–344, 2014.

[Mub10] Omer Mohsin Mubarak, Andrew G. Dempster, "Exclusion of multipath-affected satellites using early late phase", *Journal of GPS*, Vol. 9, nº 2, pags. 145–155, 2010.

[Mul04] Paolo Mulassano, Fabio Dovis, Frederic Collomb, "European projects for innovative GNSS-related applications", *GPS Solutions*, Vol. 7, nº 4, pags. 268–270, 2004.

[Nad11] Boaz Nadler, Federico Penna, Roberto Garello, "Performance of eigenvalue-based signal detectors with known and unknown noise level", *Proceedings of the IEEE International Conference onCommunications (ICC)*, pags. 1–5, IEEE, 2011.

[Nav11] Navipedia, "Critical Applications", *Navipedia webpage*, Retrieved from : `http://www.navipedia.net/index.php/Criticality_of_GNSS_Applications`, Accesed: 18-6-2017, 2011.

[Nav17] Navipedia, "Navipedia wbsite", *Navipedia*, Retrieved from : `http://www.navipedia.net/`, 2017.

[Nev75] Jacques Neveu, *Discrete Parameter Martingales*, Elsevier, 1975.

[Nik96] Igor Nikiforov, "New optimal approach to global positioning system/differential global positioning system integrity monitoring", *Journal of Guidance, Control, and Dynamics*, Vol. 19, nº 5, pags. 1023–1033, 1996.

[Nik97a] Igor Nikiforov, "Sequential, FSS and snapshot approaches to GPS/DGPS integrity monitoring", *Proceedings of the 10th International Meeting of The Satellite Division of The Institute of Navigation*, pags. 449–458, 1997.

[Nik97b] Igor Nikiforov, "Two strategies in the problem of change detection and isolation", *IEEE Transactions on Information Theory*, Vol. 43, nº 2, pags. 770–776, 1997.

[Nik02] Igor Nikiforov, Benoit Roturier, "Statistical analysis of different RAIM schemes", *Proceedings of the 15th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2002)*, pags. 1881–1892, 2002.

[Num12] Julia Numrich, Sascha Ruja, Stefan Voß, "Global navigation satellite system based tolling: State-of-the-art", *NETNOMICS: Economic Research and Electronic Networking, Springer*, Vol. 13, nº 2, pags. 93–123, 2012.

[Osk05] Taragay Osklper, H. Vincent Poor, "Quickest detection of a random signal in background noise using a sensor array", *Eurasip Journal on Applied Signal Processing*, Vol. 2005, nº 1, pags. 13–24, 2005.

[Pag54] E. S. Page, "Continuous inspection schemes", *Biometrika*, Vol. 41, pags. 100–115, 1954.

[Pag16] Jean-Baptiste Pagot, *Modeling and Monitoring of New GNSS Signal Distortions in the Context of Civil Aviation*, PhD Thesis, Universitè de Toulouse, 2016.

[Pal11] André de Palma, Robin Lindsey, "Traffic congestion pricing methodologies and technologies", *Transportation Research Part C: Emerging Technologies*, Vol. 19, nº 6, pags. 1377–1399, 2011.

[Pap02] Athanasios Papoulis, S. Unnikrishna Pillai, *Probability, Random Variables, and Stochastic Processes*, New York, 4th ed., 2002.

[Par88] Bradford W. Parkinson, Penina Axelrad, "Autonomous GPS integrity monitoring using the pseudorange residual", *Navigation*, Vol. 35, nº 2, pags. 255–274, 1988.

[Par96a] Bradford W. Parkinson, James J. Spilker, *Global Positioning System: Theory and Applications vol.1*, Aiaa, 1996.

[Par96b] Bradford W. Parkinson, James J. Spilker, *Global Positioning System: Theory and Applications vol.2*, Aiaa, 1996.

[Pel87] Lionel Pelkowitz, Stuart C. Schwartz, "Asymptotically optimum sample size for quickest detection", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-23, nº 2, pags. 263–272, 1987.

[Pic06] Andrew T. Pickford, Philip T. Blythe, *Road user Charging and Electronic Toll Collection*, Artech House London, 2006.

[Pol10] Yury Polyanskiy, H. Vincent Poor, Sergio Verdú, "Channel coding rate in the finite blocklength regime", *IEEE Transactions on Information Theory*, Vol. 56, nº 5, pags. 2307–2359, 2010.

[Pol13] Moshe Pollak, Abba M. Krieger, "Shewhart Revisited", *Sequential Analysis*, Vol. 32, pags. 230–242, 2013.

[Poo94] H. Vincent Poor, *An Introduction to Signal Detection and Estimation*, Springer, 2nd ed., 1994.

[Poo09] H. Vincent Poor, Olympia Hadjiliadis, *Quickest Detection*, Cambridge University Press, 2009.

[Por92] Boaz Porat, Benjamin Friedlander, "Performance analysis of a class of transient detection algorithms - A unified framework", *IEEE Transactions on Signal Processing*, Vol. 2, nº October, pags. 2536–2546, 1992.

[Pro07] John G. Proakis, Dimitris G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 4th ed., 2007.

[Pul12] Sam Pullen, G .Xingxin Gao, "GNSS jamming in the name of privacy-potential threat to GPS aviation", *Inside GNSS*, Vol. 7, nº 2, pags. 34–43, 2012.

[Rag08] Vasanthan Raghavan, Venugopal V. Veeravalli, "Quickest detection of a change process across a sensor array", *Proc. International Conference on Information Fusion*, pags. 1–8, 2008.

[Ray99] Jayanta Kumar Ray, M. Elizabeth Cannon, "Characterization of GPS carrier phase multipath", *Proceedings of the ION National Technical Meeting*, pags. 243–252, 1999.

[RTC06] RTCA, "Minimum operational performance standards for global positioning system/wide area augmentation system airborne equipment", Tech. rep., RTCA, 2006.

[SA12] Carlos Daniel Salós-Andrés, *Integrity Monitoring Applied to the Reception of GNSS Signals in Urban Environments*, PhD Thesis, Institut National Polytechnique de Toulouse, 2012.

[Saa04] Farida Saad, "Behavioural adaptations to new driver support systems: Some critical issues", *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Vol. 1, pags. 288–293, 2004.

[Sal10] Daniel Salós, Christophe Macabiau, Anaïs Martineau, Bernard Bonhoure, Damien Kubrak, "Analysis of GNSS integrity requirements for road user charging applications", *Proceedings of the 5th IEEE ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC)*, pags. 1–8, 2010.

[Sch09] Ron Schneiderman, "DSPs are helping to make it hard to get lost", *IEEE Signal Processing Magazine*, Vol. 26, nº 6, pags. 9–13, 2009.

[Sch11] Anna Schlösser, *Pricing and Risk Management of Synthetic CDOs*, Springer Science & Business Media, 2011.

[Seo13] Jiwon Seo, Mincheol Kim, "eLoran in Korea-Current Status and Future Plans", *Proceedings of the European Navigation Conference, Vienna*, pags. 23–27, 2013.

[SG12] Gonzalo Seco-Granados, José A. López-Salcedo, David Jiménez-Baños, Gustavo López-Risueño, "Challenges in indoor global navigation satellite systems", *IEEE Signal Processing Magazine*, Vol. 29, nº 2, pags. 108–131, 2012.

[She11] Daniel Shepard, Jahshan A. Bhatti, Todd E. Humphreys, "Drone hack: Spoofing attack demonstration on a civilian unmanned aerial vehicle", *GPS World*, Vol. 1, nº Dec., 2011.

[Shi63] Albert Shiryaev, "On the detection of disorder in a manufacturing process", *Theory of Probability & Its Applications*, Vol. 8, nº 3, pags. 247–265, 1963.

[Shi78] Albert Shiryaev, *Optimal Stopping Rules*, Springer, 1978.

[Shy14] Enik Shytenneja, Axel Garcia-Pena, Olivier Julien, "Proposed architecture for integrity monitoring of a GNSS / MEMS system with a fisheye camera in urban environment", *Proceedings of the International Conference on Localization and GNSS (ICL-GNSS)*, pags. 1–6, 2014.

[Sie85] David Siegmund, *Sequential Analysis: Tests and Confidence Intervals*, Springer, 1985.

[Sim03] Slobodan N. Simić, Shankar Sastry, "Distributed environmental monitoring using random sensor networks", *Information Processing in Sensor Networks*, pags. 582–592, Springer, 2003.

[Slo10] NDSAS (Motorways roads in Slovakia), Express, "Multi-lane free-flow electronic tolling in the Slovak Republic", Tech. rep., 2010.

[Sma07] Kenneth A. Small, Erik T. Verhoef, *The Economics of Urban Transportation*, Routledge, 2007.

[Sna91] Steven R. Snapp, James Brentano, Giham V. Dias, Terrance L. Goan, L. Todd Heberlein, Che-Lin Ho, Karl N. Levitt, Biswanath Mukherjee, Stephen E. Smaha, Tim Grance, Daniel M. Teal, Doug Mansur, "DIDS (distributed intrusion detection system)-motivation, architecture, and an early prototype", *Proceedings of the National Computer Security Conference*, pags. 167–176, Washington, DC, 1991.

[Sot08] Miguel Ángel Sotelo, José Barriga, "Blind spot detection using vision for automotive applications", *Journal of Zhejiang University: Science A, Springer*, Vol. 9, nº 10, pags. 1369–1372, 2008.

[Stk11] Czeslaw Stkepniak, "A representation of nonhomogeneous quadratic forms with application to the least squares solution", *Mathematical Problems in Engineering*, Vol. 2011, 2011.

[Str99] Roy L. Streit, Peter Willett, "Detection of random transient signals via hyperparameter estimation", *IEEE Transactions on Signal Processing*, Vol. 47, nº 7, pags. 1823–1834, 1999.

[Str16] Future Structure, "Looking ahead: The future of regulations on autonomous vehicles (industry perspective)", *govtech webpage*, Retrieved from: `http://www.govtech.com/fs/perspectives/Looking-Ahead-The-Future-of-Regulations-on-Autonomous-Vehicles.html`, Accesed: 18-6-2017, 2016.

[Stu88] Mark A. Sturza, "Navigation system integrity monitoring using redundant measurements", *Navigation*, Vol. 35, nº 4, pags. 483–501, 1988.

[Tar06] Alexander G. Tartakovsky, Boris L. Rozovskii, Rudolf B. Blazeka, Hongjoong Kim, "Detection of intrusions in information systems by sequential change-point methods", *Statistical Methodology*, Vol. 3, nº 3, pags. 252–293, 2006.

[Tar08] Alexander G. Tartakovsky, Venugopal V. Veeravalli, "Asymptotically optimal quickest change detection in distributed sensor systems", *Sequential Analysis*, Vol. 27, nº 4, pags. 441–475, 2008.

[Tar14] Alexander Tartakovsky, Igor Nikiforov, Michele Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*, CRC Press, 2014.

[TM10] Rafael Toledo-Moreo, Benito Úbeda, Jose Santa, Miguel A. Zamora-Izquierdo, Antonio F. Gómez-Skarmeta, "An analysis of positioning and map-matching issues for GNSS-based road user charging", *Proceedings of the IEEE Conference on Intelligent Transportation Systems (ITSC)*, pags. 1486–1491, 2010.

[Tri14] Trimble, "Trimble's Sky Plot Tool", *Trimble webpage*, Retrieved from: `http://www.trimble.com/GNSSPlanningOnline/#/SkyPlot`, Accesed: 19-6-2017, 2014.

[Tro11] Carmela Troncoso, George Danezis, Eleni Kosta, Josep Balasch, Bart Preneel, "PriPAYD: Privacy-friendly pay-as-you-drive insurance", *IEEE Transactions on Dependable and Secure Computing*, Vol. 8, nº 5, pags. 742–755, 2011.

[Tse09] Theodore Tsekeris, Stefan Voß, "Design and evaluation of road pricing: State-of-the-art and methodological advances", *NETNOMICS: Economic Research and Electronic Networking, Springer*, Vol. 10, nº 1, pags. 5–52, 2009.

[Tur07] Seppo Turunen, "Network assistance: What will new GNSS signals bring to it", *Inside GNSS*, Vol. 2, nº 3, pags. 35–41, 2007.

[Vah03] Ardalan Vahidi, Azim Eskandarian, "Research advances in intelligent collision avoidance and adaptive cruise control", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 4, nº 3, pags. 143–153, 2003.

[Van92] Richard D. J. Van Nee, "Multipath effects on GPS code phase measurements", *Navigation*, Vol. 39, nº 2, pags. 177–190, 1992.

[Vis17] Plate Vision, "Toll & Traffic ANPR Software", *Plate Vision webpage*, Retrieved from : `http://plate.vision/project/toll-traffic-anpr-software/`, Accesed: 19-6-2017, 2017.

[Wal45] Abraham Wald, "Sequential tests of statistical hpotheses", *The Annals of Mathematical Statistics*, Vol. 16, nº 2, pags. 117–186, 1945.

[Wal95] Todd Walter, Per Enge, "Weighted RAIM for precision approach", *Proceedings of the International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS)*, pags. 1995–2004, 1995.

[Wan01] Zhen Wang, Peter Willett, "All-purpose and plug-in power-law detectors for transient signals", *IEEE Transactions on Signal Processing*, Vol. 49, nº 11, pags. 2454–2466, 2001.

[Wan05] Zhen Wang, Peter Willett, "A variable threshold Page procedure for detection of transient signals", *IEEE Transactions on Signal Processing*, Vol. 53, nº 11, pags. 4397–4402, 2005.

[Wei10] Lu Wei, Olav Tirkkonen, "Spectrum sensing with Gaussian approximated eigenvalue ratio based detection", *Proceedings of the 7th International Symposium on Wireless Communication Systems (ISWCS)*, pags. 961–965, 2010.

[Wet91] G. Barrie Wetherill, Don W. Brown, *Statistical Process Control: Theory and Practice*, Chapman and Hall London, 1991.

[Woo06] William H. Woodall, "The use of control charts in health-care and public-health surveillance", *Journal of Quality Technology*, Vol. 38, nº 2, pags. 89–104, 2006.

[You98] Abdelrazak Younes, Bacem Bakhache, Igor Nikiforov, Abdelahad Benhallam, "Sequential RAIM: Theory and application to civil aviation needs", *Proceedings of the 11th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS)*, pags. 1993–2002, 1998.

[Zab11] Martina Zabic, *GNSS-based Road Charging Systems*, PhD Thesis, Technical Unevirsity of Denmark, 2011.

[Zar09] Sepideh Zarrin, Teng Joon Lim, "Cooperative quickest spectrum sensing in cognitive radios with unknown parameters", *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, pags. 1–6, 2009.